OPTIMAL CODING RATE AND POWER ALLOCATION FOR THE STREAMING OF SCALABLY ENCODED VIDEO OVER A WIRELESS LINK

Virgilio Rodriguez, David J. Goodman, Yao Wang

ECE Department, Polytechnic University
5 Metrotech Center, Brooklyn, NY 11201
E-mail:vr@ieee.org

ABSTRACT

Scalably encoded information results in files which can be truncated at an arbitrary point and decoded, as supported by the JPEG-2000 (image) and MPEG-4 (video) standards. This work introduces a tractable, yet flexible analytical model for resource management involving scalably encoded video. Each segment of video of a predetermined length yields a file that can be truncated and decoded independently of other segments. The problem is set up as a joint optimization of transmission power, and coding rate (where to truncate?). The analysis reveals that any one of these variables uniquely determines the other. The terminal should truncate the file at the point that maximizes quality per unit of power employed.

1. INTRODUCTION

Modern media encoders, such as those in the the JPEG 2000 (still images) and MPEG-4 (video) compression standards, support scalability. Fine granular scalability produces an "embedded" bit stream, which can be truncated at an arbitrary point, and decoded, leading to various levels of reproduced media quality. Video scalability can be achieved along various dimensions, including SNR, spatial (size), temporal (frame rate), and frequency; and these scalability modes may be combined [6, Ch.11].

In the present work, the model introduced in [4] for still images is extended to consider the transfer over a wireless link of scalably encoded video. Each T secs of video leads to a Y-bit embedded bit stream, which is independent of the other segments. For example, T may correspond to one group of pictures (GOP), or several GOPs, in video coded according to MPEG standards. An energy-limited terminal seeks to jointly optimize both the truncation point of the embedded bit stream (coding rate), and its transmission power.

Supported in part by NYSTAR through the Wireless Internet Center for Adv. Technology (WICAT), and by the National Science Foundation Grant No. 0219822

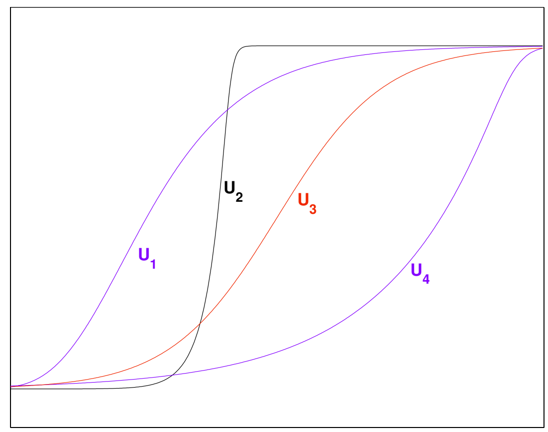

Fig. 1. Some S-curves

We postulate that *all that is known* about the function yielding the "utility" or "quality" of the resulting video segment in terms of the number of bits in the truncated file (coding rate) is that its graph is an S-curve. As shown in fig. 1, this family of curves contains as special cases ("mostly") concave curves (e.g, U_1), ("mostly") convex curves (e.g., U_4), and smoothed out "step" functions (e.g. U_2). And the "ramp" displayed by S-curves such as U_3, can express a near linear relation, over a range of interest. These shapes should accommodate most, if not all situations of interest. Other reasons for adopting this family are discussed in [4].

Another critical function is that giving the probability of success of the transmission of a data packet in terms of a signal to interference measure at the receiver. It can be safely assumed that for any physical layer, any such function has an S-shaped graph. Thus, two different S-curves are at the core of this analysis.

The scientific literature registers at least one previous use of the idea of maximizing end-user utility in video streaming in [1], later extended to [2]. But that work focuses on a wired network with renegotiable CBR services, does not consider scalability, and only considers a logarithmic utility function. There are also various works involving power allocation and the wireless transmission of video. Typically, power is minimized, and possibly other parameters are adjusted, while holding “end-to-end” distortion to an acceptable level. For instance, [7] specifically targets scalably-encoded video, while seeking an optimal power allocation, with joint source-channel coding. However, previous works seeking a joint power, and coding rate selection in order to maximize a video quality metrics within an analytical model appear unavailable.

Below, we describe the system model, and discuss more formally the key functions. Then, after formally stating the problem, we build and analytically solve an optimization model, and provide a numerical example. We conclude by discussing our results, and commenting on possible extensions.

2. CONCEPTUAL FRAMEWORK

2.1. System model

Fig. 2 shows schematically the system engaged in the wireless transmission of scalably encoded live video. Each T secs of video is encoded as a fully embedded bit stream of length Y, which may be truncated to length y. For $y \leq Y$, the reproduced video is imperfect. Its quality or utility is $u(y)$, with u an increasing function discussed below. The bit stream is broken up into packets. Each packet may have added error-control bits (error-control system *not* shown). These packets enter a large buffer prior to transmission. Packets are wirelessly transmitted at the rate of R bps. To ensure continuous video play out at the receiver, the actual transmission time alloted to the y bits corresponding to a given T-sec segment is $\Delta \leq T$ secs. (i.e., the coding rate cannot exceed $R\Delta/T$). A $\Delta < T$ may account for processing and propagation time not being modeled, and a certain “guard time”. The probability that a packet is successfully received is $f_s(x)$, with x the signal-to-interference ratio (SIR) at the receiver, which is determined by the chosen transmission power, any path loss, and the interference (noise) present at the receiver. The function f_s is discussed further below. Packets received in error which cannot be corrected result in ideal re-transmissions until correctly received and confirmed. Correctly received packets are placed in a large buffer. Other symbols shown in fig. 2 are discussed as introduced below.

2.2. Quality as a function of the compression rate

At the core of this inquiry is a function yielding the quality or utility of the decoded video as a function of the number of bits in the truncated encoded file. This function cannot

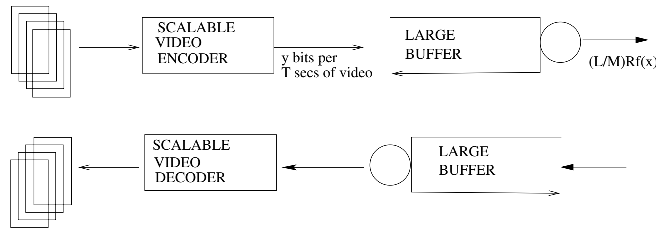

Fig. 2. Schematic of the wireless transmission of scalably encoded live video.

be derived; it is fully determined by the end-user, in the same way in which the “utility function” at the core of economic studies resides within the consumer. $u(y)$ should be obtained by psychophysical experimentation. We postulate that this function is such that its graph is an S-curve. Some of the implications of this assumption are discussed further in [4]. A fixed function could work for different video segments, in particular if the segments are sufficiently “similar” (e.g., each corresponds to different parts of the same sporting event). References [3, 5] discuss the technical characterization of a generic S-curve.

2.3. A Generalized frame-success function

The frame-success function (FSF) yields the probability that a data packet is received successfully as a function of the signal to interference ratio at the receiver. This function is determined by physical attributes of the system, including the modulation technique, the forward error detection scheme, the nature of the channel, and properties of the receiver. We assume that *all that is known* about the FSF, f_s, is that its graph exhibits a sigmoidal shape as in figure (3). For good technical reasons similar to those discussed in [3], $f(x) := f_s(x) - f_s(0)$ replaces f_s in the analysis below ($f_s(0)$ is generally very small, but not zero).

3. ANALYSIS

For our purposes, it is convenient to regard the wireless channel as if it was a deterministic channel producing the throughput that the actual channel produces on the average. Thus, we assume that, when the SIR at the receiver is x, $(L/M)Rf(x)$ information bits are received each second at the decoder buffer. The intuition is as follows. With a perfect channel, each packet would be filled with information bits (no ECC), and would be received successfully at first try. Thus, R information bits would be received each sec. However, with an imperfect channel, $M - L$ ECC bits are introduced in each packet, and still, on the average, only $nf(x)$ out of every n packets are received successfully.

Thus, an average of $(L/M)Rf(x)$ information bits are successfully transferred each second.

3.1. Problem statement

It is taken as given a (1) certain amount of energy, $\bar{E}$, available for transmission, (2) fixed transmission rate of R bits per second, (3) long sequence of files, each of length $y \leq Y$, each divided into packets of length $L \ll y$ and each corresponding to a video segment of length T secs which has been encoded scalably ($L - M$ error-control bits are added to each packet), (4) maximal time $\Delta \leq T$ secs. to complete the transmission of the y bits corresponding to a given T-sec segment (i.e., the coding rate cannot exceed $R\Delta/T$) (5) utility/quality function u as defined in section 2.2, (6) certain level of interference (noise), I, (7) frame-success function f_s as described in section 2.3.

The transmitter wants to choose optimally (i) the truncation point (coding rate) and (ii) the transmission power, in order to maximize the sum of the quality or utility of each one of the video segments that can be viewed at the receiver before energy runs out.

3.2. Objective Function

For a given level of desired quality, $\bar{u}$, there is a corresponding number of information bits, y, that produces this quality ($u(y) = \bar{u}$). Thus, the total number of information bits received successfully after Δ secs. must be not less than this y. And spending energy to exceed this level would be unwise, because it would decrease the total number of segments of quality $\bar{u}$ that are delivered before energy runs out. Thus, for given y and Δ, the terminal must choose its transmission power so that

$$\frac{L}{M}Rf(x)\Delta = y \tag{1}$$

There is one specific SIR value, $x(y)$, that satisfies eq. (1), and a specific transmitted power, $P(y)$, that yields the SIR $x(y)$ at the receiver. Thus, for a given Δ, y determines the transmission power.

The total amount of energy spent on the transmission of a video segment of quality $u(y)$ is $P(y)\Delta$. Thus, the total number of T-sec video segments of quality $u(y)$ that can be transferred with an energy budget of $\bar{E}$ is $\bar{E}/(P(y)\Delta)$. Then, the total quality viewed, which the terminal wishes to maximize, is

$$\frac{\bar{E}}{\Delta}\frac{u(y)}{P(y)} \tag{2}$$

For a fixed level of energy, $\bar{E}$, the terminal only needs to maximize $u(y)/(\Delta P(y))$ (quality per Joule), and if Δ is also fixed, just maximize $u(y)/P(y)$, the quality-to-power ratio (QPR).

3.3. Optimization Model and Solution

In view of the preceding analysis, the objective of the single user can be expressed as maximizing $u(y)/P(y)$. Assuming a CDMA technology, with a spreading gain of $G := R_c/R$ (chip rate over bit rate), channel gain of h, and interfering power I, the received SIR and the transmitted power are related as $x = GPh/I$. Thus, the terminal objective is equivalent to :

$$\begin{aligned} &\max_{x,y} \frac{u(y)}{x} & & \max_{x} \frac{u(Bf(x))}{x} \\ &\text{s.t. } y = Bf(x) & \text{OR} \quad & \text{s.t. } 0 \leq x \leq \bar{x} \\ &\quad 0 \leq x \leq \bar{x} & & \end{aligned}$$

where $B := (L/M)R\Delta$ and $\bar{x} := Gh\bar{P}/I$ with $\bar{P}$ the largest available transmission power.

With $u(Bf(x)) := h(x)$, the terminal should maximize the ratio $h(x)/x$. We expect that, as shown in fig. 3, the composite function $u(Bf(x))$ retain, in general, the S-shape of both u and f. As discussed in [3, 5], for *any* S-curve S, $S(x)/x$ is always maximized at x^*, the abscissa of the tangency point between the S-curve and a straight line that passes through the origin.

3.4. Numerical example

Fig. 3 summarizes a numerical example. We assume that it has been experimentally determined that, for this end-user, $u(y) = [1 + \exp((60 - y)/10)]^{-1}$ is the utility or quality function (plotted at the top of fig. 3). With x denoting SIR at the receiver, the frame-success function is assumed to be $f_s(x) = [1 - \frac{1}{2}\exp(x/2)]^{80}$ (whose graph is second from the top), which corresponds to non-coherent FSK modulation, no FEC and 80-bit packet size. Suppose that T secs of video can be scalably encoded, at full rate, to $Y = 100$ (in some multiple of bits). The parameters R, L, M, and Δ are such that $B = (L/M)R\Delta = 110$ (in the same unit as Y). The third subplot corresponds to the composite function $u(Bf(x)) := h(x)$, which retains the S-shape of both u and f. The terminal must choose its transmission power so that the ratio $h(x)/x$ (plotted at the bottom) is maximized. The maximizer is $x^* \approx 10.5$, and its matching truncation point is $y^* \approx 110 * f(10.5) = 88$. Thus, for this user, under this physical layer, the scalable file should be truncated to about 88% its size, leading to a per-segment video quality of about 94% that of the original.

4. DISCUSSION

We have investigated the problem faced by an energy-limited terminal transferring over a wireless link a long sequence of files, each corresponding to a segment of video which has been scalably encoded, as supported by the MPEG-4 standard. We have discussed a tractable analytical model, based

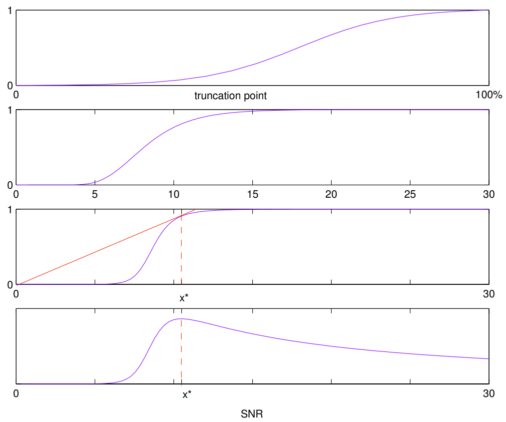

Fig. 3. From the top, (i) the S-curve $u(y)$ giving the perceptual quality of a video segment, as a function of the coding rate, (ii) $f(x)$, the probability of successful reception of a packet as a function of the SIR, (iii) the composite function $u(Bf(x)) := h(x)$, (iv) the ratio $h(x)/x$ which the terminal should maximize. For *any* S-curve S, $S(x)/x$ is always maximized at x^*, found at the tangency point between the S-curve and a straight line from the origin.

on two key functions: $u(y)$ which gives the perceptual quality or utility of a video segment as function of the coding rate, and $f(x)$, the packet success probability as function of the signal-to-interference ratio (SIR) at the receiver. By assuming that *all that is known* about these 2 functions is that they are S-curves, we are de facto allowing the possibility that ("mostly") concave, convex, "step", and linear functions play those roles (fig. 1). We have postulated that the terminal wishes to maximize the "cumulative utility" (or quality) from all the segments that reach the receiver before energy runs out. Our analysis has led us to maximize the quality-to-power ratio, which is equivalent to maximizing quality per Joule. Although we have set up the problem as a joint optimization of power and coding rate, our analysis indicates that, when the transmission time is constrained by the underlying streaming application, any one of these variables fully determines the other. The terminal should choose its transmission power so that the received SIR x maximizes the ratio $h(x)/x$, with $h(x) = u(Bf(x))$, a composite function of both S-curves. If the terminal lacks sufficient power to reach that SIR, it should operate at maximal power, unless the resulting video quality is unacceptably low.

Direct implications of our analysis include: (i) if $u(y) \approx ky$ so that the quality-coding-rate relation is nearly linear, the optimal SIR is determined by the physical layer, as the maximizer of $f(x)/x$ (which is *the same* SIR that a data-transmitting terminal would choose! [3]); (ii) if u behaves like a step function, the terminal should truncate just past the point where the step occurs; and (iii) if f behaves like a step function, then the optimal SIR is just past the point where f jumps.

Even with a fixed physical layer (f function), the optimal operating point could change due to a variation in the perception of quality (u function) at the receiver, or movement that may force the transmitter to operate at an SIR below the optimal level due to power limitations. If the streamed video has been encoded prior to transmission, scalability is essential to achieve such adaptation, via a change in the truncation point of the embedded bit stream. But if coding is being performed concurrent with transmission, a non-scalable encoder that can adapt its rate in real-time could provide a more efficient solution, at a possibly higher computational cost. We can also apply our analysis to optimally choose the coding rate of the non-scalable encoder.

A situation in which several video transmitters share a CDMA channel can be set up as a "game" in which each terminal seeks to maximize its quality-to-power index, with each terminal's "noise" including the interference caused by others. Game theory has been fruitfully applied to the wireless transmission of data, in [3] and other works.

5. REFERENCES

[1] Luna, C.E., A.K. Katsaggelos, "Maximizing user utility in video streaming applications" IEEE ICASSP, Vol. 3, pp. 1465 -1468, 2001

[2] Luna, C.E., A.K. Katsaggelos, "Maximizing user utility in video streaming applications", IEEE Trans. on Circ. and Sys. for Video Tech., Vol. 13, No. 2, Feb 2003

[3] Rodriguez, V., "Robust Modeling and Analysis for Wireless Data Resource Management", IEEE WCNC, Vol. 2, pp.717- 722, March 2003

[4] Rodriguez, V., "Resource Management for Scalably Encoded Information: The Case of Image Transmission Over Wireless Networks", IEEE ICME, Vol. 1, pp. 813-6, July 2003

[5] Rodriguez, V., "An Analytical Foundation for Resource Management in Wireless Communications", IEEE Globecom, Vol. 2, pp. 898-902, December 2003

[6] Wang, Y., J. Ostermann and Y. Zhang, "Video Processing and Communications", Upper Saddle River, New Jersey : Prentice Hall, 2002

[7] Zhang, Q., W. Zhu, Z. Ji, and Y.-Q. Zhang, "A power-optimized joint source and channel coding for scalable video streaming over wireless channels", IEEE ISCAS, vol. 5, pp. 137-140, May 2001