User Perceived Quality-of-Service for Voice-over-IP in a Heterogeneous Multi-Domain Network Environment*

[+]R.D. van der Mei[ad], H.B. Meeuwissen[b], and F. Phillipson[c]

[a]CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

[b]Bell Labs Europe, Lucent Technologies, P.O. Box 1168, 1200 BD Hilversum, The Netherlands

[c]TNO Information and Communication Technology, Brassersplein 2, 2612 CT Delft, The Netherlands

[d]Vrije Universiteit Amsterdam, Faculty of Sciences, Amsterdam, The Netherlands

Abstract: One of the main barriers to the commercial success of Voice-over-IP (VoIP) services beyond a single domain (e.g., a company intranet, a single operator's domain) as a viable alternative to the classical voice telephony services is Quality-of-Service (QoS). The ability to deliver QoS is particularly challenging in next-generation networks, where VoIP services will be accessible via a variety of end-device types and over sequences of heterogeneous network domains, each with their specific QoS-mechanisms and owned by different competing network operators that will negotiate Service Level Agreements (SLAs) amongst each other. In this paper, we provide a method to effectively realize end-to-end QoS for VoIP as perceived by the end user in an environment in which multiple business domains are involved. To this end, we first present a *quantitative* model for the relation between (1) the end-to-end QoS at the end-user perception level, (2) the QoS requirements to network infrastructure given the specifics of the end user terminal, and (3) the SLAs negotiated at the network level between the different domain owners. Second, we show how the concept of a per-domain QoS certificate can help to guarantee predictable QoS for VoIP traffic over their network domains. This enables VoIP service providers to realize the desired perceived QoS levels for both intra and inter-domain VoIP sessions. This paper provides a *simple* and *practically feasible* approach for VoIP service providers to identify which network level SLAs need to be negotiated with other parties to realize the desired QoS level to their customers.

Keywords: E-model, Quality-of-Service, Service Level Agreement, Voice-over-IP.

1. INTRODUCTION

Voice-over-IP (VoIP) services provide an alternative to the classical circuit-switched voice

---

* This work has been carried out in the context of the project End-to-end Quality of Service in Next Generation Networks (EQUANET), which is supported by the Dutch Ministry of Economic Affairs via its agency SenterNovem.

+ Part of the work was done when the author worked for TNO.

telephony services that rapidly gains momentum. Recently, a variety of commercial VoIP service offerings have been brought to the market. Today, most commercially available VoIP solutions are offered via IP Virtual Private Networks (VPN), providing VoIP connectivity between closed communities of subscribers (see for example, KPN Telecom's Epacity service [1]). One of the main barriers to the large-scale commercial success of VoIP services is Quality-of-Service (QoS). Less than acceptable QoS experienced by the end user will lead to customer churn, and hence, loss of revenue. In the next-generation networks, the ability to deliver predictable QoS is particularly challenging. In fact, the unbundling in the telecommunications market results in the involvement of a multitude of network operators and service providers. A VoIP session may traverse a sequence of heterogeneous network domains, each with their specific QoS-mechanisms (e.g., IP Diffserv, IP Intserv, MPLS, best-effort) and exploited by different network operators. Moreover, VoIP services will be accessible via heterogeneity of end-device types, each with their specific characteristics, potentially out-of-control of the network operator. Hence, to realize a predictable and desired QoS at the user level, there is a critical need to quantify the impact of the terminal characteristics on the QoS requirements at the network level. Subsequently, required end-to-end QoS levels at the network layer will need to be forced by negotiating Service Level Agreement (SLAs) between network operators. Since SLAs are typically bilateral agreements between neighboring parties, the quality of VoIP sessions that cross more than two domains may become unpredictable. Another problem is that in practice network operators are reluctant to making major modifications to up-and-running networks in production. These observations raise many questions regarding the end-to-end QoS, such as the following:

1. How can we realize desired user perceived QoS levels for VoIP sessions over multiple consecutive network domains?
2. What is the relation between user perceived QoS, the terminal characteristics, and the end-to-end QoS at the network layer?
3. What combinations of SLA's between business domains need to be negotiated to achieve a predictable and desired QoS level?
4. How can we create an incentive for business parties to cooperate and enforce predictable QoS for both intra and inter-domain VoIP sessions?

In the literature, many papers have focused on different QoS-aspects of VoIP services, such as perceived QoS, admission control, reliability, and network performance. The reader is referred to [2] (and the references therein) for an overview of the state-of-the-art in the field. The vast majority of papers in the field are focused on a single administrative domain. A limited amount of effort has been devoted to end-to-end QoS for VoIP in a multi-domain environment. The European Telecommunications Standards Institute project Telecommunications and Internet Protocol Harmonization over Networks (ETSI TIPHON) gives a complete solution for how to deliver end-to-end QoS over multiple domains by defining a call-setup procedure based on a fixed set of five QoS classes, and by defining a functional model consisting of QoS functional entities and functional relationships [3-5]. A main hurdle, however, to the large-scale adaptation of the ETSI TIPHON approach is that its implementation requires a major change of current networks and protocols to make them ETSI TIPHON-compliant, while network operators may be reluctant to make modifications to their networks in production. Another approach to the multi-domain problem is suggested by the project AQUILA (Adaptive Resource Control for QoS Using an IP-based Layered Architecture) [6], defining, evaluating and implementing an architecture for QoS on the Internet based on IP Diffserv, making use of the so-called Border Gateway Reservation Protocol (BGRP). The viability of the AQUILA approach is demonstrated by field trials.

The contribution of this paper is fourfold. First, we determine the minimal QoS requirements to the network given (1) the desired QoS as perceived by the end user (in terms of a Mean Opinion Score (MOS)), and (2) the terminal specifics such as e.g. the size of the jitter buffer at the receiving side, the codec, and the number of voice frames per IP packet. To this end, the E-model [7,8] is used to compute the end-to-end requirements, which are then partitioned into a terminal part and a network part. Hence, for given terminal specifics, the requirements to the (multi-domain) network are obtained. Second, for inter-domain VoIP sessions we explicitly identify the relation between the network level QoS in the originating domain and the QoS requirements posed on all the other domains involved in a session together. Third, we introduce service level calculus to compute which combinations of per-domain SLAs lead to the desired end-to-end QoS at the network level. This, in turn, leads to the identification of the so-called SLA negotiation space. Fourth, we show how the concept of per-domain certificates can help to enforce desired end-to-end QoS as perceived by the end user. The main advantage of this concept is that it is independent of the network technologies in the different domains. Each domain is free to meet the requirements of the certificate in its own way. As an overall result, we present a simple and practically feasible solution to realize user perceived QoS for VoIP in a heterogeneous multi-domain network environment.

The remainder of the paper is organized as follows. In Section 2, the requirements on the user perceived QoS is transformed into requirements on the multi-domain network. In Section 3, we transform these multi-domain network requirements into per-domain network requirements. In Section 4, we summarize our step-by-step approach to realize end-to-end QoS for VoIP. Finally, Section 5 contains concluding remarks and addresses a number of challenges for further research.

2. ADMISSIBLE NETWORK REGIONS

The realization of a given end-to-end QoS at the user perception level over multiple domains proceeds along two steps. First, the desired MOS-value is translated into requirements for the end-to-end QoS at the network layer by using the E-model. The translation will be referred to as *vertical integration*. Second, the requirement for end-to-end QoS at the network layer over multiple domains is translated into combinations of QoS requirements to each of the network domains. This translation will be referred to as *horizontal integration*.

Figure 1 illustrates the notions of vertical and horizontal integration for the case in which a VoIP session crosses four domains. Note that SLAs are only negotiated between domain combinations 1 and 2, 2 and 3, and 3 and 4, whereas no SLAs are negotiated between domain 1 and 3, 1 and 4, and 2 and 4. As a consequence, operator 1 does not have control over the delay-loss-jitter incurred by crossing domains 3 and 4.

2.1. Mapping perceived QoS to network QoS

The E-model [7,8] is a computational model to estimate the perceived quality of VoIP sessions. It defines the transmission rating factor R ($0 \le R \le 100$), which can be mapped one-to-one to an estimated MOS, $1 \le \text{MOS} \le 4.5$, for VoIP sessions, see [7,Annex B]. The factor R is modeled as

$$R = R\text{o} - I\text{s} - I\text{d} - I\text{e - eff} + A,$$

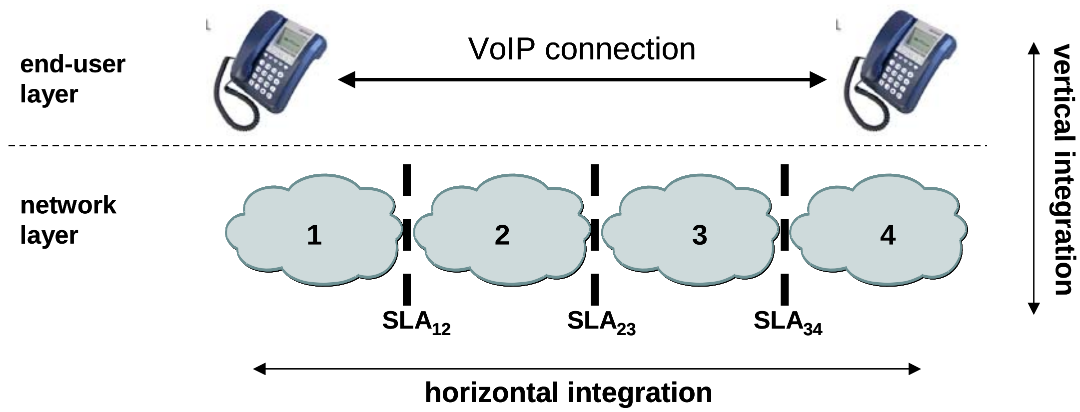

Figure 1. Multi-domain VoIP connection.

where R_0 is the basic signal-to-noise ratio, I_s is the simultaneous impairment factor, I_d is the delay impairment factor, $I_{e\text{-}eff}$ is the effective equipment impairment factor that quantifies the impact of the voice codec in combination with packet loss, and A is the advantage factor.

The E-model defines 20 different parameters, each with a default value and a permitted range. The ratio R_0 and each of the three different impairment factors are functions of a subset of the basic parameters. The four E-model parameters of particular interest for this paper are the packet-loss robustness factor Bpl [no unit, range 1…40], the equipment impairment factor I_e [no unit, range 0…40], the random packet loss probability Ppl [unit %, range 0…20], and the mean one-way delay of the echo path T [unit ms, range 0…500]. Throughout the remainder of the paper it is assumed that all the other parameters are assigned their default values. The equipment impairment factor $I_{e\text{-}eff}$ can be expressed in terms of these parameters as follows:

$$Ie\text{ - eff} = Ie + (95 - Ie)\frac{Ppl}{Ppl + Bpl}.$$

In [9], provisional values for I_e and Bpl are given for different voice codecs and codec settings, see Table 1. Note that a higher value of Bpl means that a codec is more robust to packet loss. Using the above relations, the R-value can be calculated as a function of Ppl for different voice codecs, see Figure 2 for an illustration. The graph illustrates the fact that the impact of the end-to-end packet loss on the perceived QoS strongly depends on the choice of the codec (see also for example [10] for related figures).

Table 1
Equipment impairment factor I_e and packet-loss robustness factor Bpl for random packet loss and different codecs

Codec	Rate (kbit/s)	Packet size (ms)	I_e	Bpl
G.723.1+VAD	6.3	30	15	16.1
G.729A+VAD	8	20	11	19.0
GSM-EFR	12.2	20	5	10.0
G.711	64	10	0	4.3
G.711+PLC	64	10	0	25.1

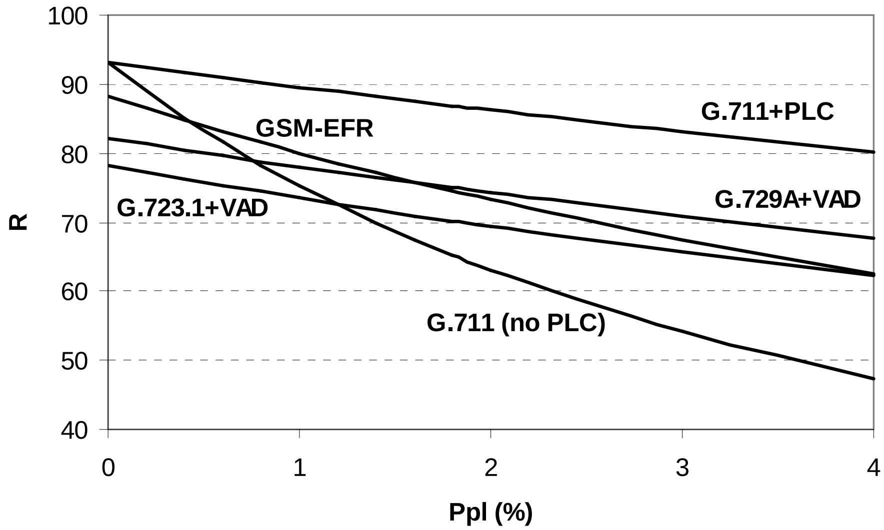

Figure 2. The perceived QoS, R, as a function of Ppl for different voice codecs.

In the next section, the E-model will be applied to determine the network level QoS requirements (in terms of the mean one-way delay T and the random packet loss probability Ppl) to achieve the desired voice quality as perceived by the end user, as quantified by the rating factor R.

2.2. Construction of admissible network regions

In this section, we give an approach for the construction of so-called admissible regions. The approach consists of two phases. In phase 1, for a given codec, we determine the set of (T, Ppl)-combinations that correspond to the desired minimal value of R. These combinations determine an *admissible region* for the combined end-to-end delay and end-to-end packet loss corresponding to the desired value of *R*. Subsequently, in phase 2, we construct the admissible region for end-to-end delay $T_{network}$ and end-to-end packet loss $P_{network}$ *at the network layer* (excluding the delay and loss induced by the terminal), by quantifying the contribution of the terminals to end-to-end delay and packet loss.

2.2.1. Phase 1: Construction of iso-R curves

We consider end-user terminals with parameters (C, J, M) characterized by their speech codec C, jitter buffer size J, and the number of voice frames per IP packet M. The jitter buffer size J is defined as the maximum packet delay variation in milliseconds that the jitter buffer can compensate. The number of voice frames per packet M is the number of coded speech frames that are encapsulated in a single IP packet. Now suppose we have an end-to-end voice quality requirement for terminals (G.729, 50, 2) of $R \geq R_{min}$.[1] Figure 3 shows the iso-R curves for $R_{min} \in \{50, 60, 70, 80\}$ that are constructed by applying the E-model, see Section 2.1. The iso-R curves show how end-to-end delay and packet loss can be traded-off to achieve the same perceived voice quality. Note that each curve bounds the region $R \geq R_{min}$.

[1] The following speech quality categories are used throughout this paper [11]: Best: 90≤*R*<100; High: 80≤*R*<90; Medium: 70≤*R*<80; Low: 60≤*R*<70; Poor: 50≤*R*<60.

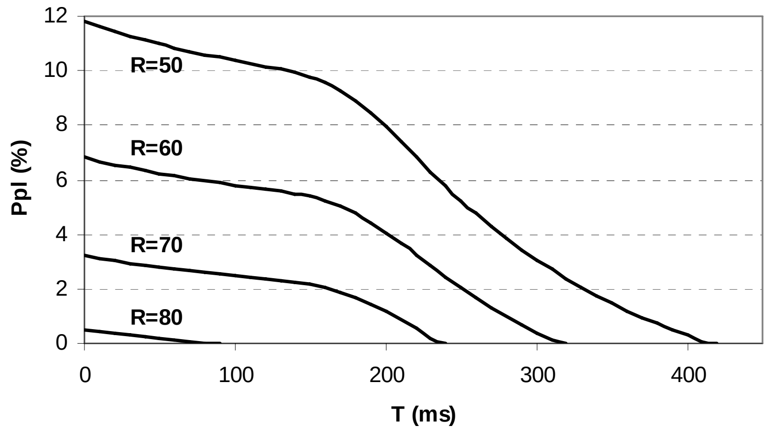

Figure 3. Iso-R curves for terminal (G.729, 50, 2).

2.2.2. Phase 2: Construction of admissible regions for network delay and packet loss

The R-values depicted in Figure 3 encompass the combined impact of both the terminal and the etwork on the perceived quality. Hence, for given terminal parameters, the requirements on the end-to-end performance indicators (T, Ppl) can be transformed into requirements on the performance of the network in terms of mean delay and packet loss ($T_{network}$, $P_{network}$). Note that in the E-model jitter is not a parameter, since it is assumed that a jitter buffer is present in the receiving terminal to compensate for end-to-end network jitter. This jitter buffer adds to the end-to-end delay, and if the jitter buffer cannot compensate for the actual jitter, it also adds to the end-to-end packet loss. In addition to jitter buffer delay, terminal delay is also caused by voice codecs and by putting (one or more) voice frames in IP packets. Below we quantify the impact of the terminal with parameters (C, J, M) on the requirements on the network delay and packet loss.

Since both T and Ppl are end-to-end parameters, we first estimate the contribution of the terminal to their values. The mean end-to-end (i.e., mouth-to-ear) delay T is the sum of the mean network delay $T_{network}$ and the mean terminal delay T_{term}, i.e. $T = T_{network} + T_{term}$. T_{term} is determined by the codec C and the jitter buffer size J. The codec delay depends on the number of voice frames per packet M, the codec frame size F (in milliseconds), and the codec look-ahead time L (in milliseconds). The values of the frame size and look-ahead time are in Table 2 for different codecs [12].

Table 2
Frame size and look-ahead time for different codecs

Codec	Rate (kbit/s)	Frame size F (ms)	Look-ahead time L (ms)
G.723.1	6.3	30	7.5
G.729	8	10	5
GSM-EFR	12.2	20	0
G.711	64	10[2]	0

[2] The value in this table differs from the frame size of 0.125 ms specified in [12]. In this table the frame size is 0.125 times the number of sample in a speech frame. A common value for this number is 80 bytes, resulting in a frame size of 10 ms.

The codec delay T_{codec} satisfies the following inequalities (cf. [12]):

$$T_{codec,\min} \le T_{codec} \le T_{codec,\max},$$

where

$$T_{codec,\min} = (M+1)F + L, \qquad T_{codec,\max} = (2M+1)F + L.$$

The jitter buffer of size J in the receiving terminal removes packet delay variation, but introduces additional delay. We assume, as recommended for planning purposes in [12], that the contribution of the jitter buffer to the mean delay equals J/2. As a result, for a given value of T,

$$T - T_{term,\max} \le T_{network} \le T - T_{term,\min},$$

where

$$T_{term,\min} = \frac{J}{2} + T_{codec,\min}, \qquad T_{term,\max} = \frac{J}{2} + T_{codec,\max}.$$

Next we assume there are two different causes for end-to-end packet loss: (1) packet loss in the network due to buffer overflow, and (2) packet loss in the terminal due to jitter buffer underflow or overflow. If their probabilities are denoted by $P_{network}$ and $P_{jitterbuffer}$, respectively, then the following relation holds for the average VoIP connection:

$$Ppl := 1 - (1 - P_{network})(1 - P_{jitterbuffer}).$$

Now if we require $0 \le P_{jitterbuffer} < \alpha$ for some $0 \le \alpha < 1$, then for a given value of *Ppl*

$$P_{network} > \frac{Ppl - \alpha}{1 - \alpha}.$$

As a result,

$$\frac{Ppl - \alpha}{1 - \alpha} < P_{network} \le Ppl.$$

The "gap" β between the upper and lower bound for $P_{networks}$ equals

$$\beta = \left(\frac{\alpha}{1 - \alpha}\right)(1 - Ppl).$$

To illustrate, Figure 4 shows the following curves for VoIP sessions with terminals with parameters (G.729, 50, 2), and for $\alpha = 0.01$: (1) the iso-R curve for $R_{min} = 70$, and (2) the ($T_{network}$, $P_{network}$)-curves *excluding* the delay and loss induced by the terminal for the optimistic and the pessimistic case. In optimistic case the terminal delay is $T_{term,min}$, and in the pessimistic case the terminal delay is $T_{term,max}$. The curve for the optimistic case has been obtained by shifting the iso-R curve $T_{terminal,\ min}$ milliseconds to the left and β down. For the pessimistic case the iso-R curve has been shifted $T_{terminal,\ max}$ milliseconds to the left, and β down. The curves depicted in Figure 4 below defines an *admissible region* that quantifies how network delay and packet loss can be traded-off in the operator's own domain to achieve a given perceived quality of R-value 70.

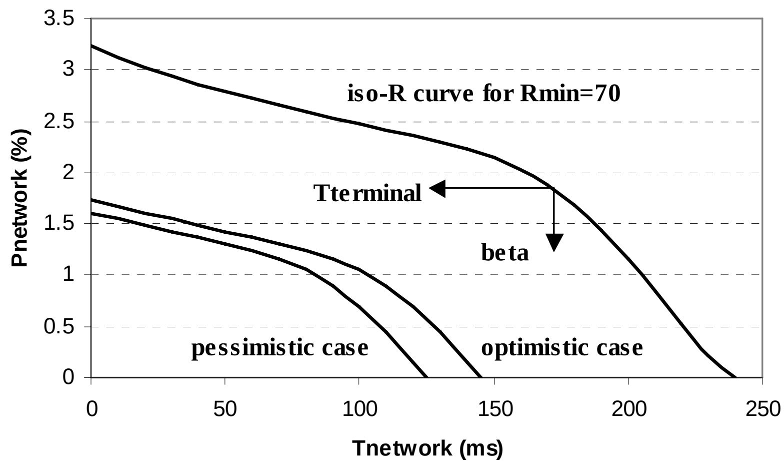

Figure 4. Network delay and loss budgets for terminals (G.729, 50, 2), R_{min}=70 and α= 0.01.

VoIP service providers can use the approach outlined above to identify the requirements to their network that must be met in order to realize desired quality levels for single-domain VoIP calls as perceived by their customers.

Note that in this approach α, the maximum packet loss in the terminal due to jitter buffer underflow or overflow, represents a degree of freedom: a particular choice of α immediately translates into a particular value of β, representing the vertical downshift of the iso-R curve in Figure 4. This, in turn, translates into requirements on the end-to-end packet loss in the network. The most economic choice of α depends on the cost involved in realizing given combinations of network-level delay, loss and jitter. A typical default value used throughout this paper is 0.01.

3. END-TO-END QoS OVER MULTIPLE DOMAINS

For VoIP sessions within a single domain, the admissible region discussed in Section 2 identifies the set of combinations of packet loss and delay that lead to the desired user perceived QoS levels. For VoIP sessions over multiple domains (e.g. international calls) these requirements have to be translated into requirements for each of the individual domains. To this end, we emphasize that in today's practice VoIP domain owners are in control of their own resources, but are not allowed to control the resources of other domains. In this context, neighboring domains usually negotiate SLAs. However, VoIP sessions originating in domain A that cross more than two domains can involve non-neighboring domains with which no bilateral SLA is negotiated. For example, in the 4-domain case illustrated in Figure 1, the connection crosses domains 1, 2, 3, and 4, whereas domain 1 has no SLA with domains 3 and 4.

3.1. VoIP calls traversing two domains

Consider the situation that operator 1 owns VoIP domain 1, which is connected to VoIP domain 2 owned by operator 2. Assume for terminals with parameters (G.711+PLC, 50, 1), operator 1 has a QoS requirement for intra-domain VoIP session (originating and terminating in domain 1) of $R \geq R_{own}$, and for inter-domain VoIP sessions (originating in 1 and terminating in 2) of $R \geq R_{own} \geq R_{min}$. Notice that in this way that operator 1 can differentiate between the QoS levels of intra- and inter-domain sessions. Now operator 1 requires at least "medium quality" for inter-domain sessions to domain 2, i.e. R_{min}=70, and at least "high quality" for sessions within its own domain 1, i.e. R_{own}=80. Let (T_1, P_1) denote the mean one-way delay and packet-loss percentage in domain 1, and similarly, let (T_2, P_2) denote these parameters in domain 2. Then, the total mean network delay $T_{network}$, and, assuming independence of packet loss in domains 1 and 2, the overall packet loss $P_{network}$ are given by the following expressions: $T_{network} = T_1 + T_2$, and $P_{network} = 1 - (1- P_1)(1- P_2)$.

If operator 1 dimensions its network in such a way that the delay-loss combination is (T_1, P_1), then by using equations for $T_{network}$ and $P_{network}$, the possible delay-loss requirements (T_2, P_2) to be negotiated with operator 2 can be determined. This way operator 1 can determine its so-called *SLA negotiation space*. Figure 5 shows an example of the SLA negotiation space in case of terminals (G.711+PLC, 50, 1), α=0.01, T_1=70 ms, and P_1=1.5%.

In addition, it is required that the end-to-end jitter is not causing jitter buffer overflow or underflow with probability larger than α. In this context, it is natural to define jitter as $J := F_T^{-1}(1-\alpha) - T$, where $F_T^{-1}(1-\alpha)$ is the 100(1-α)-percentile of the delay distribution (i.e. the value of the delay only exceeded by 100α percent of the packets), and T is the mean delay. In general, the total jitter over two domains 1 and 2, $J_{network}$, is a function of the delay distributions of both domains. $J_{network}$ can be upper-bounded or approximated by a function $g(J_1, J_2)$ of the per-domain jitter values J_1 and J_2. A detailed study of expressions for g is an active area of research (see [13,14] and references therein for studies on the propagation of jitter) and beyond the scope of this paper.

3.2. VoIP calls traversing N>2 domains

The multi-domain case in which N>2 domains, numbered 1, 2, …, N, are involved can be handled in a similar way. The calculation rules for the two-domain case discussed in Section 3.1 can be readily extended to the multi-domain case, leading to the following formulas:

$$T_{network} = \sum_{i=1}^{N} T_i , P_{network} = 1 - \prod_{i=1}^{N} (1 - P_i), J_{network} = g(J_1,..., J_N),$$

where g represents an approximation or upper bound (see Section 3.1). If operator 1 wants to dimension its network such that delay-loss-jitter combination is (T_1, P_1, J_1), then the calculation rules can be used to determine the set of *total* delay-loss-jitter combinations over domains 2, 3, …, N, taken together.

3.3.Certification of domains

One approach to control the end-to-end network performance is to define a *QoS certificate*. Such a certificate is given to a domain that satisfies a certain set of QoS requirements, including delay-loss-jitter combinations (T, P, J). Requirements on other QoS metrics such as availability and reliability are beyond the scope of this study. The SLAs are bilateral agreements between neighboring domains that include the total amount of bandwidth and the availability-related parameters, such as the Mean Time Between Failures (MTBF) and the Mean Time To Repair (MTTR). VoIP sessions are routed over certified domains only, if possible; otherwise, no QoS guarantees can be given. An important aspect of such a QoS certificate is that it creates a business incentive for operators to meet QoS requirements and obtain a certificate: the competitive edge of uncertified domains will degrade automatically.

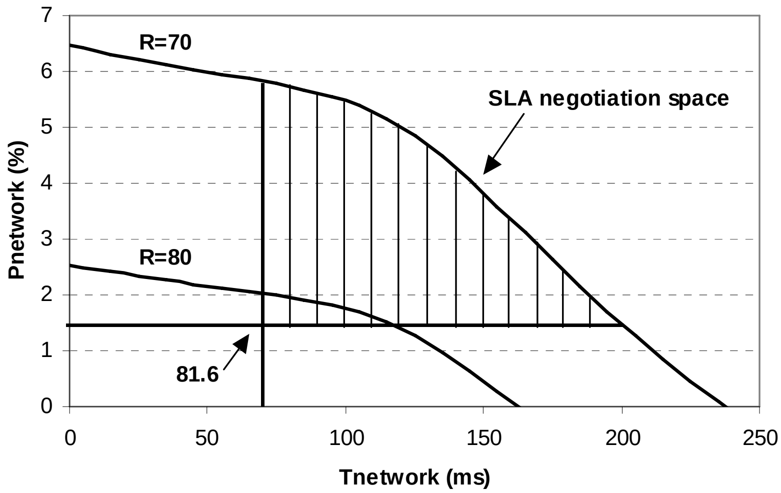

Figure 5. SLA negotiation space for (G.711+PLC, 50, 1), α=0.01, T_1=70 ms, and P_1=1.5%.

The QoS certificate may consist of two parts: (1) the QoS parameters for originating and terminating calls, and (2) the QoS parameters for transit calls. The delay-loss-jitter requirements for certification for transit traffic are parameterized by the triple ($T_{transit}$, $P_{transit}$, $J_{transit}$), and for terminating traffic by (T_{term}, P_{term}, J_{term}). Then an operator can use the above calculation rules for delay, loss and jitter to determine the *end-to-end* network-level delay, loss and jitter for any given number of (certified) transit domains. For example, if N>2 domains are involved then the triple ($T_{network}$, $P_{network}$, $J_{network}$) is given by

$$T_{network} = T_{term} + (N-2)T_{transit} + T_{term} = (N-2)T_{transit} + 2T_{term},$$

$$P_{network} = 1 - (1 - P_{term})^2 (1 - P_{transit})^{N-2},$$

and

$$J_{network} = g(J_{term}, J_{transit}, J_{transit}, \ldots, J_{transit}, J_{term}).$$

Combining these results with the E-model (see Sections 2.1 and 2.2), the relation between the perceived voice quality, the number of domains, the terminal parameters, and the QoS certificate parameters is determined. As an illustration, for the case of terminals with parameters (G.711+PLC, 50, 1), α=0.01, Figure 6 below shows the R-value of a VoIP session as a function of the number of domains N, where the per-domain QoS requirements are taken to be (T_{term}, P_{term}, J_{term}) = (50, 0.01, 10) for the terminating domains, and where two QoS classes are for transit domains: silver, with parameters ($T_{transit}$, $P_{transit}$, $J_{transit}$) = (40, 0.02, 5), and gold, with ($T_{transit}$, $P_{transit}$, $J_{transit}$) = (20, 0.01, 5), under the simplifying assumption that the per-domain delays are mutually independent and normally distributed. In this example, Figure 6 shows that in order to achieve at least medium quality (corresponding to R-value between 70 and 80) transit via “silver” domains allows for a maximum of three domains, whereas transit via “gold” domains allows for a maximum of four domains. Conversely, in order to achieve at least medium quality for multi-domain VoIP sessions crossing more than 3 domains, silver domains do not suffice, and gold subscription is required up to four domains.

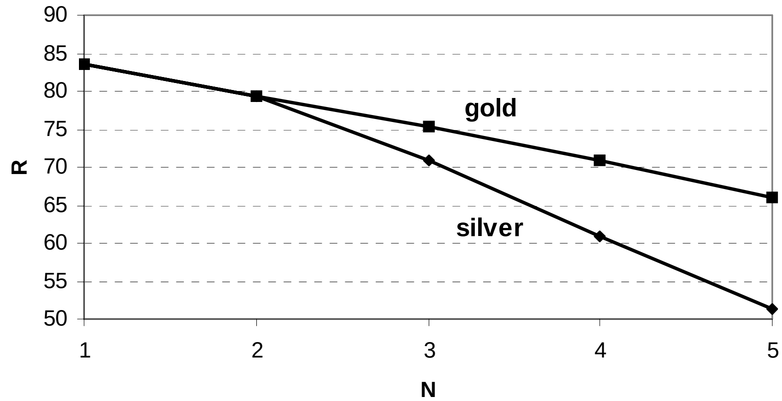

Figure 6. R as a function of N for (G.711+PLC, 50, 1) terminals, α=0.01, and “gold” and “silver” transit domains.

3.4. Computation of end-to-end performance characteristics

If a VoIP connection crosses non-certified domains, then end-to-end performance characteristics can be computed taking the following per-domain information into account: (1) content of the SLAs with neighboring domains, (2) network QoS measurements, or (3) network QoS data published, or made available, by the operator. Combining the available information on the per-domain performance characteristics, one can use the formulas in the beginning of Section 3.2 to estimate the end-to-end network level QoS. Alternatively, the end-to-end delay distribution, and therefore the end-to-end jitter $J_{network}$, can be computed based on convolving the estimated or measured delay distributions of the separate domains. An effective implementation of such a procedure can be based on the numerical procedure proposed in [15].

4. STEP-BY-STEP APPROACH TO REALIZE END-TO-END QoS

The results discussed in Sections 2 and 3 can be combined to develop a step-by-step approach to effectively realize the desired end-to-end QoS level for VoIP connections at the user-perception level. The approach is outlined below.

Step 1. Identify terminal parameters. Each terminal is identified by the following parameters: the codec (C), the jitter buffer size (J), and the number of frames per packet (M). Recall that all other terminal-related parameters in the E-model are assigned their default values [7].

Step 2. Set end-to-end QoS requirement. Identify the R-value of the end-to-end QoS requirement as perceived by the end user. This step also allows classification in end-to-end QoS for intra- and inter-domain calls.

Step 3. Construct iso-R curves. Use the E-model to calculate the combinations of end-to-end mean one-way delay and packet loss (T, Ppl) lead to desired R-values (see for example Figure 3).

Step 4. Determine end-to-end QoS requirements at the network level. For a number of values of α (e.g., in the range [0; 0.1]), determine $T_{term,min}$, $T_{term,max}$ and β to derive the admissible regions for network delay and packet loss from the iso-R curves constructed via step 3 (see for example Figure 4). For the single-domain case, the network-level requirements are determined. For the multi-domain case, N≥2, the following additional step needs to be carried out.

Step 5. Determine requirements to other domains. For calls traversing two domains, for given delay-loss-jitter combinations at an operator's domain, say operator 1, use the equations in Section 3.1, or the method discussed in Section 3.4, to identify the requirements to the combinations of delay-loss-jitter that need to be met by the other domain. Use this to determine the SLA negotiation space (see for example Figure 5). For calls traversing more than two domains, use the equations in Section 3.2, or the method discussed in Section 3.4, to identify the requirements to the combinations of delay-loss-jitter that need to be met by the other domains 2, 3, …, N, together.

Step 6. Determine the perceived quality of multi-domain calls. For each of the QoS classes defined in the QoS certificates - both for terminating and transit traffic - determine the quality for multi-domain calls, either by using the equations in Section 3.3 for certified domains, or by using the method discussed in Section 3.4. For certified domains, the quality can be estimated as a function of N (see for example Figure 6). This relation can also be used to determine which QoS classes, available within the different domains, meet the end-to-end QoS requirements.

Remarks:

1. It is important to notice that this approach can also be used to determine the relation between the *distribution* of the R-value and the distributions of packet loss and delay (over different sessions). In this way, it provides a means to realize *statistical* guarantees for the per-session QoS. This, however, requires a corresponding modification of the parameters in the certificates and the SLAs. As the complexity of certificates and SLAs increases, enforcement and monitoring by the service provider will become more complex accordingly. Strict QoS guarantees per individual session are not provided. This would require a much more complex per-session QoS enforcement mechanism (such as RSVP) as an alternative to the statistical approach described in this paper.
2. The approach presented assumes a given triple (C, J, M) of terminal parameters, but applies to different combinations of C, J, and M. In practice, VoIP sessions with many different combinations can be in progress simultaneously. In such a heterogeneous context, it is up to the service provider to decide which terminals are supported with which QoS guarantees. This decision will impact the SLAs negotiated.

5. CONCLUDING REMARKS AND TOPIC FOR FURTHER RESEARCH

In this paper, we provided a practical approach to realize user perceived QoS for VoIP in a heterogeneous multi-domain network environment. The approach, as summarized in Section 4, may serve as the basis for a Decision Support System (DSS) for operators to negotiate SLAs with neighboring domains and to choose the proper quality classes for transit domains. The performance parameters in the operator's own domains as well as the performance parameters negotiated in SLAs are input parameters for the calculation of operational costs. Conversely, the DSS can be used to calculate the most cost-effective way to realize the desired quality as perceived by the end user. Throughout the paper we considered terminals parameterized by the triple (C, J, M), resulting in the same quality perception by both end users within a VoIP session. In general, the end-user terminals in a session may be different, and parameterized by the 5-tuple (C, J_1, M_1, J_2, M_2). In this case, we define the perceived quality to be the minimum of the quality experienced by both users. In this paper we focused on requirements to delay, jitter and packet loss. Requirements on other QoS parameters such as call setup time, bandwidth and availability are beyond the scope this paper.

Finally, we address a number of topics for further research. First, in Section 3.2 we discussed the end-to-end jitter $J_{network}$. The further development and validation of approximations or upper bounds for the propagation of jitter is a challenging topic for further research. Second, we assumed static jitter buffers, whereas in practice the size of jitter buffers is often changing dynamically in response to actual network jitter. A straightforward way to cope with this problem is to assume a worst-case value of the jitter buffer size. However, this may pose unnecessary stringent network requirements, and hence, may be not cost-effective. More sophisticated methods for dealing with dynamic jitter buffers is for further study. Third, the approach discussed in this paper to deal with multi-domain QoS problems for VoIP can be extended towards different types of applications, such as video services and web browsing, amongst others. Finally, the proposed approach is a first step towards the realization of end-to-end QoS for VoIP. Validation of the approach via simulations or experiments in a test environment is for further study.

REFERENCES

1. E-pacity service, http://www.kpn.com.
2. K. Asatani, H. Ueda, and C.M. Lockhart (eds.), "Voice over IP and quality of service", IEEE Communications Magazine, Vol. 42, No. 7, July 2004.
3. ETSI TIPHON home page, http://portal.etsi.org
4. TIPHON Release 4; end-to-end QoS in TIPHON systems, Part 2: "Definition of speech QoS classes", ETSI TS 102 024-2.
5. TIPHON Release 4; end-to-end QoS in TIPHON systems, Part 3: "Signalling and control of end-to-end QoS", ETSI TS 102 024-3 V4.1.1, 2003-01.
6. AQUILA, http://www.ist-aquila.org.
7. ITU-T G.107, "The E-model, a computational model for use in transmission planning" (03/2003).
8. N.O. Johannesson, "The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks", IEEE Communications Magazine, Jan. 1997, pp. 70-79.
9. ITU-T G.113 Appendix I, "Provisional planning values for the equipment impairment factor Ie and packet-loss robustness factor Bpl" (05/2002).
10. J. Janssen, D. De Vleeschauwer and G.H. Petit, "Delay and Distortion Bounds for Packetized Voice Calls of Traditional PSTN Quality", Proc. of the 1st IP-Telephony Workshop (IPTel 2000), pp. 105-110, Berlin, Germany, 12-13 April 2000.
11. ITU-T G.109, "Definition of categories of speech transmission quality" (09/99).
12. ITU-T G.114, "One-way transmission time" (05/2003).
13. T. Bonald, A. Proutière, and J.W. Roberts, "Statistical performance guarantees for streaming flows using expedited forwarding". In Proc. IEEE Infocom 2001.
14. F. Brichet, L. Massoulié, and J.W. Roberts, "Stochastic ordering and the notion of negligible CDV". In Proc. ITC 15, 1997.
15. G.L. Choudhury and D.J. Houck, "Combined Queuing and Activity Network Based Modeling of Sojourn Time Distributions in Distributed Telecommunication Systems". In: The Fundamental Role of Teletraffic in the Evolution of Telecommunications Networks (Eds. J. Labetoulle and J.W. Roberts), Proc. ITC 14, 1994, pp. 525-534.