Journal on Mobile Networks and Applications (MONET), Special issue on Satellite-Based Information Services, Volume 7, No. 1, pp. 79-89, January 2002.

Delay bounds for voice over IP calls transported over satellite access networks

Jan Janssen[†], Danny De Vleeschauwer[†], Rudi Windey[‡], Guido H. Petit[†], Jean-Marie Leroy[‡]

[†] *Alcatel Bell, Corporate Research Center, Francis Wellesplein 1, B-2018 Antwerp, Belgium*
{jan.janssen, danny.de_vleeschauwer, guido.h.petit}@alcatel.be

[‡] *Alcatel Bell Space, Berkenrodelei 33, B-2660 Hoboken, Belgium*
{rudi.windey, jean-marie.leroy}@alcatel.be

Whether or not voice calls of traditional quality can be supported between 2 users connected to an IP backbone via satellite access systems depends largely on the mouth-to-ear delay, an important part of which is consumed by the satellite networks themselves. In this paper, a methodology is developed to calculate upper bounds for the latter delay component as a function of the used codec, the experienced packet loss ratio, the echo levels at both sides of the connection and the chosen voice packet size. Illustrations are provided.

1. Introduction

The quality of Voice over IP (VoIP) calls is largely determined by the mouth-to-ear delay, i.e., the time that elapses between the moment the talker utters the words and the moment the listener hears them. Other important factors that may impair the quality of a VoIP call are the use of compression techniques, the occurrence of packet loss and the level of echo. Given the codec used, the packet loss ratio and the echo level, we first determine the mouth-to-ear delay values below which traditional (circuit-switched) quality is attained.

Both the access networks (connecting the 2 involved parties to the IP backbone) and the IP backbone will consume a part of the tolerable mouth-to-ear delay. Here, we assume that the IP backbone is accessed via satellites at both sides of the VoIP connection. The main objective of this paper is to present a general method to calculate the portion of the mouth-to-ear delay budget that may be consumed by these satellite access systems when traditional quality is aimed for. Besides the influence of the used codec, the experienced packet loss ratio and the echo levels at both sides of the connection, the impact of the choice of the voice packet size is considered as well. The resulting effective bit rates, i.e., the rates at which bits are sent onto these access networks, are also investigated in this context.

This paper is organized as follows. We start in section 2 by a short description of the E-model, which is a standardized tool to predict the subjective quality of telephone calls. Using this model, mouth-to-ear delays are calculated below which traditional quality is attained for several standardized codecs. In section 3, we recall some general facts on satellite access systems and derive the formulae needed to determine the delay budget they may consume in a satellite-PC-to-satellite-PC VoIP call of traditional quality. Also a definition of the effective bit rate is given. In section 4, some concrete results are presented and discussed in detail. Finally, section 5 contains the conclusions.

2. The E-model

The E-model [4, 5, 9] predicts the subjective quality of a telephone call based on its characterizing transmission parameters. More precisely, it combines the impairments caused by these transmission parameters into a rating R, which can be used to forecast subjective user reactions, such as the Mean Opinion Score (MOS) or the percentage of users finding the quality Good or Better (GoB).

The R-scale was defined so that impairments are approximately additive in the R-range of interest, i.e.,

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

The first term R_0 groups the effects of noise, such as background noise and circuit noise. The second term I_s includes impairments that occur simultaneously with the voice signal, such as those caused by quantization, by too loud a connection and by too loud a side tone. The third term I_d encompasses delayed impairments, including impairments caused by (talker and listener) echo and by loss of interactivity. The fourth term I_e covers the distortion impairment caused by the use of special equipment. For example, each low bit rate codec has an associated distortion impairment value I_e that increases as the packet loss experienced by the codec increases. The fifth term is the expectation factor A, which expresses the decrease in the rating R a user is willing to tolerate because of the "access advantage" that certain systems have over traditional wire-bound telephony. For example, the expectation factor A for mobile telephony equals 10.

Consider a packetized voice call between 2 parties, referred to as party 1 and party 2 (see Figure 1). Using the E-model, we calculate how party 1 will judge the call, that is, what rating R will be assigned to it. First, we study the influence of delay (via I_d), followed by the influence of distortion (via I_e). The values of R_0 and I_s are chosen to be the default ones as they are not fundamentally different for circuit-switched and packetized voice calls. Furthermore, the influence of the expectation factor is neglected ($A = 0$) in this paper in order to make a fair comparison between the quality of packetized and traditional wire-bound telephone calls. For more details, we refer to [8].

Remark that eq. (1) implies that 2 calls with the same rating R can give a totally different subjective impression. One call might produce crystal clear, undistorted speech (e.g. $I_e = 0$) but suffer from a relatively large delay (e.g. $I_d = 10$). Another call might slightly distort the speech (e.g. $I_e = 10$), while its delay is not noticeable (e.g. $I_d = 0$). However, the E-model predicts that a judging panel will award the same MOS to both calls and the same percentage of users will find both calls GoB, albeit for different reasons.

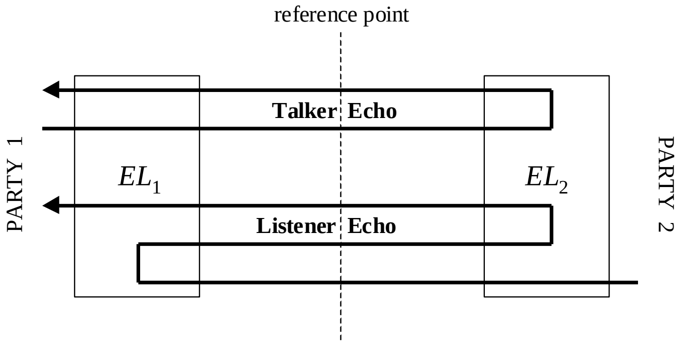


Figure 1: Talker and listener echo observed by party 1.

2.1. Influence of mouth-to-ear delay

If the voice signal party 1 hears is delayed, the rating R decreases by an amount equal to the impairment I_d associated with the mouth-to-ear delay. This impairment is the sum of 3 contributing impairments: talker echo, listener echo and loss of interactivity.

First, talker echo disturbs party 1, who hears an attenuated and delayed echo of his own voice. This echo is caused by a reflection close to party 2. The level of this echo is strongly influenced by the echo loss EL_2 close to party 2 (measured with respect to a certain reference point).

Second, listener echo also disturbs party 1, who hears the original signal from party 2 followed by an attenuated echo of the signal. This echo is determined by a reflection close to party 1 with attenuation EL_1, followed by a reflection close to party 2 with attenuation EL_2.

Echo may occur in a 4-to-2-wire hybrid if the packetized voice call is terminated over a local Public Switched Telephone Network (PSTN) or in the callers' terminal equipment. For PSTN calls from traditional handsets, where echo is mainly caused by the hybrids, a typical value for the echo loss is 21 dB [5]. The same value is valid for packetized voice calls terminated over a local PSTN to traditional handsets. Echo loss is likely to be lower for other kinds of terminals, such as personal computers and handsfree phones. The echo losses EL_1 and EL_2 can be increased by using an echo controller, which should be deployed as close to the source of echo as possible, that is, in the gateways between the PSTN and the packet-based network, or in the terminals. A simple echo controller can increase the echo loss by 30 dB. Perfect echo control, in which the echo losses EL_1 and EL_2 increase to infinity, can be achieved at moderate computational cost.

The third delay-related factor that may disturb party 1 is the loss of interactivity. If the mouth-to-ear delay is too large, an interactive conversation becomes impossible.

With respect to delay, ITU-T Recommendations G.114 [1] and G.131 [2] specify the following guidelines for traditional (undistorted) PSTN calls:

- For an echo loss value of 21 dB, echo control is needed if the mouth-to-ear delay is larger than 25 ms.
- When the echo is adequately controlled:

- a mouth-to-ear delay up to 150 ms is acceptable for most user applications,
- a mouth-to-ear delay between 150 ms and 400 ms is acceptable, provided that one is aware of the impact of delay on the interactivity, and
- a mouth-to-ear delay above 400 ms is unacceptable.

In Figure 2, the E-model is used to calculate the rating R given by party 1 in the case of undistorted voice. In the case of packetized voice calls, undistorted calls are calls transported without packet loss in the G.711@64 kb/s format. Different values of the echo loss, which is assumed to be equal at both end points ($EL = EL_1 = EL_2$), are considered.

Observe that the rating R is a non-increasing function of the mouth-to-ear delay. Moreover, the impairment associated with delay depends strongly on the echo loss value. For an echo loss value of 21 dB, the rating R drops below 70 at a mouth-to-ear delay of 25 ms, while R is larger than 70 for mouth-to-ear delays up to 400 ms if the echo is perfectly controlled ($EL = EL_1 = EL_2 = \infty$).

By comparing the latter delay bounds (obtained with the E-model for undistorted voice) with the above-mentioned standardized bounds of ITU-T Recommendations G.114 and G.131, we derive that traditional PSTN calls have a rating R of at least 70. As such, we have taken $R = 70$ as the limit for traditional or "toll quality", the latter which is an ill-used term according to ITU-T Recommendation G.109 [3].

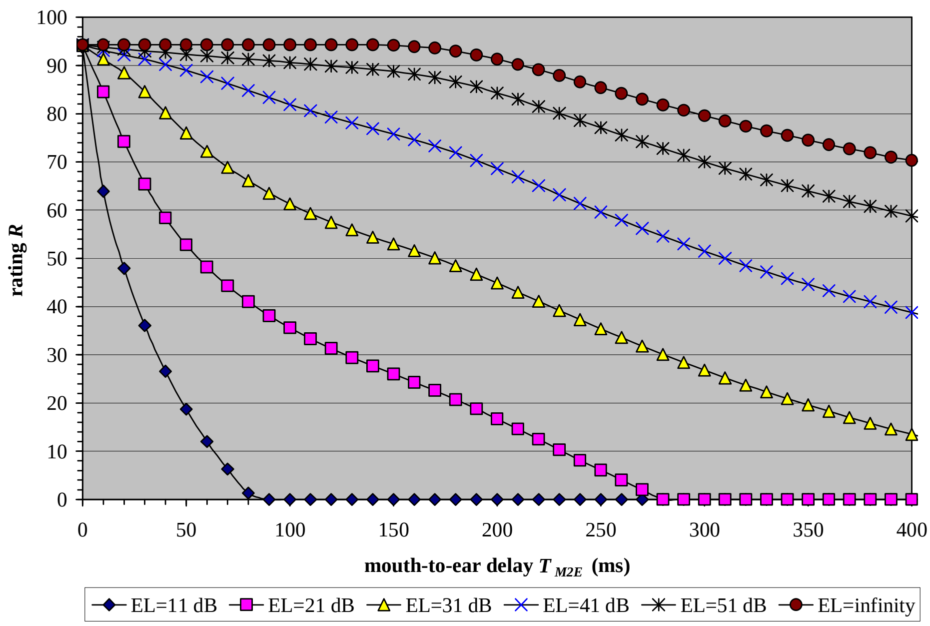


Figure 2: The rating R as a function of the mouth-to-ear delay T_{M2E} for undistorted voice and various echo loss values.

2.2. Influence of distortion

If the voice signal party 1 hears is distorted, the rating R decreases by an amount equal to the distortion impairment I_e. This impairment has 2 sources: encoding of the voice signal from party 2 and packet loss during the transport of voice packets from party 2 to party 1.

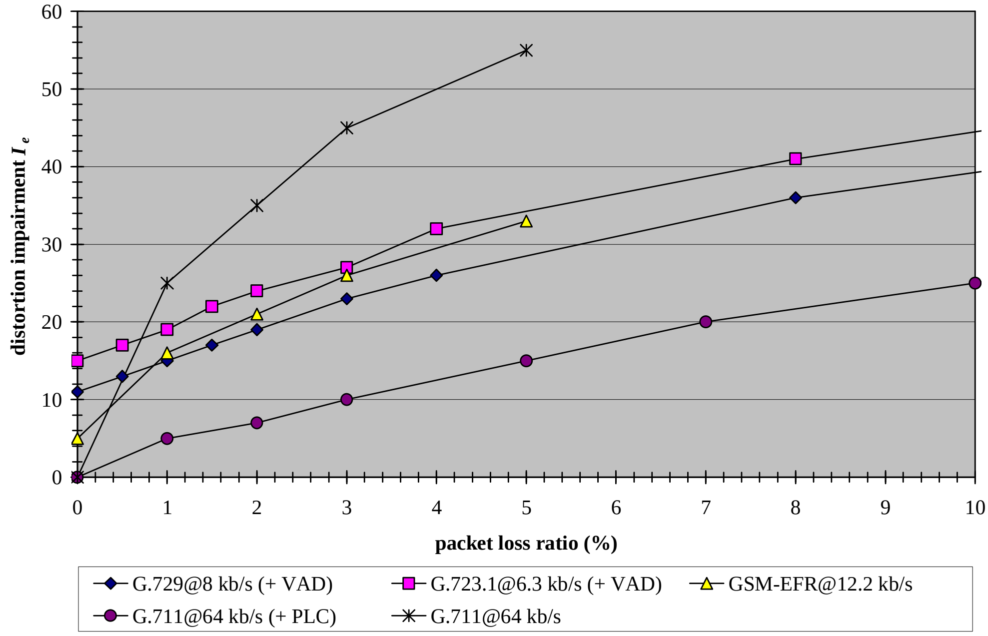


Figure 3: The distortion impairment term I_e as a function of the packet loss ratio different codecs.

	EL = 51 dB	EL = ∞
G.711@64 kb/s	282	373
G.711@64 kb/s (+ PLC)	296	397
GSM-EFR@12.2 kb/s	257	333
G.729@8 kb/s (+ VAD)	221	284
G.723.1@6.3 kb/s (+ VAD)	192	250

(a) packet loss ratio = 0.1 %

	EL = 51 dB	EL = ∞
G.711@64 kb/s		
G.711@64 kb/s (+ PLC)	264	345
GSM-EFR@12.2 kb/s	186	245
G.729@8 kb/s (+ VAD)	195	253
G.723.1@6.3 kb/s (+ VAD)	145	221

(b) packet loss ratio = 1 %

	EL = 51 dB	EL = ∞
G.711@64 kb/s		
G.711@64 kb/s (+ PLC)	195	253
GSM-EFR@12.2 kb/s		
G.729@8 kb/s (+ VAD)		
G.723.1@6.3 kb/s (+ VAD)		

(c) packet loss ratio = 5 %

Table 1: Tolerable mouth-to-ear delays T_{M2E} (in ms) below which traditional quality is obtained for different codecs, packet loss ratios of 0.1, 1 and 5 % and echo loss values of 51 dB and infinity. The colored entries denote the fact that traditional quality cannot be attained.

The distortion impairment I_e associated with a codec increases as the packet loss ratio increases. Figure 3, based on [6], shows this effect for the G.711, GSM-EFR@12.2 kb/s, G.729@8 kb/s and G.723.1@6.3 kb/s codecs, the latter 2 using a Voice Activity Detection (VAD) scheme.

This figure deals with randomly lost voice packets and a specific packetization interval per codec (10 ms for G.711, 20 ms for G.729 and GSM-EFR, 30 ms for G.723.1). Results are not yet known for other packetization intervals or other codecs.

The sensitivity to packet loss depends clearly on whether a Packet Loss Concealment (PLC) technique is used by the codec. In contrast to the G.711 codec, all considered low bit rate codecs (i.e. GSM-EFR, G.729 and G.723.1) have a built-in PLC scheme. However, a PLC scheme can be implemented on top of the G.711 codec.

In Table 1, we report the maximum mouth-to-ear delays that can be tolerated in order to obtain traditional quality, referred to as the mouth-to-ear delay budgets. We considered the codecs mentioned above, packet loss ratios of 0.1, 1 and 5 % and echo loss values of 51 dB and infinity. These results can be derived from Figure 2 and Figure 3 in a straightforward manner. First, determine the curve on Figure 2 corresponding to the correct echo loss value. Then, shift this curve downwards according to the considered value of I_e, which can be derived from Figure 3 for a given codec and packet loss ratio. Finally, determine at which delay value the rating R drops below 70.

Roughly speaking, it follows from Table 1 that for a fixed packet loss ratio and echo loss value, the mouth-to-ear delay budget increases as the bit rate of the codec increases. For a chosen codec, this mouth-to-ear delay turns out to be an increasing function of the echo loss value. On the other hand, the tolerable mouth-to-ear delay is inversely related to the experienced packet loss ratio. In particular, several codecs (from the set considered in this paper) fail to deliver traditional quality voice for large packet loss ratios.

3. The satellite delay budget

3.1. Introduction

For the satellite-PC[1]-to-satellite-PC VoIP scenario considered in this paper (and illustrated in Figure 4), the total mouth-to-ear delay T_{M2E} can be divided into several components, one of them being the delay $T_{satellites}$ that includes all delays introduced by the satellite access systems at both sides of the connection. This delay component consists of e.g. interleaving, (de)modulation, propagation, serving and queuing delays. Possibly, the voice information is transmitted over a shared medium such that in the upstream direction also a Medium Access Controller (MAC) delay can occur.

The IP packets of a voice flow leave the talker's satellite PC at a constant packet rate. In the satellite access networks and the IP backbone, each individual IP packet is delayed over a stochastic time such that the flow is jittered, that is, the voice packets do not arrive at the listener's side at a constant rate. Since the decoder needs a constant flow of packets, a dejittering buffer is necessary in the listener's satellite equipment to compensate for this difference in delays. For the VoIP scenario considered here, the dejittering delay T_{jit} (the time the first voice packet is held up in this dejittering buffer) compensates for the difference in delays encountered between the talker's and the listener's satellite PCs. In order to limit the packet loss (due to packets arriving after they were supposed to be read out), it must not be chosen too small. On the other hand, in order not to increase the mouth-to-ear delay too much, the dejittering delay must not be taken too large. In this paper, we approximated the dejittering delay T_{jit} by dividing it in 2 parts, compensating for the difference in delays in the IP backbone ($T_{jit}^{backbone}$) and the satellite access networks ($T_{jit}^{satellites}$), respectively. Although the latter component is not really consumed by the satellite access networks, it is included in the delay budget $T_{satellites}$ foreseen for them.

An important delay component in a satellite access system is the propagation delay between the satellites and the ground equipment, i.e., the satellite PC and the IP backbone network. An easy calculation shows that for Geostationary Earth Orbit (GEO) satellites with an altitude of 36000 km, the

[1] In order to emphasize the fact that an IP backbone is accessed via a satellite system, the user's equipment is referred to as a satellite PC. Obviously, the analysis presented in this paper also holds when the latter is replaced by a satellite phone with the same functionality.

propagation delay over the radio link equals at least (without taking into account possible elevation angles) 2×36000 km $\times$ 3.33 µs/km = 240 ms at one side of the connection. For Medium Earth Orbit (MEO) satellites (with an altitude of approximately 15000 km), the propagation delay at one side of the connection still equals about 100 ms. This is the reason why satellite access systems providing real-time services are likely to be designed using Low Earth Orbit (LEO) satellites (with an altitude of approximately 1500 km), whose minimum corresponding propagation delay equals about 10 ms. The delays because of the possible handovers in MEO and LEO satellite systems are also included in the delay budget $T_{satellites}$.

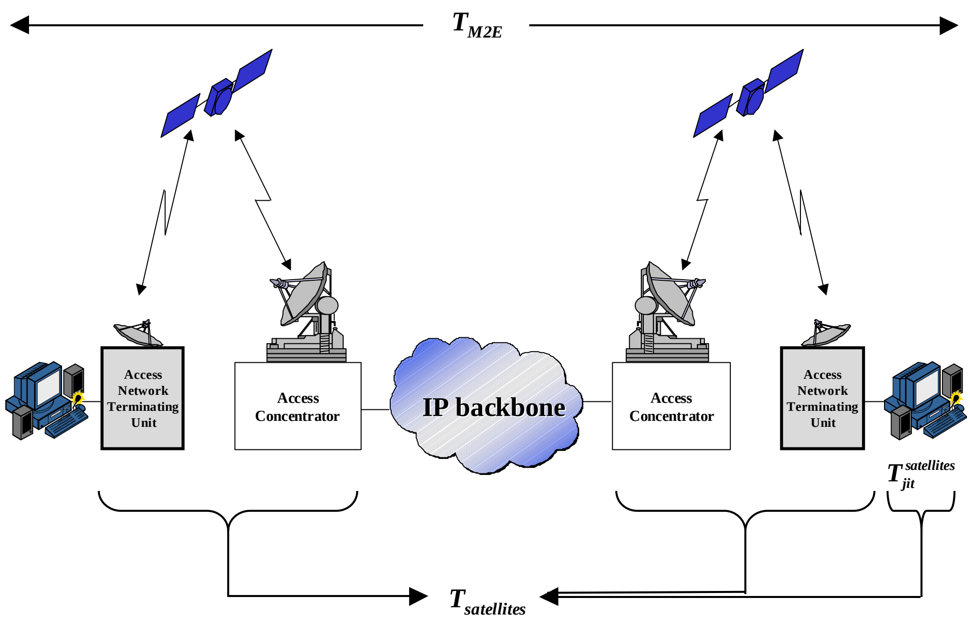


Figure 4: The satellite-PC-to-satellite-PC VoIP scenario.

Finally, we mention that in case of multi-media service offering, the size of the majority of the IP packets will be too large to send them as one single burst over the radio link. Therefore, fragmentation is needed. In the sequel of this paper, we assumed that the voice packets are fragmented into ATM cells, and sent over ATM Adaptation Layer (AAL) 5 in the satellite access networks.

3.2. Calculation methodology

In order to calculate the portion of the mouth-to-ear delay allowed to be consumed by the satellite access systems, we identify several other important components of the mouth-to-ear delay below.

All voice codecs work according to the same principle. They first collect a few quantized samples of speech, referred to as a voice frame of length T_F (in ms). Sometimes they also need some voice samples after the ones being encoded, referred to as the look-ahead of length T_{LA} (in ms), in order to better encode the samples of the current voice frame. Then the codecs calculate a code word of length B_F (in bits). This calculation takes an encoding time T_{enc}. At the receiver side the decoder uses the code word to produce a close copy of the original voice frame. Again it takes a certain time to perform this operation, referred to as the decoding time T_{dec}. We define the codec delay T_{cod} as $T_{enc} + T_{dec} + T_{LA}$. For the VoIP scenario considered in this paper, the encoding and decoding processes will be implemented in the end user's equipment (satellite PC), whose processing power is assumed to be large enough in order to perform both of them in a negligible amount of time. Hence, we have

$$T_{cod} = T_{LA} \ . \qquad (2)$$

An overview of the characterizing parameters (T_F, T_{LA} and B_F) for the codecs considered in this paper can be found in Table 2. Observe that the codec bit rate R_{cod} (in kb/s) equals B_F/T_F.

Origin	Standard	T_F (ms)	T_{LA} (ms)	B_F (bits)	R_{cod} (kb/s)
ITU-T	G.711	0.125	0	8	64
	G.723.1	30	7.5	189	6.3
	G.729	10	5	80	8
ETSI	GSM-EFR	20	0	244	12.2

Table 2: Characterizing parameters for the considered codecs.

The packetization delay T_{pack} is the time needed to fill an IP packet with N_F voice code words produced by the encoding process. Therefore, one easily derives that

$$T_{pack} = N_F T_F \ . \tag{3}$$

In order to characterize the service delay $T_{serv}^{backbone}$ in the IP backbone, we assume that this backbone network consists of N_{stag} nodes with link speed R_{link} (in kb/s). We then can write

$$T_{serv}^{backbone} = N_{stag} \frac{S_{link}}{R_{link}} = N_{stag} \frac{N_F B_F + O_{link}^{backbone}}{R_{link}} = N_{stag} \frac{N_F T_F R_{cod} + O_{link}^{backbone}}{R_{link}} \ , \tag{4}$$

with S_{link} the voice packet size (in bits) and $O_{link}^{backbone}$ the amount of overhead (in bits) on the link layer of the IP backbone. The latter equals 376 bits, stemming from 12 bytes RTP-header, 8 bytes UDP-header, 20 bytes IP-header and 7 bytes PPP-header [12]. Here (as well as in the rest of this section) we neglect the fact that IP packets are always rounded to an integer number of bytes.

The stochastic queuing delay $T_{queue}^{backbone}$ incorporates the time spent waiting behind other voice and data packets in the IP backbone network. If we assume a Quality of Service (QoS)-enabled IP backbone in which voice packets get non-pre-emptive Head Of Line (HOL) priority over all queuing data packets (arriving voice packets jump over all queuing data packets but a data packet in service is never interrupted), it is stochastically upper bounded by

$$T_{queue}^{backbone} = D(\rho, N_{HL}, P) \frac{MTU_{voice} + 56}{R_{link}} + N_{stag} \frac{MTU_{data} + 56}{R_{link}} \ . \tag{5}$$

The first term in this equation refers to the possible queuing behind voice packets in the IP backbone. It equals the (1-P)-quantile (with P very small) of the probability density function of the sum of N_{HL} statistically independent waiting times in M/D/1-queues. N_{HL} stands for the number of heavily loaded nodes and corresponds to the number of nodes at which the voice load reaches a certain maximum admissible voice load ρ. For nodes with lower voice loads, the queuing delay contributions turn out to be negligible. In order to prevent that very large voice packets monopolize the queues in the IP backbone, a maximum voice packet size MTU_{voice} (in bits) is specified. It includes 320 header bits (12 bytes RTP-header, 8 bytes UDP-header and 20 bytes IP-header). The 56 bits (7 bytes) of the PPP-header need to be added explicitly. For more details, we refer to [7, 10, 11], the first reference of which contains an explicit definition of $D(\rho, N_{HL}, P)$.

The second term of eq. (5) corresponds to the possible queuing behind a data packet in service in every node, with MTU_{data} the largest possible data packet size (in bits) on the IP backbone, including 320 header bits (20 bytes TCP-header and 20 bytes IP-header). Again, the 56 bits (7 bytes) of the PPP-header need to be added explicitly.

As mentioned already in the introductory part of this section, the total dejittering delay is approximated by the sum of 2 terms, compensating for the difference in queuing delays encountered in the satellite access networks and the IP backbone, respectively. The dejittering delay component $T_{jit}^{backbone}$ is chosen equal to the difference between the "maximum" (read: (1-P)-quantile of eq. (5)) and the minimum queuing delay (i.e. 0) in the IP backbone, that is,

$$T_{jit}^{backbone} = T_{queue}^{backbone} \ . \tag{6}$$

By doing so, we ensure that almost no (read: at most a fraction *P* of the) voice packets arrive too late in the dejittering buffer, while the contribution of the dejittering delay to the mouth-to-ear delay is as small as possible.

Finally, we accumulate all delays that do not occur in the satellite access network and that are not discussed above in the so-called other delay T_{oth}. Its main component is the propagation delay, incurred when a bit traverses a certain distance over a transmission line. Here, it is assumed that a terrestrial propagation delay of 5 μs/km is introduced. Also included in this component are e.g. digitization, switching and echo control delays, which are maximally of the order of a few ms.

Now that all delay components are specified, the tolerable mouth-to-ear delay bound can be written as

$$T_{cod} + T_{pack} + T_{serv}^{backbone} + T_{queue}^{backbone} + T_{jit}^{backbone} + T_{oth} + T_{satellites} \leq T_{M2E} \ . \tag{7}$$

By inserting eqs. (2)-(6), we can derive the delay budget that may be consumed by the satellite access networks, that is,

$$T_{satellites} \leq T_{M2E} - A - B\, N_F\, T_F \ , \tag{8}$$

with

$$A = T_{LA} + \frac{N_{stag}\, O_{link}^{backbone}}{R_{link}} + 2\left(\frac{D(\rho, N_{HL}, P)(MTU_{voice} + 56) + N_{stag}\,(MTU_{data} + 56)}{R_{link}} \right) + T_{oth} \tag{9}$$

and

$$B = 1 + \frac{N_{stag}\, R_{cod}}{R_{link}} \ . \tag{10}$$

For future IP backbone networks with high-speed (of the order of Gb/s and higher) links, it makes sense to replace R_{link} by ∞ in eqs. (8)-(10). By doing so, the influence of the service delay $T_{serv}^{backbone}$, the queuing delay $T_{queue}^{backbone}$ and the dejittering delay $T_{jit}^{backbone}$ is neglected, resulting in

$$T_{satellites}^{\infty} \leq T_{M2E} - T_{LA} - T_{oth} - N_F\, T_F \ . \tag{11}$$

For further reference, we finally define the effective bit rate of a codec in the satellite access networks (in kb/s) and the resulting IP packet size (in bytes) as

$$R_{eff} = R_{cod} \frac{\left\lceil \dfrac{N_F B_F + O_{link}^{access}}{48 \times 8} \right\rceil 53 \times 8}{N_F B_F} \tag{12}$$

and

$$S_{IP} = \left\lceil \frac{N_F B_F}{8} \right\rceil + O_{IP} \ , \tag{13}$$

respectively, with $\lceil x \rceil$ the smallest integer larger than *x*, O_{IP} = 40 bytes RTP, UDP and IP-headers and O_{link}^{access} the amount of overhead (in bits) on the link layer of the satellite access networks. It equals 384 bits (12 bytes RTP-header, 8 bytes UDP-header, 20 bytes IP-header and 8 bytes AAL5-header).

Observe that the effective bit rate equals the rate at which bits are effectively put onto the satellite access networks.

4. Results and discussion

In this section, the methodology developed in the previous section is illustrated for a "long-distance" VoIP call between 2 satellite PCs, i.e., we set $N_{stag} = 15$ and $T_{oth} = 40$ ms. The other parameters values are chosen to be $R_{link} = 33920$ kb/s (34 Mb/s), $N_{HL} = 8$, $\rho = 0.8$, $MTU_{voice} = 3200$ bits (400 bytes)[2], $MTU_{data} = 12000$ bits (1500 bytes) and $P = 10^{-5}$. For this value of P, the packet loss ratio in the dejittering buffer is negligible, i.e., practically no packets are lost in the dejittering buffer. With these parameter values, $D(\rho, N_{HL}, P) = 57.905$.

In Figure 5 to Figure 9, we plotted the delay budget $T_{satellites}$ that is left for the satellite access systems (when traditional quality is aimed for) and the resulting effective bit rate R_{eff} as a function of the number of voice code words N_F put into 1 IP packet for all considered codecs, a packet loss ratio of 0.1 % and an echo loss value of infinity. For the satellite delay budget, similar results can be easily obtained for packet loss ratios of 1 and 5 % and/or an echo loss value of 51 dB by shifting the curves of the latter figures downwards with an amount equal to the difference between the respective mouth-to-ear delay bounds of Table 1. The effective bit rate, on the other hand, is independent of the echo loss value and/or the packet loss ratio. Obviously, it makes only sense when the satellite delay budget corresponding to a specific value of N_F remains larger than 0.

For a specific codec, packet loss ratio and echo loss value, eq. (8) implies the maximum satellite delay budget to be linearly decreasing with the number of voice code words N_F that are put into 1 IP packet. Indeed, we observe that the more voice code words that are gathered into 1 IP packet, i.e., the larger the voice packet size, the less the delay budget left for the satellite access networks. By doing so, the influence of the header overhead decreases such that the effective bit rate[3] decreases as well. On the other hand, putting fewer voice code words in 1 packet allows the satellite access networks to consume a larger portion of the mouth-to-ear delay and increases the effective bit rate[3].

The latter phenomenon can also be recognized in Figure 10, where we present a complete decomposition of the mouth-to-ear delay budget for all considered codecs for a packet loss ratio of 0.1 % and an infinite echo loss value. In this figure, we depict 2 bars per considered codec. The left one corresponds to the smallest value of N_F for which the effective bit rate R_{eff} does not exceed 100 kb/s, while the right one refers to the largest possible value of N_F yielding a satellite delay budget $T_{satellites}$ above 40 ms. Obviously, these assumptions (specifying a maximum effective bit rate and a minimum satellite delay) should be tailored towards every specific satellite system. The choices made here only serve illustration purposes. We clearly recognize the above-mentioned trade-off between packetization delay (or efficient bandwidth usage) and the budget remaining for the satellite access networks. It can be explained by noting that the sum of T_{cod}, T_{oth}, $T_{serv}^{backbone}$, $T_{queue}^{backbone}$ and $T_{jit}^{backbone}$ is almost independent of the voice packet size (the value of N_F), a fact that will become even more apparent for larger link speeds as the latter 3 components then become negligible. Notice that for the G.711 codec (without or with PLC), this trade-off is the least striking. This is due to the fact that the voice packet sizes should be smaller than $MTU_{voice} = 400$ bytes, a restriction that clearly limits the size of the packetization delay T_{pack} for the rightmost bar of G.711 (corresponding to the largest possible value of N_F yielding a satellite delay budget $T_{satellites}$ above 40 ms). Consequently, more than the postulated 40 ms remain for the satellite access systems in this case.

In Table 3, satellite delay budgets are presented for all considered packet loss ratios (0.1, 1 and 5 %) and echo loss values (51 dB and infinity). For each combination of the latter values, we only considered the codecs able to attain traditional quality. Table 3 reports 2 different values of N_F for each specific codec, packet loss ratio and echo loss value. As in Figure 10, the first value of N_F is the smallest value for which the effective bit rate does not exceed 100 kb/s, while the second reported

[2] The reason for this choice is that if we assume all voice packets to consist of 400 bytes, the resulting bandwidth efficiency of voice on the IP backbone is comparable to that of ATM networks.

[3] The effective bit rate defined by eq. (12) is a decreasing function of the number of voice frames N_F put into 1 IP packet on a macroscopic scale, i.e., if the difference between the considered values of N_F is large enough. Yet, on a microscopic scale, i.e., for nearby values of N_F, some exceptions to this behaviour may occur (see Figure 5 to Figure 9).

value of N_F is the largest value for which the satellite delay budget $T_{satellites}$ stays above 40 ms. Table 3 also reports the voice packet sizes S_{IP}, the effective bit rates R_{eff} and maximum values of $T^{\infty}_{satellites}$ for all considered cases.

First, we observe that $T^{\infty}_{satellites}$ is a rather good approximation of $T_{satellites}$. Obviously, this approximation will only get better for future IP backbones whose link speeds will be much larger than 34 Mb/s.

Next, the trade-off between packetization delay and satellite delay budget can be recognized in this table for any fixed codec, packet loss ratio and echo loss value.

We also consider the influence of the packet loss ratio and the echo loss value on the results of Table 3. More precisely, when the packet loss ratio increases and/or the echo loss value decreases, the mouth-to-ear delay budget corresponding to a specific codec decreases (see Table 1). For small echo loss values and/or large packet loss ratios, some codecs can even not attain traditional quality anymore. As such, well-designed user terminals (with large echo loss values) and/or good echo controllers (increasing the echo loss value substantially) are to be used, while the experienced packet loss ratio should be kept as small as possible by using error recovery techniques or by allowing more delay to be consumed by the interleaving procedure. Error concealment can also help to diminish the effects of lost voice packets. The first reported value of N_F (the smallest value for which the effective bit rate does not exceed 100 kb/s) is independent from the mouth-to-ear delay budget. However, a larger mouth-to-ear delay budget results in a larger delay budget available for the satellite access systems in this case. The second value of N_F (the largest value for which the satellite delay budget $T_{satellites}$ stays above 40 ms) increases when the tolerable mouth-to-ear delay increases, resulting in a lower effective bit rate.

For a specific packet loss ratio and echo loss value, we refer to our observations following Table 1, where we stated that the mouth-to-ear delay below which traditional quality is attained, is related to the bit rate of the codec. The resulting satellite delay budgets obviously depend on these mouth-to-ear delay budgets, but are also largely dependent on the specific choice of the packetization delay for the different codecs. More precisely, the largest possible satellite delay budget is obtained with the G.711 with PLC and the smallest values of N_F.

Finally, we provide some general guidelines concerning the quality of VoIP calls over satellite-based access networks.

We start by mentioning that the maximum satellite access delay bounds derived above are basically also valid for VoIP scenarios with only satellite access at one side of the connection. Moreover, the results can even be easily applied to other (delay-consuming) access networks.

For the satellite-PC-to-satellite-PC scenario, VoIP calls of traditional quality cannot be supported over access systems based on GEO satellites. Indeed, the sum of the propagation delays over the radio links on both sides of the connection (480 ms) already exceeds the maximum satellite delays reported in Table 3. The former sum even exceeds the total tolerable mouth-to-ear delay bounds reported in Table 1. Even with only GEO satellite access at one side of the voice connection, the satellite delay budgets given in Table 3 are only large enough (larger than 240 ms) in some special cases.

VoIP calls accessing the IP backbone through MEO satellites at one or both sides of the connection need a propagation delay of about 100 and 200 ms, respectively. By comparing the latter values with the satellite delay budgets reported in Table 3, it follows immediately that traditional quality cannot be guaranteed for all codecs, packet loss ratios and echo loss values.

For LEO-based satellite access systems, the propagation delay is maximally 20 ms (with satellite access at both sides of the connection), which is substantially smaller than the delay budgets available for the satellite systems. Yet, in a shared medium configuration, a lot of care has to be taken in designing the MAC, which controls the time window and resources each of the user terminals can use from this shared medium. More precisely, the delay introduced by this MAC protocol should be chosen such that the total satellite access delay stay within the derived budget.

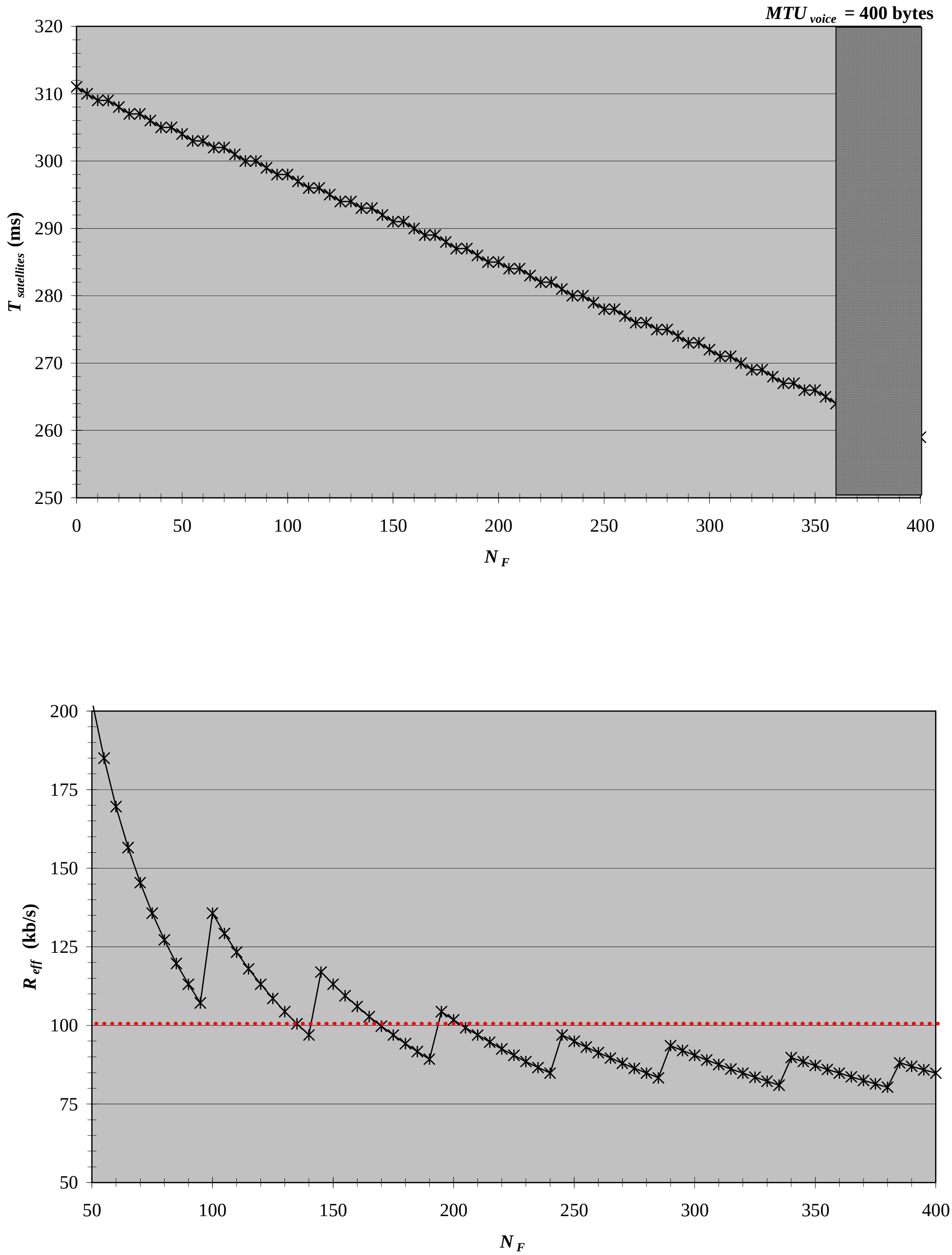
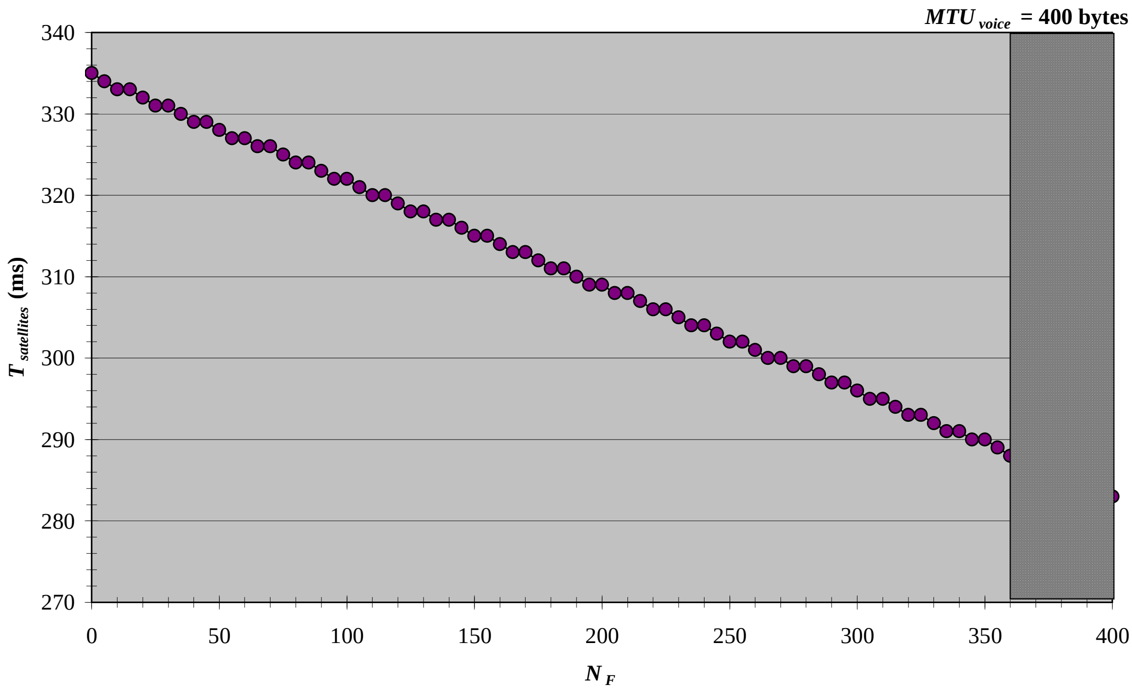


Figure 5: The satellite delay budget $T_{satellites}$ and the effective bit rate R_{eff} as a function of the number of voice code words for the G.711@64 kb/s codec, a packet loss ratio of 0.1 % and an echo loss value of infinity.



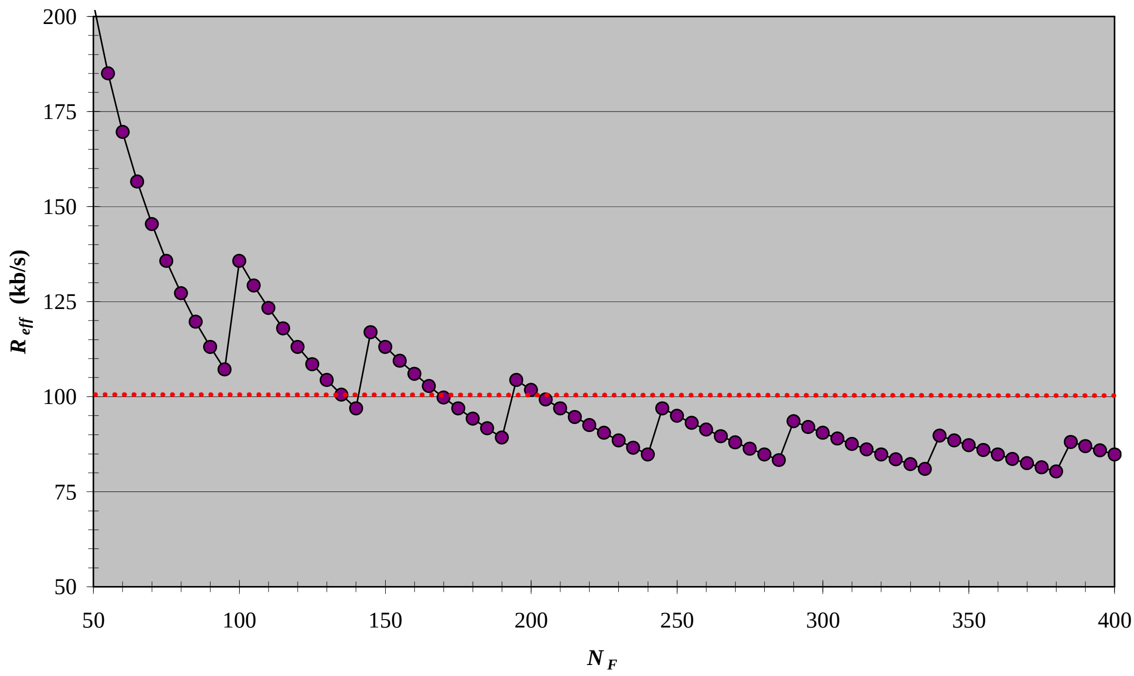
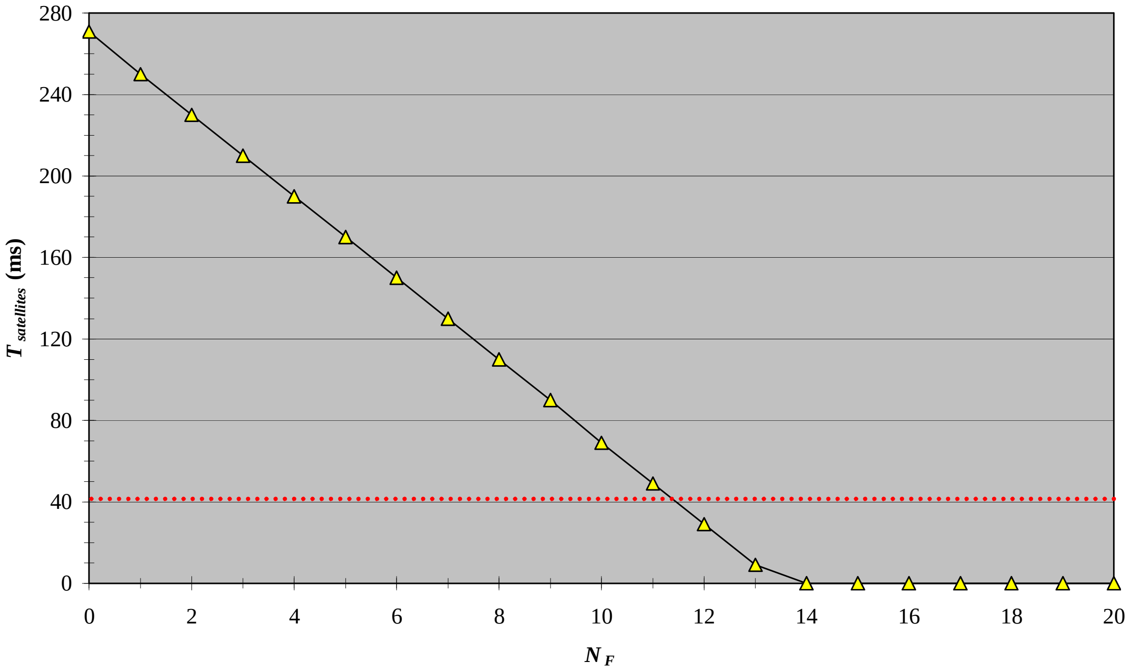


Figure 6: The satellite delay budget $T_{satellites}$ and the effective bit rate R_{eff} as a function of the number of voice code words for the G.711@64 kb/s (+PLC) codec, a packet loss ratio of 0.1 % and an echo loss value of infinity.



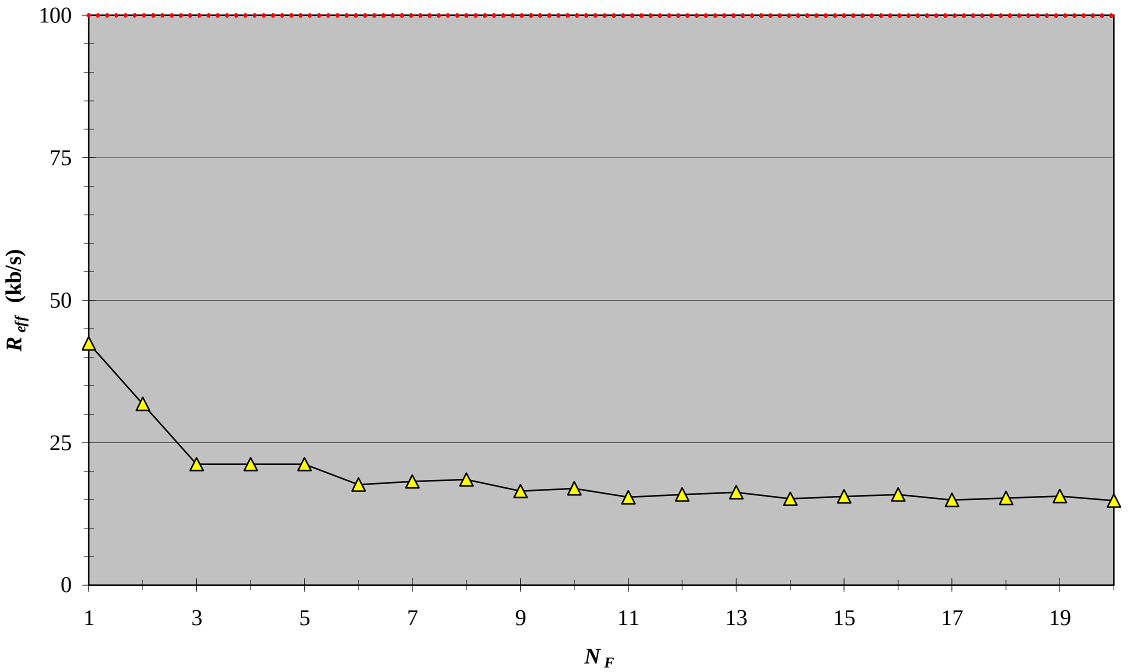


Figure 7: The satellite delay budget $T_{satellites}$ and the effective bit rate R_{eff} as a function of the number of voice code words for the GSM-EFR@12.2 kb/s codec, a packet loss ratio of 0.1 % and an echo loss value of infinity.

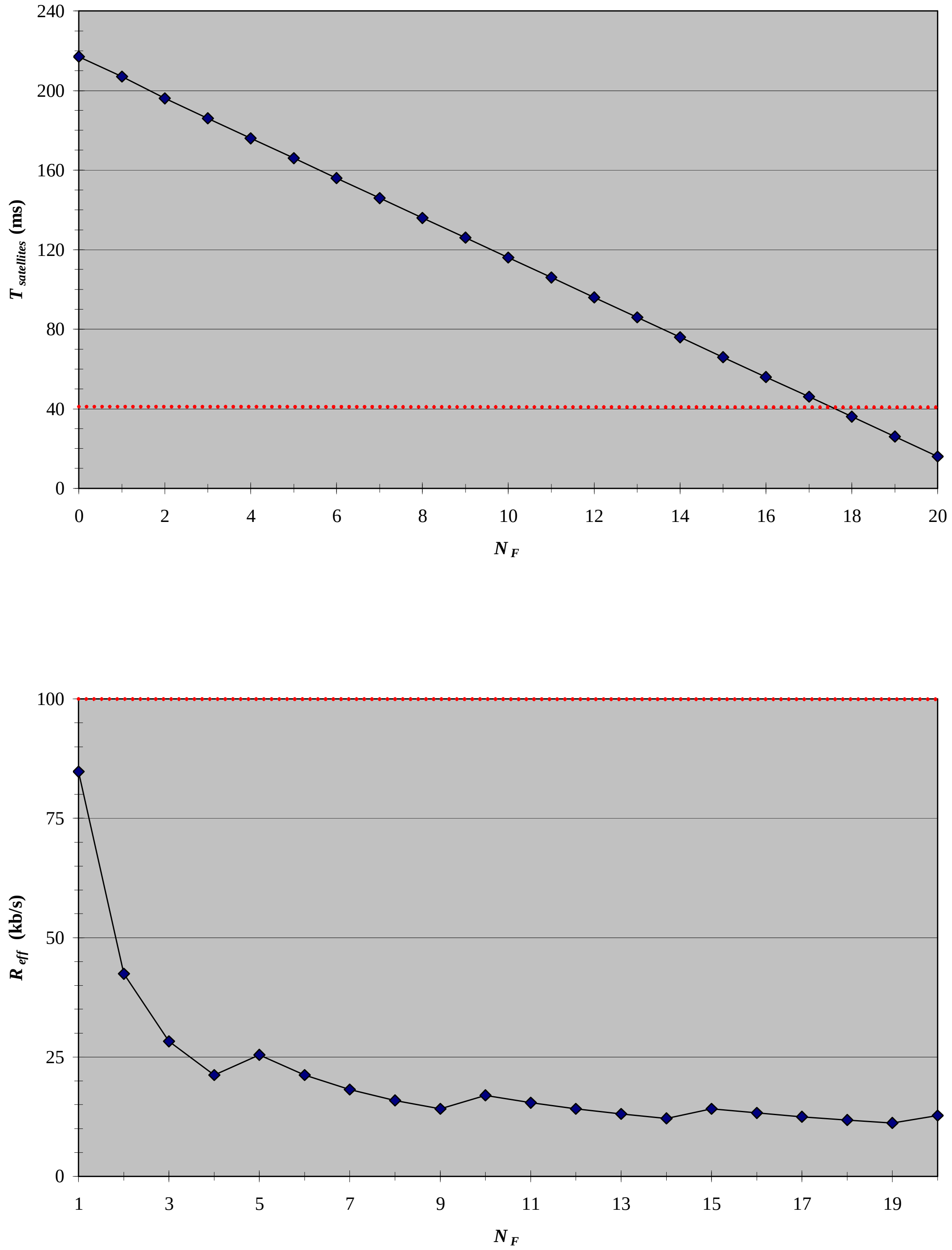


Figure 8: The satellite delay budget $T_{satellites}$ and the effective bit rate R_{eff} as a function of the number of voice code words for the $\boxed{\text{G.729@8 kb/s (+VAD)}}$ codec, a packet loss ratio of 0.1 % and an echo loss value of infinity.

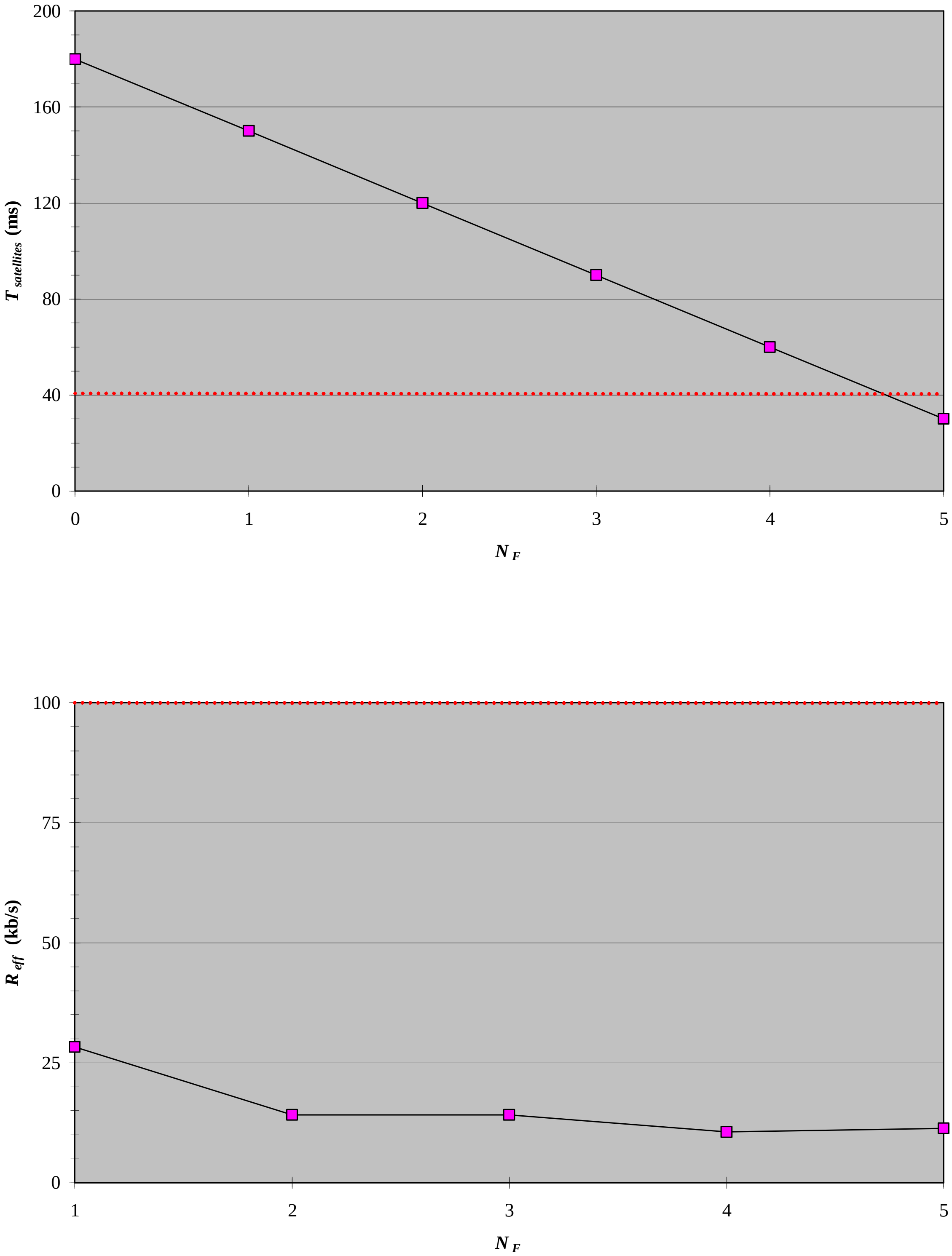


Figure 9: The satellite delay budget $T_{satellites}$ and the effective bit rate R_{eff} as a function of the number of voice code words for the G.723.1@6.3 kb/s (+VAD) codec, a packet loss ratio of 0.1 % and an echo loss value of infinity.

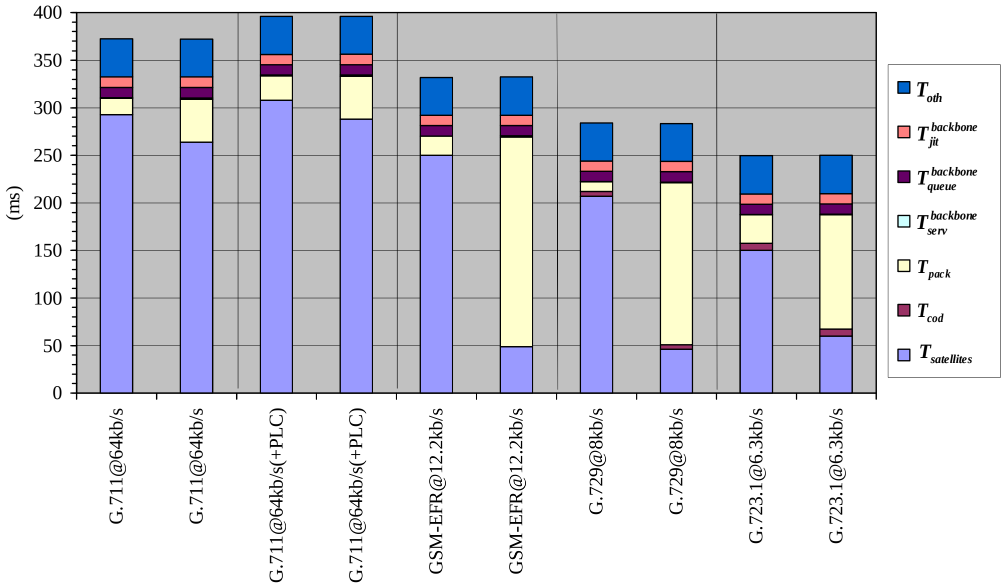


Figure 10: Decomposition of the tolerable mouth-to-ear delay budget for different codecs, a packet loss ratio of 0.1 % and an echo loss value of infinity. Two bars are depicted per considered codec. The left one corresponds to the smallest value of N_F for which the effective bit rate R_{eff} ≤ 100 kb/s; the right one refers to the largest possible value of N_F yielding a satellite delay budget $T_{satellites}$ ≥ 40 ms.

	EL = 51 dB					EL = ∞				
	N_F	S_{IP} (bytes)	R_{eff} (kb/s)	$T_{satellites}$ (ms)	$T_{satellites}^{\infty}$ (ms)	N_F	S_{IP} (bytes)	R_{eff} (kb/s)	$T_{satellites}$ (ms)	$T_{satellites}^{\infty}$ (ms)
G.711@64kb/s	136	176	99.76	202	225	136	176	99.76	293	316
	360	400	84.8	173	197	360	400	84.8	264	288
G.711@64kb/s (+ PLC)	204	244	99.76	207	230	204	244	99.76	308	331
	360	400	84.8	187	211	360	400	84.8	288	312
GSM-EFR@12.2kb/s	1	71	42.4	174	197	1	71	42.4	250	273
	7	254	18.17	54	77	11	376	15.42	49	73
G.729@8kb/s (+ VAD)	1	50	84.8	144	166	1	50	84.8	207	229
	11	150	15.42	43	66	17	210	12.47	46	69
G.723.1@6.3kb/s (+ VAD)	1	64	28.27	92	114	1	64	28.27	150	172
	2	88	14.13	62	84	4	135	10.6	60	82

(a) packet loss ratio = 0.1 %

	EL = 51 dB					EL = ∞				
	N_F	S_{IP} (bytes)	R_{eff} (kb/s)	$T_{satellites}$ (ms)	$T_{satellites}^{\infty}$ (ms)	N_F	S_{IP} (bytes)	R_{eff} (kb/s)	$T_{satellites}$ (ms)	$T_{satellites}^{\infty}$ (ms)
G.711@64kb/s (+ PLC)	204	244	99.76	175	198	204	244	99.76	256	279
	360	400	84.8	155	179	360	400	84.8	236	260
GSM-EFR@12.2kb/s	1	71	42.4	103	126	1	71	42.4	162	185
	4	162	21.2	43	66	7	254	18.17	42	65
G.729@8kb/s (+ VAD)	1	50	84.8	118	140	1	50	84.8	176	198
	8	120	15.9	47	70	14	180	12.11	45	68
G.723.1@6.3kb/s (+ VAD)	1	64	28.27	45	67	1	64	28.27	121	143
	1	64	28.27	45	67	3	111	14.13	61	83

(b) packet loss ratio = 1 %

	EL = 51 dB					EL = ∞				
	N_F	S_{IP} (bytes)	R_{eff} (kb/s)	$T_{satellites}$ (ms)	$T_{satellites}^{\infty}$ (ms)	N_F	S_{IP} (bytes)	R_{eff} (kb/s)	$T_{satellites}$ (ms)	$T_{satellites}^{\infty}$ (ms)
G.711@64kb/s (+ PLC)	204	244	99.76471	106	129	204	244	99.76	164	187
	360	400	84.8	86	110	360	400	84.8	144	168

(c) packet loss ratio = 5 %

Table 3: S_{IP}, R_{eff} and maximum values of $T_{satellites}$ and $T_{satellites}^{\infty}$ for different codecs, packet loss ratios of 0.1, 1 and 5 % and echo loss values of 51 dB and infinity. Two values of N_F are considered in each case. The first value of N_F is the smallest value for which R_{eff} ≤ 100 kb/s; the second value of N_F is the largest value for which $T_{satellites}$ ≥ 40 ms.

A last remark deals with the cases for which the foreseen satellite delay budget $T_{satellites}$ is not large enough, i.e., when more delay is introduced by the (satellite) access networks such that traditional quality (R = 70) is not attainable. More precisely, suppose that the delay experienced in the satellite access systems equals $T^{'}_{satellites}$. The resulting degradation in R-factor then can be calculated with the use of Figure 2 by adding $T^{'}_{satellites} - T_{satellites}$ to the tolerable mouth-to-ear delay budget T_{M2E} and by subtracting the R-values corresponding to T_{M2E} and the latter value, respectively. In particular, the new R-value is obtained by subtracting the result of the latter calculation from $R = 70$. A similar methodology can be used to calculate how much larger than 70 the R-factor becomes when the satellite delay budget is not completely consumed.

5. Conclusions

In this paper, we proposed a model to calculate the delay budget that can be consumed by satellite access systems for VoIP calls of traditional quality. This budget obviously depends on the type of codec, the packet loss ratio and the echo loss values of the considered scenario.

For a given codec, echo loss value and packet loss ratio, the satellite delay budget turns out to decrease as the efficiency (amount of overhead versus payload information) obtained on the network increases. That is, the larger the delay budget for the satellite access networks, the less efficient they are used, and, vice versa.

The influence of the echo loss value and packet loss ratio on the satellite delay budget was also investigated. In particular, the higher the echo loss value and/or the lower the experienced packet loss ratio, the larger the mouth-to-ear (and satellite) delay budget.

Moreover, the results presented in this paper indicate that with satellite access systems based on GEO and MEO satellites, VoIP at traditional quality is not feasible in general. Hence, LEO satellites seem to be the choice of preference to access an IP backbone when PSTN-quality voice is to be transported.

Abbreviations

AAL	ATM Adaptation Layer
ATM	Asynchronous Transfer Mode
GEO	Geostationary Earth Orbit
GoB	Good or Better
HOL	Head Of Line
IP	Internet Protocol
LEO	Low Earth Orbit
MAC	Medium Access Controller
MEO	Medium Earth Orbit
MOS	Mean Opinion Score
PLC	Packet Loss Concealment
PPP	Point to Point Protocol
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RTP	Real-time Transport Protocol
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VAD	Voice Activity Detection
VoIP	Voice over IP

Acknowledgments

This work was carried out within the framework of the project LIMSON sponsored by the Flemish Institute for the Promotion of Scientific and Technological Research in the Industry (IWT).

References

[1] "One-Way Transmission Time", ITU-T Recommendation G.114, February 1996.

[2] "Control of Talker Echo", ITU-T Recommendation G.131, August 1996.

[3] "Definition of Categories of Speech Transmission Quality", ITU-T Recommendation G.109, September 1998.

[4] "The E-model, a Computational Model for Use in Transmission Planning", ITU-T Recommendation G.107, December 1998.

[5] "Speech Processing, Transmission and Quality Aspects (STQ); Overall Transmission Plan Aspects for Telephony in a Private Network"ETSI Guide 201 050, February 1999.

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

[7] D. De Vleeschauwer, G.H. Petit, S. Wittevrongel, B. Steyaert and H. Bruneel, "An Accurate Closed-Form Formula to Calculate the Dejittering Delay in Packetised Voice Transport", Networking 2000, 14-19 May 2000, Paris.

[8] D. De Vleeschauwer, J. Janssen, G.H. Petit and F. Poppe, "Quality Bounds for Packetized Voice Transport", Alcatel Telecommunications Review, pp. 19-24, 1st quarter 2000.

[9] N.O. Johannesson, "The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks", IEEE Communications Magazine, pp. 70-79, January 1997.

[10] L. Kleinrock, "Queueing Systems, Volume 1: Theory", John Wiley and Sons, New York, 1975.

[11] L. Kleinrock, "Queueing Systems, Volume 2: Computer Applications", John Wiley and Sons, New York, 1975.

[12] W. Simpson, "PPP in HDLC-like Framing", IETF RFC 1662, July 1994.