# Cross-Layer Optimization for Wireless Video Streaming – Performance and Cost

*S. Khan*[*], *M. Sgroi*[◊], *E. Steinbach*[*], *W. Kellerer*[◊]

[*]Media Technology Group
Institute of Communication Networks
Technische Universität München

[◊]DoCoMo Communications Laboratories
Europe GmbH, Munich, Germany
Future Networking Lab

## Abstract

*Cross-Layer Design (CLD) is a new paradigm for network architecture that allows us to make better use of network resources by optimizing across the boundaries of traditional network layers. Previous work has shown that applying CLD to mobile multimedia communication systems may lead to significant performance improvements. In this paper we also consider the other side of the coin, i.e., the additional computation and communication overhead introduced by CLD. We evaluate the performance improvements and the cost of cross-layer optimization using a wireless multi-user video streaming example.*

## 1. Introduction

Next generation wireless systems will have to support applications of increasing complexity and with tighter performance requirements, such as real-time or streaming video, interactive navigation in 3D virtual worlds and ubiquitous computing. To design efficient and cost-effective network architectures the research community has recently proposed a new paradigm, called Cross-Layer Design (CLD), which is based on information exchange and joint optimization among multiple protocol layers. CLD exploits layer dependencies and therefore allows us to propagate ambient parameter changes quickly throughout the protocol stack. Hence, it is especially well-suited to mobile multimedia applications where the characteristics of the wireless medium and the application requirements vary over time.

Previous applications of CLD to mobile multimedia communication, e.g. [1],[2], have mostly focused on optimizing individual layers based on information from adjacent layers. [3] presents a cross-layer architecture for wireless streaming video that jointly optimizes the application, the data link and the physical layer. Parameters from different layers are abstracted and provided to a cross-layer optimizer which selects the values of the protocol parameters maximizing the user perceived video quality. The objective function may be selected to reflect different goals, e.g., it may maximize the quality of individual users or the average quality of all the users.

While previous work often succeeds in showing the benefits of applying CLD, it often lacks of an accurate analysis of the additional cost that is to be paid to perform the optimization and to gather the relevant parameters from multiple layers and network locations.

Multiple components contribute to the cost of CLD. First, network architectures with cross-layer optimizations are less modular and therefore more difficult to manage or upgrade [4]. Second, solving the optimization problem may result in additional delay due to a broad exploration of the parameter domain space. Third, gathering the parameters that are relevant to the optimization may result in non negligible transmission overhead.

This paper focuses on the cost due to applying CLD and exploits tradeoffs between cost and performance. We consider a distributed video streaming cross-layer architecture and trade performance expressed in terms of average Peak Signal to Noise Ratio (PSNR) versus the communication cost associated with the transmission of a rate-distortion profile from the video server. Furthermore, we exploit the dependency of the optimization computational complexity with respect to different resource allocation cases and number of users.

## 2. Video Streaming Cross-Layer Architecture

Streaming video to mobile terminals requires a highly efficient and optimized architecture able to provide each user with a good quality video. We consider a simplified single-cell architecture that delivers videos from remote servers to $K$ mobile terminals located in a cell through a base station that assigns the wireless channel resources to the different users.

The dynamic nature of the wireless channel and the diversity of frames in a video stream make it necessary to dynamically adapt the network configuration based on the current conditions of the environment. In [3] we have proposed a CLD architecture (Figure 1) with a component, called cross-layer optimizer (CLO), that periodically selects

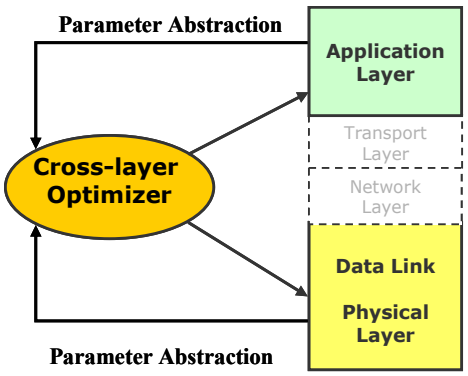


Figure 1: CLD Architecture.

*0-7803-9332-5/05/$20.00 ©2005 IEEE*

the optimal parameter settings of the different layers. The CLO uses abstractions of different layers and decides the resource assignment for each video stream.

As an abstraction of the application layer we use the rate-vector and distortion matrix (RD profile) introduced in [5] that describes the reconstruction quality, expressed in PSNR, if any of the frames in a GOP (Group of Pictures) is lost and the decoder displays the most recently decoded frame instead. The RD profile is computed at encoding time and transmitted as side information along with the video stream. As described in [3] the degrees of freedom on the data link and physical layers are abstracted into the transition probabilities of a two-state Markov burst loss (Gilbert-Elliot) model.

## 3. Cross-Layer Optimization

Let us consider a wireless video-streaming scenario with three users, each requesting a different video from the streaming server, namely Mother & Daughter (MD), Carphone (CP) and Foreman (FM). All three videos are in QCIF resolution (176×144). Each sequence has 300 frames and the frame rate is 30 fps. The videos are pre-encoded at two different target source rates of 100kbps and 200kbps, using the Xvid codec [6]. Each GOP has 15 frames, including one I-frame and 14 P-frames. Table 1 gives an overview of the main characteristics of the video sequences. The average PSNR between the encoded and the displayed video sequence is used as a performance measure.

Table 1: Main characteristics of the test video sequences

| Video Sequence | Length(s) | Frames | PSNR(dB) 100 kbps | PSNR (dB) 200 kbps |
|---|---|---|---|---|
| FM | 10 | 300 | 32.45 | 34.67 |
| CP | 10 | 300 | 33.32 | 36.78 |
| MD | 10 | 300 | 36.31 | 39.80 |

On the radio link layer, it is assumed that the total transmission symbol rate in the system is 300k symbols/s. The data packet size is equal to 54 bytes, which is the specified packet size of the IEEE802.11a or HiperLAN2 standard. The channel coherence time is assumed to be 50ms for all the three users, which approximately corresponds to pedestrian speed for 5GHz carrier frequency. The residual packet error rate can be described as a function of the average SNR [7]. User position dependent path loss and shadowing commonly observed in wireless links are taken into account by randomly choosing the corresponding average SNR for each user.

We define the total transmission bit rate constraint $R_{\max}^{m,K}$ for modulation scheme m and *K* users to be

$$R_{\max}^{m,K} = n \cdot K \cdot R \quad (1)$$

where *n* is bits per symbol for modulation scheme m, *K* is the total number of users, *R* is the average symbol rate for one user. For three users and *R*=100k symbols/sec, we have $R_{\max}^{BPSK,3} = 300\ kbit/sec$ and $R_{\max}^{QPSK,3} = 600\ kbit/sec$ for BPSK and QPSK, respectively.

We define the possible set of transmission bitrates $C_k$ for any user *k* as

$$C_k = \{0, r, 1.5r, 2r, 3r\} \quad (2)$$

where *r* in our experiments will be 100 kbit/sec. The total rate constraint, together with the set of transmission rates gives us 26 possible rate allocations (Table 2).
If the available transmission rate exceeds the source rate, the most important frames of the GOP are repeatedly transmitted.

Table 2: Possible cases of rate allocation among three users

| Case | Modulation Scheme | Transmission Data Rate (kbps) User 1 | User 2 | User 3 |
|---|---|---|---|---|
| 1 | BPSK | 100 | 100 | 100 |
| 2 | BPSK | 100 | 200 | 0 |
| 3 | BPSK | 100 | 0 | 200 |
| … | … | … | … | … |
| 13 | BPSK | 0 | 0 | 300 |
| 14 | QPSK | 200 | 200 | 200 |
| 15 | QPSK | 0 | 300 | 300 |
| 16 | QPSK | 300 | 0 | 300 |
| … | … | … | … | … |
| 26 | QPSK | 300 | 150 | 150 |

Each of the cases in Table 2 may lead to a number of operational modes depending on the available transmission rate of the users. For example, in case 2, user 2 has an available rate of 200 kbit/sec, which can be used either to send the low rate video with repetition, or the high rate video without repetition. In total, we have 72 different modes of operation (parameter tuples) among the three users.

The cross-layer optimizer selects for each GOP the optimal parameter values that maximize the user-perceived video quality. This requires computing for each user and each parameter selection the expected quality at the receiver, which can be obtained in one of the following ways:

1. Computing the expected reconstruction quality (in PSNR) given by

$$PSNR_{\exp} = \sum_{i=1}^{l} p_i D_i \quad (3)$$

where *l* is the number of different loss patterns [8], $p_i$ is the loss pattern probability, $D_i$ is the resulting reconstruction quality for loss pattern $i$ derived from the distortion matrix [5]. The probability of a particular loss pattern $p_i$ is computed from the transition probabilities of the Gilbert-Elliot model as described in [8].

2. Computing the Expected Number of Decodable Frames (ENDEF) in one GOP given by:

$$ENDEF = \sum_{i=1}^{l} p_i d_i \quad (4)$$

where $d_i$ is the number of decodable frames for a particular loss pattern. ENDEF provides an approximation of the expected PSNR values in case the distortion information $D_i$ is not available.

## 4. Performance Analysis

In this section we compare the performance gain obtained by applying cross-layer optimization for the two cases where expected PSNR (CLO PSNR) and ENDEF (CLO ENDEF) are used by the optimizer to predict video quality. The objective function is chosen to be the average PSNR of all the users:

$$F(\tilde{\mathbf{x}}) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{PSNR}_k(\tilde{\mathbf{x}}) \qquad (5)$$

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. The decision of the optimizer can be expressed as

$$\tilde{\mathbf{x}}_{opt} = \arg\max_{\tilde{\mathbf{x}} \in \tilde{X}} F(\tilde{\mathbf{x}}) \qquad (6)$$

where $\tilde{\mathbf{x}}_{opt}$ is the optimum parameter tuple which maximizes the objective function.

We analyze three different scenarios. Fig. 2 shows the CDF of the average PSNR for all the three scenarios, each one based on 1000 simulation runs. In the first scenario all the users have very bad channel conditions. The received SNR varies between 0dB and 5dB. As seen from the CDF, average PSNR increases about 2 dB for cross-layer optimization with rate-distortion side information (CLO PSNR), compared to the case of without optimization (w/o CLO). In the second scenario, simulations are performed with random user SNR in a large range (0 to 25 dB) for all the users. The curve representing CLO without RD side information lies approximately half way between the other two curves for both scenario 1 and 2. In the third scenario, all the users have very good channel conditions, with random user SNR between 20 dB and 25 dB. Also in this case we observe an average PSNR improvement of about 2 dB for cross-layer optimization with RD side information, compared to the case without optimization. In this case the optimizer can take advantage of the good channel condition by choosing the higher source rate videos. Note that the performance without RD side information is worse in this case because of the lower correlation between the number of decodable frames and the resulting PSNR.

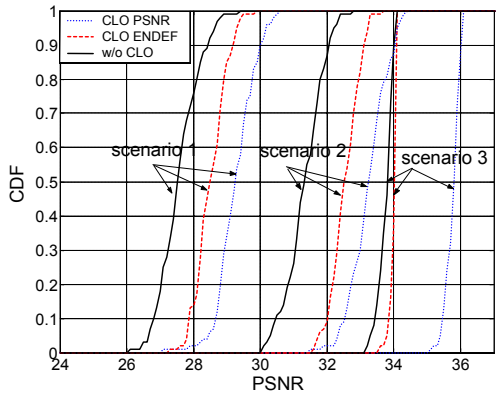


Figure 2: CDF of average PSNR for random received SNR equal to 0dB to 5dB (scenario 1), 0 to 25 dB (scenario 2) and 20 to 25 dB (scenario 3).

As the performance of the optimization depends on the accuracy of PSNR prediction, we now consider the relative PSNR prediction error for our proposed mode, which is defined as the ratio of the absolute PSNR prediction error and the actual PSNR:

$$PSNR_{relerr} = \frac{\left|PSNR_{\exp} - PSNR_{actual}\right|}{PSNR_{actual}} \qquad (7)$$

where $PSNR_{\exp}$ is the expected PSNR computed at the base station from (3), and $PSNR_{actual}$ is the actual PSNR between the original and the received video frames at the clients. We assume previous frame concealment in case of a frame loss. Figure 3 shows the CDF of relative PSNR prediction error for Packet Loss Rates (PLR) of 3% and 10%, with average burst lengths of 5 and 28, respectively. Results are based on 1000 simulation runs for each of the video sequences at a particular PLR. The source rate of all the videos is 100kbit/sec. For 3% PLR, the prediction error is less than 10% for more than 95% of the cases for all three video sequences.

From Fig. 2 and 3, we conclude that although the prediction error in (7) depends on the loss rate, the gain due to CLO PSNR compared to the case without CLO remains constant (2dB average) which can be attributed to the fact that the optimization spans across multiple users. On the other hand, the gain of applying CLO ENDEF varies with different values of the SNR due to the changing correlation between the number of decodable frames and actual PSNR at different loss rates and for different video sequences.

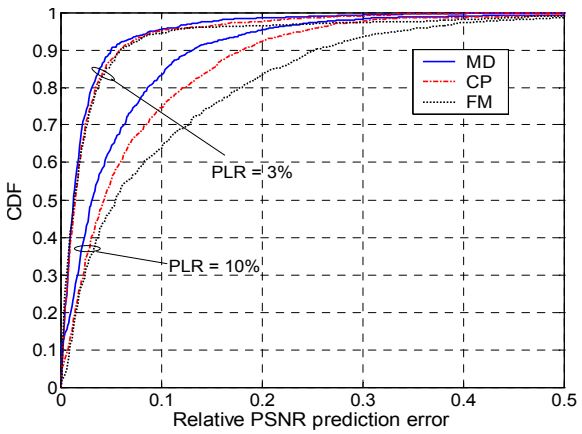


Figure 3: CDF of relative PSNR prediction error for PLR = 3% and 10%, for the MD, CP and FM video sequences.

## 5. Computational cost

The additional computational cost of cross-layer optimization is mainly due to computing the objective function (average expected PSNR or ENDEF) for all possible cases of parameter settings. In our implementation, we compute and store the possible $PSNR_{\exp}$ values into an array at the beginning of a GOP and evaluate the average $PSNR_{\exp}$ for all abstracted parameter tuples. In our experiments, the three optimization parameters: modulation

scheme, source bitrate and channel bitrate have 2, 2 and 5 possible realizations, respectively. For our complexity analysis we define the *normalized execution time* as

$$T_K = K \cdot m \cdot s \cdot r \qquad (8)$$

where $K$ is the number of users, $m$ is the number of different modulation schemes, $s$ is the number of possible cases of source bitrate and $r$ is the number of possible cases of channel bitrate. Table 3 shows the number of cases of resource allocation and normalized execution time for cross-layer optimization for different number of users.

Table 3: Number of CLO cases and normalized execution time for different number of users.

| Number of user | Number of operation modes | Normalized execution time |
|---|---|---|
| 2 | 13 | 2*2*2*5 |
| 3 | 72 | 3*2*2*5 |
| 4 | 345 | 4*2*2*5 |
| 5 | 1610 | 5*2*2*5 |
| 6 | 7811 | 6*2*2*5 |
| 7 | 36372 | 7*2*2*5 |
| 8 | 169135 | 8*2*2*5 |
| 9 | 787554 | 9*2*2*5 |
| 10 | 3507183 | 10*2*2*5 |

Although the number of operational modes increases very rapidly with the number of users, the time to evaluate the different cases increases almost linearly. This is because the normalized execution time can be approximated as the number of times we have to compute $PSNR_{\text{exp}}$ in (3) or ENDEF in (4) for a given set of parameters, as this is the computationally most expensive part of the optimization. The computational cost of the remaining task, which involves computing and comparing the objective function for different operation modes (eq. 5) can be neglected for a small number of users, e.g. $K \leq 10$. For a large number of users, however, this becomes increasingly important, as the number of operation modes increases exponentially with the number of users.

## 6. Communication cost

The communication overhead of CLO is mainly due to the transmission of parameter abstractions across the network. In particular there is an overhead due to transmitting the rate-distortion side information from the video server to the cross-layer optimizer. Fig. 4 shows the overhead for different source rates. Here we assume one GOP every half a second and every GOP consisting of only one I and else P frames. The overhead is low, but increases linearly with the number of frames in a GOP.

## 7. Conclusions

In this paper we have analyzed the tradeoff between performance and cost of CLD for a wireless multi-user video streaming application. We have compared the performance gain obtained when applying cross-layer optimization with the case where no optimization is applied for two abstractions of the application layer parameters. In one case, the CLO computes the expected PSNR using a distortion profile that is derived when the video is encoded and sent as side information. In the other case the CLO uses an approximation of the PSNR based on the expected number of decodable frames. As expected, the analysis shows that the optimization using the distortion profile provides higher gain due to the more accurate calculation of the expected video quality. However, the distortion profile must be transmitted from the server with a transmission overhead. Moreover, it is not available in applications that require real-time encoding. Our analysis shows that using the expected number of decodable frames still offers a valid gain with respect to the case without CLO especially in the case of channels with low SNR. We also observe that for a small number of users, the complexity of the system is a linear function of the number of users. As the number of user increases, however, the relationship deviates from linearity.

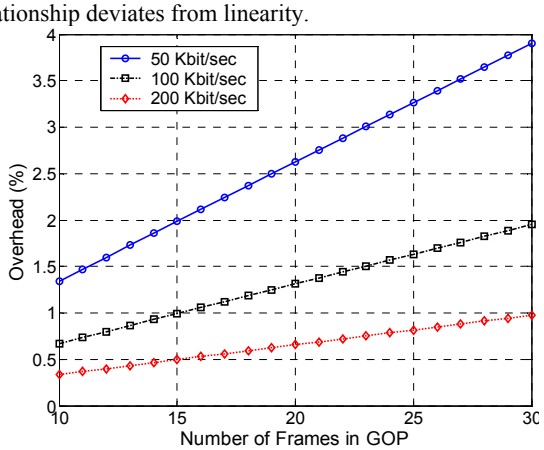


Figure 4: Traffic overhead of sending RD side information as a function of GOP size.

## References


[1] R. Tupelly, J. Zhang, E. Chong, "Opportunistic scheduling for streaming video in wireless networks," In Proceedings of the 37th Annual Conference on Information Sciences and Systems, Balitomore, Maryland, March 12-14, 2003.

[2] J. Gross, J. Klaue, H. Karl, A. Wolisz, "Cross-Layer Optimization of OFDM transmission systems for MPEG-4 video streaming", Computer Communications, vol. 27, pp. 1044-1055, 2004.

[3] 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, October 2004.

[4] V. Kawadia, P. R. Kumar, "A cautionary perspective on cross layer design," IEEE Wireless Communications, vol. 12, issue 1, pp. 3-11, February 2005.

[5] W. Tu, W. Kellerer, and E. Steinbach, "Rate-Distortion Optimized Video Frame Dropping on Active Network Nodes," Packet Video Workshop 2004, Irvine, California, December 13-14, 2004.

[6] XviD homepage, http://www.xvid.org/

[7] 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.

[8] Y. Peng, S.Khan, E. Steinbach, M. Sgroi, W. Kellerer, "Adaptive resource allocation and frame scheduling for adaptive multi-user video streaming," IEEE International Conference on Image Processing, Genova, Sept. 11-14, 2005.