12

Speech Quality Aware Resource Control for Fixed and Mobile WiMAX

Thomas Michael Bohnert, Dirk Staehle and Edmundo Monteiro

12.1 Introduction

Hardly any subjects in telecommunications caused such controversy as the Quality of Service (QoS) issue in the Internet. The persistent advance in switching capacity over bandwidth demand by real-time services is dividing those calling for Internet QoS from their counterparts. The Internet2 QoS Working Group, for instance, concluded after years-long large-scale QoS deployment that conceptual issues plus a lack of demand, that is, missing real-time applications, severely inhibits the evolution of the Internet rather than fostering it (Teitelbaum and Shalunov, 2002). They even went as far as calling for banning end-to-end Internet QoS by regulation in order to guarantee a free and therefore powerful Internet based on its original principles. Meanwhile this conviction turned into a concept known as *net neutrality* and became subject of a fierce power struggle between telecommunication and service providers (Bohnert *et al.*, 2007).

While this conclusion might possibly hold for wireline communications it is a very different story for wireless communications and WiMAX is a perfect example of this. With the roll-out of mobile WiMAX the long-awaited IP convergence comes true and means that traditional real-time services, such as voice and video, will have to be delivered over

WiMAX Evolution: Emerging Technologies and Applications Edited by Marcos Katz and Frank Fitzek
© 2009 John Wiley & Sons, Ltd

Marked proofs

the Internet infrastructure. However, subscribers will not change their expectations along with exchanging access technology and voice is therefore a very intuitive example. Mobile voice subscribers are meanwhile used to cellular-phone-like service quality and will expect a similar experience. The success of WiMAX will thus ultimately depend on the success of operators in providing competitive QoS levels.

This awareness has settled right at the beginning of IEEE 802.16 standardization and a designated feature of mobile WiMAX, based on IEEE 802.16-2005, commonly called IEEE 802.16e (IEEE, 2005) is its inherent QoS support which is largely based on the concepts of ATM's comprehensive QoS body (Eklund *et al.*, 2006). So WiMAX defines a set of services categorized and parameterized by their respective target applications. One of these services is voice and assumes standard Voice over IP (VoIP) technology at the application layer. The WiMAX QoS model in turn defines semantics adopted by VoIP to control packet delay, loss and jitter. Explicit implementation instructions, however, are not part of the IEEE 802.16 standards and the choice of which particular resource control algorithms to implement is entirely left to manufactures.

In the context of VoIP these two decisions, using the QoS model of the Asynchronous Transfer Mode (ATM) as the bottom line without implementation instructions, render themselves as a considerable challenge: the identification and implementation of efficient resource control components capable of achieving VoIP QoS support in WiMAX systems. By way of simulation we present two such resource control algorithms, Admission Control (AC) and scheduling, where they are both based on dynamic quality estimations derived from instantaneous local measurements. In order to do so we first discuss quality assessment as such in Section 12.2 before we introduce speech quality assessment as the only appropriate means for this purpose in Section 12.3. In Section 12.4 we then present our approach for dynamic and on-demand quality assessment, the fundamental resource control criterion. The first resource control algorithm leveraging this approach is a Measurement Based Admission Control (MBAC) algorithm. Its theoretical underpinning plus performance evaluation are presented in Section 12.5. In Section 12.6 we present the concept of a scheduler that uses measurements or estimations of the current speech quality for making its scheduling decisions. We present a simple R-score based scheduler and show that its performance is superior to the channel-oblivious Earliest Deadline First (EDF) scheduler and comparable to the channel-aware MaxSNR scheduler. Finally, we give a short conclusion on this chapter in Section 12.7.

12.2 Quality of Experience versus Quality of Service Assessment

Before entering into technical details, the very first question to answer is how to define QoS for VoIP. While there is a common notion attached to the term *quality*, the actual space for its definition is divided into two parts: subjective and objective. Customers define quality as the overall satisfaction based on subjective assessment in multiple dimensions while engineers tend to express quality in terms of physical and measurable parameters and thus objectively. This is a fundamental discrepancy and frequently leads to notable misunderstandings.

A comprehensive QoS definition and terminology coverage has been given by Gozdecki *et al.* (2003), from which this section is partly borrowed. In brief, and according to Hardy

Markd proofs

(2001), a general model divides QoS into three notions: *intrinsic*, *perceived* and *assessed*. Intrinsic QoS (IQ) is purely technical and evaluates measured and expected characteristics expressed by network parameters such as delay and loss. Perceived QoS (PQ) reflects user satisfaction while using a particular service. It is therefore a subjective measure and the only method to ultimately capture it is to survey human subjects. Assessed QoS (AQ) extends the notion of PQ to secondary aspects such as service price, availability, usability and reliability. Each of this definition can be considered separately but there is a tight interdependence. PQ is a function of IQ and is an element of AQ. Nevertheless, they are commonly considered in isolation and interesting enough, the IETF as well as the ITU-T and ETSI direct their focus on different definitions and neither of them considers AQ to a full extent (Gozdecki *et al.*, 2003).

The IETF lastingly coined the notion of QoS adhering to IQ with the introduction of the IntServ (Braden *et al.*, 1994) and DiffServ (Blake *et al.*, 1998) frameworks. Consequently, a later published IETF QoS definition reads ‘A set of service requirements to be met by the network while transporting a flow’ (Crawley *et al.*, 1998). In contrast, the ITU-T and ETSI jointly define QoS as ‘the collective effect of service performance which determine the degree of satisfaction of a user of the service’, expressed the first time in ITU-T (1993) and clearly compliant with PQ. On top of this, the ITU-T recently released the Quality of Experience (QoE) framework ITU-T (2004) in which an explicit distinction has been made between QoS and QoE. In this document, QoS expresses the ‘degree of objective service performance’ and QoE the ‘overall acceptability of an application or service, as perceived subjectively by the end user’. According to these definitions, QoS is equal to IQ and QoE is equal to AQ. Notwithstanding, QoE is prevailingly associated with SQ and as SQ is an element of AQ, we adopt this definition together with QoS for IQ for the remainder of this work.

So far it has been shown that looking at service quality calls for a careful distinction based on its assessment. While QoS evaluation is deemed straightforward and merely a matter of measuring physical parameters, it appears much more complicated (and tedious) for QoE as it involves humans. However, QoE is the ultimate measure and in order to harness it in systems the interdependence between QoS and QoE is to be exploited. This can be achieved by observing that QoE is a function of QoS which can be expressed by mapping physical parameters to user ratings. This quality assessment method based on mapping measured parameters, such as delay or loss, to a QoE scale, such as the Mean Opinion Score (MOS), is called Instrumental Quality Assessment (IQA) (Raake, 2006b).

IQA is central to our work and an elaborate coverage in the context of VoIP will be presented in Section 12.3. For now the reader should note that there is considerable body of work related to quality assessment and its applications, see Janssen *et al.* (2002), Raake (2006a), Takahashi *et al.* (2006) and Markopoulou *et al.* (2003). In a mobile environment, the relationship of QoS and QoE for a Skype call over a Universal Mobile Telecommunications System (UMTS) Internet access was investigated by Hoßfeld and Binzenhöfer (2008). General investigations on the exponential interdependency between QoS and QoE for different voice codecs can be found in Hoßfeld *et al.* (2008, 2007).

At last IQA paved the way for an emerging area in research and development: resource management based on QoE. Traditionally, resource management was mostly concerned with QoS but frameworks are increasingly considering QoE as the ultimate performance metric. Some examples are given by Sengupta *et al.* (2006) who present parameterization guidelines to improve VoIP quality for WiMAX.

12.3 Methods for Speech Quality Assessment

No matter how a voice service is implemented, either analog or digital, over a circuit-switched or packet network, it essentially means speech transmission and the ultimate service quality depends on how uttered information is being understood by communication participants. Henceforth, the only appropriate quality assessment method for voice services is that of subjective speech quality assessment. Given the adopted terminology it therefore relates to QoE.

12.3.1 Auditory Quality Assessment

As briefly outlined in the previous sections subjective quality assessment involves surveying human subjects. Methods falling into this category are classified as auditory methods and share a basic commonality: in controlled experiments human subjects listen to speech samples subject to varying impairment in space and time and record the perceived quality.

The outcome of an auditory method is highly individual and controlled by a multitude of features anchored in human physiology. Human perception depends on spectral and temporal processing capabilities of the auditory system, on echoic, short-term and long-term auditory memory and speech comprehension, intelligibility and communicability. Finally, it is influenced by the ability to restore missing sounds by way of analyzing context, a feature known as the 'picket fence effect' in analogy to a visual modality: whilst watching a landscape through a picket fence, the fence interrupts the view periodically but the landscape is seen to continue as the human brain is padding missing pieces from memory. Apparently beneficial at first sight, this feature can severely impair perception if padded pieces are selected from alleged context composed of the actual plus extracts of previous context information stored in memory. In an advanced state, subjects even try to anticipate future or missing information and if it does not match the perceived version, they likely classify it falsely as distorted, wrong or even entirely missing.

There is a whole science behind auditory test methods. They are divided in utilitarian and analytic tests. The former aim at directly comparing the quality of different speech communication systems while the latter try to reveal the perceptual features underlying speech quality. Utilitarian methods are subdivided into listening quality, comprehension and listening and talking tests according to different stimuli, context and other features; see Raake (2006b) for a complete treatise. Central to any of these methods is the question on how to scale ratings: absolute, relative, discrete or continuous. The well-known MOS, for example (illustrated in Figure 12.3) is a five-point Absolute Continuous Rating (ACR) scale. Its name is derived from the fact that it expresses the average over a set of individual ratings obtained in an controlled experiment. It is frequently used in methods aiming at capturing time varying quality impairment by recording the instantaneous quality over time, typically with a slider over the ACR scale; cf. Watson and Sasse (1998). As shown later in this chapter, time varying quality impairment plays a central role in speech quality assessment. Nevertheless, it is frequently neglected due to its alleged complexity.

12.3.2 Instrumental Quality Assessment

In many cases it is desirable to evaluate speech quality on demand, in real-time and without human involvement. Accordingly, much effort has been spent in developing alternatives to auditory tests, so-called instrumental methods or IQA. The principle of these methods is to correlate physical and measurable magnitudes with quality as perceived by a human subject. While such unique relationships exist in theory, it has to be noted that, so far, it remains impossible to establish them in practice, even for very simple applications.

Nevertheless, recent advances disclosed methods with considerable precisions and consistency. In particular, signal-based methods achieve accurate and reliable results based on the relatively well-understood signal processing performed by human. Examples are Perceptual Evaluation of Speech Quality (PESQ) (ITU-T, 2001) and Perceptual Speech Quality Measure (PSQM) (ITU-T, 1996). Either of these methods compares a clean reference signal with the same signal after been processed by the system under test. In a final step the estimated deviation is mapped to a rating scale, such as the five-point ACR, where this mapping is purely empirical and the result of a large number of auditory tests.

The alternative to signal-based methods are so-called parameter-based methods. These methods require instrumentally measurable magnitudes which are evaluated in a parametric model. The most popular method of this category is the E-Model (ITU-T, 1998), which is the basic component of the presented framework and is presented in great detail in Section 12.4. In addition to its distinct internal rationale, it maps the result to a subjective rating based on a purely empirically found function, just like a signal-based method.

12.4 Continuous Speech Quality Assessment for VoIP

As stated in Section 12.1, the objective of this work is to devise the ‘identification and implementation of efficient resource control components capable of achieving VoIP QoS support in WiMAX systems’. At this point in time, however, the reader should note the error in this statement. In accordance with the preceding argumentation it must be called ‘VoIP QoE support’. The immediate conclusion is that a purpose-built framework requires a component capable of continuously assessing VoIP QoE levels. Furthermore, the overall objective is to assure VoIP QoE levels by means of resource control and thus the second requirement on this framework is the delivery of precise QoE estimates in real-time as a criterion for resource assignment. Obviously this can only be an instrumental method and precludes signal-based methods since there is no space and time for transmitting reference signals over the radio interface.

12.4.1 VoIP Components and their Impact on Speech Quality

Speech is a slowly varying analog signal over time and its frequency components are limited to the lower 4 kHz band. Owing to linguistic structures a speech signal alternates between talk spurts and silence periods. The origin of talk spurts are typically syllables which themselves are phoneme sequences containing one vocalic sound. Average durations of talk spurts range from 300 to 400 ms while silence periods range from 500 to 700 ms. In order to transmit speech over a packet network several pre- and post-processing steps are required. In Figure 12.1, each of these processing steps is represented by a single logical building block and the complete set makes a basic end-to-end VoIP system.

Any of the elements of the VoIP system has an impact on the speech quality perceived by the receiver. First, the analog signal is converted into a digital signal. This step is commonly assumed to be negligible with respect to the overall quality. Additive noise is canceled

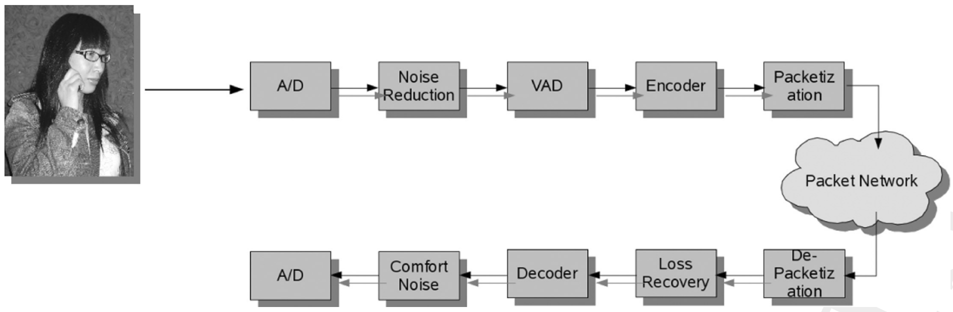


Figure 12.1 Basic Components of a VoIP System.

next before the signal enters the Voice Activity Detector (VAD). The VAD prevents the transmission of silence and thus influences the durations of talk spurts and silence periods. The sojourn times in either state are roughly exponentially distributed with a tendency to longer tails Jiang and Schulzrinne (2000) but the mean is entirely controlled by the VAD's sensitivity. Modern VADs elongate talk spurts by a so-called hangover time in order to prevent speech clipping. Small hangover times result in shorter talk spurts and vice versa. Altogether, VAD explicitly and implicitly impacts on packet loss and its distributions and thus speech quality.

Talk spurts are encoded by one of the very many voice codecs. The simplest and most well-known is Pulse Code Modulation (PCM). It is standardized by the ITU-T where it is named G.711. This encoder produces a 64 kbps digital signal and implies some level of entropy due to its discrete quantization. This is the main reason behind its inherent impact on speech quality.

The digital signal, or bit stream, is then packetized into equal sized packets. For each talk spurt the continuous bit stream therefore results in a periodic sequence of packet emissions where the period is determined by the packet length. Deciding on a proper packet size is crucial with respect to overall efficiency, that is, the tradeoff between transport overhead and the actual payload.

Voice packets are transported over the IP network using the common mechanisms. The might be treated with priority by DiffServ or IntServ implementation in some network access segments but when it comes to the Internet backbone, they most likely share the same fate as any other best-effort traffic, cf. Markopoulou *et al.* (2003). Packets might be delayed, reordered, jostled (jittered) and eventually dropped. Obviously, this part of the VoIP system therefore potentially has the major impact on the overall quality.

Once it arrives at the end system, the speech stream is extracted from the packets. Lost information is identified, for example by means of sequence numbers, and algorithms such as Forward Error Correction (FEC) or Packet Loss Concealment (PLC) recover or mask it to some extent. In a next step the modified bit stream is decoded and depending on the deployed VAD, some comfort noise is added. This is to account for the artificial silence introduced by the VAD: absolute silence is perceived by humans as odd and is likely to be falsely interpreted as the system malfunctioning. In a final step the digital system is re-converted into an analog system and played out by a speaker device.

12.4.2 Continuous Assessment of Time-varying QoE

The *de-facto* parametric IQA method is the E-Model. Started as a study by the ETSI, it has been standardized by the ITU-T (ITU-T, 1998). Its original application domain is network planning and one of the questions we answer with this work is whether it lends itself as a tool for on-demand QoE estimation as input for online resource control.

The E-Model is an IQA method for mouth-to-ear transmission quality assessment based on human perception and is defined as

$$R = R_1 - I_s - I_d - I_e + A. \tag{12.1}$$

In (12.1), R denotes the psychoacoustic quality score defined in [0, 100]. It is an additive, nonlinear quality metric based on a set of impairment factors, namely R_0, I_s, I_d, I_e, and A. It assumes that underlying sources of degradation can be transformed onto particular scales and expressed by an impairment factor. These functional relations are found prevailingly empirical. The classes of degradation are as follows.

- Noise and loudness effects are represented by R_0. These effects originate from basic environmentally inflicted signal-to-noise ratios, such as those induced by reflections and interference in rooms or noise on the line.

- The *simultaneous impairment factor* (I_s) denotes speech signal impairment such as PCM quantizing distortion or VAD hangover times.

- Impairment due to information delay, such as transmission delay or echo, is represented by the *delayed impairment factor* (I_d).

- Degradation due to information loss is expressed by the *equipment impairment factor* (I_e). It covers terminal internal information loss such as low-bit rate coding but also losses caused by lossy transport media such as IP networks.

- The *Advantage Factor* (A) quantifies the user's tolerance with respect to quality degradations if these are perceived as inherent to a feature that otherwise increased system utility or convenience. For instance, cellular-phone subscribers expose higher tolerance to noise than fixed line users as quality degradations are perceived as a natural consequence of ubiquitous telephone access over radio interfaces.

The E-Model's particular appeal lies in its simplicity. To assess the speech quality of a VoIP call, one simply has to measure parameters such as delays and losses, map them onto degradation scales, and sum all factors in order to yield the final score R. Obviously, degradation functionals are therefore central to the E-Model. In Figure 12.2 one such functional is plotted for the equipment impairment factor I_e. Recalling the definition of I_e, it has to be noted that there is a functional for each and any codec as they are differentially sensitive to losses. The depicted functional plots packet loss over quality degradation for the G.711 codec. More of these mappings can be found in Janssen *et al.* (2000), Markopoulou *et al.* (2003) and Raake (2006b).

As mentioned in Section 12.3.2, auditory tests commonly quantify speech quality on an ACR scale. This also applies to the E-Model, whose R-score scales on a 100-point ACR. The globally established ACR scale, however, is the MOS and a translation from the R-score to

Marked proofs

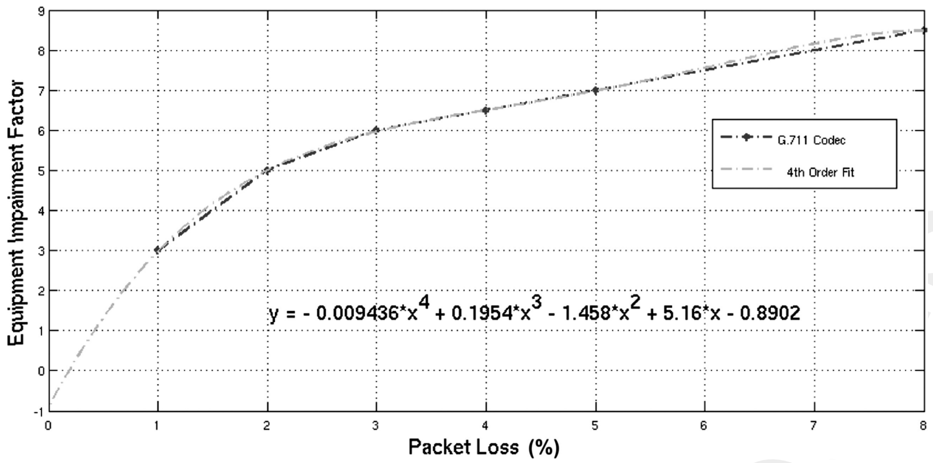


Figure 12.2 Nonlinear relation between the packet loss ratio and the equipment impairment factor (I_e).

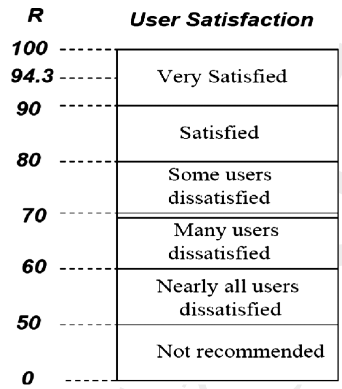


Figure 12.3 Mapping average user satisfaction (MOS) to the R-score.

MOS has been introduced as a result of extensive auditory tests. It can be found in ITU-T (1998) and is depicted in Figure 12.3.

The E-Model was used in various setups in numerous previous works; see, for example, Cole and Rosenbluth (2001), Janssen *et al.* (2002), Markopoulou *et al.* (2003), Meddahi *et al.* (2003), Raake (2006a), Sengupta *et al.* (2006), Takahashi *et al.* (2006). Many of these works discuss weaknesses and propose modifications where the two foremost points are the E-Model's additivity and that packet loss processes in IP networks are far more complex than what is captured by simple loss ratios. How to account for these phenomena is the subject of the following section.

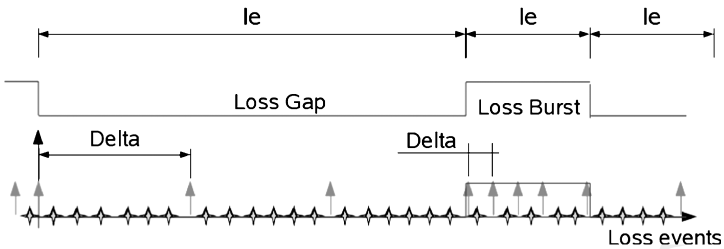


Figure 12.4 A series of consecutive periods of different microscopic loss behaviors, that is packet loss ratio and distribution, together form a macroscopic loss profile. There are two alternating microscopic loss behaviors, loss gaps and loss bursts. Stars represent VoIP packets and arrows indicate packet loss events. The distance between them is called delta. If delta is larger (smaller) than 16, and the model is in gap (burst) state, it remains is this state. Otherwise it changes from burst (gap) to gap (burst) state. In the event of an transition the impairment factor is calculated for the abandoned state. Over time, this leads to a series of I_e values.

12.4.3 Instationary Quality Distortion and Human Perception

The quality of speech is by far most influenced by information loss, that is, I_e (Kostas *et al.*, 1998, Markopoulou *et al.*, 2003, Raake, 2006b). The functional relation expressing this fact is nonlinear and the most relevant part for a G.711 codec is depicted in Figure 12.2. Using a simple fourth-order least-squares fit, the function reads

$$I_e = -0.009436x^4 + 0.1954x^3 - 1.458x^2 + 5.16x - 0.8902. \qquad (12.2)$$

However, measuring packet loss and mapping it to I_e is insufficient. In particular, if speech is delivered over IP networks by means of VoIP, the final quality is prevailingly determined by the packet loss distribution. Intuitively, single packet losses are always preferable over loss bursts. This is exactly the difference between IQ and SQ since by taking averages, as with IQ, such details are inherently ignored. Furthermore, packet loss distributions themselves are frequently instationary over a call's life time and instantaneous as well as ultimate quality assessment by humans exhibits strong correlation with this characteristic Raake (2006b, Chapter 4).

To account for this phenomenon we divide the packet loss process into periods with different loss behaviors, as proposed by Clark (2001) and refined by Markopoulou *et al.* (2003), Raake (2006a,b). In particular, we adopt the principles proposed by Clark (2001) but modified for our purpose. Essentially, this packet loss driven model defines two alternating states of microscopic loss behavior, loss gap and loss burst state, with respect to the distance of packet loss events, cf. Figure 12.4. According to Clark (2001), the model remains in the (loss) gap state as long as there is a minimum of 16 successfully received packets between two loss events (delta, δ). Otherwise there is a transition from (loss) gap to (loss) burst. The idea behind staying in a gap state under this condition is that modern loss recovery algorithms can handle isolated packet loss relatively well. In the case of a transition to burst state, the model remains in this state until 16 packets are successfully received between the latest and the previous loss event.

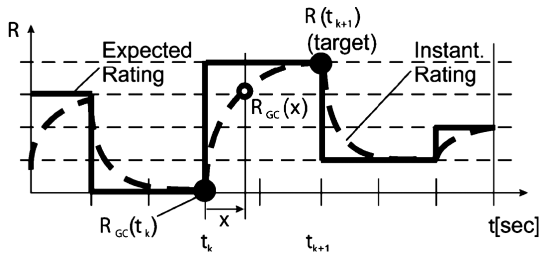


Figure 12.5 The expected rating (solid line) associated with either loss or gap state. The true, delayed perception (dashed line) by humans is indicated as an exponential decay or rise of the R-score with respect to a state transition. (Source: Raake (2006a).)

Upon the detection of any state transition the loss ratio for the previous state is used to calculate the corresponding impairment level, I_e, using the relation depicted in Figure 12.2, resulting in a time series of I_e values with respect to states. However, before these values can be used to compute R, there is another feature, inherent to human perception, which has been integrated into this model, the delayed perception (or acceptance) of quality change.

Naturally, humans tend to perceive a quality change rather continuously and not instantaneously at state transitions. A further distinction has to be made between transitions from good to bad and vice versa. So do humans confirm a change from good to bad much faster than the other way around. Generally, this feature can be modeled by an exponential function, similar to a transistor saturation curve, with specific time constants. **Q34**

Given $I_{e,g}$ and $I_{e,b}$, the impairment linked to gap or burst, I_1 is the estimated instantaneous impairment level at the change from burst to gap condition and I_2 equals the level at the return from gap to burst. In mathematical terms, I_1 and I_2 can be expressed as

$$I_1 = I_{e,b} - (I_{e,b} - I_2)e^{-b/\tau_1}, \tag{12.3}$$

$$I_2 = I_{e,g} + (I_1 - I_{e,g})e^{-g/\tau_2}. \tag{12.4}$$

Here g and b denote the sojourn time in gap or burst state and τ_1 and τ_2 are the time constants, respectively. Typical values are $\tau_1 = 9$ s and $\tau_2 = 22$ s (Raake, 2006b). A proper combination of (12.3) and (12.4) yields an expression for I_2 independent from I_1:

$$I_2 = I_{e,g}(1 - e^{-g/\tau_2}) + I_{e,b}(1 - e^{-b/\tau_1})e^{-g/\tau_2}. \tag{12.5}$$

Using (12.5) we are now in a position to calculate the average impairment level over a certain time, for example, for the life time of a call. Therefore, we first calculate *average* gap and burst length, $\bar{b}$ and $\bar{g}$, as well as the *average* impairment levels, $\bar{I}_{e,g}$ and $\bar{I}_{e,b}$. Putting these in (12.5) and integrating it over one burst and gap yields the *average* impairment level for a certain loss profile of certain length. It reads

$$\begin{aligned}\bar{I}_e = \frac{1}{\bar{b}+\bar{g}} &\times [\bar{I}_{e,b} \times \bar{b} + \bar{I}_{e,g} \times \bar{g} + \tau_1 \times (\bar{I}_{e,b} - I_2)\\ &\times (e^{\bar{b}/\tau_1} - 1) - \tau_2 \times (\bar{I}_{e,b} - (\bar{I}_{e,b} - I_2)\\ &\times e^{-\bar{b}/\tau_1} - \bar{I}_{e,g}) \times (e^{\bar{g}/\tau_2} - 1)].\end{aligned} \tag{12.6}$$

Eventually, by replacing I_e in (12.1) with $\bar{I}_e$ and using proper values for the remaining parameters, one can evaluate the subjective quality for a single call by this parametric IQA method called *integral quality by time averaging* (Raake, 2006b).

12.5 Speech Quality Aware Admission Control for Fixed IEEE 802.16 Wireless MAN

12.5.1 IEEE 802.16d Background and the Deployment Scenario

The IEEE 802.16d standard, officially called 802.16-2004 with reference to its release date, defines an air interface for fixed Broadband Wireless Access (BWA). In doing so it specifies several Physical (PHY) layers and a common Medium Access Control (MAC) layer on top of them. Target deployment is fixed Non-Line-of-Sight (NLOS) within the 2 to 11 GHz frequency band, either in Point-to-Multipoint (PMP) or in mesh mode. In PMP, a central Base Station (BS) controls all traffic interactions between Subscriber Stations (SSs) and itself and all traffic is either sent from a single SS to the BS, called the Uplink (UL), or from the BS to one or many SSs, called the Downlink (DL).

The MAC is connection oriented in order to support QoS, an essential future BWA requirement. There are several types of connections, each unidirectional and between two MAC instances. These connections serve different purposes such as MAC management and signaling with several priorities but also for data transport. Connections are identified by an unique Connection Identifier (CID).

In addition to connections, IEEE 802.16d defines the concept of Service Flows (SFs). A SF itself is defined as unidirectional transport service with predetermined QoS characteristics, that is, QoS parameters. Each SF is mapped to a single connection and has to be served by an UL (or DL) scheduler such that QoS requirements are met. This is being done by a so-called scheduling service which is related to the QoS parameters associated with the respective SF. It should be noted that the standard defines scheduling services but does not define any explicit scheduler for them. It is left to manufactures to select and implement a scheduler which meets the respective requirements. In this respect, IEEE 802.16d is in line with concepts known from DiffServ, which specifies Per Hop Behaviors (PHB); see, for example, Davie *et al.* (2002), but not how to implement them.

One of the envisioned deployments of IEEE 802.16d is to deliver VoIP services in different granularities. As IEEE 802.16 is connection oriented the spectrum ranges from a single VoIP call to VoIP aggregates. We focus on aggregates as single VoIP calls are rather typical for scenarios involving mobile terminals. As this work is in the context of the European research project 'WiMAX Extensions for Isolated Research Data Networks' (WEIRD)[1], the deployment we have in mind is a real deployment scenario defined by WEIRD. In this scenario a remote monitoring station, in the role of a SS, is connected by a BS to a central unit. In reality this is a Forest Fire Monitoring Station (FFMS) somewhere in the mountains connected to a Coordination Center (CC) in a nearby city. In this case VoIP services are used to support the personnel in the FFMS in reporting and for coordination of forest fire prevention activities by the CC. In order to do so, this scenario defines a dedicated, preprovisioned SF with a certain, fixed capacity in either direction. For more details on this

[1] See http://www.ist-weird.org.

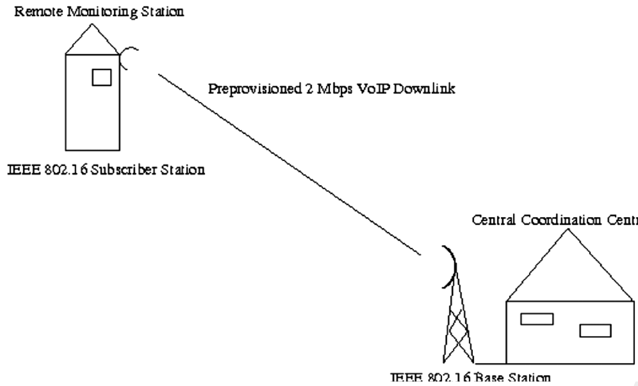


Figure 12.6 The example deployment scenario as defined in the EU IST FP6 IP Project ‘WEIRD’. A remote monitoring station (FFMS) is connected with a CC via a preprovisioned IEEE 802.16 SF for VoIP services.

scenario we point the readers to WEIRD Consortium (2007). The scenario is illustrated in Figure 12.6.

12.5.2 The Principle of Admission Control and its Application to VoIP

Admission Control (AC) is the most important mechanism for QoS provisioning on an aggregate level. In other words, if a provider decides to exploit statistical multiplexing gain within a single traffic class, AC regulates the traffic intensity by controlling the number of active flows such that a certain QoS objective is met. In the context of our VoIP scenario, this means an AC function controls the VoIP traffic arriving at the aggregation point, that is, the BS in DL direction or the SS in UL direction and destined to either the FFMS or the CC, such that a certain VoIP quality is assured.

Derived from the definition presented by Bohnert and Monteiro (2007), AC can be generally defined as

$$\chi_k \begin{cases} \geq 0 & \text{admit flow } k, \\ = 0 & \text{reject flow } k, \end{cases} \tag{12.7}$$

where χ_k denotes the admission criterion for the requesting flow k and is defined as

$$\chi_k = \max\{Q(N+1) - Q', 0\}. \tag{12.8}$$

Here we assume that $Q(n)$ expresses the level of QoS for n admitted sources, that is, the traffic aggregate, and is a monotonically decreasing function in n, while Q' is the target QoS. The computation of $Q(N+1)$ reads as

$$Q(N+1) = Q(N) - \Delta_{\text{QoS}}^{K} \tag{12.9}$$

where Δ_{QoS}^{K} denotes the QoS degradation inflicted on the aggregate by the characteristics of flow k, if the latter would be accepted.

Equipped with the expressions derived in the previous sections we can formulate an admission criterion based on QoE (speech quality). Therefore, we replace $Q(N+1)$ in (12.8) with $R(R_0, I_s, I_d, I_e, A)$ and set $R_0 - I_s = 94$, the default value with respect to inherent features of the G.711 codec. Further, I_d is set to an upper bound determined by the buffer length ω and the link capacity C; see Cox and Perkins (1999) for details. Beyond this bound, packet delay translates into packet loss and is captured by I_e. The respective equation for I_d reads

$$I_d = 4 + 1 * (\omega/C). \tag{12.10}$$

Combining all pieces and further assuming the worst case, that is, we set A to zero, we obtain

$$\bar{R}_T = 94 - 4 + 1 * (\omega/C) - \bar{I}_e(T). \tag{12.11}$$

In this equation the parameter T in $\bar{I}_e(T)$ indicates that the average impairment factor for time-varying speech quality assessment has been calculated over a window of T seconds. This is to account for an inherent feature of Measurement Based AC (MBAC) algorithms, which generally estimate a QoE/QoS criterion over a limited window. Eventually, we can express the admission criterion as follows:

$$\chi_k = \max\{\bar{R}_T - R', 0\}. \tag{12.12}$$

It has to be noted that the criterion in (12.12) slightly differs from that in (12.8) as we put $Q(N)$ ($\bar{R}_T$) in place of $Q(N+1)$. This is due to the difficulty in expressing and quantifying Δ^K_{QoS} without a precise traffic model. As we show, this has little or no impact but we are currently investigating alternatives and their merit.

Furthermore, by using this setup speech quality is assessed on an aggregate level with a method that was originally designed to assess individual call quality. Whether this makes sense is discussed in the following; see Section 12.5.4.2. At least from a model point of view there is little difference in computing $\bar{I}_e$ on an aggregate or call level. What is required in either case are the loss ratio, burst length and gap length. The single difference is the number of packets received (or lost) to trigger state transition, which is 16 for a single call, cf. Section 12.4.3.

In order to translate this trigger threshold to an aggregate level we apply a simple intuitive approach. The AC algorithm knows at any time the number of admitted flows N. By assuming that VoIP traffic can be modeled by a standard exponential on/off model with an average sojourn time in the on (talk) state of 300 ms and a mean off (silence) of 600 ms (Markopoulou *et al.*, 2003), we know that each flow is active (on) for roughly a third of its life time. We further assume that any contribution as well as impact on an individual call scales linearly with the number of calls aggregated. In other words, the contribution of any call as well as the impact on it equals in average that of any other call. On the basis of this assumptions we set the number of packets received (or lost) to trigger state transitions to $16 \times N \times 0.33$.

12.5.3 Experimental Setup and Parameterization

In order to evaluate the concept and performance of the algorithm, we implemented it in the NS-2 framework[2]. The basic scenario has been already described in Section 12.5.1 and

[2]See http://www.isi.edu/nsnam/ns.

Manuel Proofs

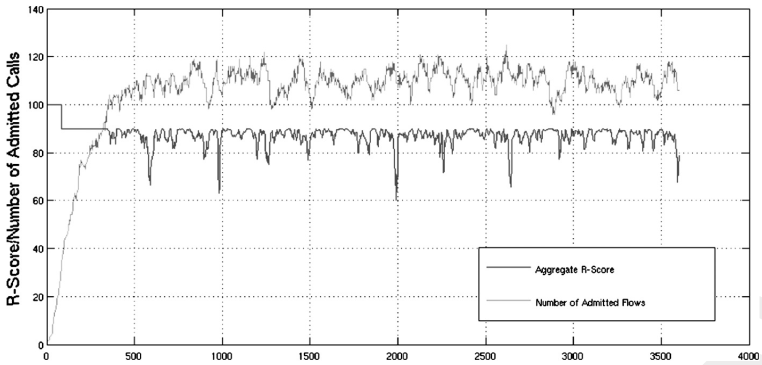


Figure 12.7 The number of admitted flows (upper curve) at the time of an admission request. The aggregate R-score is estimated (lover curve) which serves as admission criterion whenever a new call arrives. For this simulation R' in (12.12) was set to 85 (MOS: satisfied) and this target has been closely achieved over for the time the system remains in a steady state.

complies with an evaluation scenario defined by WEIRD. In this scenario a preprovisioned SF is set for VoIP. By definition this implies a contracted and assured capacity at any time and if there is any channel instability, it is accounted for by proper planning, compensated for by scheduling decisions or any other mechanism. We further assume no packet loss over the air interface by means of retransmissions or appropriate network planning. In such an scenario the UL and DL AC function, placed in the SS or BS are equivalent and allow for a reduction in the simulation setup to a single server queue with fixed capacity.

The preprovisioned link capacity of the respective SF, called the minimum reserved traffic rate in IEEE 802.16 QoS terminology, has been set to 2 Mbps and the buffer has a length of 30 packets. Call arrivals follow a Poisson process with mean arrival time of 2 s and the holding time is exponentially distributed with mean 210 s.

VoIP traffic was generated by a G.711 coder with voice fames of 20 ms length. The standard exponential on/off model is used to model talk and silent periods where average sojourn time in the on state is 300 ms and mean off time 600 ms (Markopoulou *et al.*, 2003).

Admission control is implicit, cf. Mortier *et al.* (2000), and new calls are detected at the first packet arrival. The algorithm's window length, the past time over which speech quality is assessed, is set to 300 s in order to cover calls of average length. All simulations run for 3600 simulated seconds and the first 500 s are discarded to evaluate the system in a steady state.

12.5.4 Performance Results

12.5.4.1 Admission Control Accuracy

One of fundamental problem of MBAC is precision and only a few algorithms tackle this issue. Hence, we first investigate how closely the algorithm approaches a demanded QoE objective.

Table 12.1

R'	$\bar{R}_{T,\mu}$	R_σ	$R_{\min}$	$R_{\max}$	$t^{\max}_{R<R'}$
80	84.67	4.44	63.36	89.94	31.73
82	85.55	4.16	68.37	89.95	29.66
84	86.53	3.57	66.04	89.98	28.12
86	87.64	2.77	67.05	89.97	37.76
88	88.63	1.71	77.88	89.97	43.22

For the first simulation R' in (12.12) was set to 85 and as shown in Figure 12.7, this target **Q35** was achieved for most of the time. Skipping transient state the average estimated R-score ($\bar{R}_{T,\mu}$) for the remaining time was 86.82, standard deviation $R_\sigma = 3.89$, $R_{\min} = 59.97$ and $R_{\max} = 98.89$. In addition, we computed the longest continuous period below R', $t^{\max}_{\bar{R}_T<R'}$ and found a value of 39.87 s. We repeated this simulation for different R' in the range [80, 90], which maps on MOS to 'Satisfied'. The results are listed in Table 12.1. **Q36**

While this results indicate a relative consistent performance, the AC appears a bit too conservative for lower R' values. Perhaps more important in the context of traffic aggregation and statistical QoE is that the average R-score was slightly above R' for all simulations. Among the remaining parameters, $t^{\max}_{R<R'}$ certainly holds the most interesting information. At first sight the maximum duration seems relatively large compared with an average holding time of 210 s. However, the maximum alone does not tell us much and in Figure 12.8 we plot the Cumulative Distribution Function (CDF) of the times $\bar{R}_T$ remained below R', denoted by $t_{\bar{R}_T<R'}$.

This figure further indicates consistency as the curves are very similar. For the whole range of R', the average time $\bar{R}_T$ remains below R' is approximately 10 s and the probability that $t_{\bar{R}_T<R'}$ is larger than 20 s is roughly 0.2. This qualifies the large value for $t^{\max}_{R<R'}$.

12.5.4.2 QoE Performance on the Call Level

Speaking in general terms, what has been achieved by now is an algorithm that can statistically guarantee a predefined application layer metric. However, how meaningful is this metric on the call level? Can we assume that an R-score measured and maintained on aggregate level applies to individual calls too?

In order to find this out we ran the same set of simulations as before, selected randomly 100 consecutively admitted calls and recorded their loss process. We then used the same IQA but with an adjusted state transition trigger to evaluate single call QoE, see Section 12.5.3. The question we tried to answer is how many calls receive the contracted QoE. We therefore assessed the QoE for each call's total life time and Figure 12.9 plots the CDF of these calls R-scores. The figure shows that for each QoE target R' maximally around 5% of calls are rated below $R = 80$, which is the lower threshold for 'satisfied' on the MOS scale, cf. Figure 12.3. Taking the first simulation, depicted in Figure 12.7, as an example this means that approximately 6 calls out of 110 concurrently admitted calls on average would be affected by lower QoE than contracted. Yet some of these calls fall still in the range $R = [70, 80]$ which maps to MOS 'some users dissatisfied', meaning that some of these may still be rated as 'satisfied'.

Marked Proofs

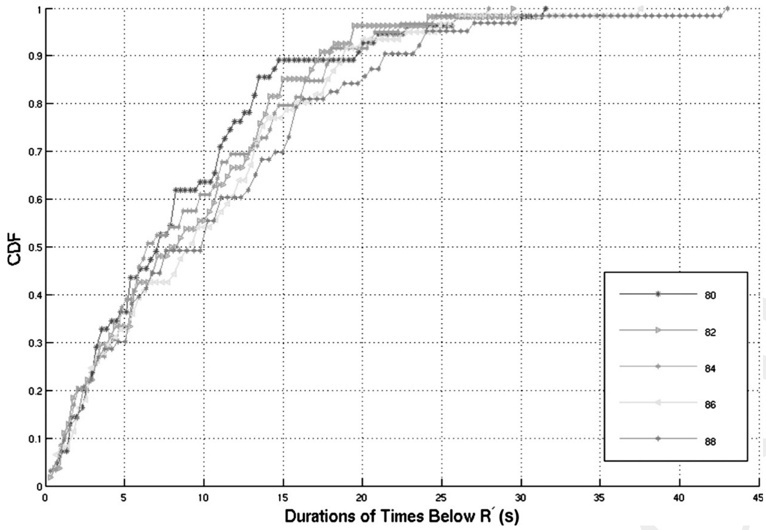


Figure 12.8 CDF of $t_{\bar{R}_T < R'}$. The probability that $t_{\bar{R}_T < R'}$ is larger than $1/10$ of the holding time, in other words that the quality is below the requested one for one-tenth of a calls life time, is approximately 0.2.

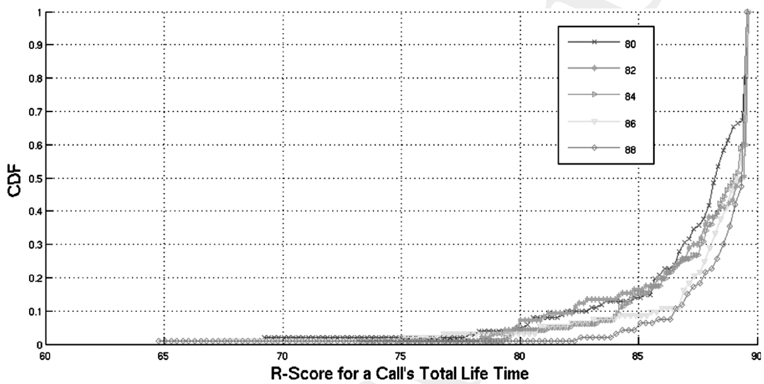


Figure 12.9 CDF of single call quality for a set of randomly recorded calls for all simulations. For each simulation, less than 5% of calls fall below $R = 80$.

Finally, from Figure 12.7 we can draw conclusions with respect to configuration and QoE versus resource utilization tradeoff. If an operator aims at making sure that less than 2% of calls fall below $R = 80$ (MOS: satisfied), it should set $R' = 88$. Obviously, the higher the QoE demands, the lower the network utilization. Hence, an operator has to tradeoff between user satisfaction and resource utilization. It appears to us that configuring $R' = 84$ seems a good tradeoff since only 5% of calls experience QoE below MOS ‘satisfied’ while roughly half of them are still in the range of MOS ‘some users dissatisfied’.

The conclusions of this work are manifold. On top of the list we found that the E-Model lends itself as a metric for QoE control by MBAC. The necessary computations are simple and do not add much burden on equipment. This opens the door to a new domain in VoIP QoS control, namely based on speech quality, the only reliable quality assessment method for VoIP. In support of this statement we found that the algorithm exhibits consistent and accurate behavior for a whole range of configurations. Probably the most intriguing conclusion, and somewhat specific to our setting, is that with a slightly modified measurement procedure we could apply the model on an aggregate level without compromising call level speech quality.

12.6 The Idea of an *R*-score-based Scheduler

In the previous part of the chapter the concept of a call admission control for VoIP services in a fixed WiMAX environment was introduced and its capability to keep the speech quality in terms of the *R*-score above a certain desired threshold was demonstrated by means of simulations. In this section, another novel concept is shortly discussed: to include measurements of the instantaneous speech quality for scheduler decisions in a mobile environment. The idea of this *R*-score-based scheduler is to handle temporary overload situations by maintaining a speech quality as good as possible. In a mobile packet-switched network, call admission control has to take into account both the on/off characteristics of VoIP calls and the temporal variations of the channel quality which again leads to a varying cell capacity. Guaranteeing a high speech quality by providing enough resources to transmit all packets even in temporary overload situations requires a rather conservative call admission control leading to a bad utilization. An aggressive call admission control, however, will achieve a high utilization while having periods of temporary overload where inevitably packets have to be dropped. The *R*-score scheduler will keep track of the instantaneous speech quality per call and consider this value in its decisions for scheduling and dropping packets. In the following, we describe and provide a very first sketch of an *R*-score-based scheduler and evaluate it for the DL scheduling of a single mobile WiMAX cell with a number of VoIP calls large enough to cause a temporary overload situation. The IEEE 802.16 MAC layer is strictly connection oriented and specifies detailed QoS parameters such as maximum latency, minimum reserved traffic rate and maximum sustained traffic rate. Hence, the first goal of the scheduler is to meet these requirements. However, we focus on the situation where the scheduler will not be able to fulfill these requirements for all ongoing VoIP connections, and since all VoIP connection have identical QoS parameters we need another metric for distinguishing the precedence of the packets. The *R*-score measurements will fill this gap and serve as an additional metric.

12.6.1 Scenario

We consider a single cell mobile WiMAX deployment without inter-cell interference. Mobiles are moving around in a 1.5 km square with the BS at the center. Sectorization is not considered. The BS operates in Frequency Division Duplex (FDD) mode with a Partially Used Subchannelization (PUSC) subchannel allocation on a 1.25 MHz band. Assuming a frame length of 5 ms, results in 104 slots subdivided into 4 subchannels with 26 slots each. Ignoring control traffic and assuming that only one-third of the slots is reserved for VoIP traffic, we end up with 35 slots for VoIP transport per frame. With the G711 speech codec

Wiley Proofs

Table 12.2

MCS	Bits/slots	Number of slots/packet	Number of blocks/packet	Required SNR
QPSK 1/2	48	35	6	8.7
QPSK 3/4	72	23	6	10.6
16QAM 1/2	96	18	6	15.8
16QAM 3/4	144	12	6	17.5
64QAM 1/2	144	12	6	21.9
64QAM 2/3	192	9	9	23.9
64QAM 3/4	216	8	8	25.1

and 20 ms framing, a VoIP packet has a size of 1648 bits including RTP, UDP, IP and MAC header. Depending on the Modulation and Coding Scheme (MCS) a VoIP packet occupies 35 slots with QPSK and 1/2 coding, and 8 slots with 64QAM and 3/4 coding. QPSK with 1/2 coding is the most robust MCS, repetition coding is not considered. The BS receives feedback on the channel quality per mobile in terms of the mean Signal-to-Noise Ratio (SNR) averaged over all subcarriers. In the simulation evaluation, we also use the mean SNR for determining frame errors on an Additive White Gaussian Noise (AWGN) channel. Consequently, in the simulation the BS has perfect knowledge of the channel without feedback. The MCS is selected in order to keep the loss probability for an entire VoIP packet below 1%. Since a VoIP packet does not fit into a single coding block, it comprises several coding blocks and, consequently, the product of the frame error rates of the coding blocks has to be below the target packet loss probability of 1%. The rather low packet loss rate of 1% is required since unacknowledged mode is chosen for VoIP traffic transport. Table 12.2 gives an overview of **Q37** the MCS with required SNR, number of coding blocks and number of slots.

The task of the scheduler is now to decide which packets to transmit in a frame. We do not support fragmentation or packing so a VoIP packet corresponds to a MAC PDU and has to be transported as a whole. Consequently, one frame is just enough to transmit a single QPSK 1/2 VoIP packet. In order to avoid the transmission of too delayed packets we drop packets after a threshold t_{drop}.

12.6.2 The Most Simple R-Score Scheduler

The term R-score scheduler relates to any scheduler that uses concurrent measurements of the R-score for making its scheduling decisions. In general, there are two different ways how the BS might obtain R-score measurements: first, the BS can measure the R-score based on its own transmissions. Second, the mobile can measure the R-score based on arriving and missing packets, and signal the R-score to the BS. Both kinds of R-score measurements are based on a series of received and lost packets as described in Section 12.4.2. In fact, the computational complexity of continuously updating the R-score is very low since it is mostly just equal to increasing a counter. Only in cases of a phase transition is a little more complex operation required.

In the following, we estimate the R-score at the BS. We assume that an intact and in order series of VoIP packets arrives at the BS. Then, we assume that a packet that is scheduled for transmission is received independent of its frame error rate. Only those packets are marked as lost that are not scheduled but dropped when the threshold t_{drop} is exceeded. Based on this sequence of transmitted and dropped packets the R-score is evaluated. Please note that this R-score is not equal to the R-score experienced at the MS but more optimistic since erroneous packet transmissions are not included.

The R-score scheduler sorts the packets in an ascending order of the measured R-score. The packets of the mobile with lowest R-score are transmitted first, that is, we could also denote the scheduler as *least R-score first*. We add two more criteria to this order of packets. First, the kth packet of one mobile receives a penalty of $k - 1$. Second, mobiles with a SNR of less than 6 dB are scheduled last since here the packet loss probability exceeds 20% even for the most robust MCS. Consider this example: mobile A has an R-score of 84.2, two packets to transmit, and an SNR of 15 dB. Mobile B has an R-score of 83.6, also two packets to transmit, and an SNR of 25.6 dB. Mobile C has an R-score of 65, one packet to transmit and an SNR of 4 dB. Then, the packets are scheduled in the following order:

Mobile	Packet	R-score	SNR	Metric
B	1	83.6	25.6	(0,83.6)
A	1	84.2	15	(0,84.2)
B	2	83.6	25.6	(0,84.6)
A	2	84.2	15	(0,85.2)
C	1	65	4.5	(1,65)

Please note that this scheduler is still a more or less channel oblivious scheduler since the only information it takes into account is that a channel is currently too bad to be used. The packets are reordered after every frame since the R-score or the SNR might have changed, new packets may have arrived or packets that are too old may have been dropped.

12.6.3 Performance Evaluation

In this section we intend to demonstrate the potential of R-score-based schedulers by comparing our *most simple R-score scheduler* with a channel-oblivious Earliest Deadline First (EDF) scheduler that in this scenario degrades to a simple First In First Out (FIFO) scheduler and a channel-aware MaxSNR scheduler that transmits the packets in order of a decreasing SNR. For the MaxSNR scheduler we also use a penalty of $k - 1$ for the kth packet of one mobile. For the EDF scheduler we also schedule packets with an SNR below 6 dB last.

In the following, we compare the three schedulers in a scenario with 25 mobiles and two different multi-path channel profiles: ITU Ped. A (PA) and ITU Ped. B (PB). The other simulation parameters are equal in the two scenarios and listed in the Table 12.3. **Q38**

For evaluating the quality of a scheduler we observe the distribution of the R-scores experienced at the subscriber stations at the end of a 60 s simulation run. A scheduler works well if no or only few VoIP calls experience bad quality. This means that we want to achieve a homogeneous R-score which is as high as possible. Let us first study the R-score in a scenario where only a single user is present and a VoIP packet is scheduled as soon as it arrives without considering the channel quality. We observe the resulting R-score with 30 independent mean SNR traces for both the ITU Ped. A and the ITU Ped. B channels. Figure 12.10 shows the CDF of the R-scores that are achieved for single users without any capacity constraints. They represent the upper bound of what is achievable by the schedulers. We can see that the PA

Table 12.3

Parameter	Value
Shadow fading	
Standard deviation	$\sigma = 8$ dB
Decorrelation distance	$X_c = 50$ m
Mobile velocity	$v = 3$ km hour^{-1}
Path loss	COST231
$PL(d) = 147.06 + 35.74 * \log_1 0(d[\text{km}])$	
Transmit power	$T_x = 20$ W
Antenna gain	$G = 10$ dB
VoIP call	
On phase	Exp(300 ms)
Off phase	Exp(600 ms)

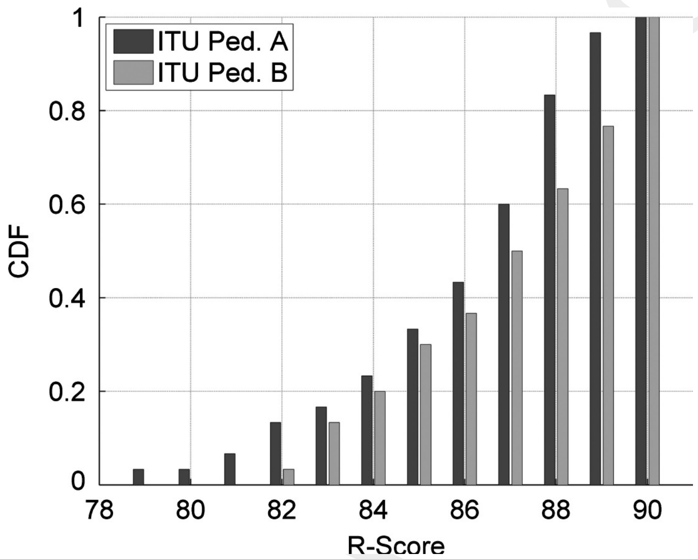
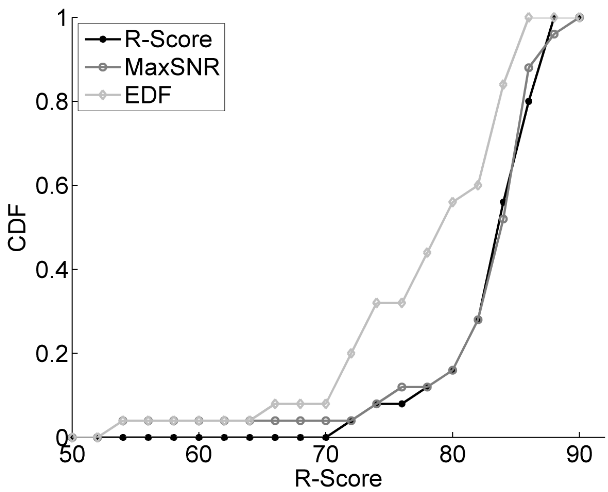


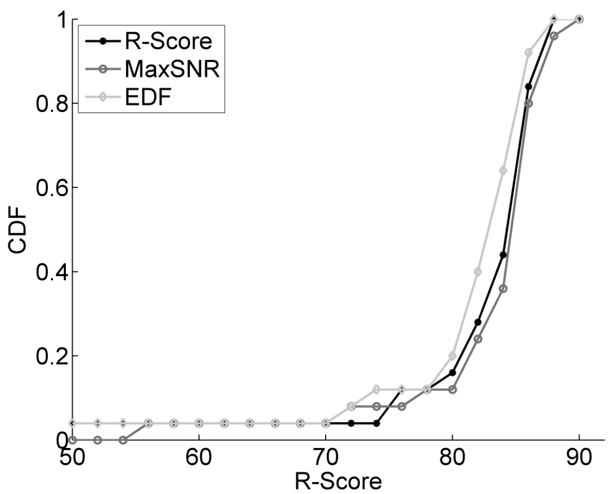
Figure 12.10 Achievable R-score without capacity constraints.

profile leads to somewhat lower R-scores than the PB profile. In particular, with PB we have in 7 of the 30 cases a maximum R-score of 90 while we achieve this R-score only once with PA. Observing the low R-score values, PB has a minimum of 82 whereas PA may lead to values as low as 79. Still, all of these values correspond to an acceptable or good speech quality.

Now, let us compare the impact of the different schedulers when we have a system with light overload, such as with 25 mobiles. Figure 12.11 shows the CDF of the R-score for the three schedulers with Figure 12.11(a) presenting the results for the PA multi-path channel



(a)



(b)

Figure 12.11 Performance of the *R*-score scheduler in comparison with channel-oblivious EDF and channel-aware MaxSNR scheduler: (a) ITU Ped. A; (b) ITU Ped. B.

profile and Figure 12.11(b) presenting the results for the PB profile. First, we can observe that in the light overload situation the R-score degrades down to 50 for some of the users while others still experience an excellent speech quality with an R-score close to 90. This holds for both channel profiles. If we observe the performance of the three schedulers in the PB profile, we can see that the three schedulers show marginal differences: only the EDF scheduler performs a little worse than the others. Figure 12.11(a) confirms this observation. Here, the EDF scheduler is clearly worse that the R-score and MaxSNR scheduler. At first glance, we also note that MaxSNR and R-score scheduler show an almost identical performance as in the PB case. If we have a closer look, we detect that for the high R-scores the performance is the same. For the low R-scores, however, the MaxSNR yields one R-score of below 55 while the R-scores obtained by the R-score scheduler are all above 70. This means that at least in this special case the R-score scheduler outperforms the MaxSNR scheduler and is able to avoid the strong degradation of the speech quality of the worst call.

In general, we can state that the 'most simple R-score scheduler' seems to be clearly better than the channel-oblivious EDF or FIFO scheduler. In addition, the performance of the channel-oblivious R-score scheduler is comparable and at least in one case better than the performance of the channel-aware MaxSNR scheduler which is quite remarkable. As a summary from the first experiences with an R-score scheduler, we think that the results are quite promising and encourage the further development of QoE aware schedulers not only for VoIP but also for other services such as Video on Demand (VoD), etc. The current simulation results are of course only examples and have to be further confirmed. In particular, the impact of different parameters such as the shadowing model, the velocity, the dropping threshold t_{drop} and many more has to be investigated. Furthermore, the simulation scenario should be more realistic in the sense that, for example, we should not assume a perfect packet arrival process at the BS, perfect knowledge of the channel or the unacknowledged mode. On the other hand, we also feel that there is still a lot of potential for improving the scheduler by combining different metrics for the scheduling decision, in particular current speech quality, current channel quality and urgency of the packet.

12.7 Conclusion

In this chapter we have presented the idea to utilize concurrent measurements of the instantaneous QoE for an improved resource control. We showed two examples of this QoE-based resource control at the scenario of the VoIP service in a WiMAX network. In the first example, a measurement-based admission control scheme for a fixed WiMAX deployment was proposed. The admission decision depends on a certain threshold for the aggregated R-score of ongoing calls. In the second example, a scheduling scheme for the DL of a mobile WiMAX BS is proposed. The scheduler uses the R-score as ordering metric and transmits the packets in a least R-score first fashion.

In both examples the proposed R-score-based resource control schemes show great promise through preliminary results. A more sophisticated evaluation of their performance is required and there is also room for optimization. Furthermore, using the R-score as a QoE metric for speech quality based resource control is only the first and probably most simple example for QoE-based resource control. Of great interest in the future will be QoE-based resource control schemes for different service composites including speech, video or gaming applications.

References

Blake S., Black D., Carlson M., Davies E., Wang Z. and Weiss W. (1998) *An Architecture for Differentiated Services, Request for Comments 2475*, IETF.

Bohnert, T.M. and Monteiro, E. (2007) Multi-class measurement based admission control for a QoS framework with dynamic resource management. *Journal of Network and Systems Management,* **15**(2), 219–240.

Bohnert, T.M., Monteiro, E., Curado, M., Fonte, A., Moltchanov, D., Koucheryavy, Y. and Ries, M. (2007) Internet quality of service: a bigger picture. *Proceedings of the First OpenNet QoS Workshop 'Service Quality and IP Network Business: Filling the Gap'*, Diegem, Belgium.

Braden, R., Clark, D. and Shenker, S. (1994) Integrated services in the Internet architecture: An Overview, *Request for Comments 1633*, IETF.

Clark, A. (2001) Modeling the effects of burst packet loss and recency on subjective voice quality. *Proceedings of the IP Telephony Workshop*.

Cole, R.G. and Rosenbluth, J.H. (2001) Voice over ip performance monitoring. *ACM SIGCOMM Computer Communication Review*, **31**(2), 9–24.

Cox, R. and Perkins, R. (1999) Results of a Subjective Listening Test for G.711 with Frame Erasure Concealment, *Committee contribution T1A1.7/99-016*.

Crawley, E., Nair, R., Rajagopalan, B., and Sandick, H. (1998) A framework for QoS-based routing in the internet, *Request for Comments 2386*, IETF.

Davie, B., Chany, A., Baker, F., Bennet, J.C.R., Benson, K., Bedeuc, J.Y.L., Chui, A., Courtney, W., Davari, S., Firoiu, V., Kalmanek, C., Ramakrishnan, K. and Stiliadis, D. (2002) An Expedited Forwarding PHB (Per-Hop Behavior), *Request for Comments 3246*, IETF.

Eklund, C., Marks, R.B., Ponnuswamy, S., Stanwood, K.L. and van Waes, N.J.M. (2006) Wireless-MAN Inside the IEEE 802.16 Standard for wireless metropolitan networks. *IEEE Standards Wireless Networks Series* (1st edn). IEEE Press, Piscataway, NJ.

Gozdecki, J., Jajszczyk, A. and Stankiewicz, R. (2003) Quality of service terminology in IP networks. *IEEE Communications Magazine* **41**(3), 153–159.

Hardy W.C. (2001) *QoS Measurement and Evaluation of Telecommunications Quality of Service*. John Wiley & Sons Ltd, Chichester.

Hoßfeld, T. and Binzenhöfer, A. (2008) Analysis of skype voIP traffic in UMTS: End-to-end QoS and QoE measurements. *Computer Networks*.

Hoßfeld, T., Hock, D., Tran-Gia, P., Tutschku, K. and Fiedler, M. (2008) Testing the IQX hypothesis for exponential interdependency between QoS and QoE of voice codecs iLBC and g.711. *Proceedings of the 18th ITC Specialist Seminar on Quality of Experience*, Karlskrona, Sweden.

Hoßfeld, T., Tran-Gia, P. and Fiedler, M. (2007) Quantification of quality of experience for edge-based applications. *Proceedings of the 20th International Teletraffic Congress (ITC20)*, Ottawa, Canada.

IEEE (2005) Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Amendment 2, *IEEE Standard for Local and Metropolitan Area Networks*, IEEE.

ITUT (1993) Terms and Definitions Related to Quality of Service and Network Performance Including Dependability, *ITU-T Rec. E.800*, ITU-T.

ITU-T (1996) *Objective Quality Measurement of Telephone-Band (300-3400 Hz) Speech Codecs*. International Telecommunication Union, Geneva, Switzerland.

ITU-T (1998) The Emodel: a computational model for use in transmission planning, *ITU-T Recommendation G.107,* ITU-T.

ITU-T (2001) *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-end Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs*, International Telecommunication Union, Geneva, Switzerland.

ITU-T (2004) Definition of quality of experience, *ITU-T Rec. D.197*, ITU-T.

Janssen, J., De Vleeschauwer, D. and Petit, G.H. (2000) Delay and distortion bounds for packetized voice calls of traditional pstn quality. *Proceedings of the First IP-Telephony Workshop (IPTel 2009)*, Berlin, pp. 105–110.

Janssen, J., De Vleeschauwer, D. D., Buechli, M. and Petit, G.H. (2002) Assessing voice quality in packet-based telephony. *IEEE Internet Computing*, p. 4856.

Jiang, W. and Schulzrinne, H. (2000) Analysis of on–off patterns in VoIP and their effect on voice traffic aggregation. *Proceedings of the Ninth IEEE International Conference on Computer Communication Networks*.

Kostas, T.J., Borella, M.S., Sidhu, I., Schuster. G.M., Brabiec, J. and Mahler, J. (1998) Real-time voice over packet-switched networks. *IEEE Network*, **12**, 1827.

Markopoulou, A.P., Tobagi, F.A. and Karam, M.J. (2003) Assessing the quality of voice communications over internet backbones. *IEEE Transactions on Networking*, **11**(5), 747–760.

Meddahi, A., Afifl, H. and Zeghlache, D. (2003) Packet-E-model: E-model for wireless VoIP quality evaluation. *Proceedings of the 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications. (PIMRC 2003)*, vol. 3, pp. 2421–2425.

Mortier, R., Pratt, I., Clark, C. and Crosby, S. (2000) Implicit admission control. *IEEE Journal on Selected Areas in Communications*, **18**(12), 2629–2639.

Raake, A. (2006a) Short- and long-term packet loss behavior: towards speech quality prediction for arbitrary loss distributions. *IEEE Transactions on Audio, Speech, and Language Processing*.

Raake, A. (2006b) *Speech Quality of VoIP*, John Wiley & Sons Ltd, Chichester.

Sengupta, S., Chatterjee, M., Ganguly, S. and Izmailov, R. (2006) Improving *R*-score of VoIP streams over WiMAX. *Proceedings IEEE International Conference on Communications (ICC '06)*, vol. 2, Istanbul, Turkey, pp. 866–871.

Takahashi. A, Kurashima, A. and Yoshino, H. (2006) Objective assessment methodology for estimating conversational quality of VoIP. *IEEE Transactions on Audio, Speech and Language Processing* **14**(6), 1984–1993.

Teitelbaum, B. and Shalunov, S. (2002) Why premium IP service has not deployed (and probably never will), *Technical Report, Internet2.*

WEIRD Consortium (2007) D.2.2 System Scenarios, Business Models and System Requirements (version 2), *Project Deliverable*, WEIRD Consortium.

Watson, A. and Sasse, M.A. (1998) Measuring perceived quality of speech and video in multimedia conferencing applications. *ACM Multimedia*, 55–60.

Q39

Marked proofs