Optimized wireless video transmission using classification

Raymond S. Wong [1], *Mihaela van der Schaar* [1] *and Deepak S. Turaga* [2]
[1] University of California, Davis, CA
[2] IBM Research, Hawthorne, NY

Abstract — *Cross protocol layer optimizations have been recently proposed for improving the performance of real-time video transmission over 802.11 WLANs. However, performing such cross-layer optimizations is difficult since the video data and channel characteristics are time-varying, and analytically deriving the relationships between quality and channel characteristics given delay and power constraints is difficult. Furthermore, these relationships are often non-linear and non-deterministic (only worst or average case values can be determined). Complex Lagrangian or multi-objective optimization problems are thus often faced. In this paper, we propose a novel framework for solving cross MAC-application layer optimization problems. More specifically, we employ classification techniques to find an optimized cross-layer strategy for wireless multimedia transmission. Our solution deploys both content- and channel-related features to select a joint application-MAC strategy from the different strategies available at the various layers. Preliminary results indicate that considerable improvements can be obtained through the proposed cross-layer techniques relying on classification as opposed to ad-hoc solutions. The improvements are especially important at high packet-loss rates (5% and higher), where deploying a judicious mixture of strategies at the various layers becomes essential.*

1 Introduction

Due to their flexible and low cost infrastructure, wireless LANs [1] are poised to enable a variety of delay-sensitive multimedia transmission applications, such as videoconferencing, emergency services, surveillance, telemedicine, remote teaching and training, augmented reality, and distributed gaming. However, existing wireless networks provide dynamically varying resources with only limited support for the Quality of Service (QoS) required by the *delay-sensitive, bandwidth-intense and loss-tolerant multimedia applications*. To address these challenges, cross-layer optimization strategies, [2] and [3], have been proposed as a solution for improving the performance of wireless multimedia streaming applications.

1.1. Cross-layer problem formulation and challenges

We formulate the cross-layer design problem as an optimization with the objective to select a joint strategy across multiple OSI layers. In this paper, we limit our discussion to MAC and Application (APP) layers only. Nevertheless, the proposed framework can easily be extended to include other layers. In particular we consider only two adaptation strategies: MAC_1, that corresponds to selecting the retransmission limit for each packet, and APP_1, that corresponds to determining the rate-adaptation and packet priority. Let N_M denote the number of available values for the retransmission limit and N_A denote the number of priorities available at the APP layer.

We define the joint cross-layer strategy S as:

$$S = \{MAC_1, APP_1\} \tag{1}$$

It is clear from the Eqn.(1) that there are $N=N_M \times N_A$ possible values for this joint strategy, for each packet. Furthermore, if we have N_P packets, and we want to perform a joint optimization, we have $(N_M \times N_A)^{N_P}$ total settings for the joint optimal strategy, which can be prohibitively large. The cross-layer optimization problem attempts to find the optimal composite strategy represented by the following equation.

$$S^{opt}(\mathbf{x}) = \arg\max_S Q(S(\mathbf{x})) \tag{2}$$

This strategy results in the best (PSNR/perceived) multimedia quality Q subject to following constraints:

$$\begin{aligned} R(S(\mathbf{x})) &\le R_{\max} \\ Delay(S(\mathbf{x})) &\le D_{\max} \end{aligned} \tag{3}$$

where $\mathbf{x} = (SNR, contention)$ is the instantaneous channel condition, $R_{\max}$ is the maximum transmission rate over the particular channel and $D_{\max}$ is the maximum tolerable delay. Hence, we need to solve Eqn.(2) subject to constraints in Eqn.(3).

Finding the optimal solution to the above cross-layer optimization problem is difficult because:

- Deriving analytical expressions for *Q, Delay,* and *Rate* as functions of channel conditions is very challenging, since these functions are non-deterministic (only worst case or average values can be determined), non-linear, and there are dependencies between the strategies MAC_1 and APP_1.
- The wireless channel conditions and multimedia content characteristics may change continuously, requiring constant updating of the joint strategies.
- Formal procedures are required to establish optimal *initialization*, *grouping* of strategies at different stages (i.e., which strategies should be optimized jointly), and *ordering* (i.e., which strategies should be optimized first) for performing the cross-layer adaptation and optimization.

Unfortunately, exhaustively trying all the possible strategies and their parameters in order to choose the composite strategy leading to the best quality performance in real-time is impractical due to the associated complexity. Instead, we propose to use classification techniques to solve this complex cross-layer optimization problem in an integrated manner.

0-7803-9332-5/05/$20.00 ©2005 IEEE

1.2 Related research

Classification techniques have already been deployed for selecting among the various encoding parameters and options based on content characteristics [4]. Moreover, in [5], classification techniques have been adopted to select among the SNR and temporal scalabilities based on a variety of content and codec specific features. However, the abovementioned efforts only consider application-layer features and optimizations.

1.3 Contributions and outline of this paper

In this paper, we avoid the intractable complexity involved in performing the abovementioned cross-layer optimization and the lack of analytical functions to describe the relationships between rate, quality and delay for various strategies by adopting classification and machine learning techniques. For this purpose, we first apply domain-specific knowledge or general unsupervised clustering to construct distinct categories of video sequences sharing similar preferred joint MAC-application layers strategies. Thereafter, a machine learning based method is applied where low level content features extracted from the compressed video streams as well as channel features such as the channel condition and maximum available bit-rate are employed to train a framework for the investigated cross-layer optimization problem.

This paper is organized as follows. We describe our proposed method for cross-layer optimization based on classification in Section 2. In Section 3, we present the simulation results based on the proposed framework. The conclusions are drawn in Section 4.

2 Joint MAC-Application optimization based on classification

As discussed in the introduction, our aim is to determine the optimal MAC retry limit (m) for the various packets based on the prioritization strategy deployed at the application layer given the maximum bit-rate ($R_{\max}$) and the packet loss rate (PLR). This is actually a conventional unequal error protection problem. To solve this problem, the basic idea of our system is to predict the multimedia quality-rate-resilience tradeoff given the channel characteristics and delay constraints for different joint MAC-application strategies. The idea is based on the observation that video sequences with similar content characteristics and under similar channel conditions have similar quality-rate-resilience tradeoffs when various cross-layer strategies are applied.

The proposed classification-based cross-layer framework is shown in Figure 1. It consists of an offline training module followed by online processing. The former includes the modules for the class definition and classifier learning, and the latter mainly involves classification and prediction.

For each training video sequence, the compressed-domain content features are extracted. Our feature sets also include the different $R_{\max}$ and PLR. The multimedia quality for the different sequences, channel conditions and cross-layer strategies is also determined. Then, the videos and channel conditions are grouped into distinct categories using unsupervised clustering. Videos in the same class are represented by a unique class label and associated with distinct cross-layer metadata. Given the class definition and labeled training data, machine learning techniques are used to train statistical classifiers for mapping video and channel features to corresponding classes. Such classifiers are then used in the online processing routine to classify the incoming video according to its content features and the channel characteristics and subsequently predict the corresponding cross-layer metadata, which will be used to select the composite strategy S. Hence, the proposed classification mechanism can predict the optimal composite strategy S^{opt} for that class.

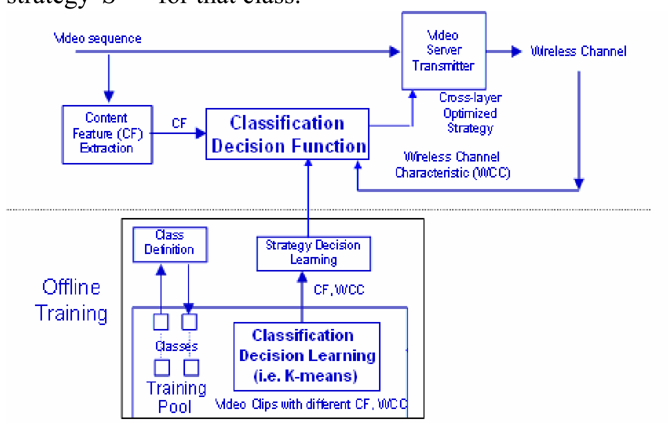


Figure 1: System architecture of proposed algorithm.

2.1 Feature selection

The features deployed in our system can be broadly divided into two categories: the video features and channel related features, respectively. In this section, we will describe why these features are selected and how they impact the cross-layer strategy selection.

Video features

In this paper, we use the SIV coder in [6], which is a fully scalable, motion-compensated wavelet coder, for the compression of the video data. Based on the coder-specific knowledge, the following features were selected.

i) Spatio-temporal band hierarchy

The various spatio-temporal bands have a different impact on the overall distortion and hence, they require different protection strategies.

ii) Motion vector magnitude

When the motion vector magnitude is high, it is likely that losing packets will result in a large impact on the quality Q and hence, the corresponding frames/packets will need to be better protected.

iii) Subband energy

Subband energy indicates the contribution to the distortion of the subband, and can distinguish sequences with different levels of spatio-temporal detail.

Channel features
i) Transmission bit-rate $R_{\max}$
ii) Packet loss rate P_L
We divided the packet loss rate into three classes: low packet loss rate (0% to 1%), medium packet loss rate (1% to 5%) and high packet loss rate (5% and above).

2.2 Classification-based cross-layer strategy selection

Summarizing, the proposed content-aware optimization strategy consists of the following major steps:

Step 1: Generate ground truth: Identify the MAC and application layer strategies available at the wireless station and the multimedia quality Q resulting when various joint strategies are deployed.

Step 2: Feature Extraction and Selection: Extract content and channel features and generate quality-resource metadata.

Step 3: Train Classifier. During training, unsupervised clustering methods (in our case K-means) are used to map content and channel features to corresponding classes represented by a unique class label and distinctive optimal strategy recommendation. The key is to determine a mapping $c_l = C(\vec{F})$ *from content and channel feature vector* $\vec{F}$ *to class label* c_l*. In our experiment, we use a set of 1687 training vectors. Each video class shares a corresponding optimal strategy recommendation* $O_s(c_l)$*.*

Step 4: Classification for optimal strategy prediction on test data set. The optimal strategy for an incoming video can be predicted based on its content and channel features $O_s(c_l)$. The selected strategy is used to determine the parameters and configurations of the cross-layer optimized system.

3 Simulation results

3.1 Retransmissions with Rate-Distortion cost

We assume packets that are not received are retransmitted up to a certain maximum number of times (*m*). For a packet loss rate P_L, and with independent packet losses, the probability that the packet is received in *n* transmissions, i.e. with *n*-1 retransmissions is $P_L^{n-1}(1-P_L)$. Hence, with a retry limit *m*, the probability that a packet is received is given by

$$P_{succ} = \sum_{n=1}^{m+1} P_L^{n-1}(1-P_L) \qquad (4)$$

and the probability that the packet is dropped is $P_{fail} = 1 - P_{succ}$. Thus, the expected number of transmissions for any packet can be obtained by equation (5).

$$\overline{m} = \sum_{n=1}^{m+1} n p_L^{n-1}(1-p_L) + (1-P_{succ})(m+1) \qquad (5)$$

The expected rate for a packet is defined as $\overline{R} = \overline{m} \cdot L$, where L is the packet length.

Each packet has a no-loss distortion D_p and a loss distortion D_p^{loss} associated with it. Thus, the expected distortion for a packet can be found as $\overline{D_p} = P_{succ} \cdot D_p + P_{fail} \cdot D_p^{loss}$. Finally, a composite cost, that captures the tradeoff between the distortion and rate associated with every packet, can be defined in equation (6).

$$C = \overline{D_p} + \lambda \cdot \overline{R} \qquad (6)$$

We exhaustively calculate the cost for every packet in our training data for different channel conditions, for a particular tradeoff between rate and distortion (a particular λ). Based on the cost, we determine which blocks should have high retry limit and which should have low retry limit. This serves as our ground truth and enable us to determine how efficient our feature sets are, and train a classifier to select the optimal strategy.

For instance, if we use only content features, e.g. the subband energy, to train our classifier, on ground truth obtained from one GOP of 16 frames, our classification performance deteriorates from 63% to 56% with increasing P_L (2% to 5%). This provides us motivation to include the packet loss rate as an additional feature in our classification scheme.

3.2 Simulation results

We used the SIV codec with 4 temporal and 4 spatial decomposition levels and treat each code block to be a separate packet. We perform the simulated packet loss experiment 20 times with a constrained total decoding bit-rate $R_{\max}$ of 256 kbps. We also repeat the experiment at additional rates 512 kbps and 1024 kbps, and present the average PSNR. We perform simulations for two CIF (352×288) video sequences @30Hz: Foreman and Coastguard.

3.3 Simulation results

We compare our classification based scheme against ad-hoc schemes for equal error protection (EEP) and unequal error protection (UEP). For equal protection (EEP), we protect all packets with a fixed retry limit of 3. For UEP, we protect the LL bands with a retry limit of 5, the LH bands and HL bands with a retry limit of 1 and the rest of the bands with retry limit 0. In all these cases we constrain the total rate (including data and redundancy) to be the same for all the schemes. In a 300 frames of sequence, we collected the first 16 frames of feature vectors to train our classifier and the rest of frames to be our test data.

Figure 2 and Figure 3 show the results of performance comparison between, Feature classification using only subband energy (FC_EN) and Feature classification (FC) under various channel conditions.

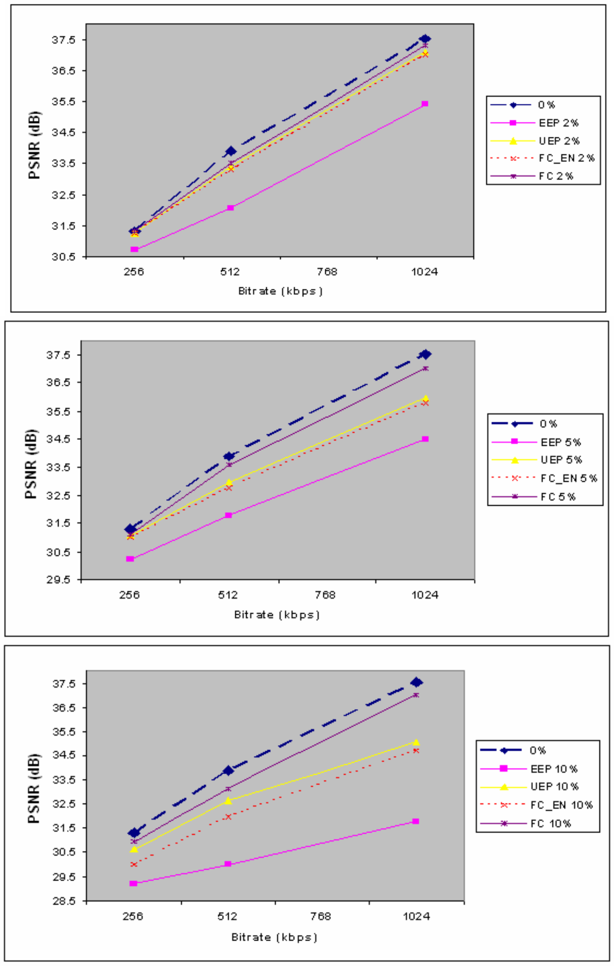
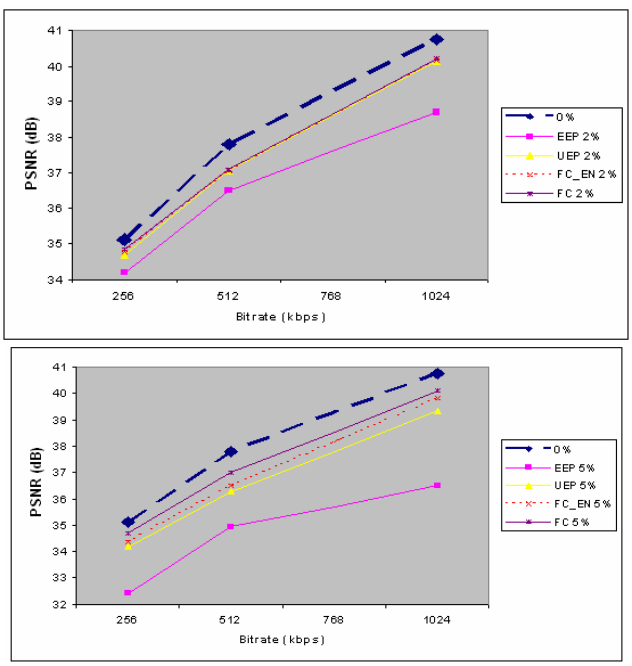


Figure 2: Foreman sequence



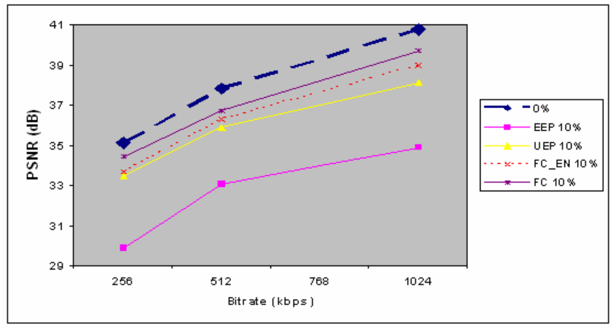


Figure 3: Coastguard sequence

From the results we can see that classification can provide a solution for cross-layer optimization that can outperform ad-hoc UEP and EEP schemes.

4. Conclusions

In this paper, we use classification based techniques to solve the MAC-APP cross-layer optimization problem. A learning based approach can significantly reduce the complexity of performing this multi-variate optimization problem. In our classification based approach, we use content and network features that can be easily computed and that are good indicators of which composite (integrated) strategy is optimal. Our preliminary results indicate a significant improvement in performance as opposed to ad-hoc cross-layer solutions.

REFERENCES

1. IEEE 802.11 WG: IEEE Std 802.11-1999, Part11: Wireless LAN MAC and Physical layer specifications. Reference number ISO/IEC 8802-11:1999 (E), (1999).
2. Leiming Qian, Douglas L Jones, Kannan Ramchandran and Swaroop Appadwdula, “A General Joint Source-Channel Matching Method for Wireless Video Transmission”, in Proc. Data Compression Conference, (Snowbird, UT), pp. 414 --423, 1999.
3. Mihaela van der Schaar, Santhana Krishnamachari, and Sunghyun Choi, "Adaptive Cross-Layer Protection Strategies for Robust Scalable Video Transmission Over 802.11 WLANS," IEEE Journal on Selected Areas of Communications, VOL. 21, NO. 10, Dec 2003.
4. Deepak S. Turaga, Tsuhan Chen, “Classification Based Mode Decisions for Video over Networks”, IEEE trans. March 2001
5. Yong Wang, Mihaela van der Schaar, Shih-Fu Chang, Alexander C. Loui, “Content-Based Optimal MDA Operation Prediction For Scalable Video Coding Systems Using Subjective Quality Evaluation”
6. David S Taubman, University of New South Wales, Sydney, Australia, SIV Codec.