Tuning the VoIP Gateways to Transport International Voice Calls over a Best-Effort IP Backbone

A. Van Moffaert, D. De Vleeschauwer, J. Janssen, M.J.C. Büchli, G.H. Petit
Alcatel Bell, Network Strategy Group
Francis Wellesplein 1
B-2018 Antwerp, Belgium
E-mail: {annelies.van_moffaert, danny.de_vleeschauwer, jan.janssen, maarten.buchli, guido.h.petit}@alcatel.be

and

P. Coppens[1]
Product Manager Internet, COLT Telecom
Zweefvliegtuigstraat 10
B-1130 Brussels, Belgium
E-mail: pcoppens@colt-telecom.be

Abstract

This paper tackles the problem of transporting long-distance voice calls over an Internet Protocol (IP) backbone. The voice calls originate from and are destined for traditional telephones connected to a Public Switched Telephone Network (PSTN). To cover the long-distance part in between, they are routed via gateways over a best-effort IP backbone. This paper describes how the gateway parameters (i.e., codec and packet size in the ingress gateway and packet loss in the dejittering buffer of the egress gateway) can be tuned such that for given (measured) network characteristics, the subjective voice quality is optimized.

Keywords: dejittering delay, packet size, subjective voice quality, voice over IP.

1 Introduction

It was the ubiquity of the Internet and its cheap flat rates that spurred the interest in transporting Voice over IP (VoIP). With some application-specific software a telephone call could be launched from a networked PC and transported over the Internet, hence reducing the price of a long-distance call to the price of a local call. Because parts of the (best-effort) Internet are sometimes congested, the quality of these calls is not always optimal (to say the least).

Therefore, some carriers deploy their own dedicated IP-based networks to transport long-distance calls. Only voice traffic can enter these networks and that only via a VoIP gateway. The limited amount of voice trunks these VoIP gateways support implicitly limits the total volume of voice traffic emerging from and destined for these gateways. To offer calls of high enough quality, these networks are usually over-provisioned.

On the other hand, an increasing number of IP backbone providers want to offer additional services on top of IP connectivity: VPN services, voice services, etc. Consequently, a mix of voice and data packets is transported over their backbones. Since data applications are mostly controlled by the Transmission Control Protocol (TCP) and since TCP increases its transmission rate until it experiences congestion, at least one router on the end-to-end route is driven into congestion (except when the complete path consists of very high-speed links, such that most TCP connections will be finished before the congestion point is reached). Because of these congested routers, also referred to as hot spots, supporting voice calls over such networks is a serious issue. There are two solutions to deal with this problem.

[1] At the time the measurements were done, P. Coppens worked at Global One Headquarters, IP Network Planning Department, Koloniënstraat 11, B-1000 Brussels, Belgium.

The first solution assumes that the routers have Quality of Service (QoS) capabilities. In that case voice packets should be given priority over data packets and the amount of voice traffic should be limited by some kind of Connection Admission Control (CAC). As in the case of dedicated VoIP networks, CAC is fairly easy if all voice traffic that is allowed on the IP backbone enters (and leaves) via VoIP gateways, because then the limited amount of trunks of the VoIP gateway limits the total volume of voice traffic. CAC is more of an issue if individual users can instigate a voice call from their networked PC, and therefore are allowed to ask for their voice packets to be treated with higher priority. In that case also some kind of policing function should be deployed. The latter problems (CAC and policing in a network where the user itself can ask for preferential treatment) are certainly not solved yet. Another predicament with this kind of solution is that all routers in the backbone need to be QoS-aware.

Therefore, we consider in this paper a second solution to the problem created by the inevitable hot spots caused by TCP. The crux is to make sure that the route taken by the voice packets avoids these hot spots. In the case of an IP backbone provider, this can be achieved as follows. In its Service Level Specifications (SLSs) with its customers the IP backbone provider knows how much data traffic enters its network through its Points Of Presence (POPs). Virtually always the amount of data traffic is physically limited by the capacity of the “pipe” (e.g., an E1) the customer has to the POP he is connected to. Most IP backbone providers use this kind of information to slightly over-provision their backbone. This makes that in this case the hot spots are situated outside the IP backbone and the backbone routers are hardly ever congested. Hence, on such a (best-effort) backbone (long-distance) voice calls can be supported as long as the VoIP gateways are located at or beyond the POP sites (i.e., beyond the hot spots caused by TCP) thus avoiding that voice flows traverse congested nodes. Figure 1 shows the network layout.

In the next section we will discuss how the quality of a voice conversation is determined by the mouth-to-ear delay and distortion of the voice signal. The latter are influenced by the parameter settings of the VoIP gateways (i.e., voice codec used, packet size and dejittering delay, or, equivalently, packet loss in the dejittering buffer) and the quality of the transport of the voice packets (i.e., delay, jitter and packet loss). In section 3 we report some measurements on the delay, jitter and packet loss of an IP backbone, which we use in section 4 to find suitable values for the packet size in the ingress gateway and the packet loss in the dejittering buffer of the egress gateway. In section 5 we discuss the major conclusions.

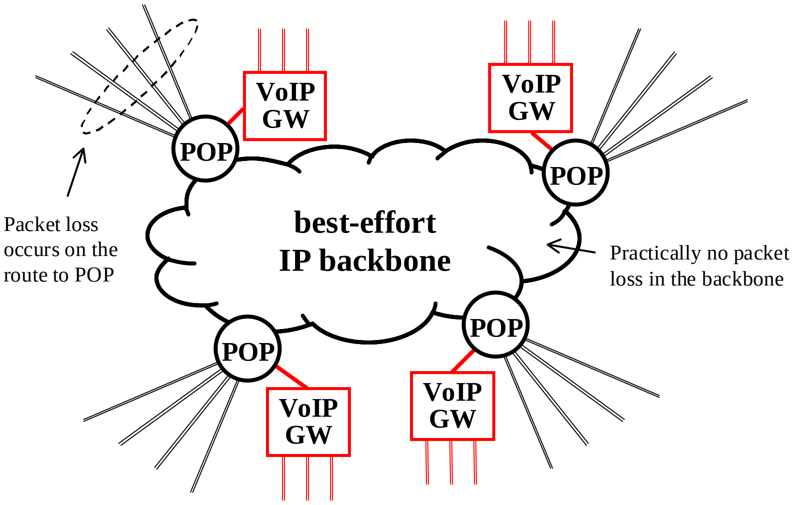

VoIP GW
VoIP GW
POP
POP
best-effort
IP backbone
Packet loss occurs on the route to POP
Practically no packet loss in the backbone
POP
POP
VoIP GW
VoIP GW

Figure 1: Transporting voice over a best-effort backbone

2 Parameters that determine the quality of a call

2.1 The influence of the mouth-to-ear delay and distortion on the quality of the conversation

Provided the end terminals (i.e., the traditional phones in the case studied in this paper) are optimally set (i.e., room noise is avoided as much as possible, loudness and side-tone are set at their optimal value), the factors that mainly determine the quality of a conversation are the mouth-to-ear delay and the distortion. They are described in the subsections below.

2.1.1 The influence of the mouth-to-ear delay

There elapses some time, referred to as mouth-to-ear delay, before the listener hears the words uttered by the talker. The mouth-to-ear delay in packetized voice transport is likely to be larger than in circuit-switched voice. Therefore, it is important to know how much delay can be tolerated.

With respect to the mouth-to-ear delay, three impairments can hamper the quality of the conversation: talker echo, listener echo and the loss of interactivity. It has been argued before that since the delay below which adequate quality can be attained is far too small without echo control, an echo controller should be switched on somewhere along the end-to-end path [2, 3, 4, 5]. Because the echo controller of most VoIP gateways is ITU-T G.168 compliant [12], it can remove practically all the echo (which is mostly generated in the hybrid in the terminating PSTN).

If the echo is perfectly controlled, the only remaining impairment associated with the mouth-to-ear delay is the loss of interactivity. In [2, 3, 4, 5], it is shown that below 150 ms, this loss of interactivity can hardly be noticed. Nevertheless, a mouth-to ear delay above 150 ms is in some cases still acceptable. How much this 150 ms bound can be exceeded, depends on the amount of distortion that is introduced [2, 3, 4, 5]. For example, if the G.711 codec (i.e., the traditional codec in digital circuit-switched voice) is used and in absence of packet loss, a delay of 400 ms is just about tolerable when the lowest quality attainable on today's PSTNs is aimed for.

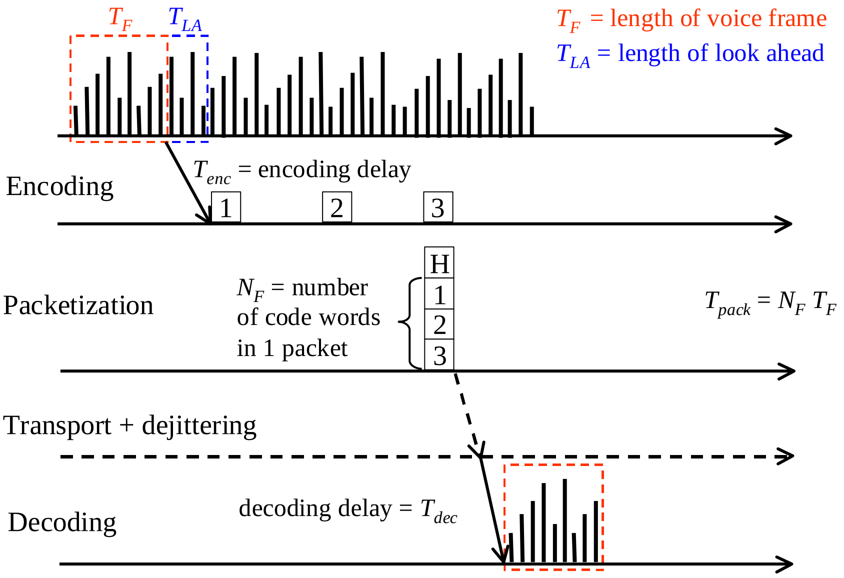
2.1.2 The influence of distortion

In the transport of an interactive conversation the speech signal heard by the listener does not need to be an exact copy of the one produced by the talker. Some distortion is tolerable. The distortion in packetized voice transport is likely to be larger than in circuit-switched (digital) voice. In packetized voice distortion can be introduced by compressing the (digitized) voice to a lower bit rate (using a low bit rate codec) and by the loss of voice packets. In [2, 3, 4, 5], it is shown that most low bit rate codecs can attain high enough quality. Moreover, it is shown that a packet loss ratio of up to 10^{-3} is hardly noticeable, and that codecs making use of an embedded Packet Loss Concealment (PLC) technique can even tolerate a packet loss ratio of a few percent. The same robustness against packet loss can be obtained with e.g. G.711 when it is enhanced with a PLC algorithm on top.

2.2 Network components determining the mouth-to-ear delay and distortion

2.2.1 Encoding and packetization stage

On the way from the originating phone to the ingress gateway the voice signal already experiences some delay. The circuit-switched voice signal has to cover a small distance and some local switches have to be traversed. These delays are not fundamentally different for VoIP and PSTN calls. The crucial difference for VoIP calls is that in the VoIP gateway encoding and packetization delay is introduced.



T_F
T_{LA}
T_F = length of voice frame
T_{LA} = length of look ahead
Encoding
T_{enc} = encoding delay
1
2
3
Packetization
N_F = number of code words in 1 packet
H
1
2
3
$T_{pack} = N_F\ T_F$
Transport + dejittering
Decoding
decoding delay = T_{dec}

Figure 2: Encoding, decoding and packetization delay

A codec works according to the principle illustrated in Figure 2. The input of the encoder is a digital voice signal sampled at 8 kHz, the resulting samples of which are quantized with a 13-bit linear quantizer. The output of the encoder is a regular stream of code words. Each code word represents a speech interval of duration T_F. Such a voice interval is referred to as a voice frame.

Sometimes the encoder also needs an interval after the one being encoded, referred to as the look-ahead. The duration of this look-ahead T_{LA} is codec-dependent. Other algorithms working on the voice signal (e.g., an algorithm to detect whether the signal is in fact a tone or a fax signal) may need a look-ahead too.

Once all necessary voice samples are collected, the encoder calculates the corresponding code word. The size B_F (expressed in bits) of this code word determines, together with the duration of the voice frame T_F, the codec bit rate

$$R_{cod} = \frac{B_F}{T_F} \quad . \tag{1}$$

To calculate the code word associated with a certain voice frame the Digital Signal Processor (DSP) of the gateway takes a certain processing time. This processing time depends on the speed of the DSP, the complexity of the codec, and also on the number and priority of the other processes running on the DSP. We define the encoding delay T_{enc} as the time that elapses between the instant the voice frame (of length T_F) is collected and the instant the corresponding voice code word emerges from the encoder. Remark that the encoding delay includes all look-aheads.

Code words need to be transported from the ingress gateway to the egress gateway. Therefore, code words are packed in voice packets. Several consecutive code words may be grouped into one voice packet. This introduces delay, referred to as the packetization delay. The packetization delay is given by $N_F T_F$, i.e., the number N_F of voice frames put into one packet times the voice frame length T_F. Notice that the packetization delay increases linearly with the number N_F of voice frames per packet and that the packetization delay has the granularity of the voice frame length T_F since the Real-time Transport Protocol (RTP) does not allow splitting code words over different IP packets.

Hence, the total delay introduced in the encoding/packetization stage (i.e., from the originating phone to just after the ingress gateway) is

$$T_{m,ingress} + N_F T_F \quad . \tag{2}$$

The first term $T_{m,ingress}$ groups all (deterministic) delays in the local PSTN and the encoding delay T_{enc}. The packetization delay is taken apart in the second term.

The voice payload size (in bits) of the packet is given by

$$P_v = N_F B_F \quad , \tag{3}$$

and the IP voice packet size (in bits) by

$$S_{IP} = P_v + O_{RTP/UDP/IP} \quad . \tag{4}$$

The overhead $O_{RTP/UDP/IP}$ is 320 bits (i.e., 40 bytes, consisting of 20 IP, 8 UDP and 12 RTP overhead bytes).

The larger the number N_F of voice frames put in a voice packet, the less the influence of the overhead. It is clear that the choice of the number N_F of voice frames per packet is a trade-off between efficiency and packetization delay.

Origin	Standard	Intrinsic quality	T_F (ms)	T_{LA} (ms)	B_F (bits)	R_{cod} (kb/s)
ITU-T	G.711	94.3	0.125	0	8	64
	G.726, G.727	44.3 69.3 87.3 92.3	0.125	0	2 3 4 5	16 24 32 40
	G.728	74.3 87.3	0.625	0	8 10	12.8 16
	G.729(A)	84.3	10	5	80	8
	G.723.1	75.3 79.3	30	7.5	158 189	5.3 6.3
ETSI	GSM-FR	74.3	20	0	260	13
	GSM-HR	71.3	20	0	112	5.6
	GSM-EFR	89.3	20	0	224	12.2

Table 1: Standardized codecs with their intrinsic quality, voice frame length T_F, look ahead length T_{LA}, code word length B_F and bit rate R_{cod}

The distortion introduced by the encoding/packetization stage depends on the codec used. For telephony the reference quality is the quality of the G.711 codec. This codec only uses a non-linear (μ-law or A-law) quantizer to requantize each 13-bit sample into a 1-byte sample. The intrinsic quality associated with the codecs as well as all other important codec parameters are listed in Table 1. The codec quality is expressed in terms of the rating *R* of the E-model, a tool standardized in ITU-T G.107 [10] to quantify the subjective quality of voice calls. The *R*-factor ranges from 0 to 100 (= perfect quality) and incorporates amongst others the effects of both delay and distortion. In particular, based on the distortion impairments given in ITU-T G.113 [11], the E-model associates an intrinsic quality rating (= rating *R* for a call without delay and packet loss) to a codec which decreases in function of delay and packet loss, see e.g. Figure 7 below for the G.711 codec. From Table 1, we observe that waveform codecs (e.g., G.726 and G.727) cannot reach bit rates below 32 kb/s without suffering too much in quality. Vocoder codecs (e.g., G.728, G.729, G.723.1 and the GSM codecs) can reach *R*-factors larger than 70 at such low bit rates as 5.3 to 16 kb/s. For comparison, the latter quality ratings, i.e., *R*-factors (largely) above 70, are considered to be typical for today's PSTN. For local GSM calls, the *R*-factors are aimed to be 60 at least.

2.2.2 Transport stage

The contribution of the network to the mouth-to-ear delay is

$$T_{m,BB} + \frac{P_v}{R_{S,BB}} + \boldsymbol{T_{q,BB}} \quad . \tag{5}$$

The first term $T_{m,BB}$, referred to as the minimal delay in the IP backbone, is the delay that even an imaginary RTP/UDP/IP packet with empty voice payload and no queuing delay would encounter. It is characterized by one value expressed in ms. It includes e.g. the propagation delay to traverse the network, the sum of all processing times needed to find the longest prefix match in the routing tables of the traversed routers, the time needed for the packet to traverse the switching fabrics of all the routers, and the time needed to put the header bits $O_{RTP/UDP/IP}$ on the links.

The second term of eq. (5) is referred to as the service delay in the IP backbone. This is the time all traversed nodes need to serve the voice payload. It is the only part of the delay through a node that depends on the payload size (expressed in bits) of the considered packet. The rate $R_{S,BB}$ with which the payload of the packet is served, is referred to as the effective service rate. Remark that the time needed to serve all overhead bits (e.g., RTP, UDP, IP and all link layer protocol headers) is included in the minimal delay.

The last term $\boldsymbol{T_{q,BB}}$ of eq. (5) is referred to as the queuing delay in the IP backbone. This queuing delay is a stochastic[2] delay that is characterized by a given probability density function. Remark that by definition the minimal queuing delay is 0. The probability density function of the queuing delay does not depend on the size of the considered voice packet. Not the considered packet itself, but all other IP packets cause the queuing delay.

The network only contributes to the distortion if there is packet loss in the network. The packet loss ratio experienced in the IP backbone is referred to as $P_{loss,BB}$.

2.2.3 Dejittering and decoding stage

We do not consider voice activity detection in this paper. As such, the packets of a voice flow leave the packetizer at a constant packet rate, i.e., every $N_F T_F$ a voice packet is produced. As explained in the previous subsection the delay during transport consists of a deterministic part and a stochastic part (i.e., the total queuing delay $\boldsymbol{T_{q,BB}}$). Because of this stochastic part the voice flow is jittered when it reaches the egress gateway, i.e., the packets of the voice flow do not reach the egress gateway at a constant packet rate. Since the decoder needs the code words at a constant rate, a dejittering mechanism is necessary.

In order to be able to dimension the dejittering buffer adequately, we need the tail distribution of the stochastic part of the delay, i.e.,

$$F(T) = \Pr\lfloor \boldsymbol{T_{q,BB}} > T \rfloor \quad . \tag{6}$$

An estimate for this tail distribution can be obtained from measurements on the network, from theoretical models, from the SLS with the backbone provider or from a combination of these. The most common dejittering mechanism artificially retains the first arriving packet of a VoIP flow a certain time, referred to as the dejittering delay T_{jit}, in the dejittering buffer before it is released to the decoder. All following packets are then read from the dejittering buffer at the original constant packet rate. If the decoder attempts to read a voice packet before it has arrived, the packet is effectively lost. The packet loss in the dejittering buffer depends on the total queuing delay $\boldsymbol{T_{q,BB,1}}$ of the first arriving packet of the VoIP flow and on the dejittering delay T_{jit} and is given by

$$\boldsymbol{P_{loss,jit}} = F(T_{jit} + \boldsymbol{T_{q,BB,1}}) \quad . \tag{7}$$

[2] As in the rest of the paper, we use **bold face** to indicate that a variable is stochastic.

Since this packet loss depends on the queuing delay of the first arriving packet, it is a stochastic variable. Remark that for a given (fixed) dejittering delay T_{jit}, the packet loss is largest if the first voice packet is a fast one (i.e., has a small queuing delay). To avoid this stochastic aspect, we assume that an adaptive dejittering algorithm is used. Such algorithms operate on a talk spurt by talk spurt basis [8]. These mechanisms delay the first payload of a talk spurt over a dejittering delay that is adjusted from talk spurt to talk spurt, based on the measured jitter of the past talk spurts. During the talk spurts the dejittering buffer is read at constant rate. When the parameters are well tuned, T_{jit} converges after a short transition time to a stable value that realizes eq. (7) for a chosen value of $P_{loss,jit}$. In other words, T_{jit} will adapt to the queuing delay that the first arriving packet happens to have such that the sum $T_{jit}+\boldsymbol{T_{q,BB,1}}$ quickly converges to a fixed (non-stochastic) value determined by $P_{loss,jit}$ through eq. (7). Under these assumptions the choice of $P_{loss,jit}$ determines the contribution of the dejittering buffer to the mouth-to-ear delay. This choice of $P_{loss,jit}$, from now on referred to as the packet loss rate tolerated in the dejittering buffer, inherently involves a trade-off between delay and packet loss. The larger the packet loss that can be tolerated, the smaller the dejittering delay is but the larger the distortion and vice versa.

After the flow of packets is dejittered the code words in the voice payloads need to be decoded. We define the decoding delay T_{dec} as the time that elapses between the instant a voice code word enters the decoder and the instant the corresponding voice frame becomes eligible to be played out.

Echo control is also performed in the egress gateway. It also introduces a small delay.

Finally, on the way from the egress gateway to the destination phone the voice signal experiences some delay, because some distance has to be covered and some switches have to be traversed.

The total delay introduced in the dejittering/decoding stage (i.e., from just before the egress gateway to the destination phone) is

$$T_{jit}+T_{m,egress} \quad , \tag{8}$$

in which the first term T_{jit} is the dejittering delay and the second term $T_{m,egress}$ groups all delays in the egress gateway except for the dejittering delay (e.g., decoding delay T_{dec} and echo control delay) and all (deterministic) delays in the local PSTN.

2.2.4 The total mouth-to-ear delay and distortion impairment

Distortion of the voice signal can be caused by the use of a (low bit rate) codec, by packet loss in the IP backbone and by packet loss in the dejittering buffer. As we will see in the next section, it is realistic to say that in a well-configured, high-speed IP backbone packet loss is quasi non-existent.

In addition, the next section will show that PSTN quality is difficult to guarantee for long-distance calls over a best-effort IP network but that for G.711 with PLC and perfect echo control, GSM quality is in most cases feasible.

The mouth-to-ear delay T_{M2E} the voice signal experiences can be split up in three terms

$$T_{M2E}=T_{m,tot}+N_F T_F\left(1+\frac{R_{cod}}{R_{S,BB}}\right)+F^{-1}(P_{loss,jit}) \quad , \tag{9}$$

in which

$$T_{m,tot}=T_{m,ingress}+T_{m,BB}+T_{m,egress} \tag{10}$$

and $F^{-1}(P_{loss,jit})$ = $\boldsymbol{T_{q,BB,1}}$ + T_{jit}. We always assume that $R_{cod}/R_{S,BB}$ *<<1* which is certainly reasonable for backbone links. $T_{m,BB}$, $R_{S,BB}$ and $F^{-1}(.)$ in eq. (9) depend on the network that is traversed. Measurements on such a backbone network are reported in the next section. $N_F T_F$ and $P_{loss,jit}$ are determined by the settings of the VoIP gateways. Their choice is discussed in section 4.

3 Measurements on the IP backbone

In this section we study the contribution of the network to the mouth-to-ear delay by way of measurements on a real backbone. For these measurements, Internet Control Message Protocol (ICMP) echo requests are sent from an ingress point to an egress point of the IP network, crossing several thousands of kilometers. Every 41 seconds[3] Ping sends 3 echo requests from ingress to egress, counts how many replies are successfully echoed back (thus measuring possible packet loss) and writes minimal, maximal and average Round Trip Times (RTTs) of the packets to a file. For a set of three packets this uniquely determines the individual RTTs. In the experiments both the ingress and the egress point (playing the role of VoIP gateways) are situated in the high-speed backbone to avoid that ICMP probe packets pass through TCP bottleneck links.

[3] On average for the traces referred to in this paper, i.e. Santiago-New York, Rio de Janeiro-Fort Worth, Sydney-Stockton and Sydney-New York.

The measurements performed on links between Santiago (Chili), Fort Worth (Texas, USA), Stockton (California, USA), New York (New York, USA), Rio de Janeiro (Brazil) and Sydney (Australia) show that a slightly over-provisioned backbone can indeed be quasi loss-free. Only 3 ICMP echo request/respond messages out of more than 15000 were lost. Hence, we can safely say that packet loss only occurs in the dejittering buffer.

As in eq. (5), we make a distinction between the deterministic part (i.e., the first two terms in eq. (5)) and the stochastic part (i.e., the last term in eq. (5)). Figure 3 shows the RTTs for a Ping from a router in Santiago to a router in New York and back over a 4.3 Mb/s link. There were no other routers in between. All indicated delays are RTTs. The minima and quantiles displayed in the curves were calculated over 90 consecutive measurements. Since 3 ICMP echo requests were sent every 41 seconds, this corresponds to a period of 21 minutes. It is obvious that the one-way deterministic delay is half the deterministic part of the RTT, the latter which is more or less equal to the half of the minimal RTT. In other words, we assume that at least one packet will have passed the network with quasi no queuing. However, it is more difficult to extract the stochastic part of the (one-way) network delay $T_{q,BB}$ from the stochastic part of the measured RTT. Since in most of the cases we measured, the load in one direction was considerably larger than the load in the other direction, we take the worst-case assumption that all the queuing comes from one direction. In other words we take $T_{q,BB}$ in eq. (5) equal to the difference between the quantile of the RTT we are interested in and the minimum RTT.

We see that the deterministic delay (min(RTT)/2 = 108 ms) is larger than the time needed for light to travel the distance Santiago-New York (8211 km) through an optical fiber (5 μs/km). This is mainly due to the fact that the cables do not follow a straight path. For this an engineering factor of 1.5 up to 2.5 has to be taken into account.

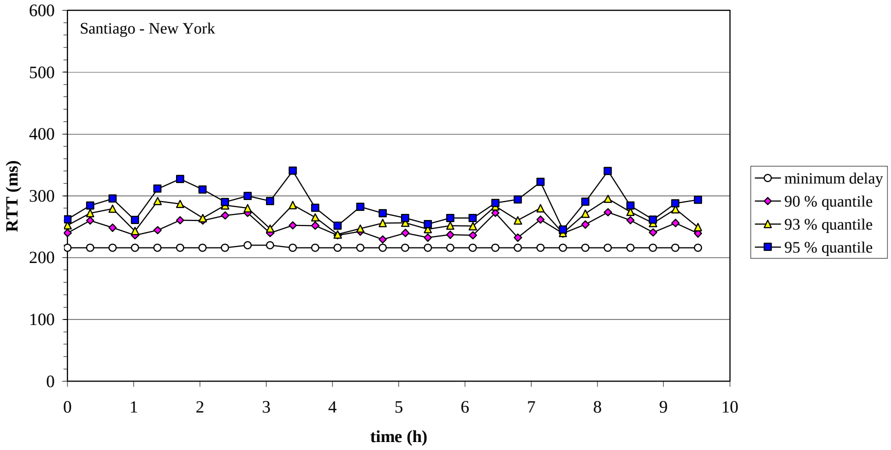


Figure 3: Measured RTTs between Santiago and New York

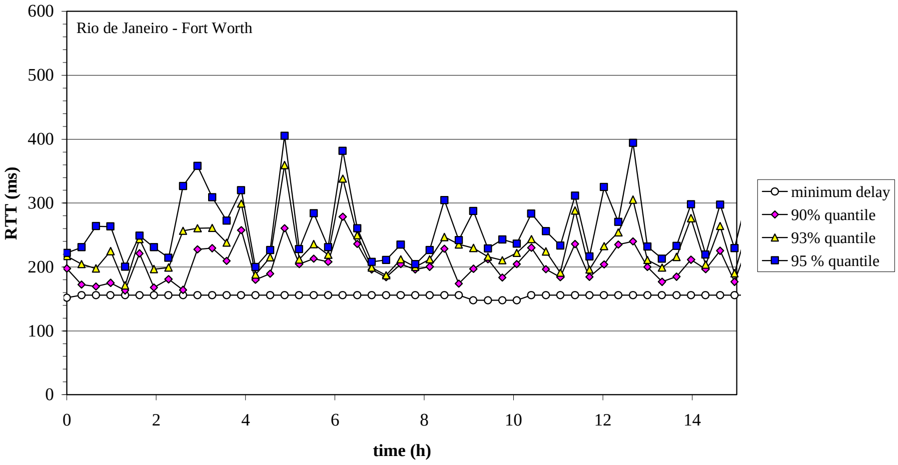


Figure 4: Measured RTTs between Rio de Janeiro and Fort Worth

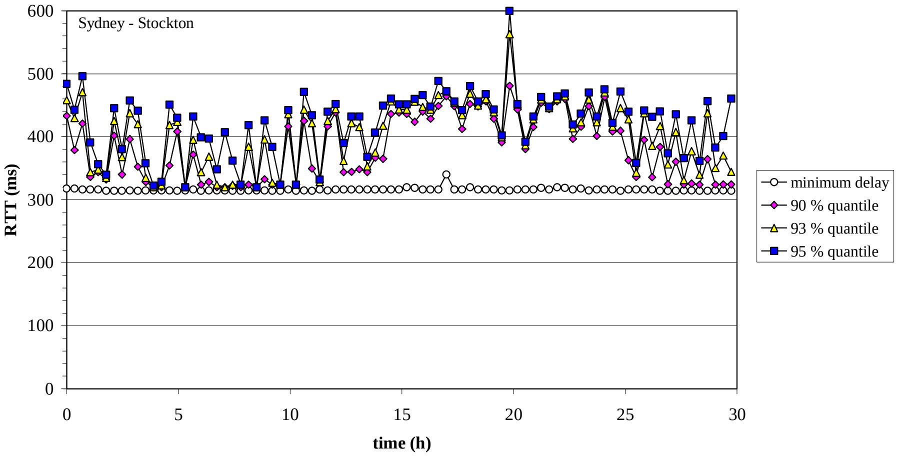


Figure 5: Measured RTTs between Sydney and Stockton

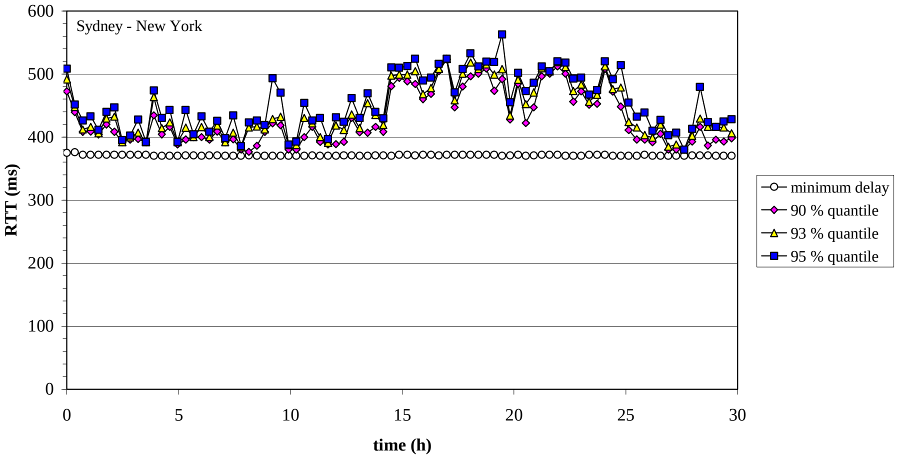


Figure 6: Measured RTTs between Sydney and New York via Stockton

Another set of Ping measurements was performed between a router in Rio de Janeiro and one in Fort Worth (8425 km as the crow flies) over a 2 Mb/s link. There are no routers in between. These results are shown in Figure 4. The deterministic part of the network delay is here 74 ms.
A third set of measurements consists of Pings between a router in Sydney and one in Stockton (12028 km in bird’s-eye view) connected via a 13 Mb/s link (Figure 5). Again no other routers were in between. The minimal network delay in this case was 157 ms. These measurements show a sudden peak in the RTTs after about 20 hours of measurement. When such a delay peak occurs voice transported over the network will inevitably be degraded. However, we will see in the next section that apart from such rare events voice can be transported with a reasonable quality (comparable to GSM quality) over these (very) long-distance (best-effort) backbone links.
Finally, Figure 6 shows the RTTs measured with Ping between the same router in Sydney and a router in New York via the router in Stockton (12028 + 4046 = 16074 km as the crow flies). The 3 routers are connected via 13 Mb/s links. Here, the deterministic network delay was 185 ms.

4 Tuning the gateways

4.1 Introduction and assumptions

In this section, we show that (even in a best-effort network) in most cases suitable values for the packetization delay $N_F T_F$ and the dejittering loss $P_{loss,jit}$ can be found to attain almost always an acceptable quality. The reasoning followed here is similar to the one in [1, 6, 7, 9]. We assume that the amount of additional voice traffic on the network is so low that the measurements of the previous section remain valid.

We restrict ourselves here to the G.711 codec with a PLC algorithm running on top. Although the latter codec has the highest intrinsic *R*-factor (see Table 1) and thus gives the best results, the same reasoning can be followed for any codec that has a good intrinsic quality. We suppose that perfect echo control is used. The latter can be performed at relatively low computational cost, and hence, should always be implemented since for the delays introduced in an IP network, echo seriously jeopardizes the voice quality.

The sum of the parameters $T_{m,ingress}$ and $T_{m,egress}$ is chosen equal to 20 ms. This value stems from twice the delay incurred from the phones to the VoIP gateways (about 2 ms each), the encoding and decoding delay (whose sum is measured to be 13 ms) and the echo control delay (of about 3 ms).

Since delay is usually much more of an issue than bandwidth and efficiency for (very) long-distance calls over a high-speed IP network, one should choose a small packetization delay $N_F T_F$. Here, we fix it to 10 ms but also 20 ms would be possible. This would increase the efficiency but decrease the quality slightly. However, the extra 10 ms (to be added to delays of 200 ms and larger) would only marginally effect the quality.

As mentioned above, we measured no packet loss in the network. Hence, all the packet loss occurs in the dejittering buffer. We further assume that the dejittering buffer is large enough to avoid buffer overflow such that packets are only lost when they arrive too late at the receiver, i.e., after their scheduled play-out time. As mentioned above, we assume that an adaptive dejittering algorithm is used that learns the relative delay of the first packet and adapts T_{jit} accordingly, such that independent of $\boldsymbol{T_{q,BB,1}}$, a chosen amount of packet loss $P_{loss,jit}$ in the dejittering buffer results.

Figure 7 below shows the quality rating *R*, defined in the E-model [10], as a function of the mouth-to-ear delay for different values of packet loss. In the remainder of this section we use these *R*-curves to deduce the optimal value for the packet loss in the dejittering buffer for the traces of Figure 3 - Figure 6.

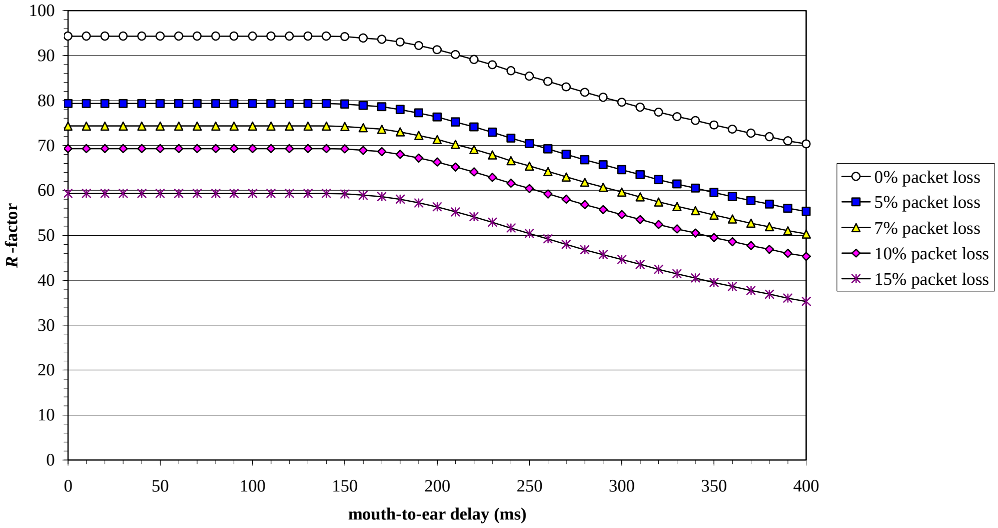


Figure 7: Quality rating *R* as a function of the mouth-to-ear delay for different packet loss ratios

4.2 Examples

The measurement results for the trace Santiago-New York are shown in Figure 3. The minimal RTT is 216 ms, i.e., the deterministic part of the network delay is 108 ms.

The worst-case 90% quantile of the RTT, i.e., the maximum value of the 90% quantiles over the measurement period, is 274 ms. According to the approach described above this gives 274 ms – 216 ms = 58 ms for the 90% quantile of the one-way queuing delay $\boldsymbol{T_{q,BB}}$. In other words, if the adaptive dejittering mechanism is set to 10% packet loss then 58 ms is a worst-case value for the sum of queuing plus dejittering delay. This value gives a total mouth-to-ear delay

$$T_{M2E} = 108 \text{ ms} + 20 \text{ ms} + 10 \text{ ms} + 58 \text{ ms} = 196 \text{ ms} \quad ,$$

where T_{pack} = 10 ms and $T_{ingress}$ + T_{egress} = 20 ms. From Figure 7 it can be seen that this mouth-to-ear delay with 10% packet loss corresponds to a quality rating *R* = 66.5.

The worst-case 93% quantile for the RTT is 295 ms. This gives a total mouth-to-ear delay

$$T_{M2E} = 108 \text{ ms} + 20 \text{ ms} + 10 \text{ ms} + 79 \text{ ms} = 217 \text{ ms} \quad .$$

The combination 7% packet loss and 217 ms mouth-to-ear delay corresponds to a rating *R* = 69.

The worst-case value of the 95% quantile for the RTT is 340 ms. Hence,

$$T_{M2E} = 108 \text{ ms} + 20 \text{ ms} + 10 \text{ ms} + 124 \text{ ms} = 262 \text{ ms}$$

and R = 68.5.

Note that these worst-case quality ratings are attained at different moments in time for the different packet loss values. The quality rating R as a function of time (by performing similar calculations as above) for the different packet loss values is displayed in Figure 8. We observe that GSM-like quality is attained with 10% packet loss, while PSTN quality can be attained at nearly all times with packet losses of 7% and 5%.

When comparing the results at measurement time 8.16 h, we obtain (T_{M2E}=196 ms, R=66.5) for 10% packet loss, (T_{M2E} = 217 ms, R = 69) for 7% packet loss and (T_{M2E} = 262 ms, R = 68.5) for 5% packet loss. More precisely, we observe that the R-ratings are about equal, while the mouth-to-ear delay is a decreasing function of the packet loss (see Figure 3). Here, the operator has to decide what he estimates most important for his customers, listening-only quality (lower packet loss but higher delays) or interactivity (lower delays but higher packet loss).

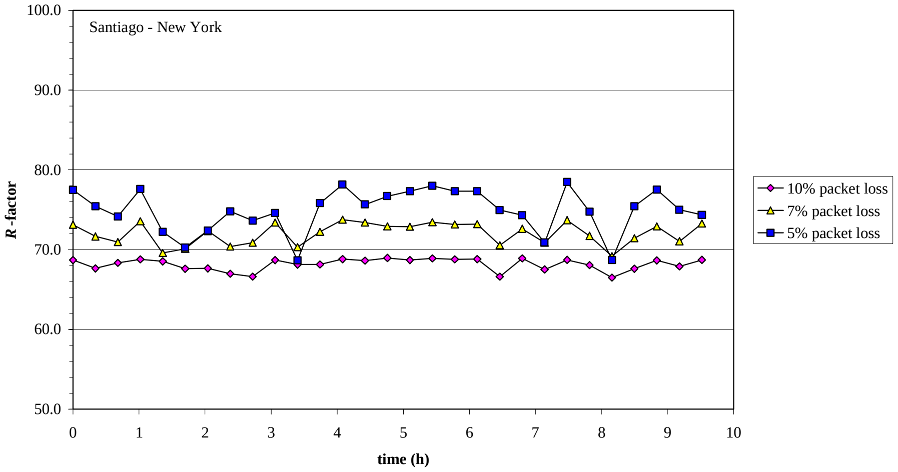


Figure 8: Quality ratings for trace between Santiago and New York

Analogous figures as above are shown in Figure 9 - Figure 11 for the traces Rio de Janeiro-Forth Worth, Sydney-Stockton and New York-Stockton, respectively. The detailed analysis is left to the interested reader. Roughly speaking, GSM quality can again be reached at all times for the trace between Rio and Fort Worth with 10% packet loss, With 5% or 7% packet loss, even PSTN quality is attainable in a lot of cases. For the very long-distance traces (Sydney-Stockton and Sydney-New York), the results turn out to be slightly worse. That is to say, even with only 5% packet loss, GSM quality cannot be guaranteed at all times. Within this context, however, the following remark should be taken into account. With a PLC algorithm in place, even packet losses up 10% are not very disturbing. The experienced quality degradation, which is mainly due to loss of interactivity, is therefore often tolerated for very long-distance calls.

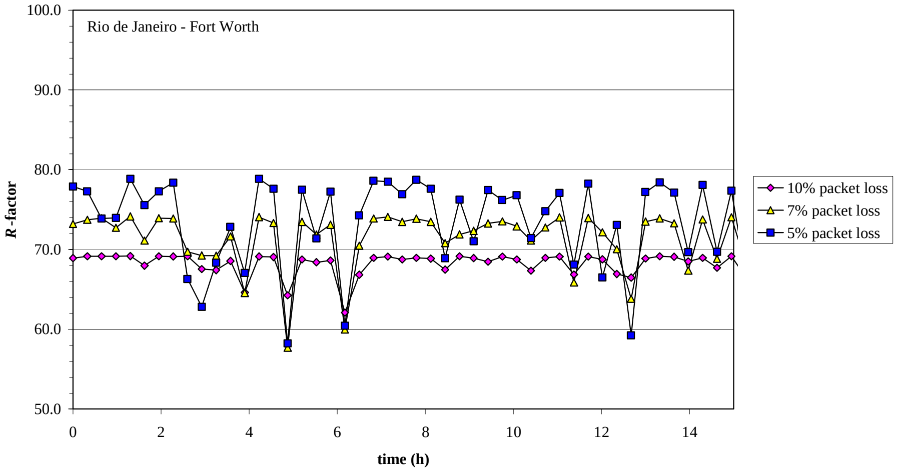


Figure 9: Quality ratings for trace between Rio de Janeiro and Fort Worth

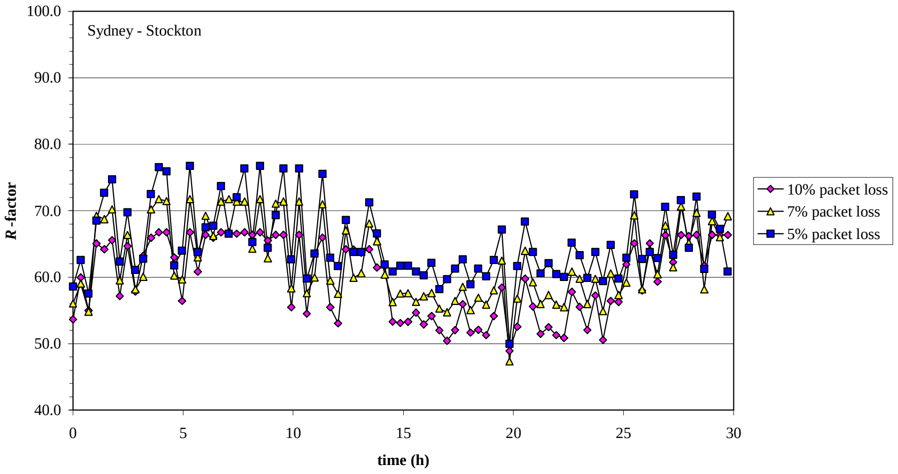


Figure 10: Quality ratings for trace between Sydney and Stockton

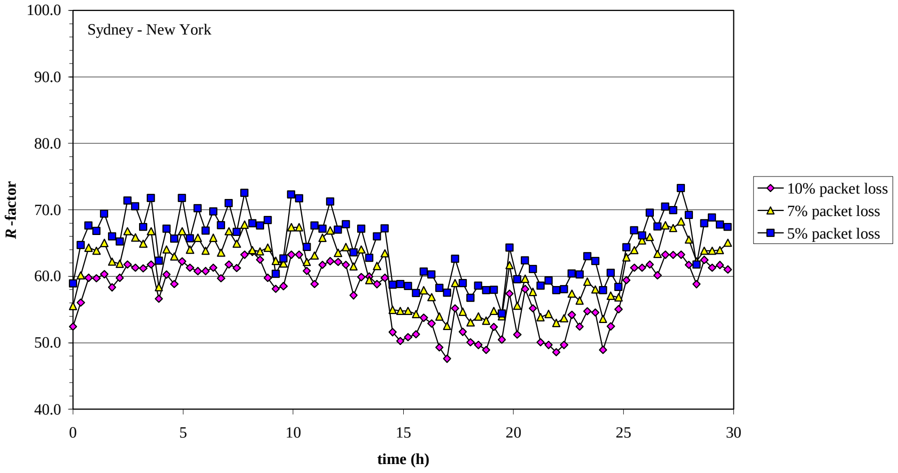


Figure 11: Quality ratings for trace between Sydney and New York via Stockton

Finally, the worst-case (over the measurement period) quality ratings and corresponding mouth-to-ear delays for the different packet loss values, as calculated above for the Santiago-New York trace, are summarized in Table 2 for all considered traces. This table could, for example, be used to tune the dejittering loss in the egress VoIP gateways. In particular, the numbers in bold indicate the amount of packet loss in the dejittering buffer leading to the best (subjective) voice quality that can always be guaranteed, independent of the time the calls are placed.

	10% packet loss		7% packet loss		5% packet loss	
	R	T_{M2E} (ms)	*R*	T_{M2E} (ms)	*R*	T_{M2E} (ms)
Santiago - New York	66.5	196	**69**	**217**	68.5	262
Rio de Janeiro - Fort Worth	**62**	**235**	57.5	315	58	361
Sydney - Stockton	50.5	337	54.5	345	**57.5**	**369**
Sydney - New York	47.5	368	52.5	369	**57**	**378**

Table 2: Worst-case quality ratings and corresponding mouth-to-ear delays for various packet loss values

5 Conclusions

Over a best-effort network hard guarantees about the obtainable quality of a VoIP call cannot be given. Unexpected high delay peaks due to exceptional queuing, router updates, etc. can always occur and in this case no reasonable voice quality can be obtained. However, most of the time an acceptable, GSM-like quality can be obtained even for (very) long-distance voice calls that are transported via a trunking gateway over an IP backbone.

In this paper, we showed how one can use network delay measurements in combination with the ITU-T E-model to deduce the optimal parameter settings for the dejittering buffer from a quality point of view, i.e., those parameters that give the best possible subjective voice quality given the network characteristics. The method described in this paper can be used by operators to configure their VoIP trunking gateways.

6 Acknowledgement

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).

7 References

[1] D. De Vleeschauwer, J. Janssen, G.H. Petit, “Voice over IP in Access Networks”, Proceedings of the 7th IFIP Workshop on Performance Modelling and Evaluation of ATM/IP Networks (IFIP ATM ‘99), Antwerp (Belgium), 28-30 June 1999.

[2] D. De Vleeschauwer, J. Janssen, G.H. Petit, “Delay Bounds for Low Bit Rate Voice Transport over IP Networks”, Proceedings of the SPIE conference on Performance and Control of Network Systems III, Vol. 3841, pp. 40-48, Boston (USA), 20-21 September 1999.

[3] D. De Vleeschauwer, J. Janssen, E. Desmet, G.H. Petit, “Tolerable Delay Bounds For Low Bit Rate Voice Transport”, Proceedings of the XVII World Telecommunications Congress (ISS2000), Birmingham (UK), 7-12 May 2000.

[4] D. De Vleeschauwer, J. Janssen, G. H. Petit, F. Poppe, “Quality Bounds for Packetized Voice Transport”, Alcatel Telecom Review, First quarter 2000, pp. 19-23, January 2000.

[5] J. Janssen, D. De Vleeschauwer, G.H. Petit, Delay and Distortion Bounds for Packetized Voice Calls of Traditional PSTN Quality”, Proceedings of the First IP Telephony Workshop (IPTEL2000), GMD Report 95, pp. 105-110, Berlin (Germany), 12-13 April 2000.

[6] M.J. Karam, F.A. Tobagi, “Delay of Voice Traffic over the Internet”, Proceedings 3845 of the SPIE conference on Multimedia Systems and Applications II, Vol. 3845, Boston (USA), 19-22 September 1999.

[7] T.J. Kostas, M.S. Borella, I. Sidhu, G.M. Schuster, J. Grabiec, J. Mahler, "Real-Time Voice over Packet-Switched Networks", IEEE Network, pp. 18-27, Jan./Feb. 1998.

[8] R. Ramdjee, J. Kurose, D. Towsley, H. Schulzrinne, “Adaptive Playout Mechanisms for Packetized Audio Applications in Wide-Area Networks”, Proceedings of IEEE Infocom 94, Toronto (Canada), pp. 680-688, June 1994.

[9] K. Van Der Wal, M. Mandjes, H. Bastiaansen, “Delay Performance Analysis of the New Internet Services with Guaranteed QoS”, Proceedings of the IEEE, Vol. 85, No. 12, pp. 1947-1957, December 1997.

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

[11] “Provisional Planning Values for the Equipment Impairment Factor I_e”, ITU-T Recommendation G.113/Appendix I, September 1999.

[12] “Digital Network Echo Cancellers”, ITU-T Recommendation G.168, April 2000.