Computer Communications 33 (2010) 1615–1622

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Cross-layer wireless video adaptation: Tradeoff between distortion and delay

Liang Zhou [a,*], Min Chen [b], Zhiwen Yu [c], Joel Rodrigues [d], Han-Chieh Chao [e]

[a] UEI, ENSTA-ParisTech, Paris, France
[b] School of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea
[c] School of Computer Science, Northwestern Polytechnical University, China
[d] Instituto de Telecomunicações, University of Beira Interior, Portugal
[e] Institute and Department of Electronic Engineering, National Ilan University, Taiwan

ARTICLE INFO

Article history:
Available online 9 May 2010

Keywords:
Cross-layer design
Wireless video communications
Delay
Distortion
Tradeoff

ABSTRACT

Adaptive scheduling is becoming an increasingly important issue in wireless video communications, which are widely used in the industry and academic organizations. However, existing scheduling schemes for real-time video services in wireless networks generally do not take into account the relationship between the transmission delay and video distortion. In this paper, we develop and evaluate a delay-distortion-aware wireless video scheduling scheme in the framework of cross-layer information adaptation. At first, we construct a general video distortion model according to the observed wireless network parameters, as well as each video sequence's rate-distortion characteristic. Then, we exploit a distortion-aware wireless video scheduling scheme and derive a bound on the asymptotic decay rate of the video distortion. Furthermore, the relationship between the delay and distortion is studied by taking into account video application delay and distortion requirements for specific wireless network environments. The proposed scheme is applied to heuristically find optimal tradeoff between the delay and distortion. Extensive examples are provided to demonstrate the effectiveness and feasibility of the proposed scheme.

1. Introduction

Rapid growth in wireless networks is fueling the demand that services traditionally available only in wired networks, such as video, be available to mobile users. However, the characteristics of wireless systems provide a major challenge for reliable transport of video since the video transmitted over wireless channels is highly sensitive to delay, interference and topology change which can cause both packet losses and bit-errors [3]. Furthermore, these errors tend to occur in bursts, which can further decrease the delivered Quality of Service (QoS). Current and future wireless systems will have to cope with this lack of QoS guarantees [1,2].

The issue of supporting error-resilient video transport over error-prone wireless networks has received considerable attention recently. In [3], a "smart" inter/intra-mode switching scheme is proposed based on an Rate-Distortion (RD) analysis, but the effectiveness of this approach with burst packet losses is not clear and it may be too complicated for implementation in the face of real-time video application requirements. A model-based packet interleaving scheme is studied in [10] which can achieve some performance gain at the cost of additional delay since the interleaving is considered within several video frames. Therefore, model-based packet interleaving scheme is not appropriate for real-time video applications due to the relatively large delay induced. [11,12] investigate the effect of different Forward Error Correction (FEC) coding schemes on reconstructed video quality and [13] proposes an adaptive Automatic Repeat Request (ARQ) approach. Specifically, in [2,4], FEC redundancy is distributed equally among all the video packets although additional performance gain can be expected when some kind of unequal protection is used [5]. Furthermore, the use of ARQ, as in [6], will cause unbounded delay which is also inappropriate for real-time video communications.

Typically, for video communications over wireless networks, there are two main factors which can greatly affect the QoS: the ***transmission delay*** and ***video distortion***. The way in which delay scales for such distortion-optimal schemes, however, has not been well-studied. Some works have researched the optimal tradeoff between the delay and throughput, i.e., [7,8]. One may make the following inference about the tradeoff between throughput and delay: a small transmission range is necessary to limit interference and hence to obtain a high throughput in the case of a fixed random wireless network. This results in multi-hop, and consequently leads to high delays. In this paper, we extend these prior works to wireless video transmission, and study the relationship between the delay and distortion. In particular, we develop and evaluate a

* Corresponding author. Tel.: +33 6 67 20 86 68.
E-mail address: liang.zhou@ieee.org (L. Zhou).

0140-3664/$ - see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.comcom.2010.05.002

delay-distortion-aware wireless video scheduling scheme in the framework of cross-layer information adaptation. The main contributions of this paper are:

- Constructing a general media distortion model according to the observed wireless network parameters, as well as each application RD characteristic.
- Exploiting a distortion-aware wireless video scheduling scheme and we derive bounds on the asymptotic decay rate of the video distortion.
- Proposing the relationship between the delay and distortion by taking into account video application delay and distortion requirements.

The remainder of this paper is organized as follows. In Section 2, we present the system model and assumption of the video distortion and wireless networks. In Section 3, we develop a distortion-aware wireless video scheduling scheme and analyze the distortion lower and upper bounds. We provide the relationship between the distortion and delay to optimize the corresponding tradeoff according to users' requirements in Section 4. At last, we give some concluding remarks and future research topics in Section 5.

The following notations will be used throughout this paper: **1** denotes a column vector with all ones. $f(n) = O(g(n))$ means that there exists a constant c and integer N such that $f(n) \leqslant cg(n)$ for $n > N$. $f(n) = o(g(n))$ means that $\lim_{n\to\infty} f(n)/g(n) = 0$. $f(n) = \Omega(g(n))$ means that $g(n) = O(f(n))$. $f(n) = \Theta(g(n))$ means that $g(n) = O(f(n))$ and $f(n) = O(g(n))$.

2. System model

We describe in this section a mathematical model, which is built to represent a system distortion framework.

2.1. Video distortion

For the distortion of wireless video transmission, we employ an additive model to capture the total video distortion as [3], and the overall distortion D_{all} can be achieved by:

$$D_{all} = D_{comp} + D_{loss}, \tag{1}$$

where the distortion introduced by source compression is denoted by D_{comp}, and the additional distortion caused by packet loss is denoted by D_{loss}. According to [3], D_{comp} can be approximated by:

$$D_{comp} = \frac{\theta}{R - R^0} + D^0, \tag{2}$$

where R is the rate of the video stream, θ, R^0 and D^0 are the parameters of the distortion model which depend on the encoded video sequence as well as on the encoding structure. They can be estimated from three or more trial encodings using non-linear regression techniques. To allow fast adaptation of the rate allocation to abrupt changes in the video content, these parameters need to be updated for each Group Of Pictures (GOP) in the encoded video sequence, typically once every 0.5 s.

The distortion D_{comp} introduced by packet loss due to transmission errors and network congestion, on the other hand, can be modeled by a linear model related to the packet loss rate P_{loss} [3]:

$$D_{loss} = 10^{\alpha} P_{loss}, \tag{3}$$

where the sensitivity factor α reflects the impact of packet losses P_{loss}, and depends on both the video content and its encoding structure. For simplicity, we assume in the rest of the paper that random packet losses due to transmission errors are remedied at the lower layers (*e.g.*, MAC-layer retransmissions and PHY-layer channel coding). In this case, P_{loss} comprises solely of packet late losses due to network congestion.

2.2. Wireless networks

We consider a wireless network of N nodes and L links. Let V be the set of nodes, E be the set of directed links between nodes, and $G(V,E)$ be the directed connectivity graph of the network. Each link $l \in E$ interferes with a set of other links in E, which we denote as ε_l. We assume that if $k \in \varepsilon_l$ then $l \in \varepsilon_k$, i.e., the interference relationship is symmetric. We also let $l \in \varepsilon_l$, i.e.,

$$\varepsilon_l = \{l\} \cup \{l' \in E : l' \quad \text{interferes with } l\}. \tag{4}$$

Let $\lambda_l(\tau)$ denote the number of packets that arrive at link l in time slot τ. We assume each link l, $\lambda_l(1)$, $\lambda_l(2)$, ... are i.i.d., and $\lambda_l = \mathbf{E}[\lambda_l(1)]$. Moreover, $\lambda_l(\tau)$ is upper bounded for all $\tau > 0$, which means the number of arrival packets is finite in each time slot.

Let $d_l(\tau)$ denote the number of packets that can be served by link l at time slot τ. Assume that the capacity of each link is a fixed number c_l. Let $s_l(\tau) = 1$ indicate that link l is scheduled in time slot τ, $s_l(\tau) = 0$ otherwise. Clearly, $d_l(\tau) = c_l s_l(\tau)$. The system state can be defined as

$$\vec{q}(\tau) = \left[q_1(\tau), q_2(\tau), \ldots, q_{|E|}(\tau)\right], \tag{5}$$

where $q_l(\tau)$ is the number of packets queued at link l at time slot τ, and the dynamics are given by

$$q_l(\tau + 1) = [q_1(\tau) + \lambda_l(\tau) - d_l(\tau)]^+, \tag{6}$$

where $[\cdot]^+$ denotes the projection to $[0, +\infty]$.

3. Distortion-aware wireless video adaptation

In this section, we propose a distortion-aware wireless video scheduling (DAWVS) scheme in the framework of cross-layer information adaptation.

3.1. Scheduling design

Each time slot is divided into two parts: a scheduling slot and a data transmission slot. The links that are to be scheduled are chosen in the scheduling slot and the chosen links transmit their packets in the data transmission slot. The scheduling slot is further divided into M mini-slots. The algorithm proceeds as follows: at the beginning of time slot t, each link l first computes

$$\rho_l = \frac{(q_l/c_l)^{\alpha_l}}{\max_{i\in\varepsilon_l}\left[\sum_{k\in\varepsilon_i}(q_k/c_k)^{\alpha_k}\right]} \cdot \log M. \tag{7}$$

Each link then picks a backoff time from $\{1,2,\ldots,M+1\}$ where picking $M+1$ implies that the link will not attempt to transmit in this time slot. The backoff time τ is chosen as follows:

$$Pr\{\tau = M + 1\} = e^{-\rho_l}, \tag{8}$$

$$Pr\{\tau = m\} = e^{-\rho_l \frac{m-1}{M}} - e^{-\rho_l \frac{m}{M}}, \quad m = 1, 2, \ldots, M. \tag{9}$$

When the backoff timer for a link expires, it begins transmission unless it has already heard a transmission from one of its interfering links. If two or more links that interfere begin transmissions simultaneously, there is a collision and none of the transmissions is successful. Further, any link that hears the collision will not attempt transmission in the rest of their time slot.

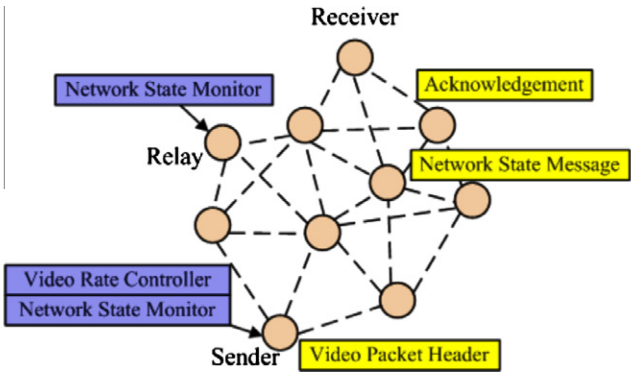

Fig. 1. Cross-layer information exchange among video scheduling scheme.

Remark 1. The proposed DAWVS can be thought of as a two-phase algorithm. In the first phase, each link l first decides whether or not it would participate in the schedule for that time slot. In our algorithm, this phase corresponds to choosing $\{1,2,\ldots,M\}$ or $(M+1)$, respectively. In the next phase, each participating link chooses a number between 1 and M and attempts to transmit starting from that mini-slot. This backoff procedure serves to reduce collision, and thus should lead to a higher capacity compared with a policy without backoff, e.g., [15]. While data transmission may start at any mini-slot, the length of each packet is assumed to be smaller than the data transmission slot so that a transmission ends within the time slot.

In order that the scheduled link rates can be adapted at the transport layer according to network states reported from the network layer, the cross-layer information exchange is needed. Fig. 1 illustrates various components in such a system [9]. At the MAC layer, a link state monitor keeps an online estimate of the effective capacity. It also records the intended rate allocation advertised by each stream, and calculates the available time slots accordingly. In addition, periodic broadcast of link state messages are used to collect the values of mini time slot from neighboring links in the same interference set. At the network layer, the routing information obtained from the routing algorithm can be used to calculate P_{loss}. At the application layer, the video rate controller at the source advertises its intended rate control in the video packet header, and calculates the value of D_{loss} accordingly. The link state monitor traversed by the stream then calculates the relevant parameters in (1) based on its local cache of capacity, utilization information of all the links within its interference set. The destination node extracts such information from the video packet header and reports back to the sender in the acknowledgment packets, so that the video rate controller can re-optimize its intended rate based on the proposed DAWVS, with updated link state information.

Definition 1. Define the Lyapunov function

$$V(\tau) = \max_{i\in E} \sum_{l\in\varepsilon_i} \left(\frac{q_l(\tau)}{c_l}\right)^{\alpha_l},$$

which denotes the largest sum of content-aware backlog in any interference neighborhood.

Assume that the offered load $\vec{\lambda} = [\lambda_1, \lambda_2, \ldots, \lambda_{|E|}]$ is such that the system is stationary and ergodic. Since the distortion caused by video compression simply depends on the given rate (please see (2)), we will focus on the distortion caused by the packet loss due to network congestion. In particular, we are interested in the probability that queue-overflow. For example, we may want to know the probability that the maximum queue length exceeds a given threshold Q. On the other hand, with the techniques developed in this paper, it is more convenient to work with the probability

$$Pr\{\mathbf{E}(V(\tau)) \geqslant Q\}. \tag{10}$$

Unfortunately, the problem of calculating the exact probability of (10) is often mathematically intractable. In this work, we are interested using large-derivation theory to compute estimates of this probability. Specifically, we use the following limits:

$$LB_0(\vec{\lambda}) \triangleq -\limsup_{Q\to\infty} \frac{1}{Q} \log Pr\{\mathbf{E}(V(\tau)) \geqslant Q\},$$

$$UB_0(\vec{\lambda}) \triangleq -\liminf_{Q\to\infty} \frac{1}{Q} \log Pr\{\mathbf{E}(V(\tau)) \geqslant Q\},$$

where $LB_0(\vec{\lambda})$ and $UB_0(\vec{\lambda})$ are the lower and upper bounds for (10) as the queue length constraint Q approaches infinity.

Proposition 1. *DAWVS guarantees that for any ω and constraints B_1, $B_2 \geqslant 0$, there exists a constant Q_C such that if $V(\tau) \geqslant Q_C$, then for any $\theta \in [0,1]$ and link $i \in E$ such that*

$$\sum_{l\in\varepsilon_i} \left(\frac{q_l(\tau)}{c_l}\right)^{\alpha_l} \geqslant \theta(\mathbf{E}(V(\tau)) - B_1 - B_2\omega), \tag{11}$$

the following holds,

$$\sum_{l\in\varepsilon_i} \Pr\{\textit{link } l \textit{ is scheduled}\} \geqslant \theta\left(1 - \frac{1+\log M}{M} - \omega\right). \tag{12}$$

Proof. Please see Appendix A. □

Remark 2. Note that although the original statement of Proposition 1 requires that $\omega, B_1, B_2 > 0$, the proof there also trivially holds for the case when $\omega, B_1, B_2 \geqslant 0$. Let $\theta = 1$, this then implies that, when $\mathbf{E}(V(\tau))$ is large, with high probability at least one link will be scheduled in those interference neighborhoods with sum of backlog close to $\mathbf{E}(V(\tau))$. It should be noted that Proposition 1 can be used to establish a negative drift of the Lyapunov function $\mathbf{E}(V(\tau))$ whenever that the offered load satisfies, for some $\omega > 0$,

$$\sum_{l\in\varepsilon_i} \left(\frac{\lambda_l}{c_l}\right)^{\alpha_l} \leqslant 1 - \frac{1+\log M}{M} - \omega, \quad \forall l \in E. \tag{13}$$

For the rest of the paper, we assume that (13) holds because otherwise we do not know the stability of the system.

3.2. Distortion bound

For any link $i \in E$, define the scaled queue length:

$$q_i^Q(\tau) = \frac{1}{Q} q_i(\lfloor Q\tau \rfloor).$$

Note that this expression represents the standard large-deviation scaling that shrinks both time and magnitude. We also define the scaled of the Lyapunov function:

$$V^Q(\tau) = V(\vec{q}^{\,Q}(\tau)).$$

The queue-overflow criterion is $\{V^Q(\tau) \geqslant 1\}$. For ease of exposition, we consider a system that starts at $\tau = 0$.

We first develop a lower bound for $LB_0(\vec{\lambda})$. For a given $\tau > 0$, we are interested in the following probability:

$$LB_0^\tau(\vec{\lambda}) \triangleq -\limsup_{Q\to\infty} \frac{1}{Q} \log Pr\left\{\mathbf{E}(V^Q(\tau)) \geqslant 1 | V^Q(0) = 0\right\}.$$

Intuitively, as $\tau \to \infty$, one would expect that $LB_0^\tau(\vec{\lambda})$ approaches $LB_0(\vec{\lambda})$, the lower bound on the decay rate of the stationary overflow probability. We will use Lemma 1 to derive a lower bound for $LB_0^\tau(\vec{\lambda})$. Note that Proposition 1 provides a lower bound on the