Application-driven Cross-layer Optimization for Mobile Multimedia Communication using a Common Application Layer Quality Metric

S. Khan, S. Duhovnikov, E. Steinbach
Media Technology Group
Technische Universität München
Munich, Germany

M. Sgroi, W. Kellerer
DoCoMo Communications Laboratories Europe GmbH
Future Networking Lab
Munich, Germany

ABSTRACT
This paper proposes a cross-layer optimization framework that provides efficient allocation of wireless network resources across multiple types of applications to maximize network capacity and user satisfaction. We define a novel optimization scheme based on the Mean Opinion Score (MOS) as the unifying metric. Our experiments, applied to scenarios where users simultaneously run three types of applications, such as realtime voice, video conferencing and file download, confirm that MOS-based optimization leads to significant improvement in terms of user perceived quality when compared to throughput-based optimization.

Categories and Subject Descriptors
C.2.1 [**Computer Communication Networks**]: Network Architecture and Design – *Network communications, Wireless communication, Packet-switching network.*

General Terms: Design

Keywords
Cross-layer optimization, wireless multimedia communication, Mean Opinion Score, multi-user resource allocation.

1. INTRODUCTION
Optimization of network architectures is critical to achieve maximal network capacity and to provide high quality services to the largest possible number of users. In common scenarios, multiple users share the wireless medium and run rather diverse applications such as video, voice and file download. Optimizing the allocation of resources across all the users and all the applications allows increasing network capacity and maximizing the satisfaction of the users.

So far application-driven cross-layer optimization (CLO) has been studied for systems supporting only one application [1] … [4]. However, in practice the users sharing the wireless medium, e.g., in a cell, usually run different applications. User satisfaction translates into a different set of requirements for each type of application. Furthermore, the impact of losses on the user-perceived quality is also very application dependent. Jointly optimizing the system for different users and applications requires: 1) defining a common metric that quantifies the satisfaction of the users for the service delivery and 2) mapping network and application parameters onto this metric.

The challenge of optimization across multiple applications has been treated mainly in the form of throughput maximization [5], [6]. Maximizing throughput leads to optimum performance only for applications which are insensitive to delay and packet loss. Multimedia applications such as video and voice are highly sensitive to changes in data rate, delay and packet loss rate. Even the importance of a packet changes dynamically depending on the transmission history of previous packets. Due to these reasons, throughput maximization leads to performance which is usually not optimal with respect to user perceived quality for multimedia applications.

Mean Opinion Score (MOS) was originally proposed for voice quality assessment and provides a numerical measure of the quality of human speech at the destination end of the circuit. The scheme uses subjective tests (opinionated scores) that are mathematically averaged to obtain a quantitative indicator of the system performance. To determine MOS, a number of listeners rate the quality of test sentences read aloud over the communications circuit by a speaker. A listener gives each sentence a rating as follows: (1) bad; (2) poor; (3) fair; (4) good; (5) excellent. The MOS is the arithmetic mean of all the individual scores.

The multi-application CLO approach proposed in this paper extends the use of MOS as a user-perceived quality metric to other applications, such as video, web browsing and file download. This enables us to optimize across applications using a common optimization metric. The objective function can be chosen, e.g., to be the average MOS of all the users competing for the wireless resources:

$$F(\tilde{\mathbf{x}}) = \frac{1}{K}\sum_{k=1}^{K} \lambda_k \cdot \mathrm{MOS}_k(\tilde{\mathbf{x}}) \qquad (1)$$

where $F(\tilde{\mathbf{x}})$ is the objective function with the cross-layer parameter tuple $\tilde{\mathbf{x}} \in \tilde{X}$. $\tilde{X}$ is the set of all possible parameter tuples abstracted from the protocol layers representing a set of candidate operation modes. λ_k is the relative importance of the user as determined by the service agreement between the user and the service provider. Although the MOS functions for different applications can be different, a linear combination, as in equation (1),

IWCMC'06, July 3–6, 2006, Vancouver, Brithis Columbia, Canada.

can be used because the range of the functions is the same, i.e., from 1 to 5. The decision of the optimizer can be expressed as:

$$\tilde{\mathbf{x}}_{opt} = \arg \max_{\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}} F(\tilde{\mathbf{x}}) \qquad (2)$$

where $\tilde{\mathbf{x}}_{opt}$ is the parameter tuple which maximizes the objective function. Once the optimizer has selected the optimal values of the parameters, it distributes them to all the individual layers which are responsible for translating them back into actual modes of operation.

In this work the abstracted parameters for the physical and data link layers are transmission rate and packet error probability for all users for all candidate modes of operation. For a detailed description of the principle of parameter abstraction and the formulation of objective functions for multi-user cross-layer optimization please refer to [1] … [3].

The proposed MOS-based optimization approach has several advantages with respect to previous work. First, compared to traditional techniques for multi-user diversity [7] it allows us to directly relate network parameters, such as rate (R), packet error probability (PEP) and delay to a user-perceived application quality metric such as MOS. Second, compared to the application-driven cross-layer optimization described in [2], [3] it allows us to further maximize the optimization gain taking advantage of the diversity not only across multiple users running the same application, but also across users running different applications. Our experiments applied to scenarios including multiple concurrent video, voice and file download applications show that MOS-based optimization significantly outperforms throughput-based optimization.

This paper is arranged as follows. In Section 2 we describe MOS functions for three different applications, namely voice, FTP and video conferencing. In Section 3 we give a detailed description of our multi-application cross-layer optimization framework. Section 4 gives an overview of our simulation setup that is used to compare our approach with throughput maximization. Section 5 presents our experimental results.

2. MEAN OPINION SCORE (MOS)

The objective function in (1) requires the mapping of transmission characteristics (in our case transmission rate and packet error probability) to MOS for different applications. We now describe this mapping for voice, FTP and video conferencing applications.

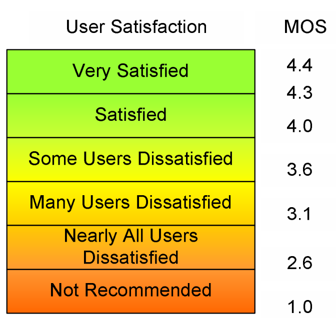

Figure 1: Relation between MOS and user satisfaction

2.1 Voice

The traditional method of determining voice quality is to conduct subjective tests with panels of human listeners. The results of these tests are averaged to give MOS but such tests are expensive and are impractical for online voice quality assessment. For this reason the ITU has standardized a new model, Perceptual Evaluation of Speech Quality (PESQ) [8], an algorithm that predicts with high correlation the quality scores that would be given in a typical subjective test. This is done by making an intrusive test and processing the test signals through PESQ.

PESQ measures one-way voice quality: a signal is injected into the system under test, and the degraded output is compared by PESQ with the input (reference) signal. Mapping between MOS and user satisfaction is presented in Figure 1.

The PESQ algorithm is computationally too expensive to be used in real-time scenarios. To solve this problem we propose a model to estimate MOS as a function of R and PEP. The available rate determines the voice codec that can be used. In Figure 2 we show experimental curves for MOS estimation as a function of PEP for different voice codecs. The curves are drawn using an average over a large number of voice samples and channel realizations (packet loss patterns). These curves can be stored in the base station for every codec that is supported. If transcoding from an unsupported codec is required, such curves have to be signaled to the base station as side information.

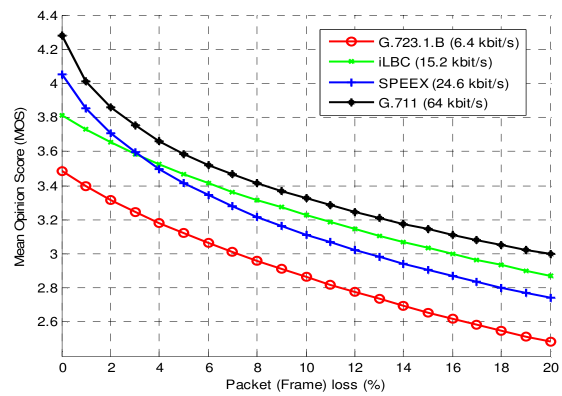

Figure 2: PESQ-based MOS vs. packet error probability (PEP) for different voice codecs (different transmission rate)

2.2 FTP

To estimate FTP user satisfaction we use the logarithmic MOS-throughput relationship introduced in [9] which results from the assumption that the utility of an elastic traffic (e.g. FTP) is an increasing, strictly concave and continuously differentiable function of throughput. We assume that every user has subscribed for a given data rate and his satisfaction is characterized by the real rate he receives. The MOS is estimated based on the current rate R offered to the user by the system and packet error probability PEP:

$$MOS = a * \log_{10}[b * R * (1 - PEP)] \quad (3)$$

where a and b are determined from the maximum and minimum user perceived quality. If a user has subscribed for a specific rate $R_{service}$ and receives $R_{service}$, then in case of no packet loss user satisfaction on the MOS scale should be maximum, i.e. 4.5. On the other end, we define a minimum transmission rate (e.g. 0) and assign to it a MOS value of 1. Using the parameters a and b, we fit a logarithmic curve for the estimated MOS. Varying the actual transmission rate R and packet error probability PEP, this model results in the MOS estimation surface of Figure 3 for every user.

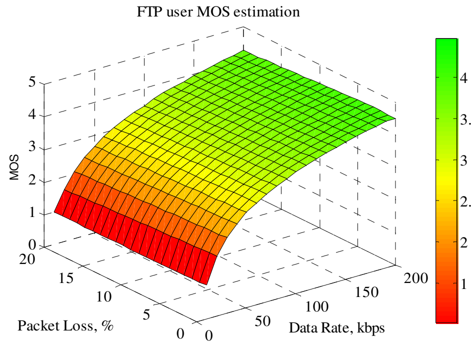

Figure 3: FTP user MOS estimation surface

2.3 Realtime video

To support videoconferencing in the wireless multimedia network we propose a simple model for evaluating the quality of a video material. We assume that we have all the information about the distortion caused by a slice loss and we evaluate the Peak Signal to Noise Ratio (PSNR) for different slice loss percentages. The model in this example is constructed for the Foreman video sequence, but can be easily extended for different videos and has to be sent as side information along with the video bitstream.

Encoding and decoding is performed with the H.264 JM 8.4 codec. The encoder is set to encode the first frame as an I-frame and all the following frames as P-frames. We assume nine slices per frame and in every frame, the macro-blocks of a single slice are intra coded (Figure 4).

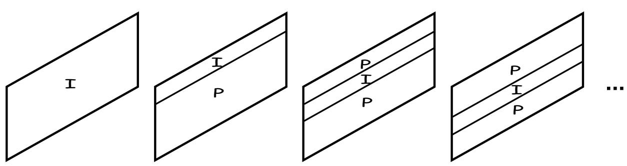

Figure 4: H.264 based encoding of video sequences for conversational video applications

This results in higher bit-rate, but also gives higher resilience against lost packets (slices). If a slice is lost, the effect of this loss will be washed out after a maximum of nine frames. The resulting average PSNR over all 400 frames for zero percent packet loss is 35.30 dB in our experiment.

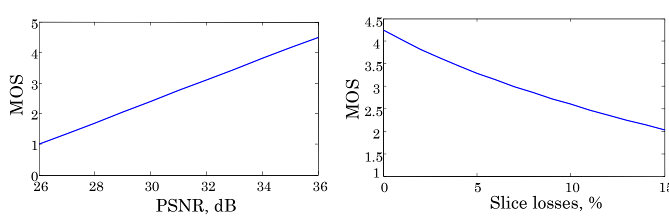

Figure 5: (a) MOS vs. PSNR (left) and (b) MOS vs. slice loss rate (right) for the Foreman video sequence

As shown in [10], PSNR has a reasonably high correlation with video quality. Figure 5(a) shows our assumed linear relation between the decoded average PSNR and user satisfaction measured with the metric MOS. Figure 5(b) shows the average MOS in case we have packet losses over the wireless channel. Every slice is encapsulated into one packet. Every % slice loss is simulated 1000 times with random slice loss patterns. The average decoded PSNR is computed over all the decoded frames.

3. MULTI-APPLICATION CROSS-LAYER OPTIMIZATION

3.1 Architecture

In [1], [2] we have proposed a cross-layer design architecture (Figure 6) with a component, called cross-layer optimizer (CLO), that periodically selects the optimal parameter settings of the different layers. The CLO uses abstractions of different layers and optimizes the assignment of resources to each user. In our work, the abstracted parameters from the lower layers are R and PEP for every user for all possible modes of operation.

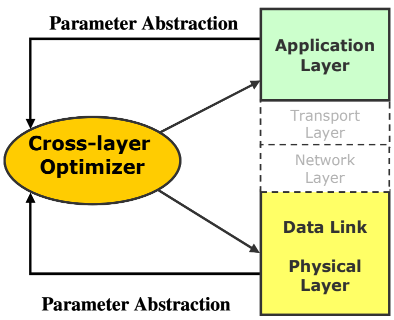

Figure 6: CLD Architecture

3.2 Optimization policy

As an example, we consider three types of users: U – requesting voice service, V - file download and W – videoconference. Depending on the type of application, the mobile users require different resources over the wireless channel. The available transmission rate depends on the modulation scheme, the channel code rate and the assigned share of the medium access. In our example a user requesting voice service may be served with different voice codecs (G.711, Speex, iLBC or G.723.1.B), his data may be encoded with different channel code rate 1/2, 1/3, or 1/4 and DBPSK or DQPSK modulation can be used. Every transmission policy gives different quality of service to the user and requires different amount of channel resources.

We create sets of transmission policies for every service. T_U is the set of transmission policies for voice service, T_V is the set of transmission policies for the file download service and T_W is the set of transmission policies for the video service.

3.3 Mean Opinion Score maximization

The goal of this optimization is to achieve maximum user satisfaction and fairness among the users. For every user, depending on the service, we define a decision variable for every transmission policy – whether this user is served with a given transmission policy or not. Consequently these decision variables are of boolean type, i.e., either the user transmits its information using this policy or not. For the voice users, we have decision variables u_{ij}, where "i" denotes the i-th user and "j" denotes the transmission policies available for the voice users. The next step is to associate an expected user QoS defined with MOS.

Mobile users in the wireless network have time-varying position, which results in variable SNR at the receiver. Based on the SNR, we compute an estimate of the PEP [11] for different modulation schemes (DBPSK and DQPSK) and different channel code rates, i.e., for all candidate transmission policies. A channel realization is generated and the estimation of the PEP is performed for all the transmission policies given the particular SNR at the receiver.

Our objective function for multi-user multi-application cross-layer optimization is defined in equation (4). A maximization of the sum of the MOS perceived by every user in our multimedia wireless network has to be achieved. The parameter λ is used to give higher priority to a given user and it is up to the network operator to choose its value.

Maximize

$$\sum_{i\in U}\sum_{j\in T_U}\lambda_{ui}u_{ij}E[MOS_{ij}]+\sum_{i\in V}\sum_{j\in T_V}\lambda_{vi}v_{ij}E[MOS_{ij}]+\sum_{i\in W}\sum_{j\in T_W}\lambda_{wi}w_{ij}E[MOS_{ij}] \quad (4)$$

Subject to:

$$\sum_{j\in T_U}u_{ij}=1\,,\quad \forall i\in U \quad (5)$$

$$\sum_{j\in T_V}v_{ij}=1\,,\quad \forall i\in V \quad (6)$$

$$\sum_{j\in T_W}w_{ij}=1\,,\quad \forall i\in W \quad (7)$$

$$\sum_{i\in U}\sum_{j\in T_U}r_{ij}u_{ij}+\sum_{i\in V}\sum_{j\in T_V}r_{ij}v_{ij}+\sum_{i\in W}\sum_{j\in T_W}r_{ij}w_{ij}\le TotalSymbolRate \quad (8)$$

In our example, every user must be associated with only one transmission rate, channel code rate and modulation scheme. The decision variables u_{ij}, v_{ij} and w_{ij} are of boolean type which leads to constraints (5), (6) and (7). The total available symbol rate for all the users is constrained to be less than the total symbol rate of the system. Every transmission policy has an associated symbol rate r_{ij} and the sum of all the chosen symbol rates of all the users must be less or equal to the total symbol rate. The above problem can be solved with a full search through the possible parameter space which has the worst case number of searches of $|T_U|^{K_U}|T_V|^{K_V}|T_W|^{K_W}$ where $|T_U|$, $|T_V|$ and $|T_W|$ are the number of transmission policies and K_U, K_V, K_W are the number of users of user classes U, V and W respectively.

The parameters λ_{ui}, λ_{vi}, λ_{wi} in (4) are inserted to ensure a fair allocation of resources. The optimizer finds a resource allocation which maximizes the user satisfaction based on MOS. In this case there is a possibility that even though the system performance is maximized, a given user is not satisfied. This could be caused by low receiver SNR and the optimizer can decide to allocate the resources to the other users. This contradicts with the fairness we are trying to offer to the users independent of their location. To solve this problem we propose to select the scaling coefficients based on the history of the user estimated MOS. On every rate allocation procedure, we find the user with the maximum average of the estimated MOS for the previous steps. Let us assume that we are at rate allocation step "N" and we have K users in the system. The value of the maximum perceived MOS by a single user is found by

$$MaxMOS_N=\frac{1}{N-1}\max\left(\sum_{n=1}^{N-1}MOS_{1n};\sum_{n=1}^{N-1}MOS_{2n};\ldots;\sum_{n=1}^{N-1}MOS_{Kn}\right) \quad (9)$$

The scaling coefficient for every user is calculated with

$$\lambda_{kN}=\frac{MaxMOS_N}{\frac{1}{N-1}\sum_{n=1}^{N-1}MOS_{kn}},\quad k=1\ldots K \quad (10)$$

The user with the maximum perceived MOS has a scaling coefficient of one. The other users have scaling coefficients in the range [1; 4.5], because the denominator is also bounded in the interval [1; $MaxMOS_N$]. This is important for preserving the stability of the optimization algorithm. Since these λ values scale the estimated MOS for every transmission policy and we maximize the sum of the MOS of all the users, the optimizer assigns transmission policies with high estimated MOS to the users with higher λ. This gives higher priority to the users which have been receiving lower MOS up to the time of the current optimization step.

3.4 Throughput maximization

A common network performance metric is the throughput of the system. Traditionally, the goal of the network operator is to maximize the network throughput. By throughput we consider the effective rate (goodput) G_{ij} of a given user i at time j:

$$G_{ij}=R_{ij}*(1-PEP) \quad (11)$$

with R_{ij} the actual transmission rate and PEP the packet error probability. The objective function for such an optimization model is to maximize the sum of the goodput allocated to all the users in the system and is given with equation (12). Here the optimizer is not aware of the user perceived quality. The assumption is that if a user receives more data rate, then he also has a higher QoS.

For throughput maximization we have the same set of decision variables as in equation (4)-(8). The difference is the absence of the scaling parameter λ. Here we do not need scaling of the allocated transmission rate, because the transmission rates required by different applications are not comparable.

Maximize

$$\sum_{i\in U}\sum_{j\in T_U}u_{ij}G_{ij}+\sum_{i\in V}\sum_{j\in T_V}v_{ij}G_{ij}+\sum_{i\in W}\sum_{j\in T_W}w_{ij}G_{ij}$$

(12)

Subject to:

$$\sum_{j\in T_U}u_{ij}=1\,,\quad \forall i\in U \quad (13)$$

$$\sum_{j\in T_V}v_{ij}=1\,,\quad \forall i\in V \quad (14)$$

$$\sum_{j\in T_W}w_{ij}=1\,,\quad \forall i\in W \quad (15)$$

$$\sum_{i\in U}\sum_{j\in T_U}r_{ij}u_{ij}+\sum_{i\in V}\sum_{j\in T_V}r_{ij}v_{ij}+\sum_{i\in W}\sum_{j\in T_W}r_{ij}w_{ij}\le TotalSymbolRate \quad (16)$$

4. SIMULATION SETUP

The simulations shown in this paper are performed with the following parameter settings. We assume a total of seven simultaneous users in the wireless network. Four voice users, two male and two female voices, are used. The voice samples are 30 seconds long. The voice signal comes from the backbone network encoded with G.711 voice codec at 64kbps. In the base station, following the optimization output, the signal could be transcoded to 6.4kbps with G.723.1 codec, 15.2kbps with iLBC codec, 24.6kbps with Speex or it can be transmitted without transcoding at 64 kbps.

Two users subscribe for file download using FTP. Both of them have subscribed for a service with maximum offered transmission rate of 192kbps.

One user is requesting videoconferencing. The video sequence used is Foreman, encoded with H.264 encoder. The frame sequence is I-P-P-P-…-P, which is the appropriate format for real-time video.

The λ values in (4) are all set to 1 in our experiments. The total available system rate is constant and we have examined three different cases: 500ksymbols/s, 700ksymbols/s and 900ksymbols/s. The supported modulation schemes are DBPSK and DQPSK. Channel code rates of one-half, one-third and one-fourth are supported, using convolutional code.

To reflect user mobility, the receiver SNR for every optimization step is drawn randomly for every user from a uniform distribution from 7 dB to 25 dB. The system is active for 30 seconds and we assume that the average channel characteristics remain constant for 1.2 seconds, which results in 25 optimization loops.

The wireless system we have implemented in this work does not refer to any particular physical layer interface. We kept it intentionally simple, as the main goal of our work was to demonstrate the potential gain for any wireless system considering joint optimization across multiple different applications.

For the voice users, the signal samples are partitioned into 1.2 seconds and every sample is encoded using the voice codec determined by the optimization algorithm. At the end of the optimization loops, these voice samples are assembled into a single file and the perceived quality (MOS) is computed by comparing the original signal and the distorted one.

For the video user, if a slice is lost, it is not written in the bit stream, which tells the decoder to invoke the error concealment algorithm. The PSNR of every frame and the resulting average PSNR are computed. The average PSNR is converted to MOS value using the relationship shown in Figure 5. For file download we compute the MOS using the relationship given in (3).

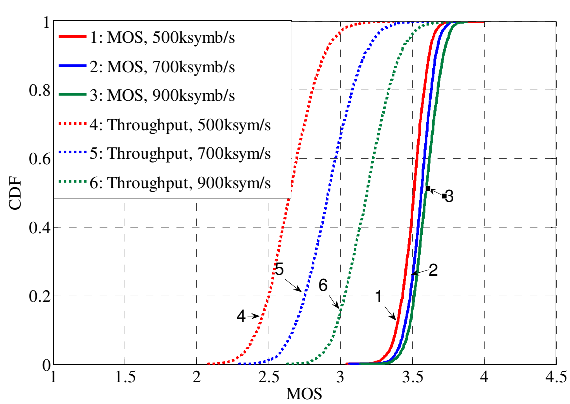

Figure 7: MOS of voice users

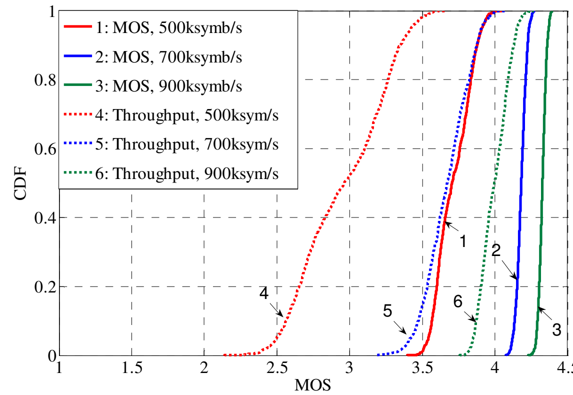

Figure 8: MOS of FTP users

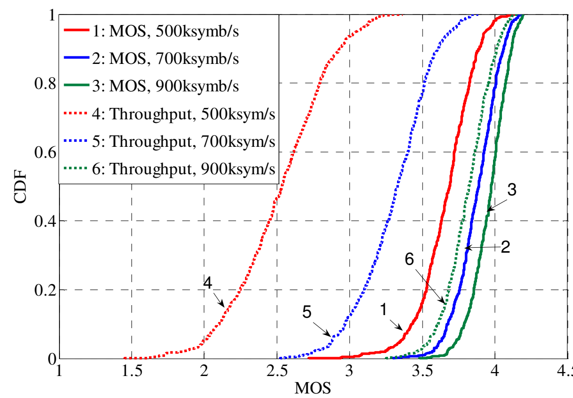

Figure 9: MOS of video conferencing users

5. RESULTS

In this section a comparison between both investigated optimization approaches (MOS maximization and throughput maximization) is done. We use the setup described in the previous section and we run each simulation 600 times.

Figure 7 presents the cumulative density function (CDF) of the voice user satisfaction for the MOS maximization and throughput maximization schemes. At a total system rate of 500ksymbols/s the average gain in terms of MOS is 0.85. At 700ksymbols/s the gain is still significant - 0.6 and for 900ksymbols/s it is around 0.4. The observed gains are hence biggest for scarce system resources.

Figure 8 shows the gain for the FTP users. The MOS maximization approach outperforms again the throughput maximization approach. Here the gain is lower, but it is still significant. For 500ksymbols/s the gain is 0.7 MOS on the average, for 700ksymbols/s it is 0.45 and for 900ksymbols/s it is 0.3.

Figure 9 shows the videoconferencing quality improvement. The gains in terms of MOS are similar to the case of the voice users and with the increase of the available transmission rate, the gain decreases.

For all the cases presented in the figures 7 – 9, the MOS maximization has the advantage of offering lower spread of the QoS offered to the users. For example if we consider Figure 7 for the case of a total system symbol rate of 500ksymbols/s, the resulting MOS in 90% of the cases for the throughput maximization varies between MOS of 2 and 3.5, i.e., a spread of 1.5 MOS. On the other hand the MOS maximization results in MOS variations between 3.4 and 4.1, i.e., a spread of only 0.7 MOS.

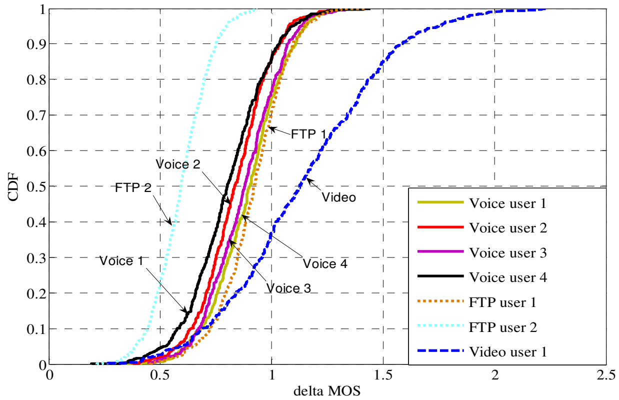

Figure 10: MOS gain per user, system symbol rate of 500ksymbols/s

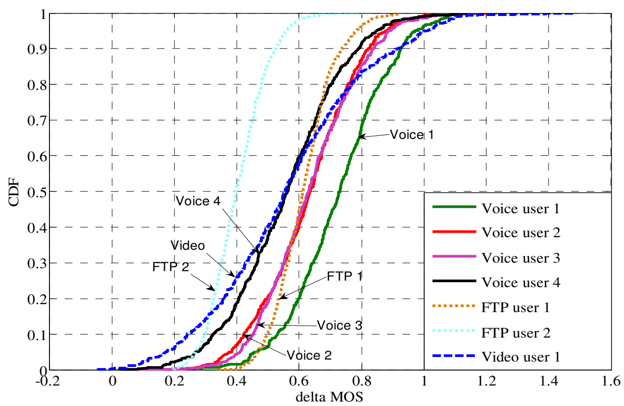

Figure 11: MOS gain per user, system symbol rate of 700ksymbols/s

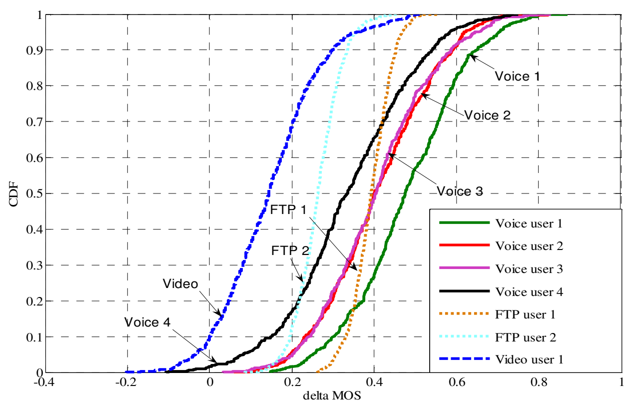

Figure 12: MOS gain per user, system symbol rate of 900ksymbols/s

Figures 10 – 12 present the gain per user in the system. The curves are produced as a difference between the mean MOS computed with MOS maximization and throughput maximization. Starting with a system symbol rate of 500ksymbols/s (Fig. 10), in 50% of the simulations, the average gain for all users is 0.8. Exceptions are the videoconferencing user, who has even higher MOS gain and the FTP user 2, who has a lower gain. In the system with 700ksymbols/s (figure 11)) there are cases (1% for the user having videoconferencing) where the throughput maximization gives better results for a given user. This is even more visible in the case with system symbol rate 900ksymbols/s (figure 12) when two users (the user having videoconferencing and the fourth voice user) have better performance in case of throughput maximization (10% of the cases for the video user). The mentioned users are the ones with the best channel with respect to SNR at the receiver. In case of MOS maximization, the optimizer takes resources from them to increase the MOS of the users who have worse channels.

6. CONCLUSION

In this paper we propose a novel cross-layer optimization approach across multiple applications using MOS as a common application layer performance metric. To our knowledge, this is the first attempt to use a common application level performance metric for efficient wireless resource allocation. We compare our approach to a traditional approach where allocation is done based only on physical layer parameters (throughput). Our simulation results show significant improvements in terms of user perceived quality for a variety of circumstances.

REFERENCES

[1] W. Kellerer, Lai-U Choi, and E. Steinbach, “Cross-Layer Adaptation for Optimized B3G Service Provisioning,” Proc. 6th International Symposium on Wireless Personal Multimedia Communications (WPMC), Yokosuka, Japan, October 2003.

[2] Lai-U Choi, W. Kellerer, and E. Steinbach, “Cross-Layer Optimization for wireless multi-user video streaming,” IEEE International Conference on Image Processing, ICIP 2004, Singapore, 24-27 Oct. 2004, pp. 2047-2050.

[3] S. Khan, M. Sgroi, Y. Peng, E. Steinbach, and W. Kellerer, “Application-driven cross-layer optimization for video streaming over wireless networks,” IEEE Commun. Magazine, vol. 44, no. 1, Jan. 2006, pp. 122-130.

[4] M. van der Schaar, S. Shankar N., “Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms,” IEEE Wireless Communications, Aug. 2005, pp. 50-58.

[5] V. Tsibonis, L. Georgiadis, L. Tassiulas, “Exploiting wireless channel state information for throughput maximization,” IEEE Infocom 2003, San Francisco, Mar. 30 – Apr. 3, 2003, pp. 301-310.

[6] X. Liu, E. Chong, N. Shroff, “Transmission scheduling for efficient wireless utilization,” IEEE Infocom 2001, Alaska, 22–26 April 2001, pp. 776-785.

[7] S. Shakkottai, T. S. Rappaport and P. C. Karlsson, “Cross-layer design for wireless networks,” IEEE Commun. Magazine, vol. 41, no. 10, Oct. 2003, pp. 74-80.

[8] ITU-T P.862, “PESQ: An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.”

[9] F. P. Kelly, “Charging and rate control for elastic traffic,” Eur. Trans. Telecommun., vol. 8, pp. 33--37, Jan. 1997.

[10] C. Rohaly et al., “Video Quality Experts Group: current results and future directions,” Visual Communicationas and Image Processing 2000, Proc. SPIE, 4067, pp. 742-753.

[11] M. T. Ivrlac, “Parameter selection for the Gilbert-Elliott model,” Technical Report TUM-LNS-TR-03-05, Institute for Circuit Theory and Signal Processing, Munich University of Technology, May 2003.