RATE ALLOCATION FOR MULTI-CAMERA SURVEILLANCE
OVER AN AD HOC WIRELESS NETWORK

(Invited Paper)

Xiaoqing Zhu, Eric Setton and Bernd Girod
Information Systems Laboratory
Stanford University, Stanford, CA 94305
{esetton, zhuxq, bgirod}@stanford.edu

Abstract—We propose a dynamic rate allocation scheme among multiple video sources over an ad hoc wireless network. With a total rate constraint capturing the media access characteristic of an 802.11 network and the selected route for each sender, our rate allocation algorithm minimizes the weighted sum of video distortion for all sources. The optimization is performed at the common receiver of the video streams and the decision is based on a simple rate-distortion model of video and a fast algorithm to classify the activity level of the raw video content. Experimental results for a video surveillance scenario demonstrate that the proposed scheme outperforms a simple fixed quality scheme and the conventional TCP-friendly rate control. For most cameras, the average PSNR of received video quality in the active periods from the proposed scheme exceeds those from the two competing schemes by a margin of $1.2 \times 7.1 \, \text{dB}$.

I. INTRODUCTION

Recently, there has been a growing interest in ad hoc wireless networks, where nodes communicate with each other without the support of a fixed infrastructure. In such a network, each node can act as a source, a destination, or as a relay for the traffic of other nodes. While this provides appealing features of rapid deployment and flexible configuration for many applications, technical challenges also arise when the network is required to support real-time media applications such as video streaming. The demanding rate and delay constraints are difficult to accommodate in a network where bandwidth is scarce and shared among users, node mobility may require frequent route updates and energy is limited at each node. The combination of these challenges calls for novel design paradigms.

Unlike in the conventional network structure with protocols independently designed for each layer, cross-layer design allows information sharing across the different layers, and considers adaptive power control, media access control, routing and source coding jointly for efficient utilization of network resources. This idea has been explored both in its theoretical aspects, with convex optimization formulations [1][2], and for more practical systems concerns [3][4].

We propose to exploit the advantages of cross-layer information exchange in the case where multiple video sources share the wireless medium of a multi-hop wireless network to transmit content to a central location via an ad hoc routing protocol. The aim of the paper is to allocate rate optimally between the different sources. In this situation, instead of adjusting the rate of each user independently at the transport layer, as the congestion control in TCP [5] or TCP-Friendly Rate Control (TFRC) [6], one can do so according to the importance of the transmitted content. For video streaming specifically, this importance may be reflected in the rate-distortion (RD) characteristics of the encoded video, as well as in the relevance of the content itself. In addition, the rate allocation also needs to function as a rate control scheme to avoid excessive network congestion. For this purpose, it should take into consideration routing and media access as well.

In this paper, we describe a dynamic cross-layer rate allocation scheme, based on fast classification of video content and simple RD models extracted for each video stream. We provide a convex optimization formulation of the rate allocation problem, with additional constraints reflecting the carrier sense mechanism of the 802.11b protocol. The optimization can be solved in closed form, and the rate allocation procedure can be integrated with the routing protocol.

The rest of the paper is organized as follows. We briefly explain the network model based on the IEEE 802.11b protocol in Section II. A low complexity classification algorithm is provided in Section III to distinguish different levels of activity of the video content. A rate-distortion model of video is described in Section IV. The algorithm for rate allocation among multiple video sources is explained in Section V. In Section VI, we apply the proposed scheme to a video surveillance scenario and compare its performance with two conventional approaches: fixed quality encoding and TFRC rate control.
II. NETWORK MODEL

In this section, we provide the capacity region achievable by time sharing for a set of $N$ wireless nodes communicating via the 802.11b protocol in ad hoc mode at a single rate. A more general analysis is presented in [4].

In a wireless network, at one time instant, only a limited subset of nodes are generally transmitting simultaneously.

In the following, we describe formally the actions at this time instant and denote elementary transmission scheme a given arrangement of senders and receivers. Consider an elementary transmission scheme $k$ in which there are $n$ sender / receiver pairs. We denote $\{t_{k_1}, \ldots, t_{k_n}\} \in \{1, \ldots, N\}^n$ the set of senders, and $\{r_{k_1}, \ldots, r_{k_n}\} \in \{1, \ldots, N\}^n$ the set of receivers. If these nodes communicate via the 802.11 protocol at rate $R_{tx}$, the protocol imposes that the distance between each sender / receiver pair be less than the transmission range for this rate. In addition, through the exchange of RTS (Request To Send) / CTS (Clear To Send) packets before data transmission, the protocol ensures that each source-destination pair is separated from the other transmitting and receiving nodes by more than one carrier sense range.

In this case, the exchanged data rate between a given sender / receiver pair $(i,j)$ is $R_{tx}$. The rate transmitted for the elementary transmission scheme $k$ between each pair of nodes can be described by the matrix $R_k$ for which each coefficient $R_{k,ij}$ indicates the rate transmitted between node $i$ and $j$:

$$R_{k,ij} = \begin{cases} R_{tx}, & (i,j) \in \{(t_{k_1},r_{k_1}), \ldots, (t_{k_n},r_{k_n})\}, \\ 0, & \text{otherwise} \end{cases}$$

Different matrices $R_k$ should be combined to reflect the rate exchanged between the nodes of the network averaged over different time instants. To determine whether a set of rates between different wireless nodes may be supported by the network, one must find a convex linear combination of these matrices which gives higher or equal data rates for all sender / receiver pairs. The space spanned by the time sharing combinations of the matrices $R_k$ describes the capacity region $C$:

$$C = \left\{ \sum_{k=1}^{L} \lambda_k R_k : \lambda_k \geq 0, \sum_{k=1}^{L} \lambda_k \leq 1 \right\}.$$  

In this equation, $\lambda_k$’s indicates the fraction of time the elementary transmission scheme $k$ is used.

In a small network where all nodes are within one carrier sense range, only one sender can transmit at each time instant. In this case, the capacity region reduces to time-sharing among single-link transmissions.

The maximum data rate supportable between each sender / receiver pair separated by less than a transmission range can be expressed as:

$$R_{max} = \alpha \times R_{tx}$$

where the scaling factor $\alpha < 1$ accounts for the overhead of MAC layer control packets including RTS, CTS and ACK. Although this scaling factor should depend in reality on factors such as packet size, packet loss rate, exponential backoff, we will assume in the following that all active links are equivalent and that this parameter is constant.

In this case, it is easy to show that the classic minimum hop routing provides highest throughput for each user, as each data packet occupies the transmission medium for the minimum amount of time. Note that we also assume single-rate transmission. Routing in the multi-rate transmission case is analyzed in related work [7], [8].

III. VIDEO CONTENT CLASSIFICATION

As video content changes over time, its RD tradeoff is difficult to capture with one single RD model. For surveillance applications in particular, the RD characteristic of the active scenes containing a moving foreground can be considerably different from that of the static scenes.

We therefore pre-process the captured video frames before encoding and classify them according to the level of activity. The classifier divides each frame into $16 \times 16$ blocks and computes the sum of absolute differences (SAD) between the blocks at the same location on the current and previous frame. The computed SAD is normalized by the mean of the block and compared with a given threshold. For each video frame, the percentage of all the constituent blocks exceeding the threshold is in turn compared to a cutoff percentage to determine whether the frame is active or not.\(^1\)

Figure 1 illustrates the result of the classification for a short video sequence in which a car passes in front of a static background. The size of compressed video frames encoded at a fixed quantization level is also represented for two different qualities. All the frames are encoded as P-frames. When the car is moving in the foreground, the size of the video frames increases as more bits are required to represent the video signal. The correspondence between the level of activity and the size of the encoded frames confirms the effectiveness of the classification which can be used to predict activity periods.

IV. VIDEO DISTORTION MODEL

For different levels of activity of a captured scene the encoder distortion may be fitted by the following model:

$$D_{enc} = D_0 + \theta/(R - R_0),$$

\(^1\)More sophisticated classification techniques which would separate the video into more states would lead to an even better characterization of the video content. They are not analyzed in this work. In the rest of the paper, we use a binary classifier, however the rate allocation presented in the next section is considered for an arbitrary number of activity levels.
where $R$ is the rate of the video stream, and the parameters $D_0$, $\theta$ and $R_0$ are estimated from empirical RD curves via regression techniques [9], [10].

Figure 2 shows the fit obtained for different video sequences recorded by different cameras. The RD fit for the two different levels of activity is significantly different. For the same video quality, the difference in terms of bit rate for the two different levels sometimes exceeds one order of magnitude. Note that the curves are also very different for both cameras which indicates that the RD characterization should be extracted for each video source and each activity level separately.

The rate allocation among $N$ video sources, each at activity level $l_n$, can be formulated as the following constrained optimization problem:

$$
\min_{R_n} \sum_{n=1}^{N} w_{l_n}^n D_n^i (R_n^i)
$$

subject to

$$
\sum_{n=1}^{N} \alpha_n^l R_n^l \leq R_{max}
$$

In (6), the linear constraint stems from the medium access characteristic discussed in Section II. Note that the rate contribution from each video source is weighted by the number of hops $\alpha_n$ required to relay that stream.

The other set of weights, $w_{l_n}^n$, can be adjusted so that streams of active contents are treated with higher priority than the rest, and are allocated an increased rate. In the experiments in Section VI, the weights for active streams is twice that of the inactive streams.

Applying the Lagrangian multiplier technique to the above problem, a closed-form solution can be expressed as:

$$
R_n^l = R_{0,n}^l + \sqrt{\frac{w_{l_n}^n \theta_{l_n}^n}{\alpha_n^l \lambda}}
$$

where the Lagrangian multiplier $\lambda$ can be calculated as:

$$
\lambda = \left( \frac{\sum_{n=1}^{N} \sqrt{\alpha_n^l w_{l_n}^n \theta_{l_n}^n}}{R_{max} - \sum_{n=1}^{N} \alpha_n^l R_{0,n}^l} \right)^2
$$

After simplification, the rate allocation can be expressed as:

$$
R_n^l = R_{0,n}^l + \sqrt{\frac{\theta_{l_n}^n}{\alpha_n^l \lambda}} \sum_{n=1}^{N} \alpha_n^l w_{l_n}^n \theta_{l_n}^n - \sum_{n=1}^{N} \alpha_n^l R_{0,n}^l
$$

depending on the weights $w_{l_n}^n$, the RD parameters for each camera $\theta_{l_n}^n$, $\alpha_n^l$, $R_{0,n}^l$, and the total rate constraint $R_{max}$.

VI. NETWORK SIMULATION

A. Scenario Description

We evaluate the performance of the rate allocation scheme in a surveillance application where cameras self-organize into a wireless network to transmit video to a central location. Video sequences were recorded at 7 different locations around a residential area on the campus of Stanford University. The layout of the area and the camera positions are illustrated in Fig. 3, where Node 0 acts as the central location for video collection. The
available transmission links in the network are marked with dashed lines. The recorded videos show an event of a car and a bike riding in opposite directions on the path around the residential area. Figure 4, shows some examples of the captured scenes.

![Surveillance Camera Layout](image)

**Fig. 3.** Camera layout and network topology

![Frames from 4 different captured sequences](image)

**Fig. 4.** Frames from 4 different captured sequences

Video sequences are captured at the resolution of 352 × 240 pixels and 15 frames per second. They are then compressed off-line with the H.264 codec. Each frame is a P-frame. The encoding is performed at different quantization parameters in order to fit the RD characterization for each camera at both activity levels.

Network communication between the wireless nodes is simulated in NS-2. The topology of the simulated network is matched with the real camera positions by hand. To simulate live encoding, each node is also provided with a complete video description, including the classification result and the packet sizes of each encoded frame at different quantization parameters. The wireless nodes follow the 802.11b protocol in ad hoc mode, and operate at 5.5 Mbps for each link. With a two-way ground propagation model, the nodes are within one carrier sense range (approximately 1 km), therefore only one transmitter is active at a time. However, this does not imply a fully connected network as the transmission range is much shorter (approximately 200 m). As a consequence, direct transmission to the collecting center is impossible for most sources as can be seen in Fig. 3 which shows all the links of the network. This explains why multi-hop routing is necessary rather than a more simple polling scheme managed by the central location as used for example in the 802.11 PCF mode.

Packets are routed using the DSR (dynamic source routing) protocol [11] which finds the shortest path upon request. At the transport layer, video frames are sent over UDP which handles fragmentation and re-assembly. When a new camera starts or the activity level of an existing camera changes, the RD characteristic of the current activity level is sent in a special request message to the central location, which re-allocates the source rate among all cameras according to (7) and notifies them by means of a special reply message. At each camera, the encoder switches the quantization parameters according to the RD model to conform to the allocated rate. As the periods of activity typically span several seconds, the delay in the rate switch does not significantly impact the performance. For robust transmission, a simple ARQ scheme is also implemented to retransmit unacknowledged packets once a gap is discovered in the acknowledgment sequence numbers. The playout deadline is set to 2.0 seconds.

B. Competing Schemes

In the experiments we also compare the performance of the proposed scheme with two other conventional rate control methods without information from the network layer incorporated into the application layer.

The first scheme imposes a fixed quantization parameter for all encoders. This is chosen to be equal at all senders and allows satisfactory delivery of all the video content. This fixed encoding is easy to implement and is commonly used in commercial applications.

In the second scheme, each sender adjusts the rate according to the end-to-end delay and packet loss rate estimated via application layer acknowledgements independently. The rate control uses the TCP-friendly rate-control (TFRC) equation:

$$ R = \frac{kS}{RTT \sqrt{p}} $$

(10)

This is true when all nodes communicate using the same frequency band. The results described in this paper can be readily extended to the case where parallel transmissions using pre-assigned non-overlapping frequency bands are possible. The potentially more interesting case of dynamic frequency allocation is not treated in this paper.
where $S, RTT, p$ and $k$ denote the packet size, the round-trip-time estimate, the loss rate estimate and a scaling constant respectively [12]. This rate control is used as an upper bound on the average rate transmitted by the sender. Note that this scheme is oblivious of the video content variation and would not allocate more rate to more active video cameras. In addition, since the algorithm relies on statistics collected from previous packets there may be more latency in the rate adaptation than for the proposed scheme. In the latter, the centralized decision is obtained within one round trip time following an activity level change for one of the cameras.

**C. Experimental Results**

In this section, we present the experimental results from the video streaming experiments, using the three rate allocation schemes described above. In general, the fixed quality scheme yields a very low constant rate at each video source, except for the natural rate increase and slight quality degradation when the video content is active. The TFRC scheme introduces more rate fluctuations at each sender, and responds to the rate changes with a lag. The proposed rate allocation scheme, on the other hand, increases the rates for active cameras and reduces the rates for cameras recording background only. The reaction time for rate re-allocation from the central location is usually within one round trip time of the video packet, typically on the order of 20ms.

As an illustration, Fig. 5 presents the instantaneous rates for the three schemes at Camera 2 during the active period. Unlike the TFRC and the fixed quality scheme, where the rate increase is purely due to the more active video content\(^3\), the proposed scheme allocates additional rate to the active camera while reducing rate of the inactive ones. This strategy results in enhanced quality of the active part of the video sequences. Note however, the allocated rate to the active region may not always be higher for the proposed scheme, as cameras with more hops are weighted more heavily in the total rate constraint of (6). During the active period for Camera 5, for instance, the allocated rate from the TFRC scheme is the higher, as it starts from a higher rate allocation for the background scenes in Camera 5 as well (Fig. 6).

Table I shows the average PSNR value of the received video during the active period for each camera. The proposed scheme outperforms the fixed quality scheme by a margin of 0.7 - 4.2 dB. When compared to the TFRC scheme, for most cameras, the proposed scheme achieves better video quality by 1.2 - 7.1 dB. The video quality at cameras 4 and 5 are inferior by 0.6 and 0.2 dB.

\(^3\)The rate variation for TFRC is due to the fact that the rate control is only considered for the average transmitted rate.

**TABLE I**

<table>
<thead>
<tr>
<th>Camera</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>31.7</td>
<td>32.2</td>
<td>31.8</td>
<td>33.5</td>
<td>29.9</td>
<td>30.5</td>
<td>33.5</td>
</tr>
<tr>
<td>TFRC</td>
<td>24.9</td>
<td>27.7</td>
<td>24.7</td>
<td>34.1</td>
<td>30.1</td>
<td>29.3</td>
<td>28.7</td>
</tr>
<tr>
<td>Fixed</td>
<td>28.5</td>
<td>29.0</td>
<td>28.4</td>
<td>29.3</td>
<td>29.2</td>
<td>28.8</td>
<td>29.4</td>
</tr>
</tbody>
</table>

![Camera 2](image1)

**Fig. 5.** Video source rate at Camera 2 during the active period

![Camera 5](image2)

**Fig. 6.** Video source rate at Camera 5 during the active period

**VII. Conclusion**

We investigate dynamic rate allocation for multiple video sources in a wireless ad hoc network. Based on a model of an 802.11 wireless network, and a RD model tuned to the video activity level at each camera, the rate allocation algorithm minimizes the weighted sum of video distortions at all cameras, while conforming to a weighted rate sum constraint. A simple classification method is used to determine the activity level of un-compressed video frames.
The performance of the system is evaluated for a video surveillance network of 7 cameras on wireless nodes using the 802.11 protocol in ad hoc mode. Network simulations incorporating real video sequences demonstrate the effectiveness of centralized rate allocation. For most cameras, the proposed algorithm outperforms the fixed quality scheme and a dynamic TCP-friendly rate control scheme by a wide margin in terms of average PSNR of the encoded active video content. Higher rate is allocated to the active periods of the video while more static scenes are encoded at a reduced quality.

ACKNOWLEDGMENT

This work has been supported, in part, by NSF grant No. CCR-0325639.

REFERENCES


