

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,

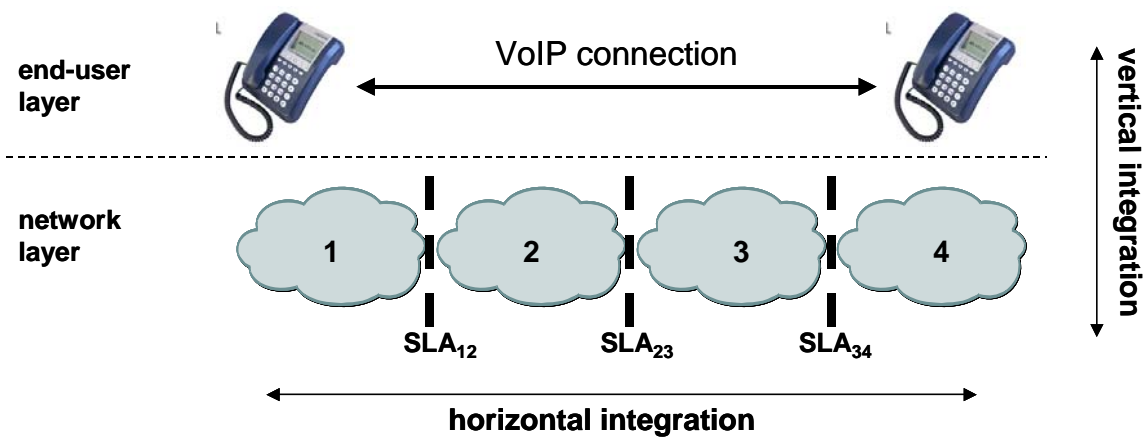


Figure 1. Multi-domain VoIP connection.

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:

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

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

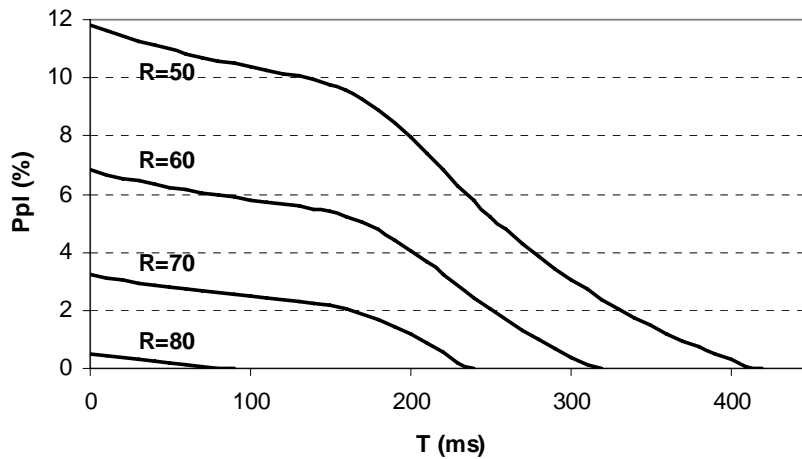


Figure 3. Iso-R curves for terminal (G.729, 50, 2).

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 network 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. $T = T_{\text{network}} + T_{\text{term}}$. 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.

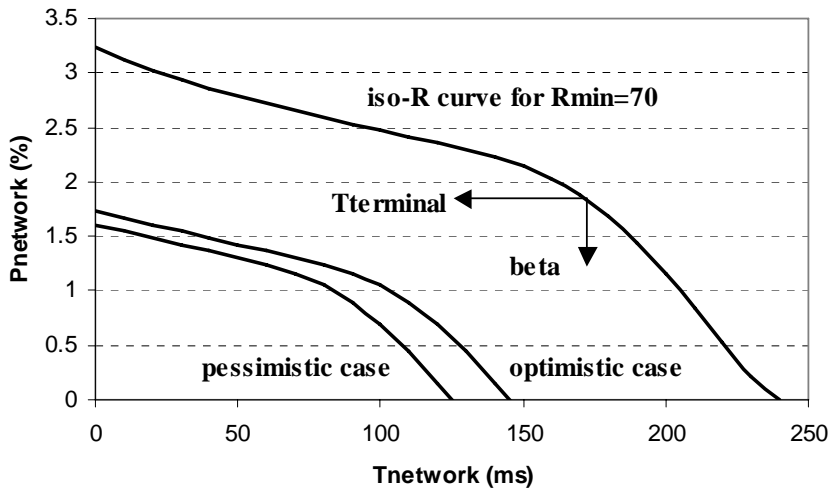


Figure 4. Network delay and loss budgets for terminals (G.729, 50, 2), $R_{\min}=70$ and $\alpha=0.01$.

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 $R \geq R_{\text{own}}$, and for inter-domain VoIP sessions (originating in 1 and terminating in 2) of $R \geq R_{\text{own}} \geq R_{\min}$. Notice that in

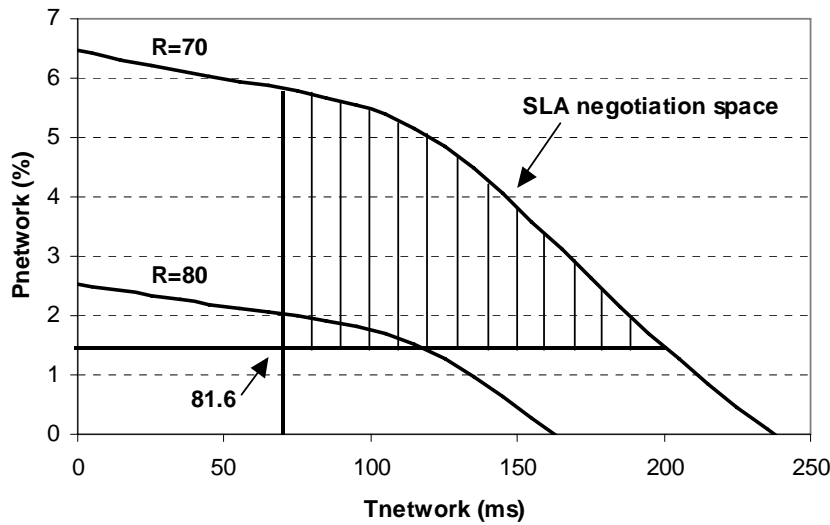


Figure 5. SLA negotiation space for (G.711+PLC, 50, 1), $\alpha=0.01$, $T_1=70$ ms, and $P_1=1.5\%$.

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

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

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), $\alpha=0.01$, 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.

$N \geq 2$, 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, J₁, M₁, J₂, M₂). 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.

