Energy-Aware MPEG-4 FGS Streaming

Kihwan Choi[1], Kwanho Kim[2], and Massoud Pedram[1]
[1]Dept. of EE-Systems, Univ. of Southern California, Los Angeles, CA 90089, {kihwanch, pedram}@usc.edu
[2] School of Computer Science and Engineering, Seoul National University, Korea, khkim@cselab.snu.ac.kr

Abstract

In this paper, we propose an energy-aware MPEG-4 FGS video streaming system with client feedback. In this client-server system, the battery-powered mobile client sends its maximum decoding capability (i.e., its decoding aptitude) to the server in order to help the server determine the additional amount of data (in the form of enhancement layers on top of the base layer) per frame that it sends to the client, and thereby, set its data rate. On the client side, a dynamic voltage and frequency scaling technique is used to adjust the decoding aptitude of the client while meeting a constraint on the minimum achieved video quality. As a measure of energy efficiency of the video streamer, the notion of a normalized decoding load is introduced. It is shown that a video streaming system that maintains this normalized load at unity produces the optimum video quality with no energy waste. We implemented an MPEG-4 FGS video streaming system on an XScale-based testbed in which a server and a mobile client are wirelessly connected by a feedback channel. Based on the actual current measurements in this testbed, we obtain an average of 20% communication energy reduction in the client by making the MPEG-4 FGS streamer energy-aware.

Categories and Subject Descriptors

H.4 [Information systems]: Information systems applications.

General Terms

Algorithms, Measurement, Experimentations.

Keywords

Low energy, video streaming, MPEG4-FGS, dynamic voltage and frequency scaling.

1 Introduction

With the availability of mobile, communication and computing systems, we have seen an explosive demand for wireless multimedia, e.g., streaming video. This trend in turn poses two challenges: (1) establishing and maintaining a stable channel for real-time operation and (2) power-aware operation so as to increase the lifetime of the battery-powered mobile system while meeting a minimum quality of service (QoS) requirement. Furthermore, it is desirable to provide a mechanism for a graceful degradation in the video quality as the system trades off QoS for higher energy efficiency. Fine Granularity Scalability (FGS) coding technique [1], which was adopted as the standard in MPEG-4, indeed provides an effective mechanism for graceful video quality degradation based on its hierarchical layer structure, which consists of a base layer and one or more enhancement layers. Although extensive studies have been conducted on the hierarchical layer structure of MPEG-4 and its error resiliency under fluctuations in the channel bandwidth [2][3][4], energy efficiency in server-client system has received little attention.

Indeed, this topic is the focus of the present paper, where we show how to use FGS feature of MPEG-4 standard to achieve high energy efficiency subject to meeting a minimum video quality constraint.

For video streaming application, there are two sources of energy consumption in a wireless mobile client: communication energy for transmitting and receiving data and computation energy for processing the received data at the client. These energy consumptions are proportional to the amount of data in a video frame. When a mobile client is prepared to accept (or tolerate) a low video quality (a condition that may arise due to low levels of remaining energy in the client, lack of a need to receive high quality video due, etc.), then it should be able to communicate this looser performance requirement to the server, which will in turn reduce the amount of data that is sent for a given video frame. Consequently, a significant amount of energy can be saved in the mobile client who need not receive and decode additional packets sent by the server. An energy-efficient wireless video streaming system can thus be achieved by creating feedback from the mobile client to the server and by enabling the server to adjust the amount of information that it puts into the channel.

The *decoding aptitude* of a mobile client is defined as the amount of data that its decoder can decode in a given deadline. It is frequently the case that the client can change its decoding aptitude by employing methods such as dynamic voltage and frequency scaling (DVFS) or dynamic power management (DPM) [5]. Many of the state-of-the-art processors that are designed for mobile application are equipped with DVFS for low-power operation [10].

In this paper, we propose a client-feedback wireless video streaming system in which the decoding aptitude of the client is considered by the server when it determines the amount of data that is sent for a given frame. This then results in trading video quality for lower energy consumption in the client. The streaming system must also guarantee a minimum achieved video quality. The proposed technique is made possible by the FGS feature of the MPEG-4 standard as will be described later. The proposed energy-saving technique has been implemented on a XScale-based platform and has resulted in 20% reduction in communication energy in the client.

The remainder of this paper is organized as follows. Related works on energy consideration of streaming system and MPEG-4 FGS are described in Section 2. In Section 3, a client-feedback streaming system is presented. Details of the implementation, including both hardware and software, are described in Section 4. Experimental results and conclusion are given in Sections 5 and 6, respectively.

2 Background

2.1 Fine Granularity Scalability (FGS)

To adapt to the time-varying channel capacity (which is due to dynamic channel condition change such as congestion or fading), scalable video coding schemes are used. Example techniques include SNR scalability, temporal scalability, and spatial scalability in MPEG-2 and MPEG-4. In these layered scalable coding technique, total encoded bit-streams consist of a base layer and several enhancement layers. The bit-rate of base layer is determined

This research was supported in part by DARPA PAC/C program under contract DAAB07-02-C-P302 and by NSF under grant no. 9988441.

DAC 2003, June 2-6, 2003, Anaheim, California, USA.
Copyright 2003 ACM 1-58113-688-9/03/0006…$5.00.

by the minimum channel bandwidth and is sufficient to ensure a minimum achievable video quality. The enhancement layer provides higher video quality when the channel has extra bandwidth for the transmission of this layer. The FGS video coding technique, which was adopted as the standard in MPEG-4, provides for a very smooth video quality variation compared to other scalable coding technique because any number of bits in the enhancement layers may be truncated according to the channel condition. In FGS, the bit-plane coding of the DCT coefficients difference between the original frame and the reconstructed frame using base layer only is used to generate an enhancement layer [1]. The maximum number of bit-planes in a frame is given by $\log_2$ ($|\text{DCTcoef}|_{\max}$)+1 where $\text{DCTcoef}_{\max}$ is the maximum DCT coefficient difference in a frame. In wired (or wireless) video streaming with a server and a client, the data rate (in bits/sec), R_{send}, at the server is determined by the available bandwidth in the network (or channel) as $R_{send} = R_b + R_e$ where R_b denotes the base layer bit-rate and R_e denotes the enhancement layer bit-rate. R_b should be less than the minimum allowed bandwidth while R_e varies according to channel conditions. R_{send} is set to provide the optimum video quality by transferring the maximum amount of video data to the client subject to the existing channel conditions. The higher the bandwidth of the channel is, the better the quality of the video is.

2.2 Wireless Environment in Video Streaming

For energy-efficient video streaming in a wireless environment, two important factors should be considered: the energy consumption, which impacts the system lifetime and the channel bandwidth, which impacts the video quality. Traditional low-power design techniques have mainly focused on minimizing the energy consumption due to charging/discharging of capacitive loads in a CMOS circuit which is performing computations or making decisions. As the battery-powered mobile devices became connected through a wireless network, the communication energy cost for transmitting and receiving data may dominate the computation energy cost. The video streaming application, which is one of the popular applications for mobile devices, results in rather large communication energy cost. To achieve a low energy video streaming system, more consideration must be given to optimizing the energy consumption due to the video data transmit and receive over the wireless channel.

The communication energy is the sum of two energy dissipations: the transmit energy (E_{TX}) and the receive energy (E_{RX}), which are given by [6]: $E_{TX} = K_p*(S*\alpha_{TX} + \beta_{TX})$ ad $E_{RX} = K_p*(S*\alpha_{RX} + \beta_{RX})$ where K_p denotes the number of transferred packets, S denotes the packet size, and α_{TX}, β_{TX}, α_{RX}, and β_{RX} are regression coefficients for the transmit and receive energy dissipation costs, respectively. Notice that E_{TX} and E_{RX} in the above equations are proportional to the amount of data that is sent and received. E_{SERVER} and E_{CLIENT} for video streaming can be written as:

$$\begin{aligned} E_{SERVER} &= E_{COMM\_SERVER} + E_{COMP\_SERVER} \\ &= Kp*(S*\alpha_{TX} + \beta_{TX}) + C_S\, V_S^2\, f_{CPU\_S}\, T \\ E_{CLIENT} &= E_{COMM\_CLIENT} + E_{COMP\_CLIENT} \\ &= Kp*(S*\alpha_{RX} + \beta_{RX}) + C_C\, V_C^2\, f_{CPU\_C}\, T \end{aligned}$$

where C_S and C_C denote the switched capacitance per clock cycle time, V_S and V_C are the supply voltage level (assuming full swing transitions), f_{CPU_S} and f_{CPU_C} are the clock frequency of the server and client, and T is the total streaming time needed to transmit/receive K_p packets.

Let M denotes the decoding aptitude of the client. M can be calculated as: $M = k*f_{CPU\_C}*D$, where k is a constant conversion factor and deadline D denotes the inverse of the target frame rate of the video streamer. Assuming that f_{CPU_C} scales linearly with the operating voltage V_C, the computation energy consumption at the client is given as: $E_{COMP\_CLIENT} = \gamma*M^3$ where γ denotes a hardware-dependent constant. Thus, a client can greatly reduce its energy consumption by decreasing its M.

In a wireless environment, the channel conditions vary widely over time and space in several ways due to factors such as noise, interference and multi-path fading. These factors can lead to errors in packets resulting in variation in throughput of a wireless channel. Commonly used methods for reliable packet transmission through a wireless channel are *Automatic Retransmission reQuest* (ARQ) and *Forward Error Correction* (FEC). ARQ basically means retransmission of corrupted packets where the expected number of retransmissions for a packet is 1/(1-p) with p denoting the packet loss rate. In FEC, some redundant information about each packet is added by the encoder so that some of packets that incur errors may be recovered based on the redundant information. In general ARQ is more suitable for uni-cast operation with low error rate whereas FEC is more effective in multi-cast operation with many nodes. By using these two methods, a certain degree of variation in channel throughput depending on error occurrence can be tolerated. Using ARQ, channel throughput variation due to packet error rate may be calculated based on IEEE 802.11b protocol that is widely used in wireless LAN [7]. Let t_{TX} denotes the time to transmit a packet without error. The expected time required for reliable transmission of a packet with an average error rate of p is given by t_{TX}/(1-p). In video streaming application with a deadline D for every frame, an increase in the data communication time due to potential retransmissions causes a reduction in the channel throughput and finally leads to degradation of the video quality. If the server sends K packets during D, the number of correctly arrived packets at the client, A, is given by:

$$A = \begin{cases} K & if\ \dfrac{t_{TX}}{1-p}K \le D \\ \dfrac{D}{t_{TX}}(1-p) & otherwise \end{cases}$$

3 A Client-feedback Video Server

When designing an energy-aware video streaming system, communication and computation energy and the video quality must be considered. We examine a video streaming system, in which a server and a mobile client exist and the client can control its decoding aptitude M according to its energy management policy. The server employs no such policy because we assume that it is AC-powered and therefore, it has an inexhaustible energy source. The mobile client is battery-powered, and therefore, E_{CLIENT} is dependent on the energy management policy of the client. For example, depending on the remaining battery capacity, a client may lower its operating frequency to extend its battery lifetime. So, depending on the requirements imposed, a client can employ a variety of energy saving means such as dynamic voltage scaling and power mode change. In a typical video streaming system between a server and a mobile client, both high video quality and low energy consumption on the client side are required. Now when the client changes its M value, the ratio between A and M becomes quite important. When A is larger than M, the communication energy for handling a total of A-M packets is wasted. On the other hand, when A is smaller than M, the server can send more packets to improve the video quality if the channel conditions permit.

Definition: *The normalized decoding load at time instance i (denoted by N_i) for the client is defined as the ratio A_i/M_i. Notice that N_i captures the amount of wasted energy at instance i.*

Theorem: *For an energy-efficient streaming system, N_i should be set to one such that the* optimal *video quality can be provided with no energy wasted handling useless data in the client.*

Let B = D/ t_{TX} denote the maximum bandwidth. To estimate how much energy is wasted when $N_i > 1$, we calculated the wasted energy while changing the M_i/B ratio from 0.1 to 1 for different p values: 0.02, 0.06, 0.1, 0.13 corresponding to bit error rates (BER) of 1e-5, 3e-5, 5e-5 and 7e-5, respectively. We used the regression coefficients values for communication energy calculation from reference [6] at 11Mbps data rate and assumed that 30 packets with size of 256-byte are sent in 100msec. Results are shown in Fig.1 which shows that a significant energy waste occurs as M_i becomes smaller than B for a given packet error rate.

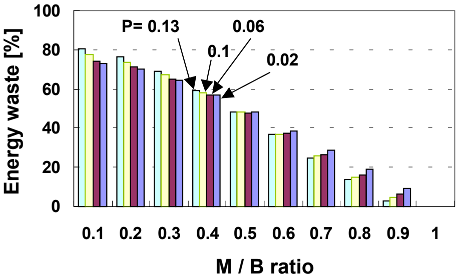

Figure 1. Communication energy waste when $N_i > 1$.

The mismatch between A_i and M_i usually arises from lack of information exchange between the server and the mobile client. To gain better video quality without wasting energy, a server must send the optimal amount of data in a given deadline considering both the channel throughput and the decoding aptitude of the client. However, while channel throughput can be determined by using network parameters such as the packet loss rate, BER, and transmission delay [9], it is difficult for a server to figure out the decoding aptitude of the client. One way to achieve this goal is that the client periodically sends a small packet that contains information about its internal state (e.g., high performance or low power) to the server. In Fig.2 a video streaming system with a feedback path from the mobile client to the server is shown. A status packet is periodically sent to the server at regular time intervals.

If the client's decoding aptitude is changed during the i^th^ frame, the server should also change A_{i+1} to min[$M_{i,}$ B*(1-p_i)] such that energy waste can be avoided while enabling the client to process the data in the i+1^st^ frame. A_{i+1} is set to M_i when there is enough bandwidth in the channel, but if M_i is larger than the current maximum bandwidth, then A_{i+1} is set to B*(1-p_i). A_{i+1} calculated in this way has some error since M_i from the client cannot give exact values for the i+1^st^ frame time. As a heuristic to reduce average error, we can predict M_{avg} based on the history. Using an exponentially weighted average technique with a weighting factor α, A_{i+1} is calculated as follows:

$$A_{i+1} = \min[M_{avg},\ B_i \cdot (1 - p_{i+1})]$$

$$where,\ M_{avg} = \alpha \cdot \sum (1-\alpha)^{\tau} \cdot M_{i-1}$$

In order to verify the effectiveness of our proposed streaming policy we performed a simulation. The wireless channel is modeled as a Gilbert-Elliot model [8], with two states, good and bad, which represents the channel conditions during data transmission. BER of good and bad state are set to 1e-5 and 1e-4, respectively. Packet size of 256 bytes and frame count of 500 are assumed. For the maximum bandwidth of the channel model, we assumed 11Mbps IEEE 802.11b network protocol. For varying M, we generated a trace in which M is set to 0.8, 0.4, 1.2, 0.5, 0.7, and 1.4 of the maximum bandwidth, B. Fig.3 shows variation in N for different M values over the 500 frames. From this figure, feedback from the mobile client is essential to avoid energy waste by keeping N to close 1. N changes abruptly as M changes since M_{avg} is predicted using the weighted average method. If there are no severe fluctuations in M, calculating A_{i+1} using M_i may be better as explained before. Energy waste with M variation is summarized in Table.1. When M < B, a significant amount of energy is wasted while there is no video quality improvement because low M limits the video quality. Although these results may be different for different M values and channel conditions, they are representative of the general case.

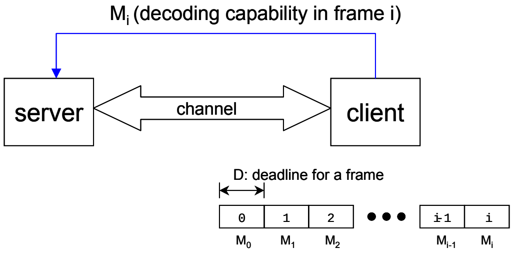

Figure 2. A video streaming system with feedback.

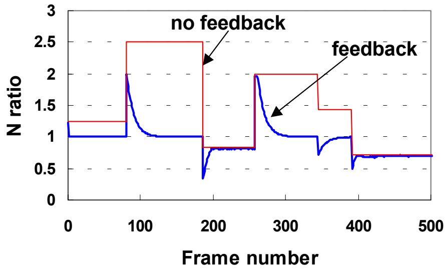

Figure 3. Comparison of the N ratio.

Table 1. Comparison of energy waste.

	Energy waste	
	No FB	FB
0.8B	18.74%	0.21%
0.4B	57.35%	2.57%
1.2B	0%	0%
0.5B	48.49%	6.27%
0.7B	28.69%	0%
1.4B	0%	0%

4 Implementation

We implemented an MPEG-4 FGS streaming system on a high-performance test bed. The hardware used is the Intel's Xscale processor which supports 9 different frequencies from 200MHz to 733MHz. A D/A converter was used as a variable operating voltage generator to control the reference input voltage to a DC-DC converter that supplies operating voltage to the CPU. Inputs to the D/A converter were generated using customized CPLD logic. When the CPU clock speed is changed, a minimum operating voltage level should be applied at each frequency to avoid a system crash due to increased gate delays. In our implementation, these minimum voltages are measured and stored in a table so that these values are automatically sent to the variable voltage generator when the clock speed changes. Voltage levels mapped to each frequency are distributed from 0.9V @200MHz to 1.5V @733MHz. For the software work, Microsoft reference MPEG-4 FGS encoder/decoder was modified to fit our purpose. Two generated bit-streams of QCIF video sequence with 150 frames, a base layer and a FGS enhancement layer with 5 bit-planes (bp0~bp4), are split into packets with size of 256-byte. RTP/RTCP on UDP was used as a network protocol between server and client. We set our Linux machine as the AC-powered server and our test bed as the mobile battery-powered client, which is equipped with an IEEE 802.11b wireless LAN (WLAN) card. Energy consumption of the WLAN interface was measured using our data acquisition (DAQ) system.

5 Experimental Results

We considered six different scenarios as follows: (I) base only, (II) base + bp0, (III) base + bp0 + bp1, (IV) base + bp0 + bp1 + bp2, (V) base + bp0 + bp1 + bp2 + bp3, and (VI) base + bp0 + bp1 + bp2 + bp3 + bp4, where bpX denotes bit-plane X. Each test case represents a different decoding aptitude for the client. In Fig. 4, the video qualities for each case were shown using average peak signal noise ratio (PSNR), which is a widely used estimate of the quality of a reconstructed image compared with the original image. The video quality increases as the number of decoded bit-planes increases. In theory, FGS provides continuous levels in image quality by truncating any number of bits in an enhancement layer, but a bit-plane is considered one level of image quality for practical purpose. Fig.5 shows the current drawn from 5V supply of the WLAN when receiving packets for a frame. If more data are sent in a given deadline, more receiving energy is required. For example, with frame rate of 10, 125.2mJ and 133.4mJ for a frame were consumed for cases (V) and (VI), respectively. We measured decoding aptitude variation depending on the CPU's operating frequency, but less than 5 fps rate was achieved even at the highest CPU frequency (733MHz). This is because the MPEG-4 decoder that we used is not optimized to run on the Xscale hardware. So we measured the CPU energy consumption at 733MHz and normalized it to 10 fps rate. Fig.6 shows the energy consumptions of the WLAN card and the CPU, which capture the bulk of the communication and computation energies in the client, respectively. Fig.7 shows energy waste in each test case. Let VQ(i) denotes the video quality for case (i). VQ(I) is set by the client; in this case, for example, the wasted communication energy is 25.4mJ, which is about 40% of the CPU energy for VQ(I). Using our proposed method, about 20% of the communication energy may be saved.

6 Conclusion

A client feedback scheme is proposed to reduce redundant energy consumption in a wireless video streaming system consisting of a server and a client. The server does not send more data to the client when it cannot handle more than a given data rate, which is in turn determined by the client based on its own local energy management policy. From analysis and based on actual measurements under different channel conditions and decoding aptitudes for the mobile client, it was demonstrated that a video streaming system that maintains its normalized decoding load at unity produces the optimum video quality with no energy waste.

Acknowledgement: The authors would like to thank Prof. N. Chang of Seoul National University for his assistance.

References

[1] W. Li, "Overview of Fine Granularity Scalability in MPEG-4 Video Standard," IEEE Trans. On Circuits and Systems for Video Technology, Vol.11, No. 3, March 2001.

[2] M. van der Schaar, H. Radha, and C. Dufour, "Scalable MPEG-4 Video Coding with Graceful Packet-loss Resilience over Bandwidth-varying Networks," ICME 2000, vol.3, pp.1487-1490, 2000.

[3] R. Cohen and H. Radha, "Streaming Fine-Grained Scalable Video over Packet-Based Networks," GLOBECOM'00. IEEE, pp.288-292, 2000.

[4] R. Yan, F. Wu, S. Li, and R. Tao, "Error resilience methods for FGS video enhancement bitstream," The First IEEE Pacific-Rim Conference on Multimedia (IEEE-PCM 2000),Dec. 13-15, 2000 Sydney, Australia.

[5] T. Pering, T. Burd, and R. Broderson, "The simulation and evaluation of dynamic voltage scaling algorithms," International Symposium on Low Power Electronics and Design, pp.76-81, 1998.

[6] L. M. Feeney and M. Nilson, "Investigating the Energy Consumption of a Wireless Network Interface in an Ad Hoc Networking Environment," IEEE INFOCOM 2001, vol.3, pp.1548-1557, 2001.

[7] IEEE, IEEE Std 802.11b – Wireless LAN Medium Access Control (MAC) and Physical layer (PHY) Specifications: High Speed Physical Layer (PHY) in the 2.4 GHz Band, 1999.

[8] J. R. Yee and E. J. Weldon, "Evaluation of the performance of error correcting codes on a Gilbert channel," IEEE trans. Comm. vol.43, no.8, pp.2316-2323, 1995.

[9] G. Wang, Q. Zhang, and Wenwu Zhu, "Channel-adaptive error protection for scalable video over channels with bit errors and packet erasures," GLOBECOM'01. IEEE, vol.3, pp.2045-2049, 2001.

[10] http://www.transmeta.com .

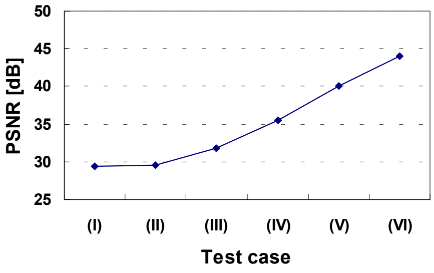

Figure 4. Video quality as a function of no. of bit-planes.

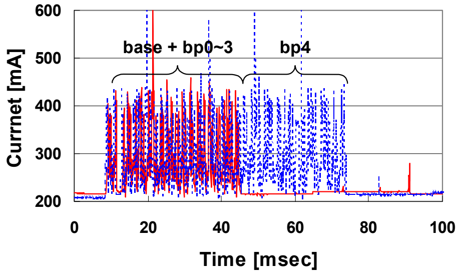

Figure 5. Current in WLAN interface when receiving packets.

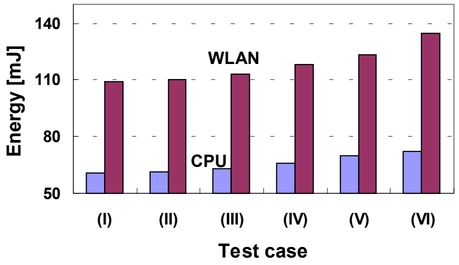

Figure 6. Energy consumption of the client.

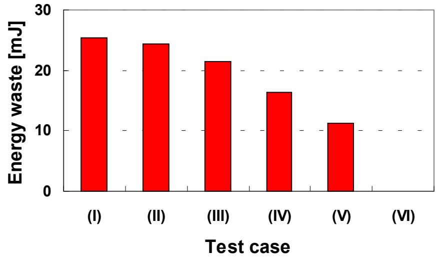

Figure 7. Energy waste with different decoding aptitudes.