Cross-Layer Wireless Video Adaptation: 
Tradeoff between Distortion and Delay

Liang Zhou, Min Chen, Zhiwen Yu, Joel Rodrigues, and Han-Chieh Chao

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.

Index Terms

cross-layer design; wireless video communications; delay; distortion; trade-off

I. 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] and [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] and [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] and [8]. One may make the following inference about the trade-off 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 extent these prior works to wireless video transmission, and study the relationship between the delay and distortion. In particular, we develop and evaluate a 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 II, we present the system model and assumption of the video distortion and wireless networks. In section III, 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 IV. At last, we give some concluding remarks and future research topics in Section V.

The following notations will be used throughout this paper: \( \mathbf{1} \) denotes a column vector with all ones. \( f(n) = O(g(n)) \) means that there exists a constant \( c \) and integer \( N \) such that \( f(n) \leq cg(n) \) 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 \( f(n) = O(g(n)) \).

## II. System Model

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

### A. 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 \( D_{\text{all}} \) can be achieved by:

\[
D_{\text{all}} = D_{\text{comp}} + D_{\text{loss}},
\]

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

\[
D_{\text{comp}} = \frac{\theta}{R - R^0} + D^0,
\]

where \( R \) is the rate of the video stream, \( \theta, R^0 \) 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 second.

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]:

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

(3)

where the sensitivity factor $\alpha$ 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.

B. 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 $\varepsilon_l$. 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.,

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

Let $\lambda_l(\tau)$ denote the number of packets that arrive at link $l$ in time slot $\tau$. We assume each link $l$, $\lambda_l(1), \lambda_l(2),...$ are i.i.d., and $\lambda_l = E[\lambda_l(1)]$. 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 $\tau$. 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 $\tau$, $s_l(\tau) = 0$ otherwise. Clearly, $d_l(\tau) = c_l s_l(\tau)$. The system state can be defined as

$$\overrightarrow{q}(\tau) = [q_1(\tau), q_2(\tau), ..., q_{|E|}(\tau)], \quad (5)$$
where \( q_l(\tau) \) is the number of packets queued at link \( l \) at time slot \( \tau \), and the dynamics are given by

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

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

III. 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.

A. 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

\[
\rho_l = \max_{j \in \epsilon_l} \frac{(q_l/c_l)_{\alpha_l}}{\max\{\sum_{k \in \epsilon_l} (q_k/c_k)_{\alpha_k}\} \cdot \log M}. \tag{7}
\]

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

\[
Pr\{\tau = M+1\} = e^{-\rho_l},
\]

\[
Pr\{\tau = m\} = e^{-\rho_l \frac{m-1}{M}} - e^{-\rho_l \frac{m}{M}}, m = 1, 2, ..., M. \tag{9}
\]

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.

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 \( \{1, 2, ..., M\} \) 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
\[ V(\tau) = \max_{i \in E} \sum_{l \in \epsilon_i} \left( \frac{q_l(\tau)}{c_l} \right)^{\alpha_l}, \]
which denotes the largest sum of content-aware backlog in any interference neighborhood.

Assume that the offered load \( \lambda = [\lambda_1, \lambda_2, ..., \lambda_{|E|}] \) 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
\[ \Pr \{ \mathbf{E}(V(\tau)) \geq Q \}. \]  

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:
\[
LB_0(\lambda) \triangleq -\limsup_{Q \to \infty} \frac{1}{Q} \log \Pr \{ \mathbf{E}(V(\tau)) \geq Q \},
\
UB_0(\lambda) \triangleq -\liminf_{Q \to \infty} \frac{1}{Q} \log \Pr \{ \mathbf{E}(V(\tau)) \geq Q \},
\]
where \( LB_0(\lambda) \) and \( UB_0(\lambda) \) are the lower and upper bounds for (10) as the queue length constraint \( Q \) approaches infinity.

Proposition 1: DAWVS guarantees that for any \( \omega \) and constraints \( B_1, B_2 \geq 0 \), there exists a constant \( Q_C \) such that if \( V(\tau) \geq Q_C \), then for any \( \theta \in [0, 1] \) and link \( i \in E \) such that
\[
\sum_{l \in \epsilon_i} \left( \frac{q_l(\tau)}{c_l} \right)^{\alpha_l} \geq \theta(\mathbf{E}(V(\tau)) - B_1 - B_2 \omega),
\]
the following holds,
\[
\sum_{l \in \epsilon_i} \Pr \{ \text{link } l \text{ is scheduled} \} \geq \theta \left( 1 - \frac{1 + \log M}{M} - \omega \right).
\]

Proof: Please see Appendix I.

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 \( \omega, B_1, B_2 \geq 0 \). Let \( \theta = 1 \), this then
implies that, when \( E(V(\tau)) \) is large, with high probability at least one link will be scheduled in those interference neighborhood with sum of backlog close to \( E(V(\tau)) \). It should be noted that Proposition 1 can be used to establish a negative drift of the Lyapunov function \( E(V(\tau)) \) whenever that the offered load satisfies, for some \( \omega > 0 \),

\[
\sum_{l \in \epsilon_i} \left( \frac{\lambda_l}{c_l} \right)^{\alpha_l} \leq 1 - \frac{1 + \log M}{M} - \omega, \forall \tau \in E. \tag{13}
\]

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

\[\text{B. Distortion Bound}\]

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-derivation scaling that shrinks both time and magnitude. We also define the scaled of the Lyapunov function:

\[
V^{Q}(\tau) = V(\overline{q}^{Q}(\tau)).
\]

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

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

\[
LB_0^{\tau}(\overline{\lambda}) \triangleq - \limsup_{Q \to \infty} \frac{1}{Q} \log P_r \left\{ E(V^Q(\tau)) \geq 1 \mid V^Q(0) = 0 \right\}.
\]

Intuitively, as \( \tau \to \infty \), one would expect that \( LB_0^{\tau}(\overline{\lambda}) \) approaches \( LB_0(\overline{\lambda}) \), the lower bound on the decay rate of the stationary overflow probability. We will use Lemma 1 to derive a lower bound for \( LB_0^{\tau}(\overline{\lambda}) \). Note that Proposition 1 provides a lower bound on the service rate of those interference sets whose backlogs are almost the largest. However, these interference sets with largest backlog can change from time to time, which makes it difficult for us to track the system dynamics directly by Proposition 1. To address the problem, in the following derivation we divide the entire scaled time into many small intervals. In each small interval, the interference sets that have almost the largest backlog do not change and therefore we are able to use Proposition 1 to estimate \( LB_0^{\tau}(\overline{\lambda}) \).
In order to provide an exact estimation for $LB_0^Q(\hat{\lambda})$, motivated by [16], we employ the concept of Local Rate Function (LRF), and define the corresponding LRF for $V^Q$ as follows:

**Definition 2:** For a fixed $\tau$, let $\delta > 0$ be a small number. Let $\Delta V^Q(\delta, \tau) = V^Q(\delta + \tau) - V^Q(\tau)$ denote the drift of the scaled Lyapunov function. The LRF for $V^Q$ can be defined as follows given $\overline{q} \neq \overline{0}$ and $W > 0$.

$$
\lim_{Q \to \infty} \frac{1}{Q} \log \sup_{\overline{q}} \Pr \{ \Delta V^Q(\delta, \tau) \geq \delta W | \overline{q}^Q(\tau) = \overline{q} \}. \quad (14)
$$

**Remark 3:** (14) can be viewed as the asymptotic decay rate of the probability that the growth rate of $V^Q$ is not smaller than $W$, conditioned on $\overline{q}^Q(\tau) = \overline{q}$.

**Lemma 1:** Assume that for some $\omega > 0$, (13) holds. For any small and positive $\xi$ such that $0 < \xi < \omega$, given $\tau$, there exists $\delta_0$ such that for all $0 < \delta < \delta_0$,

$$
\lim_{Q \to \infty} \frac{1}{Q} \log \sup_{\overline{q}} \Pr \{ \Delta V^Q(\delta, \tau) \geq \delta W | \overline{q}^Q(\tau) = \overline{q} \} \leq -\delta \min_{i \in E} \inf_{0 \leq d \leq 1 - \omega} (I^A_i(a) + I^D_i(d)). \quad (15)
$$

**Proof:** Since we assume finite arrive rate and service rate, from Theorem 2 in [17], for any $W > 0$, there exist upper bound rate function $I^AD_i(a, d)$ such that

$$
\limsup_{Q \to \infty} \frac{1}{Q} \log \Pr \{ V^Q(\delta) - V^Q(\tau) \geq \delta W \} \leq -\delta \inf_{|a-d| \leq W} I^AD_i(a, d), \quad (16)
$$

where

$$
I^AD_i(a, d) = \sup_{(\theta_1, \theta_2) \in \mathbb{R}^2} \left\{ \theta_1 a + \theta_2 d - \limsup_{Q \to \infty} \frac{1}{Q} \log \mathbb{E} \left( \exp \left( Q \theta_1 \lambda_i(\delta) - Q \theta_2 d_i(\delta) \right) \right) \right\}, \quad (17)
$$

Since $\lambda_i$ and $d_i$ are independent, we have

$$
I^AD_i(a, d) = I^A_i(a) + I^D_i(d), \quad (18)
$$

where

$$
I^D_i(d) = \sup_{\theta \in \mathbb{R}} \left\{ \theta d - \log \left( \theta_1 + (1 - \theta_1) e^{\theta_2} \right) \right\}. \quad (19)
$$

Choose $\theta_1$ and $\theta_1$ the same in [17], we have

$$
\lim_{Q \to \infty} \frac{1}{Q} \log \sup_{\overline{q}} \Pr \{ \Delta V^Q(\delta, \tau) \geq \delta W | \overline{q}^Q(\tau) = \overline{q} \} \leq -\delta \min_{i \in E} \inf_{|a-d| \leq W} (I^A_i(a) + I^D_i(d)). \quad (20)
$$

According to the Theorem 1 in [18], we can claim that

$$
\inf_{|a-d| \leq W} (I^A_i(a) + I^D_i(d)) = \inf_{0 \leq d \leq 1 - \omega} (I^A_i(d + W) + I^D_i(d)). \quad (21)
$$
Hence the LRF can be bounded as (15).

**Theorem 1:** For any $\tau$, the lower bound on the decay rate function satisfies
\[
\text{LB}_s^L(\lambda) \geq \inf_{W \geq 0} \min_{\delta \leq 1} \frac{\inf_{d \leq (1-\epsilon)} (I^A(d + W) + I^D(d))}{W} \leq L. \tag{22}
\]

**Proof:** By the assumption that the system is stable, for any $Q$, there exists $V_{\max} > 0$, such that $V^Q(\tau) \leq V_{\max}$ for all $\tau \in [0, \tau]$. Fix $\zeta > 0$, given $V_0 = 0$, $V_n = 1$, and $0 \leq V_1, ..., V_{n-1} \leq V_{\max}$, define $\Gamma_k(V_k) = \{V_k - \zeta \leq V^Q(k\delta) \leq V_k + \zeta\}$, for $k = 1, 2, ..., n-1$, $\Gamma_0(V_0) = \{V^Q(0) = V_0\}$ and $\Gamma_n(V_n) = \{V^Q(n\delta) \geq V_n\}$. Let $m_V$, $0 \leq m_V \leq n$, be the largest integer $m$ such that $V_{m-1} < V_m$. For any $\zeta > 0$, there exists a finite set of $V$ of vectors $(V_0, ..., V_n)$, such that
\[
\bigcup_{(V_0, ..., V_n) \in V} \Gamma_{n-1}(V_{n-1}) \times ... \times \Gamma_1(V_1) \supseteq \{(V^Q((n-1)\delta), ..., V^Q(\delta)) | 0 \leq V^Q((n-1)\delta) \leq V_{\max}, ..., 0 \leq V^Q(\delta) \leq V_{\max}\},
\]
where $\times$ denotes the Cartesian product. Then
\[
Pr\{V^Q(\tau) \geq 1|V^Q(0) = V_0\} \leq \sum_{(V_0, ..., V_n) \in V} Pr\left\{\bigcap_{k=1}^{n} \Gamma_k(V_k) | \Gamma_0(V_0)\right\} \leq \sum_{(V_0, ..., V_n) \in V} Pr\left\{\bigcap_{k=m_V}^{n} \Gamma_k(V_k) | \Gamma_0(V_0)\right\} \leq \sum_{(V_0, ..., V_n) \in V} \prod_{k=m_V+1}^{n} \phi_k^Q(\delta, \zeta).
\]
The last inequality comes from [18]. If we take the log and let $Q$ goes to infinity, we have
\[
\limsup_{Q \to \infty} \frac{1}{Q} \log Pr\{V^Q(\tau) \geq 1|V^Q(0) = V_0\} \leq \min_{(V_0, ..., V_n) \in V} \limsup_{Q \to \infty} \frac{1}{Q} \sum_{k=m_V+1}^{n} \log \phi_k^Q(\delta, \zeta).
\]
To estimate the limit in the above inequality, we use the local rate. From the definition of $\phi_k^Q(\delta, \zeta)$, since $m_V < k \leq n$, we have
\[
\limsup_{Q \to \infty} \frac{1}{Q} \log \phi_k^Q(\delta, \zeta) \leq -(V_k - V_{k-1} - 2\zeta) L
\]
Therefore taking the sum from $k = m_V + 1$ to $n$, we can get
\[
\min_{(V_0, ..., V_n) \in V} \limsup_{Q \to \infty} \frac{1}{Q} \sum_{k=m_V+1}^{n} \log \phi_k^Q(\delta, \zeta) \leq \min_{(V_0, ..., V_n) \in V} \left(2(n - m_V)\zeta - 1\right) L.
\]
let $\zeta \to 0$, we have
\[
\limsup_{Q \to \infty} \frac{1}{Q} \log P\left\{ E(V_Q(\tau)) \geq 1 | V_Q(0) = 0 \right\} \leq -L.
\]
Therefore, we can get
\[
LB_0^* = \limsup_{Q \to \infty} \frac{1}{Q} \log P\left\{ E(V_Q(\tau)) \geq 1 | V_Q(0) = 0 \right\} \geq L.
\]

Note that the bound in Theorem 1 is independent from $\tau$. As $\tau \to \infty$ we could get that $LB_0^* \to LB_0(\lambda)$. Hence we could expect that
\[
LB_0(\lambda) \geq L. \tag{23}
\]

We then develop the upper bound for $UB_0(\lambda)$ under the node-exclusive interference model. In the node-exclusive model, each interference set may have at most two links scheduled in the same slot. According to [17], the over flow function for each link $i$ is given by
\[
\lim_{Q \to \infty} \frac{1}{Q} \log P\left\{ \sum_{l \in \epsilon_i} \left( q_i^Q l(0) c_l^Q \right)^{\alpha_l} \right\} = -\inf_{x > 0} I_{A_i}(x + 2). \tag{24}
\]
where $I_{A_i}(x + 2)$ can be viewed as the rate function. Hence, the decay-rate of the queue-overflow probability is given by
\[
\lim_{Q \to \infty} \frac{1}{Q} \log P\{ V_Q(0) > 1 \} = -\min_{a > 0} \frac{I_{A_i}(a + 2)}{a}. \tag{25}
\]
Then, we have the upper bound:
\[
UB_0(\lambda) \leq I_{ub} \triangleq \min_{i \in E} \inf_{a > 0} \frac{I_{A_i}(a + 2)}{a}. \tag{26}
\]
We now pose a constraint on this decay rate function. Suppose that we want to guarantee that $I_{ub} \geq \chi_0$. Define the limit of the moment generating function of the arrival process as:
\[
\Lambda_i^A(\chi) = \limsup_{Q \to \infty} \frac{1}{Q} \log E(e^{(q_i^Q(\delta))^{\alpha_i} Q \chi}). \tag{27}
\]
The rate function for arrival could be written as
\[
I_{A_i}(a) = \sup_{\chi \in \mathbb{R}} \{ \chi a - \Lambda_i^A \}. \tag{28}
\]
In other words, $I_i^A$ is the Legendre transform of $\Lambda_i^A$. Since $I_i^A$ is convex, $\Lambda_i^A$ is also the Legendre transform of $I_i^A$. Therefore the following holds

$$I_{ub} \geq \chi_0 \iff \min_{i \in E} \inf_{a > 0} \frac{I_i^A(a + 2)}{a} \geq \chi_0 \iff \inf_{a > 0} \frac{I_i^A(a + 2)}{a} \geq \chi_0, \forall i \in E$$

$$\iff \sup_a \{ \chi_0 a - I_i^A(a) \} - 2\chi_0 \leq 0, \forall i \in E$$

$$\iff \Lambda_i^A(\chi_0) - 2\chi_0 \leq 0, \forall i \in E$$

$$\iff \max_{i \in E} \frac{\Lambda_i^A(\chi_0)}{\chi_0} \leq 2. \quad (29)$$

**Remark 4:** The quantity $\frac{\Lambda_i^A(\chi_0)}{\chi_0}$ in (29) is often called effective bandwidth of the arrival process. The inequality (29) implies that the maximum possible effective capacity region of the system under the proposed algorithm is such that the effective bandwidth in every interference range must be no great than 2.

## IV. General Delay-Distortion Tradeoff

In order to realize the adaptation of the video scheduling in the wireless networks. In this section, we present a general delay-distortion tradeoff in the framework of the proposed DAWVS scheme.

### A. Relationship between Delay and Distortion

We consider a random wireless network model similar to that introduced by [14]. There are $n$ nodes distributed uniformly at random on a unit torus and each node has a randomly chosen destination. Each node transmits at $W$ bits per second, which is a constant, independent of $n$.

We assume slotted time for transmission. For successful transmission, we assume a model similar to the Protocol model as defined [14]. Under our *Relaxed Protocol* model, a transmission from node $i$ to node $j$ is successful if for any other node $k$ that is transmitting simultaneously,

$$d(k, j) \geq (1 + \Delta)d(i, j), \quad \text{for } \Delta > 0 \quad (30)$$

where $d(i, j)$ is the distance between nodes $i$ and $j$. This is a slightly more general version of the model presented in [14] in the sense that nodes do not require a common range of transmission.

We now present a parametrized communication scheme and show that it achieves the optimal trade-off between throughput and delay. This scheme is a generalization of the Gupta-Kumar random network scheme [14].
Divide the unit torus into a number of square grids, each of area $\mathcal{A}(N)$.

**Theorem 2**: For DAWVS, with $L \geq N \log N$, $\mathcal{A}(N) \geq 2 \log N/N$,

$$T(L, N) = \Theta \left( \frac{1}{L \sqrt{\mathcal{A}(N)}} \right) \quad \text{and} \quad D(L, N) = \Theta \left( L^2 \sqrt{\mathcal{A}(N)} \right),$$

i.e., the achievable distortion-delay trade-off is

$$D(L, N) = \Theta \left( \frac{L}{T(L, N)} \right). \quad (31)$$

To prove **Theorem 2**, we need the following lemmas. Lemma shows that each unit will have at least one scheduled link, thus guaranteeing successful transmission. Lemma shows that each unit can be active for a constant fraction of time, independent of the node number $N$. Lemma bounds the maximum number of scheduled links passing through any unit. Combining these results yields a proof of **Theorem 2**.

**Lemma 2**: (i) For $\sqrt{\mathcal{A}(N)} \geq 2 \log n/n$, each unit has at least one node.

(ii) For $\sqrt{\mathcal{A}(N)} \Omega(\log n/n)$, each unit has $N\mathcal{A}(N) \pm \sqrt{2 N \mathcal{A}(N)} \log N$. For $L \geq N \log N$ then each unit has $N\mathcal{A}(N) \pm o(N\mathcal{A}(N))$ links.

(iii) Let $L = N \log N$ and let $c_k(N)$, $k \geq 0$, be the fraction of units with $k$ links. Then $c_k(N) = e^{-1}/k!$.

**Proof**: This lemma can be proved using well-known results (for example, see [19], Chapter 3). Due to space constraints, we do not repeat the proof here.

**Lemma 3**: Under the wireless network model for DAWVS, the number of units that interfere with any given unit is bounded above a constant $c_1$, independent of $\mathcal{A}(N)$.

**Proof**: Consider a link in a unit transmitting to another link in one of its 8 neighboring units. Since each unit has area of $\mathcal{A}(N)$, the distance between the transmitting and receiving nodes can not be more than $r = \sqrt{L \mathcal{A}(N)}/8$. Under the wireless network model for DAWVS, data is successfully received if no link in its interference set transmits simultaneously. Therefore, the number of interference units, $c_1$, is at most

$$c_1 \leq 2 \frac{r^2}{\mathcal{A}(N)} = L/4, \quad (32)$$

which for a constant $L$, is a constant, independent of $\mathcal{A}(N)$.

**Remark 5**: A consequence of **Lemma 3** is that interference-free scheduling among all units is possible, where each unit becomes active once in very $1 + c_1$ slots. In other words, each unit can have a constant network performance.
Lemma 4: The number of scheduled links in any unit is $O(L^2 \sqrt{A(N)})$.

Proof: Please see Appendix II.

We now ready to prove Theorem 2.

Proof: (of Theorem 2) From Lemma 3, it follows that each unit can be active for a guaranteed fraction of time, i.e., it can have a constant distortion. Lemma 4 suggests that if each unit divides its unit time-slot into $\Theta(L^2 \sqrt{A(N)})$ packet time-slots, each source-destination pair hopping through it can use one packet time-slot. Equivalently, each source-destination pair can experience with distortion for $\Theta(L^2 \sqrt{A(N)})$ fraction of time. That is, the expected distortion per source-destination pair is $D(L, N) = \Theta(L^2 \sqrt{A(N)})$.

Next we compute the average packet delay $T(L, N)$. As defined earlier, packet delay is the sum of the amount of time spent in each hop. We first bound the average number of the hops then show that the time spent in each hop is constant.

Since each hop covers a distance of $\Theta(\sqrt{A(N)})$, the number of hops per packet for source-destination pair $i$ is $\Theta(d_i \sqrt{A(N)})$, where $d_i$ is the length of source-destination pair $i$. Thus the number of hops taken by a packet averaged over all source-destination pairs is $\Theta(\frac{1}{N} \sum_{i=1}^{N} d_i \sqrt{A(N)})$. Since for large $N$, the average distance between source-destination pairs is $\frac{1}{N} \sum_{i=1}^{N} d_i = \Theta(L)$, the average number of hops is $\Theta(L \sqrt{A(N)})$.

Now we note that by Lemma 3 each unit can be active once every constant number of unit time-slots and by Lemma 4 each source-destination pair passing through a unit can have its own packet time-slot within that unit’s time-slot. Since the packet size scales inverse proportion to the distortion $D(L, N)$, each packet arriving at a node in the unit departs within a constant time.

From the above discussion, we conclude that the delay $T(L, N) = \Theta(L \sqrt{A(N)})$. Given the $D(L, N)$ and $T(L, N)$, we can find that the relationship between them is that:

$$D(L, N) = \Theta \left( \frac{L}{T(L, N)} \right).$$

This concludes the proof of Theorem 2.

B. Examples and Results

To simulate the video applications, two HD (High-Definition) sequences (City and Tennis) are used to represent video with dramatically different levels of motion activities. In terms of HD video, the sequence has spatial resolution of $1280 \times 720$ pixels, and the frame rate of 60 frames.
per second. Video stream is encoded using a fast implementation of H.264/AVC codec at various quantization step sizes, with GOP (Group Of Pictures) length of 25 and \textit{IBBP}... structure similar to that often used in MPEG-2 bitstreams. Encoded video frames are segmented into packets with maximum size of 1500 bytes, and the transmission intervals of each packet in the entire GOP are spread out evenly, so as to avoid unnecessary queuing delay due to the large sizes of intra coded frames.

We validate our proposed delay-distortion tradeoff model. Fig. 2 shows the delay-distortion tradeoff when two user streams the given video sequences (300 frames of each) over the simulated 802.11b wireless network. The model is fit to experimental data for two cases: in the first case, the only losses considered are due to late arrivals; in the second case, an additional end-to-
end random loss rate of 5% is considered. The curves illustrates that The proposed delay-
distortion tradeoff model matches closely the real experimental data. Generally speaking, the
fewer distortion is obtained when the larger transmission delayer is allowed. That is to say, the
user can achieve the optimal tradeoff between transmission delay and video distortion according
to its own requirement.

V. Conclusions

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
media distortion model according to the observed parameters in wireless network, as well as each
application’s characteristic. After that, we exploit a distortion-aware wireless video scheduling
scheme and we derive a bound on the asymptotic decay rate of the video distortion. Furthermore,
the relationship between the delay and distortion is proposed by taking into account video
application delay and distortion requirements in addition to wireless networks conditions. 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.

For future work, we plan to study some practical issues for implementing the proposed scheme.
Note that in real wireless video communications, additional works need to be developed to: i)
reduce the dependence of the media content or scheduling scheme to automatically adapt the
original media content; ii) simplify the adaptive scheduling scheme, especially the network and
source information exchange and feedback; iii) extend the results to more practical wireless
networks (e.g., wireless multimedia sensor networks). In our ongoing work, we plan to carefully
address these open problems and study their impacts on the actual wireless systems.

APPENDIX I

PROOF OF PROPOSITION 1

Proof: Fix any link $k$ such that inequality (11) holds. Consider a link $j \in \varepsilon_k$. We first find
a lower bound on the probability that $j$ is scheduled.

Link $j$ gets scheduled when it attempts transmission and each of the other attempting links in
its interference set choose a bigger backoff time. Let $S_j$ be the event that $j$ is scheduled and let
\( Y_l \) be the backoff time chosen by a link \( l \). Then we get
\[
Pr\{S_j\} \geq \sum_{m=1}^{M} Pr\{Y_j = m\} \prod_{h \in \varepsilon_j, h \neq j} Pr\{Y_h > m\} \\
= \sum_{m=1}^{M} \left( e^{-\rho_j \frac{m+1}{M}} - e^{-\rho_j \frac{m}{M}} \right) \prod_{h \in \varepsilon_j, h \neq j} e^{-\rho_h \frac{m}{M}} \\
= (e^{\rho_j - 1}) \sum_{m=1}^{M} e^{-\rho_j \frac{m}{M}} \prod_{h \in \varepsilon_j, h \neq j} e^{-\rho_h \frac{m}{M}} \\
= (e^{\rho_j - 1}) \sum_{m=1}^{M} e^{-\frac{m}{M} \sum_{h \in \varepsilon_j} \rho_h} \tag{33)
\]

We now find an upper bound on the term \( \sum_{h \in \varepsilon_j} \rho_h \) that appears in (33).
\[
\sum_{h \in \varepsilon_j} \rho_h = \sum_{h \in \varepsilon_j} \frac{(q_h/c_h)^{\alpha_h}}{\max_l \in \varepsilon_h \sum_l (q_l/c_l)^{\alpha_l}} \cdot \log M \\
\leq \sum_{h \in \varepsilon_j} \frac{(q_h/c_h)^{\alpha_h}}{\sum_{l \in \varepsilon_j} (q_l/c_l)^{\alpha_l}} \cdot \log M \leq \log M \tag{34)
\]

where in (34) we have assumption that if \( h \in \varepsilon_j \), then \( j \in \varepsilon_h \). This implies that the denominator in (34) is never less than \( \sum_{l \in \varepsilon_j} (q_l/c_l)^{\alpha_l} \). Using (33) and (34), we get
\[
Pr\{S_j\} \geq (e^{\rho_j - 1}) \sum_{m=1}^{M} e^{-\frac{m}{M} \log M} \geq \frac{\rho_j}{M} \sum_{m=1}^{M} e^{-\frac{m}{M} \log M}. \tag{35)
\]

Hence, summing over all \( j \in \varepsilon_k \), we have
\[
\sum_{j \in \varepsilon_k} Pr\{S_j\} \geq \sum_{m=1}^{M} e^{-\frac{m}{M} \log M} \sum_{j \in \varepsilon_k} \frac{\rho_j}{M}. \tag{36)
\]

Therefore, we can get
\[
\sum_{j \in \varepsilon_k} \rho_j \geq \log M \cdot \frac{\sum_{j \in \varepsilon_k} (q_j/c_j)^{\alpha_j}}{E(V(\tau))} \geq \log M \cdot \theta \frac{E(V(\tau)) - B_1 - B_2 \omega}{E(V(\tau))} \\
= \log M \cdot \theta \left( 1 - \frac{B_1 + B_2 \omega}{E(V(\tau))} \right), \tag{37)
\]

where (37) follows from the assumption that \( \sum_{l \in \varepsilon_k} (q_l/c_l)^{\alpha_l} \geq \theta (E(V(\tau)) - B_1 - B_2 \omega) \). Using (36) we get
\[
\sum_{j \in \varepsilon_k} Pr\{S_j\} \geq \frac{\log M}{M} \theta \sum_{m=1}^{M} e^{-\frac{m}{M} \log M} \left( 1 - \frac{B_1 + B_2 \omega}{E(V(\tau))} \right) \\
= \frac{\log M}{M} \theta \frac{1 - e^{-\log M} e^{-\frac{m}{M}} \left( 1 - \frac{B_1 + B_2 \omega}{E(V(\tau))} \right) \cdot (1 - e^{-\log M} e^{-\frac{m}{M}} \left( 1 - \frac{B_1 + B_2 \omega}{E(V(\tau))} \right))}{1 - e^{-\log M} e^{-\frac{m}{M}} \left( 1 - \frac{B_1 + B_2 \omega}{E(V(\tau))} \right)}. \tag{38)
\]
Since \( M > 1 \), we can get that \( \frac{\log M}{M} / (1 - e^{-\log M}) \geq 1 \), \( e^{-\log M} \geq 1 - \log(M)/M \) and \( 1 - e^{-\log M} = 1 - 1/M \). Hence,

\[
\sum_{j \in \delta_k} P_r \{ S_j \} \geq \theta \left( 1 - \frac{1 + \log M}{M} \right) \left( 1 - \frac{B_1 + B_2\omega}{E(V(\tau))} \right). \tag{39}
\]

Now if \( E(V(\tau)) \geq R \) then \( \frac{B_1 + B_2\omega}{E(V(\tau))} \leq \frac{B_1 + B_2\omega}{R} \). Thus, for sufficient large \( R \), we have \( \frac{B_1 + B_2\omega}{E(V(\tau))} \leq \omega \) and this gives

\[
\sum_{j \in \delta_k} P_r \{ S_j \} \geq \theta \left( 1 - \frac{1 + \log M}{M} \right) (1 - \omega) \geq \theta \left( 1 - \frac{1 + \log M}{M} \right) \omega. \tag{40}
\]

This ends the proof.

**APPENDIX II**

**PROOF OF LEMMA 4**

*Proof:* Consider \( n \) source-destination pairs. Let \( d_i \) be the distance between the source-destination pair \( i \). Let \( h_i \) be the number of the number of links per packet for source-destination pair \( i \). Then \( h_i = d_i / \sqrt{A(N)} \). Let \( H = \sum_{i=1}^{L} h_i \), i.e., the total number of links required to send one packet from each send to its corresponding destination.

Now consider a particular unit and define the Bernoulli random variables \( Y_k^i \), for source-destination pair \( 1 \leq i \leq N \) and links \( 1 \leq k \leq h_i \), to be equal to 1 if link \( k \) of source-destination pair \( i \)'s packet originates from a link in the unit. Hence, the total number of the source-destination pair passing through the unit is \( Y = \sum_{i=1}^{n} \sum_{k=1}^{h_i} Y_k^i \). Note that since the nodes are randomly distributed, the \( Y_k^i \)'s are identically distributed. For any \( 1 \leq i \neq j \leq n \), \( Y_k^i \) and \( Y_l^j \) (for any \( 1 \leq k \leq h_i \), \( 1 \leq l \leq h_j \)) are independent. However, for any given \( 1 \leq i \leq n \), \( Y_k^i \) and \( Y_l^i \) (for any \( 1 \leq k \neq l \leq h_i \)) are independent and in fact the event \( \{ Y_k^i = 1, Y_l^i = 1 \} \) is not possible, as source-destination pair \( i \) can intersect the unit at most once.

First consider the random variable \( H = \sum_{i=1}^{n} d_i / \sqrt{A(N)} \). Since, for all \( i, d_i \in [0, 1/\sqrt{2}] \), \( H = O(L/\sqrt{A(N)}) \). Now, we use this result to find a bound on \( E[Y] \) as follows:

\[
E[Y] = E_H[E[Y|H]] = E_H \left[ \sum_{i=1}^{n} \sum_{k=1}^{h_i} E[Y_k^i|H] \right] \\
= E_H[H E[Y_1^i]] = \Theta(L^2 \sqrt{A(N)}), \tag{41}
\]

where (41) follows from the fact that, by the symmetry of the torus, any hop is equally likely to originate from any of the \( 1/\sqrt{A(N)} \) units.
consider a random variable $\tilde{Y} = \sum_{l=1}^{h} \tilde{Y}_l$, where $\tilde{Y}_l$ are i.i.d. Bernoulli random variables with $Pr(\tilde{Y}_l = 1) = Pr(\tilde{Y}_l^i = 1) = \mathcal{A}(N)$. Because of the particular dependence of $Y_k^i$ and $Y_l^i$ (for any $1 \leq k \neq l \leq h_i$), it can be shown that, for any $m$,

$$E[Y^m] \leq E[\tilde{Y}^m].$$

This implies that, for any $\phi > 0$,

$$E[\exp(\phi Y)] \leq E[\exp(\phi \tilde{Y})].$$

(42)

For any $\delta_0$, define $P(\tilde{Y}, \delta_0) \triangleq Pr(\tilde{Y} \geq (1 + \delta_0)E[\tilde{Y}])$. By the Chernoff bound i.i.d. Bernoulli random variables,

$$P(\tilde{Y}, \delta_0) \leq \exp(-\delta_0^2 E[\tilde{Y}]/2).$$

(43)

(44)

Considering the following:

$$P(Y, \delta_0) \triangleq Pr(Y \geq (1 + \delta_0)E[Y]) = Pr(\exp(\phi Y) \geq \exp(\phi(1 + \delta_0)E[Y]))$$

$$\leq \frac{E[\exp(\phi Y)]}{\exp(\phi(1 + \delta_0)E[Y])}$$

$$\leq \frac{E[\exp(\phi \tilde{Y})]}{\exp(\phi(1 + \delta_0)E[Y])}$$

(45)

(46)

where (45) follows by the Markov inequality and (46) follows from (43) and the fact that $\exp(\phi(1 + \delta_0)E[Y]) = \exp(\phi(1 + \delta_0)E[\tilde{Y}])$. From (46) and the proof of the Chernoff bound (for example, see [8]) it follows that $P(Y, \delta_0)$ can be bounded above by the bound on $P(\tilde{Y}, \delta_0)$ as given in (44).

By taking $\delta_0 = 2\sqrt{\log L/E[Y]}$, we obtain

$$Pr(Y \geq EY + 2\sqrt{\log LE[Y]}) \leq 1/N^2.$$

(47)

Thus, for any unit, the number of links originating from it are bounded above $L^2 \sqrt{\mathcal{A}(N)} + o(L^2 \sqrt{\mathcal{A}(N)})$ with probability $\geq 1 - 1/N^2$. Since there are at most $N$ units, by the union of events bound, the above bound holds for all units with probability $\geq 1 - 1/N$. This completes the proof of the lemma.
REFERENCES


