ENERGY-EFFICIENT BANDWIDTH ALLOCATION FOR MULTI-USER VIDEO STREAMING OVER WLAN

Xin Ji (abc), Sofie Pollin (abc), Gauthier Lafruit (b), Iole Moccagatta (b), Antoine Dejonghe (bc), Francky Catthoor (ab)

[a]Katholieke Universiteit Leuven, Leuven, Belgium
[b]IMEC, Leuven, Belgium [c]IBBT, Gent, Belgium

ABSTRACT

We consider the problem of packet scheduling for the transmission of multiple video streams over a wireless local area network (WLAN). A cross-layer optimization framework is proposed to minimize the wireless transceiver energy consumption while reaching the user required visual quality. The framework relies on the IEEE 802.11 standard and on a wavelet-based scalable video coding scheme. It extends our previous work on energy-efficient scheduling by introducing an application-level video quality metric as QoS constraint (instead of a quality metric at the level of the communication layers) and by reformulating the energy minimization problem subject to the QoS constraint in order to also consider the fairness among users. Simulation results demonstrate significant additional energy gains by means of these extensions.

***Index Terms*—**QoS control and scheduling, Energy optimal control, Scalable video coding, Multi-user WLAN, Cross-layer

1. INTRODUCTION

The demand for multimedia transmission over wireless networks is continuously growing. Transmission of multiple video streams over a wireless local area network (WLAN) is a typical example of this evolution. In this context, Quality of Service (QoS) provision for real-time applications is becoming more and more critical, as wireless networks are affected by highly error-prone and time-varying conditions, especially when a lot of users interact. Besides this QoS challenge, ensuring low-power consumption is becoming imperative in battery-operated portable devices.

Performing high-quality and energy-efficient video packet selection and scheduling for such wireless networks is a challenging task. Most of the WLAN transmission studies consider throughput as performance measurement, while it can be shown that throughput may not be the most appropriate metric for video traffic. Some recent studies try to improve the performance by also exploring the specificities of video traffic. For instance, considering scalable video coding techniques [6,7] (which provide an inherent prioritization among the compressed data and offer a natural method for selecting different portions of the data stream under different network conditions), different retransmission limits were defined for different priority queues at the medium access control (MAC) layer in [8]. In [10], a solution for scheduling transmission opportunities (referred to as TXOP in the remainder of the present paper) according to the data type is proposed. Considering energy efficiency, a substantial body of prior work focuses on wireless transmission energy, and different approaches exist at MAC and physical (PHY) layers [1, 2]. There is however very few work considering the real video quality and the total system energy cost, with a focus on the whole protocol stack. Furthermore, few of the literature articles provide complexity analysis when solving their formulated optimization problem. We claim that by designing a two-phase cross-layer framework, a low complexity run-time solution can be provided to optimize energy consumption while meeting QoS and fairness requirements.

In [11, 12], we have introduced a cross-layer optimization methodology (MEERA) enabling the energy-efficient and reliable delivery of delay-sensitive network flows over a WLAN. In this context, a two-phase systematic approach for optimally allocating the network resources and controlling the system configuration was proposed. A first contribution of the present paper is the addition of a practical application-level video quality metric as QoS constraint in the optimization system, instead of using a packet loss probability at the communication layers. Doing so, a true cross-layer quality and energy optimization method is obtained. The resulting global solution enables to further minimize the wireless transceiver energy consumption by a factor 2 with respect to our prior work without degrading the visual quality. Next, we reformulate the optimization problem as a min-max problem in this paper, which enables to increase the energy cost fairness among all users.

The remainder of this paper is organized as follows. Section II briefly reviews the IEEE 802.11 WLAN standards and the deployed 3D wavelet motion-compensated temporal filtering (MCTF) scalable video coding scheme. Section III introduces the proposed energy-efficient video scheduling strategy with rate-distortion awareness. Appropriate system models are used to instantiate the proposed cross-layer optimization framework given the aforementioned standards. In Section V, we examine the performance of our framework through simulations. Finally, concluding remarks are provided in Section VI

2. WLAN VIDEO STREAMING SYSTEM OVERVIEW

The considered setup in the present work consists of the downlink transmission of several pre-encoded video streams to different mobile terminals in a WLAN. The content server is assumed to be connected to the access point. In this section, we introduce the wavelet-based scalable video coding scheme and the IEEE 802.11 standard that are considered for the test case. However, it is important to emphasize that the cross-layer algorithms proposed in this paper can be deployed with any video coding scheme where the bitstream can be organized into data units with scalability. Additionally, any schedule-based protocol can be used. A system model is introduced to calculate energy consumption, transmission delay and the expected quality of the coded video after transmission.

2.1. MAC and PHY channel model

The IEEE 802.11a PHY layer is based on Orthogonal Frequency Division Multiplexing (OFDM), and provides eight different PHY modes offering Data transmission Rates (DR), ranging from 6 Mbps to 54 Mbps. Our system modeling is based on an 802.11a direct conversion transceiver implementation with turbo coding [3]. Four control parameters have significant impacts on energy and performance for the OFDM transceivers: the modulation order N_{mod} (number of bits per transmitted symbol), the code rate B_c (amount of redundancy introduced in the signal), the power amplifier (PA) transmit power P_{TX} and the back-off b characterizing the linearity of the amplification [4]. Let us represent a possible transmission configuration as a vector K (each specific transmission parameter corresponds to an entry in this vector). For reliably transmitting on a wireless network, a long application layer packet *p* is usually further fragmented into smaller data units. In this paper, we consider link layer fragmentation only. The energy and time needed to send a Mac Service Data Units (MSDU) is functions of this configuration vector: $E_{MSDU}(K)$ and $TXOP_{MSDU}(K)$ [11, 12]. The energy cost and time of transmitting an application layer packet is defined as $E_p(K)$, $TXOP_p(K)$, and these values depend on the number of fragments that need to be transmitted or retransmitted for successful packet transmission. To determine the channel impact on the loss probability, the fading channel was discretized into 8 classes, corresponding to a 2dB difference in received SNIR (Signal-to-Noise-and- Interference-Ratio) for reaching a given turbo code block error rate target. In order to derive a time-varying link-layer error model, we associate every channel class to a Markov state, each with a probability of occurrence based on the channel fading statistics [4]. The loss probability of a MSDU resulting from the model is denoted by $P_{MSDU}(K)$. To obtain the corresponding loss statistics at application layer, we compute the Packet Error Rate (PER). The $PER_y^m(K)$ of an application data unit (assuming it is further fragmented into m MSDU packets and y retransmissions are allowed) can be calculated according to Eq. (1) – (3):

$$P_{ey}^m(K) = \sum_{i=1}^{\min(m,y)} C_i^m (P_{MSDU})^i (1-P_{MSDU})^{(m-i)} P_{y-i}^i(K) \quad (1)$$

$$P_{e0}^m(K) = (1-P_{MSDU})^m \quad (2)$$

$$1-PER_y^m(K) = \sum_{j=0}^{y} P_{ej}^m(K) \quad (3)$$

Energy cost and time is linearly addable, hence $E_p(K)$ and $TXOP_p(K)$ are respectively given as:

$$E_p(K) = (m+y)E_{MSDU}(K) \quad (4)$$

$$TXOP_p(K) = (m+y)TXOP_{MSDU}(K) \quad (5)$$

We refer the interested reader to [12] for more details of the wireless channel model and the link layer scaling (adapting the modulation order and code ration to spread the transmission over time) and sleeping optimization (introducing as much as possible transmission idle periods). In [12] it is shown how to obtain a schedule to optimize the communication energy cost by leveraging these scaling and sleeping techniques, while working in the dimensions Energy, TXOP and PER. And the sleeping is achieved by piggyback the wakeup time info onto contention free polls defined in IEEE 802.11e protocol.

2.2. Architecture of the deployed scalable video coder

We consider a scalable video codec based on Motion Compensated Temporal Filtering (MCTF). After the removal of the temporal redundancies, the frames are decomposed spatially by performing the Discrete Wavelet Transform (DWT). In a typical MCTF-based video compression, the rate allocation of the scalable bitstream is performed with a maximum granularity of one Group of Pictures (GOP). This creates natural independent data units.

Taking only quality (SNR) scalability into account and assuming a stable channel during one GOP time period, it is possible to calculate the expected distortion contribution of each quality layer on a GOP-per-GOP basis. The embedded bitstream of 3D MCTF wavelet coding has a sequential dependency, and each layer can be decoded only under the condition that all the former layers have been received, assuming there is no error concealment used at the decoder side. Suppose each GOP is encoded into *l* quality layers and the quality layers are the smallest application data units. Let D_i denote the distortion corresponding to the reception of layers 1 to i, and D_0 denote the distortion associated with losing the first layer. Denoting the error probability of layer *i* under configuration K_i as PER_{K_i} , the probability of correctly receiving the quality layers until layer *i* then writes $\prod_{j=1}^{i}(1-PER_{K_j})$. Relying on the sequential dependency of the embedded sub-streams' structure, the expected average distortion D_e over one GOP can then be calculated as:

$$D_e = PER_{K_1}D_0 + \sum_{i=1}^{l-1}\prod_{j=1}^{i}(1-PER_{K_j})PER_{K_i+1}D_i + \prod_{i=1}^{l}(1-PER_{K_i})D_l \quad (6)$$

Depending on the length of layer *i*, the associated energy cost under configuration K_i is $E_{p_i}(K_i)$, yielding the energy E_{GOP}, transmission time $TXOP_{GOP}$ of the whole GOP as: $E_{GOP} = \sum_{i=1}^{l} E_{p_i}(K_i)$ and $TXOP_{GOP} = \sum_{i=1}^{l} TXOP_{p_i}(K_i)$

3. PROBLEM FORMUATION OF ENERGY EFFICIENT MULTI-USER CROSS-LAYER OPTIMIZATION

Most of the former network allocation researches focused on maximizing the throughput, the number of flow admitted and rate etc. These solutions are not scalable to the rate-distortion properties brought by video bitstreams, and therefore often lead to inferior network efficiency and suboptimal resulting qualities for the video users. From the former analysis, and under the assumption that every video users can require their own end-to-end quality, we reformulated the optimization problem with video quality to be one of the constraints. For n users inside the network, the optimization problem is formulated as a min-max problem to, find for each of the user *i*, the configuration K_i such that:

$$K_i = \arg\min(\max E_{GOP_i}(K_i)), i = 1,...,n$$

Subject to: $D_e(i) \le D^r$, *i=1,..., n*

$$\sum_{i=1}^{n} TXOP_{GOP_i} \le T^{\max},$$

Where D^r and $T^{\max}$ denotes the distortion and time constraint respectively. Each user experiences different channel and application dynamics, resulting in different system states over time. It is this important characteristic which makes it possible to exploit multi-user diversity for energy efficiency. To decrease the run time complexity, we propose a two-phase solution approach. At ***design time***, for each possible system state, the optimal operating points are determined according to their minimal energy cost and resource (TXOP) consumption, for a given distortion constraint. To that end, we introduce the Pareto concept for multi-objective optimization from microeconomics [5]. Compared to a convex hull approach, this *Pareto Frontier* enables more feasible settings at run-time.

a) Initialization: Allocate to each of the n user the smallest cost possible for the given state, $E_{GOP_i}^{\min}$. Construct an n-value energy level vector, each of these values corresponding to one of the users' energy cost.
b) If $\sum_{i=1}^{n} TXOP_i^0 > T^{\max}$, for the user which require the smallest energy cost in this step, sorting out its Energy-TXOP trade-off curve, until finding a setting which energy cost exceeds the second small energy cost level or the resource constraint is satisfied.
c) While resource constraint not satisfied, update the energy level vector and repeat b until resource constraint satisfied.

Table I Run-time greedy water-filling algorithm

After the system state of all the users is known at runtime, a lightweight scheme is proposed relying on the Pareto property to assign the best system configuration for each user. First, convert the 3D Pareto frontier to a 2D Pareto curve by pruning those settings that cannot satisfy the QoS constraint. The remaining Energy-TXOP trade-offs are further explored to make a Pareto-curve with energy cost in ascending order. A greedy water-filling algorithm is proposed to solve the run-time searching for the best bandwidth allocation. The implementation of the algorithm is shown in Table I. The resulting outputs are the optimal settings over all users. Assuming each of the n users maintains N Energy-TXOP Pareto settings, the complexity of the water-filling algorithm is $O(nN^2)$. In our experiments, it turned out after Pareto pruning, N is normally smaller than 10, which makes the complexity of this step almost negligible.

4. NUMERICAL RESULTS

4.1. Simulation setup

In the experiments, a GOP size of 16 was used. Four sequences (Bus, City, Foreman, Mother and daughter) are presented here as examples of video with various motion activities (i.e., various rate-distortion properties). All the sequences are at CIF (352x288, 4:2:0) resolution and 30 frames per second. We encoded every sequence around 36 dB for the full-length bitstream, and the number of quality layers was set to 5. The intermediate bitstreams' rates (quality layers) are as follows: 256, 384, 448, 512, 1028 kbps for bus sequence; 64, 128, 256, 384, 448 kbps for City sequence; 96, 112, 192, 288, 384 kbps for Foreman sequence; 64, 80, 96, 112, 128 kbps for Mother and Daughter sequence. These quality layers also composed the truncation points of the bitstream at run time transmission. Since network congestion's influence is not the exploration focus of current paper, we limit the number of users so that every user's requirement can be satisfied with the available network bandwidth (notice congestion can be solved with the same optimization by considering drop packets as one of the transmission strategies that with energy cost and transmission time equal to zero). MSDU size is set to 1500 Bytes. The maximal MSDU retransmission time is limited to 10. The mobile devices are uniformly distributed from the AP with the radius of 10m. And the transceiver energy consumption unit is shown in Joule.

In order to evaluate the relative performance of the proposed approach, we provide results for the three following transmission strategies: ***(SoA reference point):*** The transceiver uses the highest feasible modulation and code rate that will successfully deliver the packets. After that, it switches to sleep. This approach is proposed by commercial 802.11 interfaces [9], which aims to maximize the sleep duration. ***(Constant PER, our previous work):*** in this transmission strategy, we use our approach to transmit every video packet with a configuration resulting in a PER smaller than 1e-2 (experiments results show that with the transmission PER smaller than 1e-2, the video transmission can be regarded as error free). In both the aforementioned strategies, we transmit the packet in each users' GOP until the transmitted bitstream reaches the quality required by the user. Significant energy decrease can be observed with respect to the SoA approach. ***(Expected PSNR, proposed method):*** in this transmission strategy, we introduce the expected visual distortion into the design time Pareto frontier calculation. From the results, we see that the transceiver energy is always smaller than the aforementioned two transmission strategies. When simulating on a channel environment similar to a dynamically varying one, another factor 2 can be achieved compared to the constant PER approach.

4.2. Result analysis

Our simulations use the Markov channel model [3] with increasing channel numbers corresponding to worse channel conditions. In this first set of simulations, fixed channel states have been used. Figure 1 shows the energy cost with one user down streaming the Foreman sequence with quality PSNR 35 dB. Clearly, in any channel conditions, our proposed Expected PSNR schedule outperforms all other schedules by at least a factor 2. For good channel states (low number channel states), the energy gains more than factor 8, while over bad channel states (high number ones) the energy gains around factor 2.

Moreover, Expected PSNR and Constant PER schedules behave similarly for good channel states, while Constant PER and SoA schedules behave similarly worse than Expected PSNR for bad channel states: to maintain a low packet loss probability, the bad channels require more transceiver energy consumption (e.g., more MAC layer retransmissions are needed). With the Expected PSNR approach, by reducing the low PER requirements of low importance video packets, the energy cost naturally lowers down.

Figure 2 shows the energy costs for four video sequences transmitted simultaneously over the time variant channel. Obviously, a weighted averaging phenomenon occurs over the different static channel states, achieving an overall average energy gain of Expected PSNR of a factor 2 over Constant PER, which itself is a factor 2 better than the SoA approach.

The impact of the different user QoS requirements' on the overall energy cost is shown in Figure 3. The four different sequences with QoS requirements 35dB, 33dB and 31dB are tested simultaneously on the time variant channel. Clearly, lower PSNR QoS requirements lead not only to lower energy requirements, but also to larger relative energy gains of the Expected PSNR scheduler, compared to others, yielding a larger scalability in energy gains at lower requested quality.

6. CONCLUSIONS

We have introduced an application+wireless cross-layer optimization framework to fairly minimize the wireless transceiver energy consumption for simultaneously multiple video streams downloading over a WLAN. Relying on the IEEE 802.11 standard and scalable video coding, the proposed solution optimally schedules the packets by both optimizing the link layer scaling and sleeping scheme and introducing rate-distortion properties of video sequences into the scheduling optimization framework. Results have shown that, in comparison with a state-of-the-art approach and with our previous work, the proposed PSNR target approach achieves the energy cost with at least a factor 2 according to the user requirements.

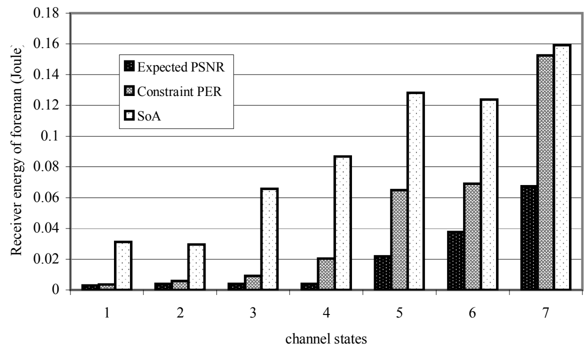

Figure 1 Impact of channel states to transceiver energy

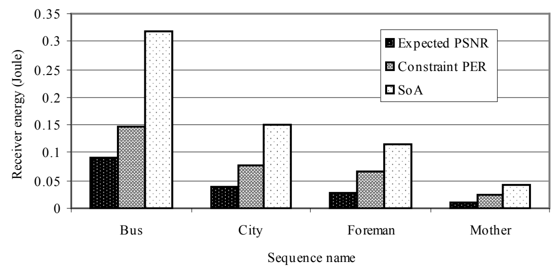

Figure 2 Impact of time variant channel to multi-user scheduling

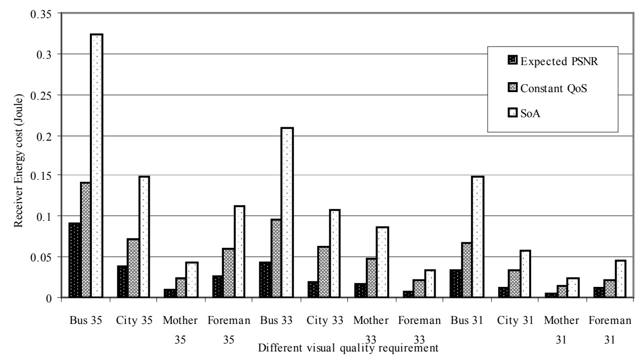

Figure 3 Impact of time variant channel and multi-user to transceiver energy of different schedulers

REFERENCES

[1] A. Sinha et al., "Energy Scalable System Design," *Trans. On VLSI Systems*, pp. 135-145, Apr 2002

[2] A. J. Goldsmith, "The Capacity of Downlink Fading Channels with Variable Rate and Power," *IEEE Trans. On Veh. Tech.*, Vol.46, No.3, 1997.

[3] B. Bougard, et al, "Energy-Scalability Enhancement of Wireless Local Area Network Transceivers," *IEEE SPAWC*, Lisboa, July 11-14, 2004.

[4] B. Bougard, et al., "A scalable 8.7-nJ/bit 75.6-Mb/s Parallel Concatenated Convolutional Turbo Codec," *IEEE ISSC*, Feb. 2003

[5] Giguan G. Lin, "Multiple-objective Problems: Pareto-Optimal Solutions by Method of Proper Equality Constraints," *IEEE Trans. Automatic Control*, Vol. AC-21, No. 5, G Oct. 1976

Figure 1 Impact of channel states to transceiver energy

[6] J. Xu, et al, "Three-dimensional embedded subband coding with optimized truncation (3D ESCOT)," *Applied and Computational Harmonic Analysis*, vol.10, pp.290-315, 2001.

[7] H. Radha, et al, "Scalable Internet video using MPEG-4," *Signal Processing: Image Communication.*, vol. 15, no. 1-2, pp. 95-126, Sep. 1999

[8] Q. Li, et al, "Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation," *IEEE Trans. Multimedia*, vol. 6, no.2, pp. 278-290, Apr. 2004

[9] Atheros White Paper, "Power Consumption & Energy Efficiency Comparisons of WLAN Products", 2003.

[10] Skyrianoglou, D, et al, "Traffic scheduling for multimedia QoS over wireless LANs," *IEEE ICC*, 2005.

[11] R. Mangharam, et al, " Optimal fixed and scalable energy management for wireless networks," *(INFOCOM'05)*, Miami, Fl, USA, March 2005.

[12] S. Pollin, et al, "MEERA: cross-layer methodology for energy-efficient resource allocateion for wireless networks" to appear in *IEEE Transactions on Wireless Communications*.