Noname manuscript No.
(will be inserted by the editor)

Cross-Layer Optimization for Many-to-One Wireless Video Streaming Systems

Mohammad A. Alsmirat and Nabil J. Sarhan

Received: date / Accepted: date

Abstract This paper develops a cross-layer optimization solution for video streaming from multiple sources to a central proxy station over a wireless network. The proposed solution manages the application rates and transmission opportunities of various video sources based on the dynamic network conditions in such a way that minimizes the overall video distortion. The solution includes a new online approach for estimating the effective network airtime and a new video bitrate-distortion model. We demonstrate the effectiveness of the proposed solution through extensive experiments. The results show that the proposed solution substantially enhances the perceptual video quality while reducing the power consumed by the video sources and that the solution is highly adaptable to the existence of interfering network traffic.

Keywords Bandwidth allocation · cross-layer optimization · effective airtime estimation · video streaming · wireless networks · WLAN

1 Introduction

This paper considers video streaming from multiple video sources to a central station over a single-hop IEEE 802.11 wireless LAN (WLAN). This application, referred to here as *many-to-one video streaming*, is typical in many

Mohammad A. Alsmirat
Department of Computer Science
Jordan University of Science and Technology
Irbid, 22110, Jordan.
E-mail: masmirat@just.edu.jo

Nabil J. Sarhan
Department of Electrical and Computer Engineering
Wayne State University
Detroit, MI 48202, US
E-mail: nabil@wayne.edu

systems, including video surveillance. As shown in Figure 1, the wireless video sources, such as video cameras and/or video sensors, share the same medium and can be either battery-powered or outlet-powered. The central proxy station is connected with a high-bandwidth link to the access point, and thus this link is not deemed as a bottleneck in the system. Large systems may be composed of multiple such systems or cells.

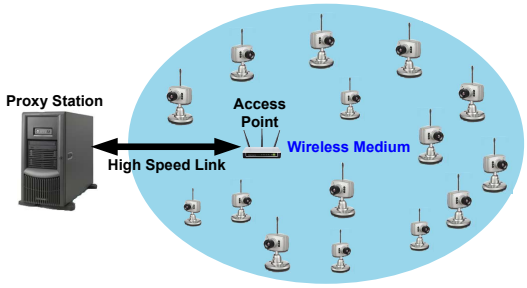


Fig. 1: A Wireless Video Surveillance System

The main challenge in many-to-one video streaming is that the wireless network has limited available bandwidth, which changes dynamically and should be estimated accurately and distributed efficiently among various video sources to maximize the quality of the video streams received by the proxy station. Although numerous studies considered the cross-layer approach for allocating bandwidth [1–14], only [15, 16] considered many-to-one video streaming. As discussed in Subsection 2.2, the solutions in the latter studies are unfortunately highly limited and cannot be used in practice.

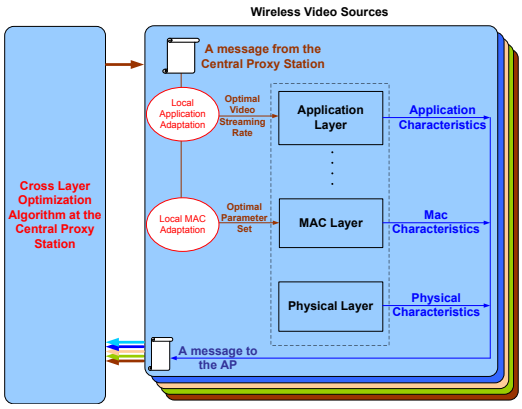


Fig. 2: Clarification of the Cross-Layer Solution

This paper develops a cross-layer optimization solution for many-to-one video streaming systems. The solution dynamically distributes and allocates the available network bandwidth among various video sources in such a way that minimizes the overall video distortion. As clarified in Figure 2, the solution utilizes and dynamically controls parameters in the application, data link (MAC), and physical layers. The solution includes a new approach for dynamically estimating the effective network airtime (defined as the fraction of the network time that is used for delivering useful data). For better estimation accuracy, the proposed approach computes the effective airtime when the packet loss is below a specified threshold. In addition, we develop an accurate video bitrate-distortion model as well as a new model for adapting the link-layer parameters.

In contrast with prior studies [15, 16], we provide extensive evaluations using a more realistic methodology, experimenting with a variety of datasets and analyzing a comprehensive set of metrics, including power consumption. Although we use OPNET as in [16], we stream real video frames (and not just an abstract bit streams) over a simulated network. Additionally, we implement an error concealment algorithm at the proxy station to mitigate the impact of packet loss on perceptual video quality. Moreover, we analyze the performance in terms of power consumption as well as perceptual video quality. Power consumption is a primary concern, especially when the video sources are battery-powered. We consider the aggregate power consumed by video sources due to capturing, encoding, and transmission. Furthermore, we use a variety of datasets and consider several important parameters.

The **main unique contributions** of this paper can be summarized as follows. (1) We propose a new complete cross-layer optimization solution for many-to-one video streaming systems. (2) We develop an accurate video bitrate-distortion model. (3) We propose an online and dynamic approach for estimating the effective network airtime. (4) We develop a new model for adapting the link layer parameters in the video sources. (5) We analyze the effectiveness of the proposed solution in terms of both video quality and power consumption using a variety of datasets.

The results show that the proposed solution enhances substantially the perceptual quality of received video streams and that the proposed effective airtime estimation algorithm is accurate and converges quickly. The proposed solution also significantly reduces the power consumed by the video sources and is highly adaptable with interfering traffic in the network.

The rest of this paper is organized as follows. Section 2 present background information and some related work. Subsequently, Section 3 presents the proposed cross-layer optimization solution. Section 4 discusses the performance evaluation methodology. Finally, Section 5 presents and analyzes the main results.

2 Background Information and Related Work

2.1 IEEE 802.11e Standard

The 802.11e standard enables the provision of different quality-of-service (QoS) levels among different access categories (AC) in the same station, thereby enhancing the support of multimedia applications. These access categories include Voice, Video, Best Effort, and Background. The IEEE 802.11e MAC layer provides two methods for managing the access to the wireless channel: *Hybrid Coordination Function Controlled Channel Access* (HCCA) and *Enhanced Distributed Channel Access* (EDCA). In contrast with HCCA, EDCA provides reduced complexity and better flexibility by providing a distributed coordination function [16, 17].

With EDCA, priorities are implemented using four EDCA parameters: *Arbitration Inter Frame Space (AIFS)*, *Minimum Contention Window (CW_{min})*, *Maximum Contention Window (CW_{max})*, and *Transmission Opportunity Time (TXOP)*. AIFS controls the waiting time before an AC starts the transmission when the medium is not busy. In case of a collision, the AC will back off for a random time between 0 and CW, where CW is a variable that is initialized to CW_{min}, is incremented after every transmission failure until it reaches CW_{max}, and is reset to CW_{min} after a successful transmission. The backoff timer is decremented every time the medium is sensed to be idle for at least $AIFS$ seconds. Finally, the TXOP limit controls the period during which the AC keeps transmitting when it gains access to the medium.

2.2 Cross-Layer Optimization for Video Streaming

Wireless video transmission has received much attention [18, 19] (and references within), and numerous studies have discussed cross-layer optimization in video streaming over wireless networks. Studies [1–8] consider a system in which only one station streams a video at a time, whereas [9–14] consider a system in which multiple stations receive video streams form a central video server, and [15, 16] consider systems in which multiple stations deliver video streams to a central station. The latter studies are more related to this work.

Study [15] optimizes video streaming over 2G wireless network. The solution proposed in that study adapts the video streams by using video summarization techniques, such as frame skipping, which are not practical in video surveillance because of the system's sensitivity to losing video frames.

Study [16] formulates and solves an optimization problem that minimizes the sum of distortion in all video streams. It used the formulation in [20] to develop an analytical model for the effective airtime, but the formulation is problematic, as discussed in Subsection 2.3. In addition, that paper assumed a *p-persistent* EDCA, which differs from the standard EDCA in the backoff timer selection process. Moreover, it ignored the packetization overhead of the transport and application layers when determining the optimal application

rate and link layer parameters. We address these limitations in that study and improve its link-layer adaptation model, which was based on the formulation in [17, 21].

2.3 Effective Airtime Estimation

The effective airtime is the fraction of the network time that is used for delivering useful data. As will be discussed later, solving the optimization problem requires an accurate estimation of the effective airtime. In [17], the effective airtime for ad-hoc networks was simply determined as the total throughput divided by the physical rate, assuming that all stations in the network have the same physical rate. Study [16] developed an analytical model for the effective airtime for video streaming from multiple stations to a proxy, based on the formulation in [20]. These two studies involve significant simplifications, approximations, and assumptions. The developed airtime model was simply given in terms of only CW_{min} and the number of stations in the network. Furthermore, according to the model, the effective airtime increases with the number of nodes and yields a value close to 1 (i.e. 100%) in networks with 30 stations or more. Our detailed empirical study proves a different behavior, which is also logically evident. As will be shown in Section 5, this model leads to significant dropping after the optimization and gives relatively high distortion.

Other studies [22–30] sought to determine other related parameters, such as the saturation bandwidth and network capacity for ad-hoc networks. None of these studies is directly applicable to finding the effective airtime in our considered system.

3 Proposed Cross-Layer Optimization Solution

This study presents an optimal cross-layer solution that dynamically distributes and allocates the available network bandwidth among various video sources in many-to-one wireless video streaming systems. In the considered system, multiple video sources stream videos to a central station over a single-hop IEEE 802.11 WLAN in EDCA mode, as shown in Figure 1. The system has S video sources, where each source s has physical rate (y_s) and streams an encoded video at rate r_s. As illustrated in Figure 2, the proposed solution utilizes and dynamically controls parameters in three layers in the network stack: Application, Link, and Physical. Table 1 summarizes and explains the used notations.

3.1 Cross-Layer Optimization Problem Formulation

We formulate the problem as a cross-layer optimization problem of the sum of the distortion of all video streams received by the central proxy station,

Table 1: Description of Notations

Notation	**Definition**
CW_{min}	Minimum Contention Window
CW_{max}	Maximum Contention Window
AP	Access Point
AC	Access Category
AIFS	Arbitration Inter Frame Space
TXOP	Transmission Opportunity Time
S	Number of video sources
r_s	Rate of the encoded video sent by source s
y_s	Source s physical rate
A_{eff}	The effective airtime of the medium
f_s	Airtime fraction of source s
τ	Frame rate
Z	Video frame size
t_s	Throughput of source s video stream as received by the application layer of the proxy station
d_s	Source s data dropping rate
A_Δ	The overall average dropping ratio, $A_\Delta = \sum_{s=1}^{S} d_s/y_s$.
A_{thresh}	A threshold that controls the allowable dropping in the network
λ	Lagrangian constant
a_s, b_s, c_s	Distortion curve constants
x_s	Optimized TXOP limit
R_s	Maximum load rate measured by the MAC layer at source s
L_s	Maximum video frame size at source s
N_s	Number of MAC layer data frames per maximum video frame at source s
l_s	Average data load per MAC layer data frame at source s
O_s	Average MAC and physical layers overhead at source s
t_s	Short Interframe Space (SIFS) time
t_a	Time required to send an acknowledgment

following initially the general formulation in [16] and then adapting it to include the packetization overhead of both the transport and application layers. Thus, we seek to find the optimal fraction of the airtime allocation $F^* = \{f_s^* | s = 1, 2, 3, ..., S\}$ for various video sources that minimizes the total distortion ($\sum_{s=1}^{S} Distortion_s(r_s)$), where r_s is the application layer transfer rate for video source s, and S is the number of video sources. Specifically, the problem

can be stated as follows:

$$Find \quad F^* = \arg\min_{F} \sum_{s=1}^{S} Distortion(r_s) \tag{1a}$$

$$s.t. \quad \sum_{s=1}^{S} f_s = A_{eff} \tag{1b}$$

$$r_s = f_s \times y_s \tag{1c}$$

$$0 \leq f_s \leq 1 \tag{1d}$$

$$s = 1, 2, 3, ..., S, \tag{1e}$$

where F^* is the set of optimal fractions (f_s^*) of the airtime of all sources, r_s^* is the optimal application-layer rate of video source s, y_s is the physical-layer rate of video source s, and A_{eff} is the effective airtime. This optimization is subject to the following constraints. (1) The total airtime of all video sources is equal to the effective airtime. (2) The application layer transfer rate of source s is the product of the its airtime (f_s) and the physical layer transfer rate (y_s) for video source s. (3) The airtime of each source is between 0 and 1 (inclusive).

The solution for the problem formulated in Equation (1) requires characterizing the distortion function and assessing the effective airtime of the network.

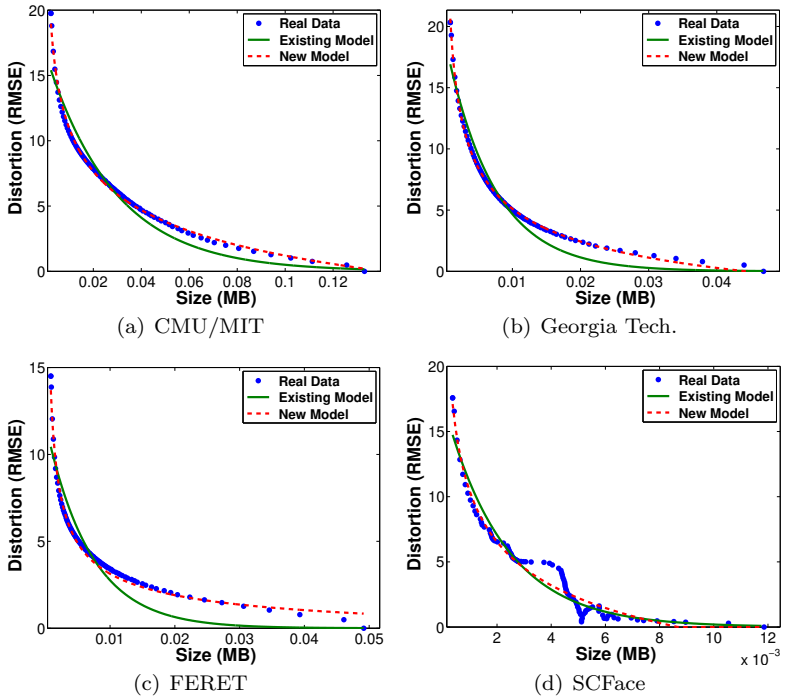
3.2 Distortion Function Characterization

We seek to develop a model of the relationship between the video bitrate (or alternatively the frame size in the case of MJPEG) and video distortion, using a variety of datasets, specifically CMU/MIT [31], Georgia Tech. [32], FERET [33], and SCFace [34]. For each image set, we use the IJG JPEG library to compress each picture in the set with quality factors from 1 to 100, with 1 being the lowest, and then assess the distortion of each image against the original image using the Root Mean Square Error (RMSE) metric, which is one of the most used distortion measures [35]. Finally, we find the average size and the average distortion of all images with the same quality factor. Our results indicate that the model formulated in [36] (referred to here as "Existing Model") is not accurate. We determine that the distortion can be better characterized by

$$Distortion = a \times Z^b + c, \tag{2}$$

where Z is the frame size and a, b, and c are constants. This model is referred to here as "New Model". Figure 3 shows the comparative results of the two models for four different image sets and illustrates the accuracy of the new model. For MJPEG, the image size Z and video bitrate r are related by $r = Z \times \tau$, where τ is the frame rate. The model constants depend on the system (including the

specific video sources used) and the environment (e.g., surveillance site). These constants can be determined empirically by the system through capturing a short video by each source and then encoding each video at different bitrates and assessing the distortion relative to the original video. This process can be done upon system calibration and occasional re-calibration.



(a) CMU/MIT (b) Georgia Tech.

(c) FERET (d) SCFace

Fig. 3: Video Bitrate-Distortion Characterization

3.3 Enhanced Effective Airtime Estimation

We propose a novel online and dynamic effective airtime estimation algorithm for wireless networks in infrastructure configuration. Being "online" means that the algorithm operates during the normal operation of the system. In contrast with existing analytical models, the algorithm uses complete information of the network and involves the cooperation of the access point and all video sources, including various layers in each source. This algorithm needs to be executed only when significant changes in the network (such as the physical rates) happen.

Figure 4 shows the proposed estimation algorithm. The algorithm can be summarized as follows. First, each video source s starts with a moderate initial value of 0.5 for A_{eff}. Subsequently, during a period of time, called estimation period, each video source s assesses its own data dropping rate (d_s) while sending its video stream, and then conveys this information to the access point (AP). Meanwhile, the AP determines the overall average dropping ratio as follows: $A_\Delta = \sum_{s=1}^{S} d_s/y_s$. A_Δ is then used to adjust the current value of A_{eff} at the end of the current estimation period. If A_Δ is zero, A_{eff} is incremented by A_{thresh} . If A_Δ, however, is greater than a certain threshold A_{thresh}, the AP reduces A_{eff} by $C \times (A_\Delta - A_{thresh})$, where C is a positive constant. In contrast, if A_Δ is less than A_{thresh}, the AP increases A_{eff} by $C \times (A_{thresh} - A_\Delta)$. The estimation period and the constant C should be chosen based on the best tradeoff between convergence and stability.

We note that the estimation algorithm needs only the dropping rate and the physical rate of each video source. These data are already included in each state report packet sent once every state report interval.

```
Initialize A_eff with 0.5
At the end of each estimation period{
    A_Δ = Σ_{s=1}^{S} d_s/y_s;
    if (A_Δ == 0)
        A_eff = A_eff + A_thresh;
    else if (A_Δ < A_thresh)
        A_eff = A_eff + C × (A_thresh − A_Δ);
    else if (A_Δ > A_thresh)
        A_eff = A_eff − C × (A_Δ − A_thresh);
}
```

Fig. 4: Enhanced Effective Airtime Estimation Algorithm

3.4 Cross-Layer Optimization Solution

After determining the distortion function and the effective airtime, we can solve the optimization problem. We first prove that the formulated problem is a convex programming problem by determining that all constraints ((1b)-(1e)) in the problem are linear and thus convex and that the optimization function ($\sum_{s=1}^{S}(a_s(f_s y_s/\tau_s)^{b_s} + c_s)$) is also convex. The latter is valid because the derivative of the sum term is monotonically non-decreasing and convex. Since the problem is a budget-constrained convex programming problem, it can be solved by the Lagrangian relaxation technique [37], using the following Lagrangian-relaxed formula:

$$L(F^*, \lambda) = \sum_{s=1}^{S}(a_s(f_s y_s/\tau_s)^{b_s} + c_s) + \lambda(\sum_{s=1}^{S} f_s - A_{eff}), \quad (3)$$

where $0 \leq f_s \leq 1$, and $s = 1, 2, 3, ..., S$. Next, the Lagrangian conditions are formulated as follows:

$$\frac{\partial L}{\partial f_s} = 0 \text{ and } \frac{\partial L}{\partial \lambda} = 0. \tag{4}$$

Assuming that all video sources have the same constant b_s, which is empirically valid, solving these two equations yields the following solution:

$$f_s^* = \left(\frac{-\lambda^* \tau_s}{a_s b_s y_s (y_s/\tau_s)^{(b_s-1)}}\right)^{(1/(b_s-1))}, \tag{5}$$

where

$$\lambda^* = \left(\frac{A_{eff}}{\sum_{s=1}^{S} \left(\frac{-\tau_s}{a_s b_s y_s (y_s/\tau_s)^{(b_s-1)}}\right)^{(1/(b_s-1))}}\right)^{(b_s-1)}. \tag{6}$$

With this solution the AP determines λ^* and announces it to each video source s, which in turn determines its airtime fraction f_s^* and thus its encoding rate as $r_s^* = f_s^* \times y_s$.

3.5 Enforcing the Optimization Results

The link-layer parameters should be determined based on the allocated airtime for each source. They can be either the transmission opportunity duration limit (TXOP limit) or alternatively the frequency of the transmission opportunity. The control of the TXOP limit is more preferred since it is only one parameter, whereas the transmission frequency involves three parameters (AIFS, CW_{min}, and CW_{max}).

In [16], the TXOP limit (x_s^*) for source s is determined as the time required to transmit the number of MAC data frames that source s can transmit during one beacon interval time (t_b). As demonstrated in Subsection 5.3, our results show that choosing intervals other the beacon interval can achieve higher performance. In particular, the received video quality improves with the chosen time interval up to a certain point, and then it starts to worsen. In addition, the time interval that achieves the best results varies with the network size. Furthermore, the model in [16] does not incorporate the packetization overhead of the transport and application layers.

In our solution, we address these limitations as follows. Each video source determines its TXOP limit as the time required to transmit the packets belonging to a single video frame along with all associated overhead. Because of the cross-layer approach, the MAC layer in source s can know the frame rate τ_s of the video sent by the application layer. The MAC layer can also determine the maximum load rate R_s coming from the source's upper layers. Using this information, we can determine the maximum video frame size L_s (with overhead) as R_s/τ_s and the number of MAC layer data frames per maximum video frame N_s as L_s/l_s, where l_s is the average data load per MAC layer data frame. Given that O_s is the average MAC and physical layers overhead, t_s is

Short Interframe Space (SIFS) time, and t_a is the time required to send an acknowledgment, The TXOP limit can be found by

$$x_s^* = [\frac{L_s}{y_s}] + [\frac{O_s N_s}{y_s}] + [(2N_s - 1)t_s] + [N_s t_a], \quad (7)$$

where $\frac{L_s}{y_s}$ is the time required for transmitting the data of a single video frame and the overhead of the upper layers associated with that video frame, $\frac{O_s N_s}{y_s}$ is the time required for transmitting the associated MAC and physical layers overhead, $(2Ns - 1)t_s$ is the sum of the *SIFS* periods needed for transmitting all the packets of the video frame, and $N_s t_a$ is the time required for receiving all the acknowledgment packets of the video frame packets.

4 Performance Evaluation Methodology

We use OPNET to conduct various experiments. Table 2 summarizes the main simulation parameters. In contrast with prior studies, which deal with video streams as abstract data streams with certain bitrates, we implement a traffic source in OPNET that streams MJPEG videos as Real-time Transport Protocol (RTP) packets. Moreover, we implement a realistic video streaming client at the application layer of the proxy station. This client receives and reassembles the RTP packets from various video streams, and then carries out error concealment [38] to mitigate the impact of packet loss. The use of MJPEG enables the use of standard image data sets that are suitable for surveillance applications because the MJPEG video stream is a set of JPEG images. We use the following image sets to assemble the video streams: CMU/MIT [31], Georgia Tech. [32], and FERET [33]. The streamer at each source takes a bitrate, a frame rate, and an image set as inputs and produces a corresponding MJPEG video stream. Intuitively, fair evaluations of various bandwidth allocation solutions dictate that we enable each video source to send the video stream at a bitrate that closely matches the one determined by the corresponding solution. The Georgia Tech. image set, however, produces a highly limited range of bitrates for the studied network, thereby preventing the video sources from complying with the desired bitrates. To address this issue, we experiment with this image set at 50% of the original resolution (in addition to the original resolution).

As shown in Figure 2, the implementation of the proposed solution is distributed between the video sources and the AP. Each video source sends periodically its state information (physical rate, data dropping rate, and Rate-Distortion curve characteristics) to the AP using a new management packet. This packet is called *state report* [16]. The AP then calculates λ^* using Equation (6) and distributes it using the beacon packet. Each video source s then uses λ^* from the received beacon packet to calculate f_s^* according to Equation 5 and thus its optimal application rate according to Equation (1c) and its optimal TXOP limit time according to Equation (7).

For comparative purposes, we implement the cross-layer algorithm proposed in [16], referred to here as *Distortion Optimization (DO)*. Our proposed solution is referred to as *Enhanced Distortion Optimization (EDO)*. We also compare the results with standard EDCA. In addition, we implement a variation of the standard EDCA, called *Adaptive EDCA*, in which the application layer in each video source adapts its video rate according to the physical rate of that source, and thus the rate is set as y_s/S.

We analyze the following **performance metrics**: perceptual video quality, average packet delay, percentage of complete frames, percentage of incomplete frames, percentage of missed frames, overall network load rate, overall buffer dropping rate, overall retransmission dropping rate, and video source power consumption. The perceptual video quality is the main metric and is measured in terms of the overall Peak Signal to Noise Ratio (PSNR). We set the PSNR of missed frames to 0. The average packet delay is an essential metric due to the real-time nature of the application and is defined as the average time needed by the AP to receive a video packet from any source. The overall network load rate is defined as the total load sent from the application layers of all sources. The power consumption is the average power consumed by each video source for capturing, encoding, and transmission and is determined using the power consumption model in [39].

Since the proposed solution utilizes a dynamic and online effective airtime estimation, the solution adapts the video streaming bitrates to achieve the best performance with the existence of *cross traffic* (also called interfering traffic) in the network. To analyze the impact of cross traffic on the performance of the proposed solution, we conduct various experiments with 8 none-video sources sharing the same wireless medium as the video sources. These none-video sources send best-effort traffic to the AP at the same pre-specified rate. We consider three types of cross traffic, called *Cross Traffic 1*, *Cross Traffic 2*, and *Cross Traffic 3*, with packet inter-arrival times of 0.005, 0.003, and 0.001 second, respectively.

5 Results Presentation and Analysis

5.1 Effectiveness of the Cross-Layer Approach

Let us start by demonstrating the benefits of utilizing information from different network layers in bandwidth allocation. Figure 5 compares standard EDCA and adaptive EDCA in terms of PSNR. These results indicate that adapting the application rate according to the physical rate in each video source (as done in Adaptive EDCA) improves the PSNR by at least 50%.

5.2 Analysis of the Proposed Effective Airtime Estimation Algorithm

Extensive analysis of the proposed effective airtime estimation algorithm indicates that the values of the design parameters are best to be set as fol-

Table 2: Summary of Simulation Parameters

Parameter	**Model/Value(s)**
Number of video sources	10-44
Simulation Time	10 min
Packet Size	1024 bytes
Application Rate	Optimized, Default = Max Physical Rate / No. of Sources
Video Frame Rate	20 frames/sec
Physical characteristics	Extended Rate (802.11g)
Physical Data Rate	Random from {12Mb/s, 18Mb/s, 24Mb/s, 36Mb/s, 48Mb/s, 54Mb/s}
Buffer size	256 Kb
Video TXOP limit	Optimized, Default = $3008\mu s$
Video CW_{min}	15
Video CW_{max}	31
Video AIFS	2
Short Retry Limit	7
Long Retry Limit	4
Beacon Interval	0.02 second
State Report Interval	1 second

lows to enhance the overall performance in terms of stability and convergence: $A_{thresh} = 0.0075$, $C = 0.2$, and *Estimation Period* = 5 seconds. Figure 6 shows the effective airtime over the whole time of running the EDO solution using the FERET image set. The results for the other image sets follow the same trend and thus are not shown. The figure demonstrates that the algorithm converges quickly in all studied network sizes. Note that the different trend in effective airtime with 10 nodes is attributed to network underutilization.

Figure 7 shows the impact of estimation period on the convergence and the stability of the proposed effective airtime estimation algorithm. In this figure, the EDO solution uses the Georgia Tech. set at 50% resolution with three different values of estimation period: 2, 5, and 10 seconds. As expected, the estimation algorithm with 10-second estimation period achieves the best stability and the longest convergence time, whereas a 2-second period achieves the fastest convergence and the worst stability. From now on, we choose 5 seconds for the estimation period because it serves as a good tradeoff.

The impact of A_{thresh}, which determines the allowable dropping in the network, on perceptual video quality is shown in Figure 8. This figure depicts the results of running the proposed EDO solution for two different network sizes: 28 and 32 stations. The results demonstrate that the quality of the received video streams improves with A_{thresh} up to a point and then the quality starts to worsen. The peak happens when A_{thresh} is smaller than 0.01, suggesting that the optimal perceptual video quality is achieved when packet dropping is very small.

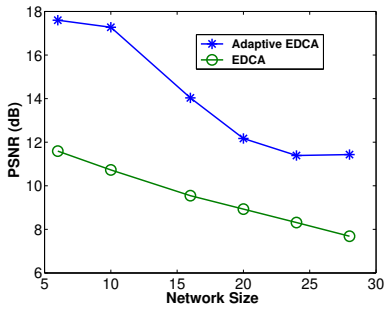


Fig. 5: Comparing EDCA with Adaptive EDCA [CMU/MIT Image Set]

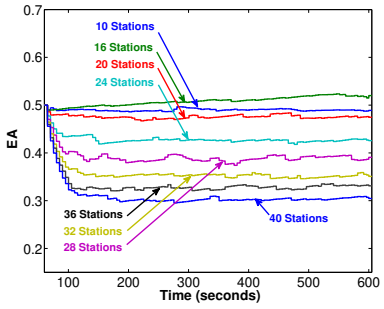


Fig. 6: Effectiveness of the Enhanced Effective Airtime Estimation [FERET Image Set]

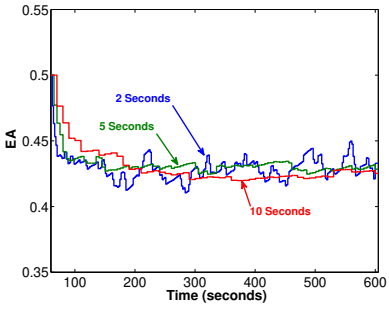


Fig. 7: Impact of Estimation Period on Enhanced Effective Airtime Estimation [Georgia Tech. Set at 50% Resolution, 24 Stations]

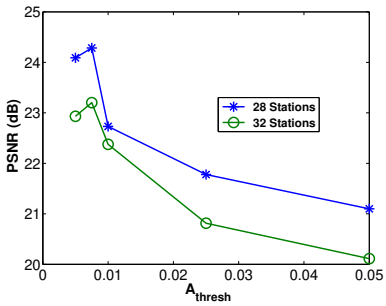


Fig. 8: Impact of Dropping Ratio Threshold (A_{thresh}) on the Perceptual Video Quality [EDO, CMU/MIT Image Set]

5.3 Analysis of Link-Layer Adaptation

As discussed in Subsection 3.5, study [16] determines the TXOP limit (x_s^*) for source s as the time required to transmit the number of MAC data frames that source s can transmit during one beacon interval time (t_b). Let us now discuss the impact of choosing time intervals other than the beacon interval. Table 3 shows the PSNR when running the solution proposed in [16] (i.e. DO) for different time intervals and network sizes. These results indicate that perceptual video quality improves with increasing the chosen time interval until a peak is reached, and then it starts to worsen. Furthermore, the best value of the time interval varies with the network size.

Table 3: Impact of the Time Interval Selected to Determine TXOP Limit on Perceptual Video Quality [PSNR (dB)]

Time Interval \ Network Size	**10**	**16**	**20**	**24**
0.02	19.15	12.30	10.25	8.81
0.5	24.46	18.77	16.17	15.69
1	24.46	19.00	16.40	15.66
2	24.46	19.00	16.46	15.66
3	24.46	19.00	16.46	15.65

5.4 Comparing Various Bandwidth Allocation Solutions

Figure 9 compares the performance of various solutions (EDCA, DO, and EDO) using CMU/MIT image set in terms of (a) perceptual video quality, (b) average packet delay, (c) percentage of complete frames, (d) percentage of incomplete frames, (e) percentage of missed frames, (f) overall network load rate, (g) overall buffer dropping rate, (h) overall retransmission dropping rate, and (i) average source power consumption. Figure 10 shows the same results when using Georgia Tech. at 50% resolution. The results for FERET and SCFace image sets follow the same trends and thus are not shown.

The results demonstrate that EDO significantly outperforms other studied solutions. In particular, it improves PSNR by more than 100% compared with standard EDCA and by 20% to 100% compared with DO. In addition, EDO greatly reduces the video packet delay. Moreover, EDO yields the highest percentage of completely received video frames and the lowest percentage of incomplete and missed frames. This behavior is attributed to both the proposed effective airtime estimation algorithm (which minimizes packet dropping) and the link-layer adaptation model. Therefore, EDO improves significantly the overall network load rate, the overall buffer dropping rate, and the overall retransmission dropping rate. These results suggest that EDO consumes much less energy for processing at the proxy station by sending and dropping much less data. Finally, EDO significantly enhances the power consumption at the video sources by 20% on the average.

5.5 Impact of Cross Traffic on the Performance of the Proposed Bandwidth Allocation Solution

In this subsection, we analyze the impact of cross traffic with different data rates on the performance of various bandwidth allocation solutions. Figures 11 and 12 show the results in terms of PSNR, Packet Delay, and Normalized Power Consumption with Georgia Tech. at 50% resolution and FERET image sets, respectively, when using the proposed EDO solution. Only these three main metrics are reported because they are the most important overall metrics. The results demonstrate that the proposed solution is highly adaptable to

interfering traffic in the network. In particular, the degradation in PSNR and the average packet delay is almost negligible even with the cross traffic of the highest data rate (Cross Traffic 3). Power consumption is reduced with cross traffic because the proposed solution adapts the video sources, thereby making them send video streams at lower bitrates.

Figures 13 and 14 compare the effectiveness of EDO, DO, and EDCA, all with Cross Traffic 3 when using Georgia Tech. at 50% resolution and FERET image sets, respectively. These results indicate that even in the presence of cross traffic, EDO greatly outperforms DO and EDCA in terms of all metrics.

Finally, Figure 15 compares the effectiveness of EDO under Cross Traffic 3 with the other two solutions applied without any interfering when the FERET image set is used. The results with other image sets as they follow the same trend. Despite that EDO runs here with the worst interfering traffic, it still outperforms DO and EDCA in terms of all metrics.

6 Conclusions and Future Work

We have proposed a cross-layer optimization solution that manages the bandwidth allocated to each video source in such a way that minimizes the total distortion in video streams. The proposed solution includes a new online network effective airtime estimation algorithm. In addition, we have developed an accurate video bitrate-distortion model as well as a new model for adapting the link-layer parameters.

The main results can be summarized as follows. (1) The proposed solution outperforms existing solutions in the perceptual quality of video streams, packet delay, and power consumed by the video sources. (2) The proposed effective airtime estimation algorithm is accurate and converges quickly. (3) Optimal perceptual video quality is achieved when the packet dropping is very small. (4) The proposed solution is highly adaptable to the existence of any interfering traffic in the network.

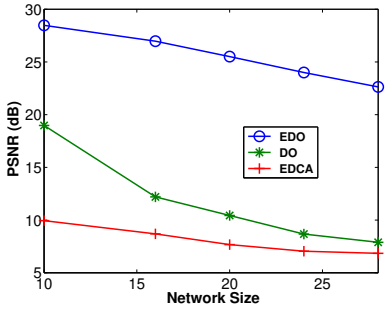
In future work, we plan to develop a cross-layer optimization approach of the computer vision accuracy.

References

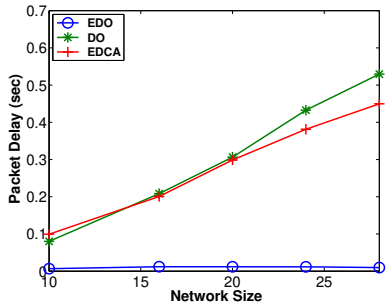
1. Chenglin Li, Dapeng Wu, and Hongkai Xiong. Delay-power-rate-distortion model for wireless video communication under delay and energy constraints. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(7):1170–1183, July 2014.
2. Rui Deng, Guizhong Liu, and Jian Yang. Utility-based optimized cross-layer scheme for real-time video transmission over HSDPA. *IEEE Transactions on Multimedia*, 17(9):1495–1507, Sept 2015.
3. Yun Ye, Song Ci, Ni Lin, and Yi Qian. Cross-layer design for delay- and energy-constrained multimedia delivery in mobile terminals. *IEEE Wireless Communications*, 21(4):62–69, August 2014.
4. Xiaoan Lu, Elza Erkip, Yao Wang, and David Goodman. Power efficient multimedia communication over wireless channels. *IEEE Journal on Selected Areas in Communications*, 21:1738–1751, 2003.

5. Zhihai He and Dapeng Wu. Resource allocation and performance analysis of wireless video sensors. *IEEE Transactions on Circuits and Systems for Video Technology*, 16:590–599, 2006.
6. Y. Andreopoulos, N. Mastronarde, and M. van der Schaar. Cross-Layer Optimized Video Streaming Over Wireless Multihop Mesh Networks. *IEEE Journal on Selected Areas in Communications*, 24(11):2104–2115, 2006.
7. Liang Zhou, Rose Qingyang Hu, Yi Qian, and Hsiao-Hwa Chen.
8. Liang Zhou. Mobile device-to-device video distribution: Theory and application. *ACM Trans. Multimedia Comput. Commun. Appl.*, 12(3):38:1–38:23, March 2016.
9. Mincheng Zhao, Xiangyang Gong, Jie Liang, Wendong Wang, Xirong Que, and Shiduan Cheng. Qoe-driven cross-layer optimization for wireless dynamic adaptive streaming of scalable videos over http. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(3):451–465, March 2015.
10. Zan Yang and Xiaodong Wang. Scalable video broadcast over downlink MIMO-OFDM systems. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(2):212–223, Feb 2013.
11. S. Cicalo and V. Tralli. Distortion-fair cross-layer resource allocation for scalable video transmission in ofdma wireless networks. *IEEE Transactions on Multimedia*, 16(3):848–863, April 2014.
12. S. Khan, J. Brehmer, W. Kellerer, W. Utschick, and E. Steinbach. Application-driven cross-layer optimization for video streaming over wireless networks. *IEEE Communications Magazine*, 44:122–130, 2006.
13. Honghai Zhang, Yanyan Zheng, Mohammad Ali Khojastepour, and Sampath Rangarajan. Cross-layer optimization for streaming scalable video over fading wireless networks. *IEEE Journal on Selected Areas in Communications*, 28(3):344–353, 2010.
14. Mihaela van der Schaar, Yiannis Andreopoulos, and Zhiping Hu. Optimized scalable video streaming over IEEE 802.11a/e hcca wireless networks under delay constraints. *IEEE Transactions on Mobile Computing*, 5:755–768, June 2006.
15. Jian-Wei Huang, Zhu Li, Mung Chiang, and Aggelos K. Katsaggelos. Pricing-based rate control and joint packet scheduling for multi-user wireless uplink video streaming. In *Proc. 15th International Packet Video Workshop (PV2006)*, 2006.
16. Cheng-Hsin Hsu and Mohamed Hefeeda. A framework for cross-layer optimization of video streaming in wireless networks. *ACM Trans. Multimedia Comput. Commun. Appl.*, 7(1):5:1–5:28, February 2011.
17. N.S. Shankar and M. van der Schaar. Performance analysis of video transmission over IEEE 802.11a/e wlans. *IEEE Transactions on Vehicular Technology*, 56(4):2346 –2362, July 2007.
18. George Kokkonis, Kostas E. Psannis, Manos Roumeliotis, and Dan Schonfeld. Real-time wireless multisensory smart surveillance with 3d-hevc streams for internet-of-things (iot). *The Journal of Supercomputing*, 73(3):1044–1062, Mar 2017.
19. Guan-Ming Su, Xiao Su, Yan Bai, Mea Wang, Athanasios V. Vasilakos, and Haohong Wang. Qoe in video streaming over wireless networks: perspectives and research challenges. *Wireless Networks*, 22(5):1571–1593, Jul 2016.
20. Ye Ge, Jennifer C. Hou, and Sunghyun Choi. An analytic study of tuning systems parameters in IEEE 802.11e enhanced distributed channel access. *Computer Networks*, 51(8):1955–1980, 2007.
21. Chun-Ting Chou, S. N. Shankar, and Kang G. Shin. Achieving per-stream qos with distributed airtime allocation and admission control in IEEE 802.11e wireless lans. In *INFOCOM*, pages 1584–1595, 2005.
22. Jie Hui and M. Devetsikiotis. A unified model for the performance analysis of IEEE 802.11e EDCA. *IEEE Transactions on Communications*, 53(9):1498–1510, 2005.
23. Lixiang Xiong and Guoqiang Mao. Saturated throughput analysis of IEEE 802.11e EDCA. *Computer Networks*, 51(11):3047–3068, 2007.
24. Zhenyu Yang and Klara Nahrstedt. A bandwidth management framework for wireless camera array. In *NOSSDAV*, pages 147–152, 2005.
25. Samarth H. Shah, Kai Chen, and Klara Nahrstedt. Dynamic bandwidth management in single-hop ad hoc wireless networks. *MONET*, 10(1-2):199–217, 2005.

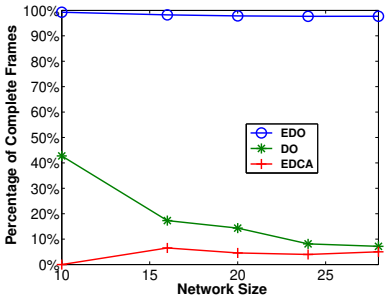
26. Cheikh Sarr, Claude Chaudet, Guillaume Chelius, and Isabelle Guerin Lassous. Bandwidth estimation for IEEE 802.11-based ad hoc networks. *IEEE Trans. Mob. Comput.*, 7(10):1228–1241, 2008.
27. Guoqiang Mao, Zihuai Lin, Xiaohu Ge, and Yang Yang. Towards a simple relationship to estimate the capacity of static and mobile wireless networks. *IEEE Transactions on Wireless Communications*, 12(8):3883–3895, August 2013.
28. Weihuang Fu and D.P. Agrawal. Capacity of hybrid wireless mesh networks with random aps. *IEEE Transactions on Mobile Computing*, 12(1):136–150, Jan 2013.
29. L. Budzisz, F. Ganji, G. Rizzo, M.A. Marsan, M. Meo, Yi Zhang, G. Koutitas, L. Tassiulas, S. Lambert, B. Lannoo, M. Pickavet, A. Conte, I. Haratcherev, and A. Wolisz. Dynamic resource provisioning for energy efficiency in wireless access networks: A survey and an outlook. *IEEE Communications Surveys Tutorials*, 16(4):2259–2285, Fourthquarter 2014.
30. R. Laufer and L. Kleinrock. The capacity of wireless csma/ca networks. *IEEE/ACM Transactions on Networking*, 24(3):1518–1532, June 2016.
31. CMU/MIT image set. http://vasc.ri.cmu.edu/idb/html/face/frontal_images.
32. Georgia tech face database. http://www.anefian.com/research/face_reco.htm.
33. P. Jonathon Phillips, Harry Wechsler, Jeffrey Huang, and Patrick J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vision Comput.*, 16(5):295–306, 1998.
34. Mislav Grgic, Kresimir Delac, and Sonja Grgic. Scface - surveillance cameras face database. *Multimedia Tools Appl.*, 51(3):863–879, 2011.
35. Ze-Nian Li, Mark S. Drew, and Jiangchuan Liu. *Fundamentals of Multimedia*. Springer, 2014.
36. Wenye Cheng, Xi Chen, and Zhihai He. Doubling of the operational lifetime of portable video communication devices using power-rate-distortion analysis and control. In *ICIP*, pages 2473–2476, 2006.
37. A Ortega and K Ramchandran. Rate-distortion methods for image and video compression. *IEEE Signal Processing Magazine*, 15(6):23–50, 1998.
38. S. Shirani, F. Kossentini, and R. Ward. Reconstruction of baseline jpeg coded images in error prone environments. *IEEE Transactions on Image Processing*, 9(7):1292–1299, July 2000.
39. Yousef Sharrab and Nabil J. Sarhan. Aggregate power consumption modeling of live video streaming systems. In *ACM Multimedia Systems (MMSys 2013)*, Oslo, Norway, February 2013.



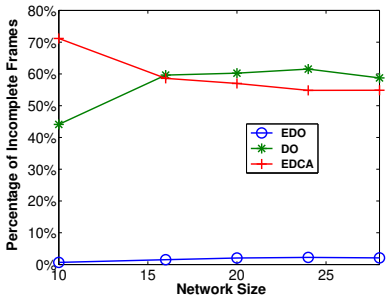
(a) Perceptual Video Quality



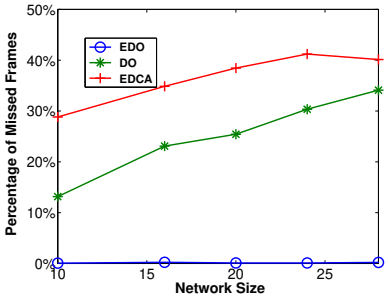
(b) Packet Delay



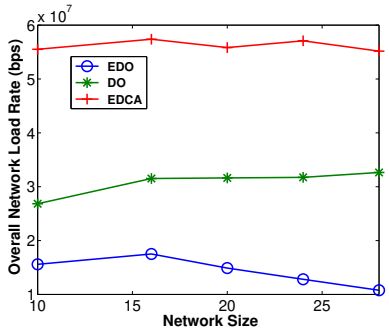
(c) Percentage of Complete Frames



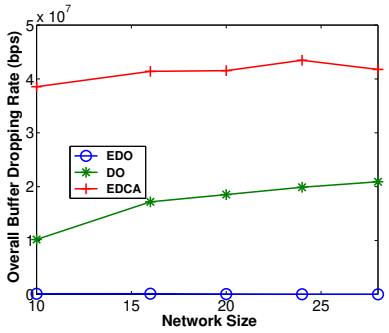
(d) Percentage of Incomplete Frames



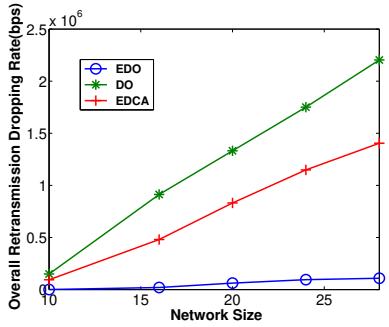
(e) Percentage of Missed Frames



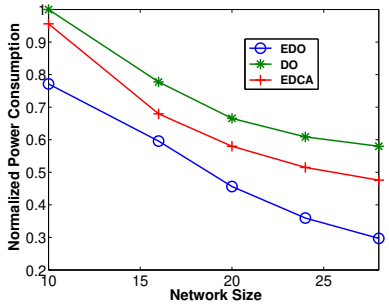
(f) Overall Network Load Rate



(g) Overall Buffer Dropping Rate

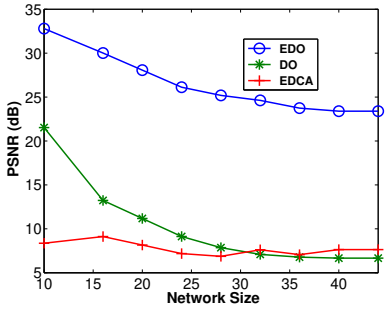


(h) Overall Retransmission Dropping Rate

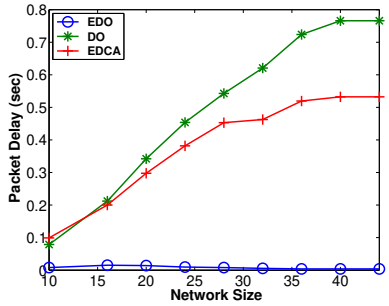


(i) Normalized Power Consumption

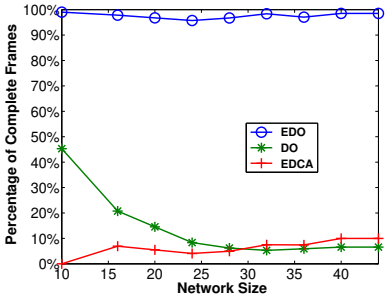
Fig. 9: Comparing Various Bandwidth Allocation Solutions [CMU/MIT Image Set]



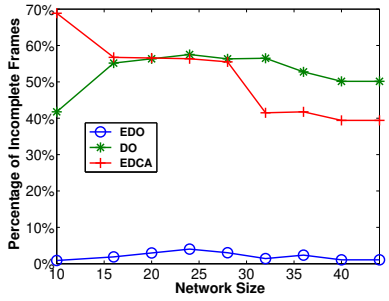
(a) Perceptual Video Quality



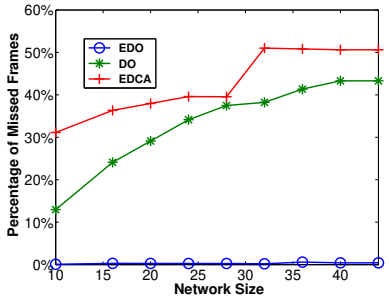
(b) Packet Delay



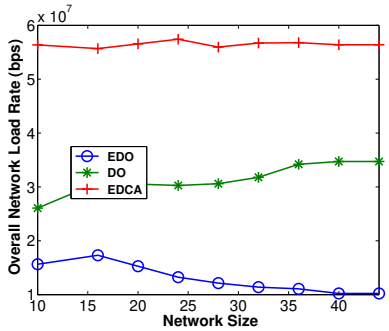
(c) Percentage Complete Video Frames



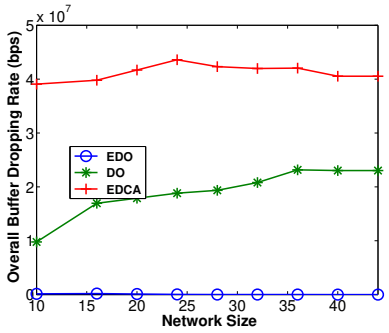
(d) Percentage Incomplete Video Frames



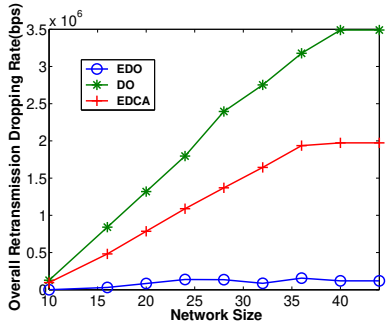
(e) Percentage Missed Video Frames



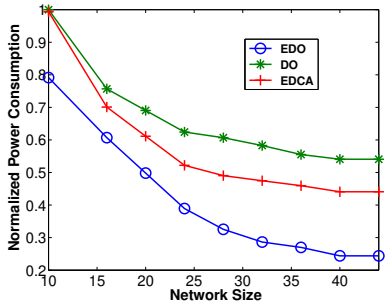
(f) Overall Network Load



(g) Overall Buffer Dropping Rate

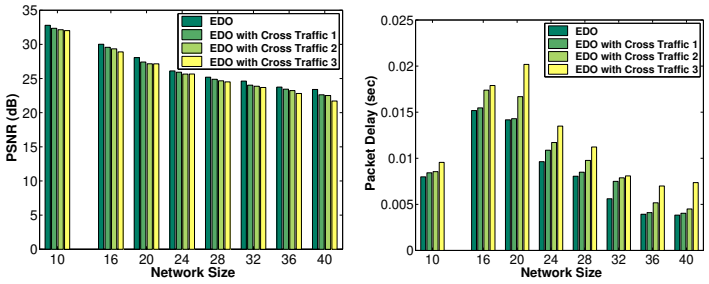


(h) Overall Retransmission Dropping Rate



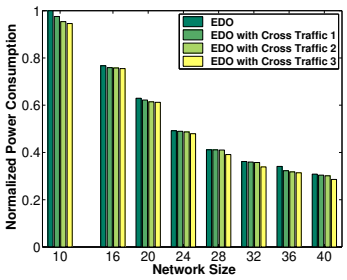
(i) Normalized Power Consumption

Fig. 10: Comparing Various Bandwidth Allocation Solutions [Georgia Tech Image Set At 50% Resolution]



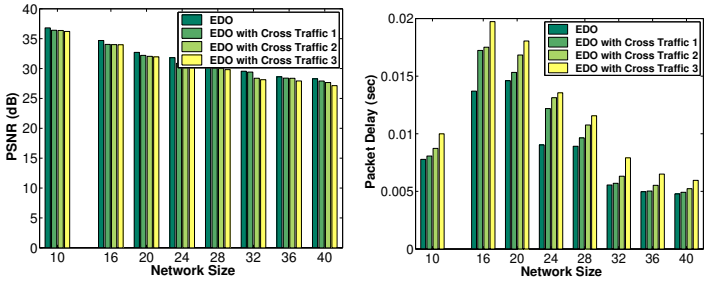
(a) Perceptual Video Quality

(b) Packet Delay



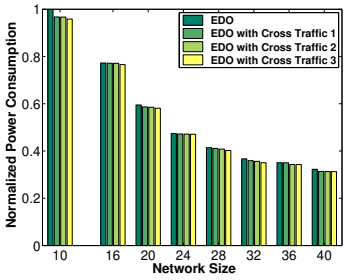
(c) Normalized Power Consumption

Fig. 11: Impact of Cross Traffic on EDO [Georgia Tech. Set at 50% Resolution]



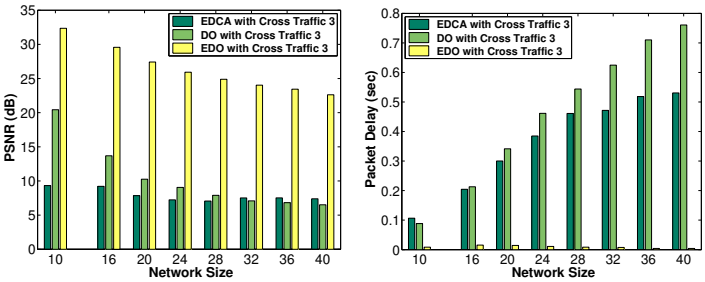
(a) Perceptual Video Quality

(b) Packet Delay

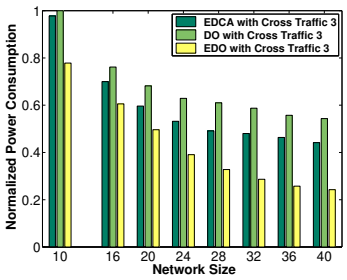


(c) Normalized Power Consumption

Fig. 12: Impact of Cross Traffic on EDO [FERET Set]

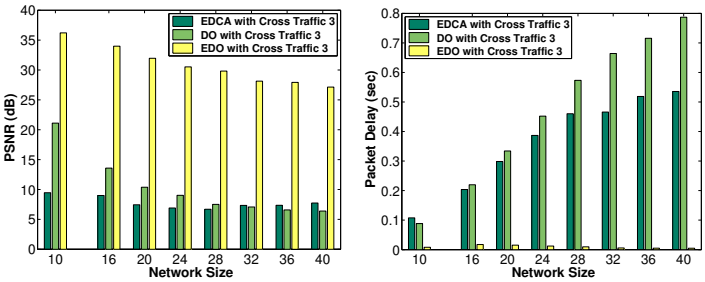


(a) Perceptual Video Quality (b) Packet Delay



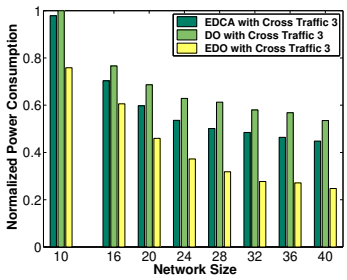
(c) Normalized Power Consumption

Fig. 13: Comparing EDCA, DO, and EDO, all with Cross Traffic 3 [Georgia Tech. Set at 50% Resolution]



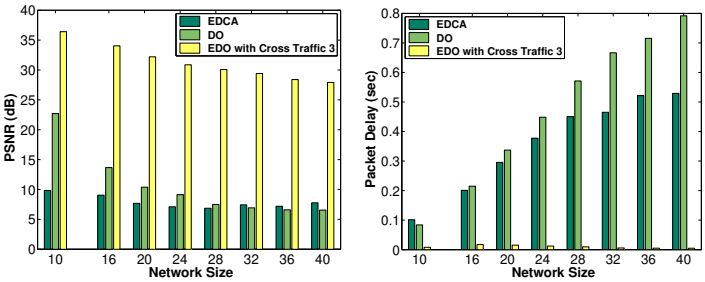
(a) Perceptual Video Quality

(b) Packet Delay

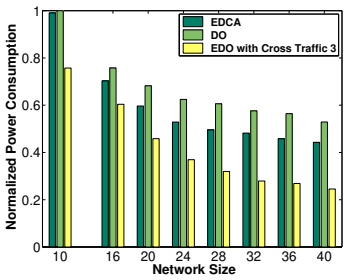


(c) Normalized Power Consumption

Fig. 14: Comparing EDCA, DO, and EDO, all with Cross Traffic 3 [FERET Image Set]



(a) Perceptual Video Quality (b) Packet Delay



(c) Normalized Power Consumption

Fig. 15: Comparing EDO under Cross Traffic 3 with EDCA and DO [FERET Image Set]