CHANNEL-AWARE SCHEDULERS FOR VOIP AND MPEG4 FOR WIRELESS COMMUNICATIONS

Ana Cristina Costa Aguiar (1), Christian Hoene (2), Jirka Klaue (3), Holger Karl (4), Horst Miesmer (5) and Adam Wolisz (6)

(1 – 4, 6) Telecommunication Networks Group, Technical University of Berlin, Einsteinufer 25, 10587 Berlin, Germany, Email: aaguiar, hoene, jklaue, karl, wolisz@tkn.tu-berlin.de
(5) ICM N PG SP RC TM, Siemens AG, Mü nchen, Germany, horst.miesmer@siemens.com

Abstract

When transmitting multimedia applications over a wireless channel, multimedia traffic is faced with the error-prone, time-varying nature of wireless communication. One means to improve this situation is to adapt the transmission decisions, e.g., time to transmit or parameters like transmission power or FEC to the current, actual state of the wireless channel. Also, semantic-aware adaptation techniques have proven to achieve important quality gains. However, these approaches require modifications to existing applications.
In this paper, we investigate how the combination of channel knowledge with different deadlines impacts the perceived quality of voice over IP and MPEG4 encoded video streams. We show that such channel-aware scheduling alone, without requiring any changes to existing applications or protocols, can considerably improve multimedia quality while simultaneously reducing the load on the wireless channel. For voice over IP in particular, we characterise a trade-off between delay and loss rate that optimises the perceived quality.

1 Introduction

Multimedia applications enjoy increasing popularity among wired and wireless users. These applications require the timely and correct delivery of data over a communication channel. But the *wireless* channel is error-prone and varies with time. Multimedia applications, which were developed for wired channels, cannot easily cope with such channel conditions. As a consequence, the user-perceived quality is low when the user is connected over a wireless channel. This necessarily leads to the question of how to improve the quality of multimedia applications over wireless channels without redesigning existing applications.
A popular approach to improve the quality of wireless communications is to increase the reliability of the wireless links using conservative modulations or error correction, resulting in low data rates. Since the error characteristics of the wireless channel show high variability, static techniques either are too conservative and waste bandwidth or are not strong enough to maintain a satisfying quality level when the channel is in bad condition. The use of channel-adaptive techniques is thus becoming more and more popular: information about the state of the channel is used to adapt transmission parameters like modulation or FEC, so that only the currently necessary amount of overhead is used.
Since multimedia traffic is loss-tolerant up to a certain degree but it is sensitive to delay, control of packet loss and delay can be used to improve perceived quality.
Furthermore, for multimedia traffic channel-adaptive techniques can be combined with semantic-aware adaptation to the varying bandwidth demands (Chou & Miao, 2001). In this work, however, we are interested in the gains that can be obtained fom using channel information alone. The advantage of this approach is that it is completely transparent; it requires no changes to legacy applications.
This paper looks at the possibilities of leveraging knowledge about the (estimated) future channel state to improve the *perceived* quality of multimedia applications by postponing packets until they can be correctly transmitted. As case studies, we use the two most important multimedia applications, voice over IP and MPEG4 video. We will use two different scheduling policies that adapt the transmission of a packet to the channel state. We will then investigate how these scheduling policies impact the perceived quality of VoIP and MPEG data streams.

2 Channel-aware Scheduling for Multimedia Applications

Channel-aware scheduling is a technique that adapts the transmission start time of a packet to the channel condition.[1] Sending packets over a channel in a bad state is avoided, as the data would get lost anyway. To make such decisions, knowledge about future channel state is needed. This knowledge can be obtained from channel prediction algorithms based on

[1] Other adaptations (e.g., transmission power) are also possible but not investigated here.

mathematical models (Hu et al., 2000) (Sternad et al., 2001) or on heuristics (Fragouli et al., 1998) (Gomez et al., 1999) (Lu et al., 2000). In this paper, we simplify this problem by assuming perfect knowledge about *future* channel state to choose the transmission start time of the packet so that packets do not suffer errors.

2.1 System Model

The system model used for the study of possible schedulers is illustrated in Figure 1. A source generates packets that travel through the Internet where they are randomly delayed but not dropped. When the packets arrive at the node before the wireless link they are put into a queue and later handed on to the scheduler. The scheduler requests information about future channel state from the predictor and uses this information to decide, according to the scheduling policy, when to send the packet over the wireless channel.

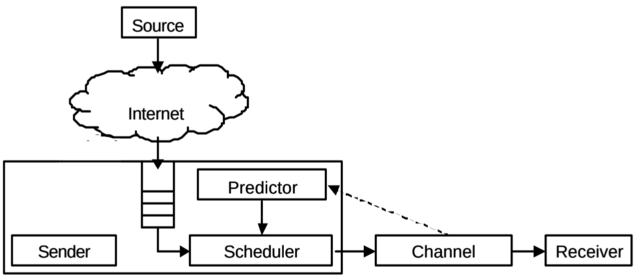


Figure 1 System Model

2.2 Scheduling Policies

As we are dealing with multimedia applications, meeting deadlines is a more important issue than not having any losses. This can be achieved by using deadlines for the transmission of a packet. We therefore propose the following scheduling approaches:

- The scheduler ignores deadlines and simply waits for a transmission start time which does not cause any errors in the packet— the *never drop* policy.
- The scheduler looks for a perfect (error-free) transmission start time up to a deadline and drops the packet if none is found; otherwise, the earliest such time is used — the *drop at deadline* policy.

The drop at deadline policy has two variants:

- *Fixed deadline*, where a fixed value for the deadline is used for every packet;
- *Adaptive deadline*, where a new link-local deadline is calculated for every packet, trying to compensate for the delay it suffered on the Internet (we assume that the delay on the Internet can be calculated). This adaptive link-local deadline is then used for scheduling and is calculated as the difference between the maximum admissible delay minus the actual delay a packet has encountered in the internet backbone.

As a reference, we additionally use a *Blind* scheduler, which is oblivious of channel-state predictions and always sends a packet immediately when it arrives at the scheduler (or immediately when the channel is free again).

3 Evaluation

3.1 Load Models

We investigated the behaviour of the schedulers introduced in Section 2.2 for typical interactive multimedia applications: voice over IP (VoIP) and video conferencing respective video streaming (MPEG4 encoded).

3.1.1 Voice over IP

A speech sample with a length of 95,1s was transmitted. The sample is encoded by the reference implementation of ITU G.729, which is a low-rate speech coding. This reference implementation includes an encoder, decoder and loss concealment algorithm. G.729 has a coding rate of 8 kbps and generates a frame each 10 ms. Two speech frames form one VoIP packet. The VoIP packets are transmitted using RTP and UDP, so that the packets at the link layer include the headers of these protocols as well as the 20 Byte IP headers.

3.1.2 MPEG4

Two typical video scenarios were studied:

- Video conference: 3 minute low motion video, encoded at constant bit rate of 256 kbps with 256 Byte packets using only I- and P-frames;
- Streaming live-event: 3 minute high motion video, encoded at variable bit rate (average of 920 kbps) with 1024 Byte packets using I-, P-, and B- frames.

These packet lengths include the UDP and IP headers.

3.2 Parameters

The evaluation-relevant parts of the system model are the Internet delay, the wireless channel, and the prediction model.

First, the **Internet delay** of each packet was randomly generated from a shifted Gamma distribution with shape factor 2.5 and scale factor 1 (Mukherjee, 1992) (Corlett & Pullin, 2002). The shift of the Gamma distribution depends on the distance between the sending application and the wireless node, and several values were used, from 50ms to 300ms. Packets are not lost on the Internet and they arrive at the wireless

sender in order. This may be unrealistic, but the effects of loss and reordering on the backbone do not have an effect on the performance of the link layer, which is the subject of this paper.
Second, the **error-prone channel** is a bit-level Gilbert-Elliot model generated from a Rayleigh fading channel according to (Wang & Moayeri, 1995), assuming a raw channel capacity of 2 Mbps (parameters in Table 1).

Table 1 Parameters of the Gilbert-Elliot channel and of a corresponding Rayleigh fading channel (e_G, e_B: error probability in the good/bad channel state; μ_G, μ_B: average duration of the good/bad channel state).

Rayleigh fading model		Gilbert-Elliot Model	
F	2,4 GHz	e_G	0
V	6 Km/h	e_B	0.02848
SNR_{avg}	12 dB	μ_G	53.21 ms
$SNR_{threshold}$	10 dB	μ_B	17.29 ms
Modulation	BPSK	p_e	$20.76\ 10^{-3}$

Third, using a real channel prediction algorithm is both cumbersome and resource-intensive. As a simplification, we used instead a **prediction model** (Aguiar et al., 2003). This model can express the inherent inaccuracies of a real predictor; however, as this paper is mostly concerned with the impact channel-adaptability on perceived quality, we use a perfectly accurate prediction of future channel state.

3.3 Metrics

The metric used for the performance evaluation was perceived user quality.

3.3.1 VoIP

We used psycho-acoustic models that evaluate the perceptual telephone and speech quality and take into account speech quality and delay to estimate the performance of VoIP transmission.
More precisely, perceptual quality assessment has to take into account the complete end-to-end transmission path. For example, the quality largely depends on the playout buffer scheme, which is based on a particular system implementation. Until now no standard approach has been agreed upon. Therefore, we simulated multiple fixed playout buffer schemes, each with a different absolute, maximum delay threshold. Packets that arrived too late to be played out after this maximum acceptable transmission delay were dropped. Packets dropped by either the wireless link or the playout buffer decrease the subjective speech quality.
To approximate the subjective speech quality (which could only be derived by cost-intensive experiments with many people) with an objective metric, we applied the PESQ algorithm (ITU P.862).
PESQ compares an original speech sample with its degraded, transmitted version to estimate the Mean Opinion Score (MOS). The MOS value ranges from 1 (bad) to 5 (excellent speech quality). Our original speech sample had a MOS value of 3.77 due to coding distortions.
Finally, we estimated the quality of a telephony call with the ITU E-Model (ITU G.107). The E-Model's primary output is the R-Factor, which ranges from 0=bad over 70=toll-quality to 100=excellent. To calculate the R-Factor, the E-Model takes into account the speech quality, the mouth-to-ear delay and other factors that influence the quality (e.g. echoes and loudness).

3.3.2 MPEG4

Decisive for the video quality assessment is, like for voice, the quality perceived by the human observer — the subjective video quality. Although it is standardised how to determine the subjective quality — the MOS (ITU-T, 1996) (ISO-IEC, 1996) (ANSI, 1996) — it is extremely costly: highly time consuming, high manpower requirements and special equipment needed. Therefore, again akin to voice, other methods have been developed to emulate the subjective impression for digital video quality assessment (Berts & Persson, 1998) (Wolf and Pinson, 2002). The most widespread of these objective methods is the PSNR. This metric is calculated image by image and can be mapped to MOS using a heuristic (Klaue et al., 2002), (ITU-T, 1996), (Ohm, 1999). We used the overall percentage of frames within each MOS category as a metric for quality of video transmission. To illustrate the tradeoff between loss and delay on the wireless channel we use the cumulative distribution function (CDF) of the end-to-end MPEG-4 frame delays.

4 Results

4.1 Voice

The perceived quality (R Factor) results obtained for the VoIP call are shown in Figure 2. It displays the relation between the size of a playout buffer and the quality of telephony with respect to different deadline strategies. An additional fixed system delay, which includes Internet backbone and terminal delays, of 300 ms is added to the transmission delay.
Interestingly, the quality increases with the length of the deadline. The best possible quality is achieved at a link-local deadline between 90 and 100 ms. Thus, even though telephony has tight real-time constraints, the trade-off between loss and delay leads to the requirement to avoid losses instead of reducing delays. However, we can also see in
Figure 3 that using no transmission deadline is not always beneficial. The fixed or adaptive strategies perform equally well.

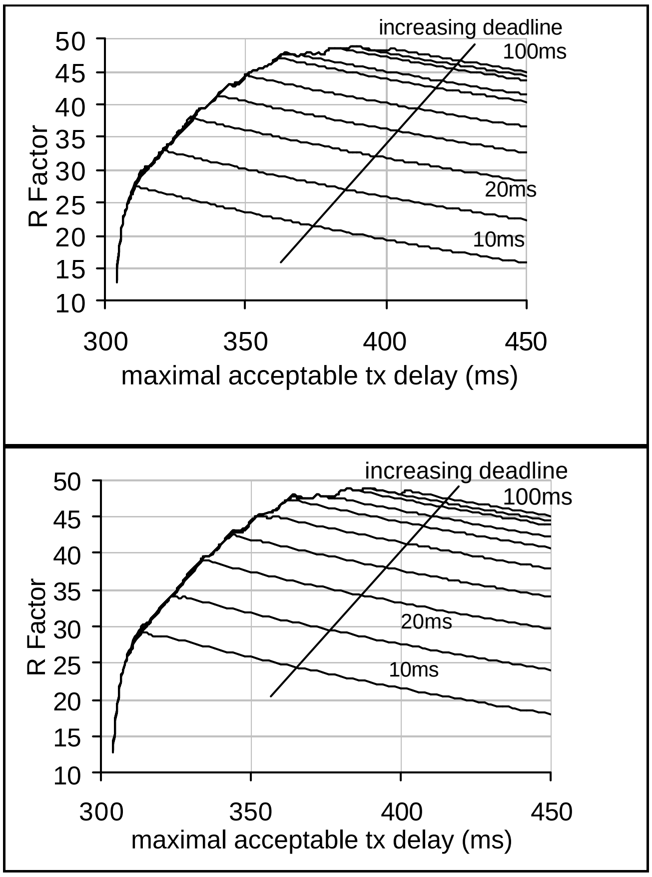


Figure 2 Telephone quality vs. maximal acceptable transmission delay using adaptive (top) and fixed scheduling (bottom) with different deadlines

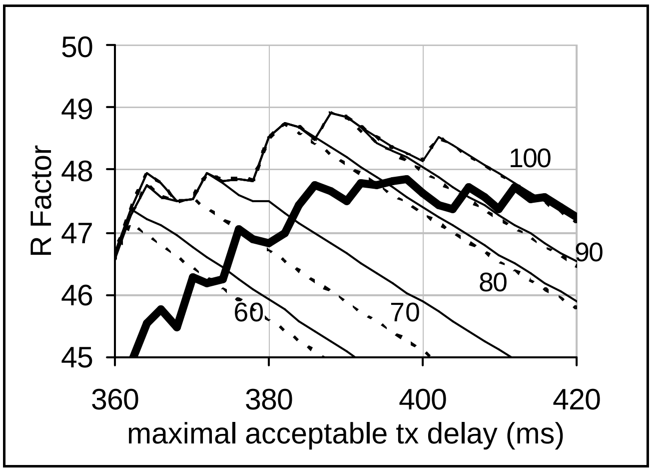


Figure 3 Zoom view of figure 2: Fixed, adaptive and no-deadline scheduling (normal, dashed, and bold line respectively)

4.2 Video

The (approximated) MOS results obtained for the MPEG4 video transmissions are shown in Figure 4. This figure also shows the quality achieved by the blind scheduler and the quality of the encoded video before transmission. These two quality grades should be used as comparison.

A big improvement in the amount of frames with excellent and good MOS, decisive for the perceived quality, is achieved using channel-aware scheduling (compared with the blind policy). The improvement is the bigger the longer the deadline is. This is due to the fact that the longer the deadline, the more often a packet can be postponed past the end of a bad channel state and thus be correctly transmitted.

The low motion video achieves an average MOS similar to the encoded video at a link-local deadline of 40 ms, whereas the high motion video only reaches a quality similar to the original encoded video at deadline above 90 ms. The reason is that the low motion video uses shorter packets, for which error-free transmission times can be found more easily.

The biggest gain is achieved by increasing very tight deadlines; for longer deadlines, there is a diminishing return.

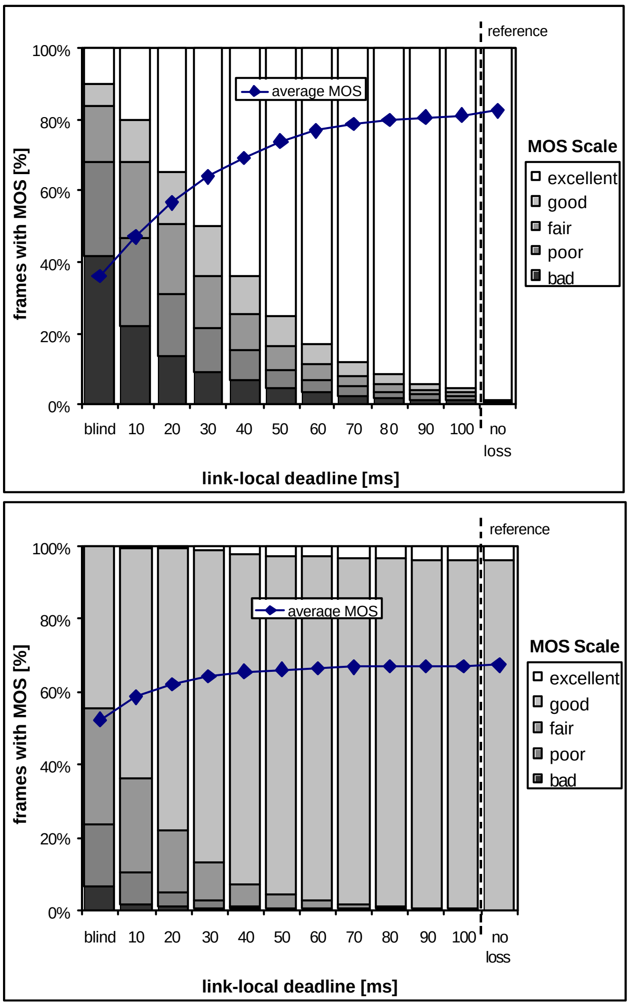
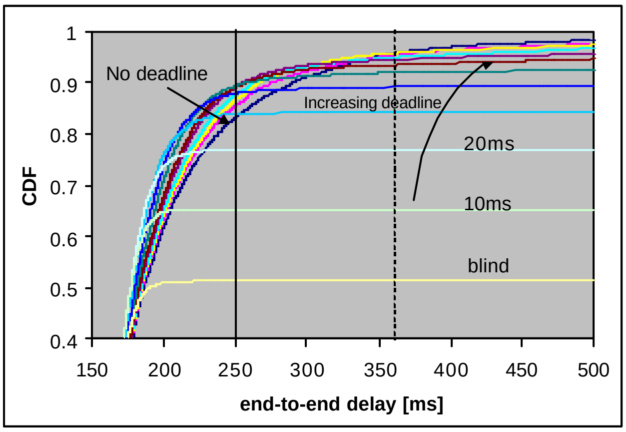


Figure 4 MOS of video transmissions with blind and fixed-deadline scheduling (top: high-motion video, bottom: low-motion video)

Figure 5 shows the cumulative distribution functions of the end-to-end *video frame* delays (an internet backbone delay of 150ms was used). Thus far, we have not taken into account the length of the playout buffer of the application at the receiver, which will drop encoded *video frames* (and not only single packets), which arrive (completely or in part) too late. This means that the playout buffer will discard frames, which arrive too late, changing the loss results seen before.



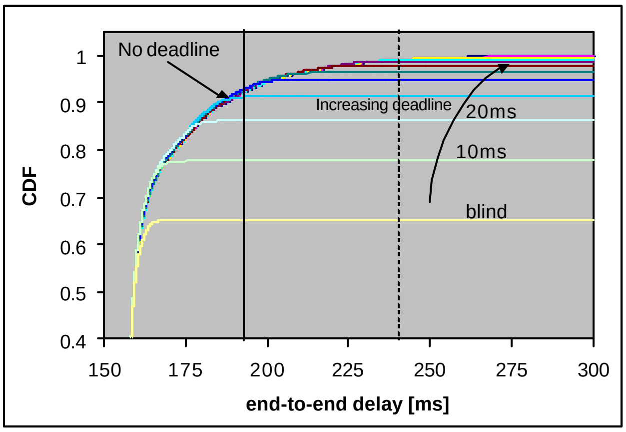


Figure 5 Cumulative Distribution Function (CDF) of the video frame end-to-end delays (top: high motion, bottom: low motion)

The CDF results for the high motion video show that for a 250 ms playout buffer, the no-deadline scheduler would cause more frames to be dropped off the buffer, due to the longer waiting times, resulting in a worse perceived quality. In fact, this would happen for every deadline beyond 50 ms at this playout buffer length. On the other hand, if the length of the playout buffer is bigger than 360 ms (dashed line), longer deadlines achieve the best results. The reason is that the long packets of the high motion video sometimes take very long to be transmitted error-free.

The CDF results for the low motion video show a similar behaviour: for a playout buffer length of 190 ms, increasing the deadline beyond 40 ms does not improve the number of lost frames; for playout buffer values longer than 240 ms, a longer deadline is always better.

5 Conclusions

Channel-adaptive scheduling policies based on channel prediction alone can greatly improve the user-perceived quality of multimedia applications, compared to not taking channel condition into account. These gains were achieved without using any ARQ or coding overhead.

For VoIP applications, the telephony quality (expressed by the R-factor) can be substantially increased. It is also clear that, although delay is of great importance, dropping a greater amount of frames for keeping low delays does not necessarily lead to an improvement in the perceived telephony quality. Since the perceived quality for a certain link-local deadline value depends on the length of the playout buffer, it seems to be advisable to set the scheduler deadline according to that value, what can be done at the time the VoIP telephony flow is established.

For video applications, the possibilities for improvement depend on how the video stream to be transmitted has been encoded, especially on packet size, but the percentage of received frames with good or excellent MOS can be substantially increased with channel-adaptive scheduling. Using a playout buffer of fixed length will cause dropping of video frames that have been subjected at the scheduler to long waiting for error-free transmission, limiting the gain, which can be achieved by using longer, and longer deadlines. Similarly to VoIP telephony, setting the link-local deadline of the scheduler according to the length of the playout buffer and size of the packets is advisable. How this can be done needs further research.

The use of adaptive playout buffers is becoming state-of-the-art, and this will change the dropping behaviour of the receiving applications. Further studies are needed to assess the gain, which can be achieved by channel-aware schedulers in these cases, as well as how the link-local deadline influences the perceived quality.

As was mentioned in Section 1, the additional use of semantic information should allow further improvements. Further work focuses on combining the use of channel knowledge and semantic information. How these two approaches can be optimally joined is, however, a very challenging issue.

6 Acknowledgements

This work has been partially supported by Siemens AG ICM, by the Portuguese government under the grant SFRH/BD/4610/2001, and by Deutsche Forschungsgemeinschaft (DFG) via the priority program "Adaptability in heterogeneous communication networks with wireless access".

We would also like to thank the reviewers for their constructive comments.

7 References

(Berts & Persson, 1998) Berts, J. and Persson, A. (1998). *Objective and subjective quality assessment of compressed digital video sequences.* Master's thesis, Chalmers University of Technology.

(Fragouli et al., 1998) Fragouli, C. and Sivaraman, V. and Srivastava, M. (1998). *Controlled multimedia wireless link sharing via enhanced class-based queueing with channel-state-dependent packet scheduling*, in *INFOCOM*. IEEE, 1998.

(Gomez et al., 1999) Gomez, J. and Campbell, A.T. and Morikawa, H. (1999). *The havana framework: Supporting application and channel dependent qos in wireless networks*, on *Proceedings of the 7th International Conference on Network Protocols (ICNP'99)*.

(Hu et al., 2000) Hu, S. and Duel-Hallen, A. and Hallen, H. (2000). *Long-range prediction of fading signals*. IEEE Signal Processing Magazine, 17, 62–75.

(ISO-IEC, 1996) ISO-IEC (1996) SO-IEC/JTC1/SC29/WG11. *Evaluation methods and procedures for july mpeg-4 tests*.

(Lu et al., 2000) Lu, S. and Nandagopal, T. and Bharghavan, V. (2000). *Design and analysis of an algorithm for fair service in error prone channels*. ACM/Baltzer Wireless Networks Journal, 6(4), 323-343.

(Ohm, 1999) Ohm, J.-R. (1999). *Skript zur Vorlesung Bildsignalverarbeitung für Multimedia-Systeme*. Institut für Nachrichtentechnik und Theoretische Elektrotechnik, TU Berlin.

(ITU-T, 1996) ITU-T (1996). *Subjective video quality assessment methods for multimedia applications*, ITU-T Recommendation P.910.

(Sternad et al., 2001) Sternad, M. and Ekman, T. and Ahln, A. (2001). *Power prediction on broadband channels*, in *Proceedings of the 53rd IEEE Vehicular Technology Conference — VTC 2001 Spring*, vol. 4, 2328–2332.

(ANSI, 1996) ANSI (1996). *Digital transport of video teleconferencing / video telephony signals - video test scenes for subjective and objective performance assessment*, in ANSI T1.801.01-1996.

(Wang & Moayeri, 1995) Wang, H.S. and Moayeri, N. (1995). *Finite-state markov channel—a useful model for radio communication channels*, in *Transactions on Vehicular Technology*, 44,163–171.

(Wolf and Pinson, 2002) Wolf, S. and Pinson, M. (2002). *Video quality measurement techniques*. In Technical Report 02-392, U.S. Department of Commerce, NTIA.

(Mukherjee, 1992) Mukherjee, A. (1992). *On the dynamics and significance of low frequency components of internet load*, in Technical Report MS-CIS-92-83/DSL-12, Computer and Information Science Department, University of Pennsylvania.

(Corlett & Pullin, 2002) Corlett, A. and Pullin, D. (2002). *Statistics of one-way Internet packet delays*, in IETF Networking Group Internet Draft (work in progress).

(Aguiar et al, 2003) Aguiar, A. and Karl, H. and Miesmer, H. and Wolisz, A. (2003). *A framework for evaluating the effects of prediction accuracy on the performance of channel-adaptive techniques*, in Proceedings of International Conference on Wireless Communications.

(Klaue et al, 2003) Klaue, J. and Gross, J. and Karl, H. and Wolisz, A. (2003). *Semantic-aware link layer scheduling of MPEG-4 video streams in wireless systems*, in Proceedings of Application and Services in Wireless Networks.

(Chou & Miao, 2001) Chou, P. A. and Miao, Z (2001). *Rate-distortion optimized streaming of packetized media*, in Microsoft Research Technical Report MSR-TR-2001-35.