Investigation of H.264 Video Streaming over an IEEE 802.11e EDCA Wireless Testbed

Richard Haywood, Saty Mukherjee and Xiao-Hong Peng
Electronic Engineering
Aston University
Birmingham, United Kingdom
haywoorj,mukherjs,x-h.peng@aston.ac.uk

Abstract—Although a number of investigations have been conducted using IEEE 802.11e enabled networks to stream class differentiated video, very few reports are available based on a real testbed. In our work, we set up a wireless testbed for H.264 video streaming through assigning the partitioned video packets onto the DCF MAC layer and different access classes of the EDCA MAC layer. We investigate three assignment schemes: 1) DCF is used and all the traffic is treated equally; 2) video traffic is assigned to each of the access classes in turn; and 3) the packets are assigned according to their importance and the class priority. In addition to the video stream we introduce TCP traffic from three clients in the best effort class. We show that video quality can be improved through properly assigning packets to wireless access classes compared to the standard best effort scheme. Importantly, we show, based on our testbed results, that the single class assignment can achieve better performance than the multi-class assignment suggested by other researchers. Finally we show that virtual contention between traffic classes at the access point is an important issue to address.

I. INTRODUCTION

The rapid growth of IEEE 802.11 wireless networks in airports, coffee shops, homes and offices combined with the growth of services such as the BBC iPlayer[1] and YouTube[2], shows that the demand for wireless Internet access and video content are both increasing. It is very common that the last link of the connection between the Internet and end users is wireless. The fact that the wireless network has inherently higher loss rate than a traditional wired equivalent network must be taken into account for the transportation of real time video streams. The work conducted by the authors of [3] demonstrated that with a loss of just 3% of the packets can result in errors being displayed in 30% of a decoded video. Therefore, greater care needs to be taken when streaming packets across wireless links.

In this paper, we show, using a real testbed, that different assignment schemes will have diverse performances in the quality of a video stream received by a client in WLANs. We also compare the multi-class assignment scheme proposed by the authors of [4] with the single class assignment scheme and find that the latter can produce better results in the tested environment. Additionally we show how virtual contention at the access point can, in some circumstances, cause a greater number of packet losses than the wireless medium itself.

Section II explains the background of the 802.11 distributed access methods and H.264 data partitioning. We then discuss the related work reported by other researchers in Section III and present a cross-layer approach for the assignment of video packets to access classes in Section IV. In Section V we explain our testbed set up. Finally in Section VI we present the results from our experiment then we draw conclusions in Section VII.

II. BACKGROUND

A. 802.11 Wireless MAC

In our investigation we concentrate on the following distributed access methods:

- Distributed Coordination Function (DCF) from legacy 802.11 [5]
- Enhanced Distributed Channel Access (EDCA) from 802.11e [6]

In addition to DCF and EDCA, other methods, such as Point Coordination Function (PCF) [5] and Hybrid Controlled Channel Access (HCCA) [6], exist. However due to the added complexity, these access methods are not as widely deployed as DCF and EDCA. As such, in this work, we focus on DCF and EDCA.

DCF is a random access scheme which is based on carrier sense multiple access with collision avoidance (CSMA/CA). A binary exponential back-off mechanism is deployed for contention[7]. All flows and stations are treated equally when they contend for access to the medium. EDCA builds on the legacy DCF process by providing prioritisation through four access classes (AC’s) or traffic classes for service differentiation at the MAC layer. Service differentiation is provided by the following methods.

a) Arbitrary Interframe Space (AIFS): This is similar to the distributed interframe space used in DCF, except that the AIFS is varied according to the access class. Classes with smaller AIFS’s can contend for the medium before those with larger AIFS values.

b) Variable Contention Window: By giving high priority traffic smaller contention windows \((CW_{min} \text{ and } CW_{max})\), less time is spent in the back-off state, resulting in more frequent access to the medium. Lower priority classes are given larger contention windows, giving rise to less frequent access to the wireless medium.
c) Transmission Opportunity (TxOP): This allows a station that has access to the medium to transmit a number of data units without having to contend for access to the medium. In effect this is a form of frame bursting. The TxOP limit is defined in microseconds per traffic class.

Multiple access class queues can exist on a single station, contending between each class for access to the physical medium is regarded as virtual contention. The EDCA parameter set used by MADWiFi is shown in Table I.

<table>
<thead>
<tr>
<th>Access Class</th>
<th>Priority</th>
<th>CW_{min}</th>
<th>CW_{max}</th>
<th>AIFS</th>
<th>TxOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC_BK</td>
<td>Lowest</td>
<td>15</td>
<td>1023</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>AC_BE</td>
<td>Low</td>
<td>15</td>
<td>1023</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>AC_VI</td>
<td>High</td>
<td>7</td>
<td>15</td>
<td>2</td>
<td>3008 μs</td>
</tr>
<tr>
<td>AC_VO</td>
<td>Highest</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>1504 μs</td>
</tr>
</tbody>
</table>

### B. H.264 Data Partitions

The H.264 standard [8][9] is starting to be used more widely than just for research, and as such this provides a great motivator to see how it operates in a real world environment. In this work we use an 802.11e EDCA wireless network with H.264 Data Partitioning[10] that is present in the extended profile of the H.264 standard.

To encode a video sequence we first split it down into a number of raw frames, which we arrange together as a group of pictures (GOP). We then compress each frame, using H.264, into a slice. During the encoding process two different encoded slice types are produced, termed Intra and Predicted. Instantaneous Decoding Refresh (IDR) pictures contain only intra coded slices, this means that it only requires the successful reception of its own packets to be able to reconstruct the entire frame. Every GOP starts with an IDR picture which is then followed by predicted slices. During the compression of frames into predicted slices, repetition is removed from the video stream. This means that for a frame to be fully reconstructable its related intra (for frames encoded into intra slices) or both intra and predicted (for frames encoded into predicted slices) slices need to be received by the client.

When H.264 video is encoded without Data Partitioning, each slice of a frame is encoded into one Network Abstraction Layer (NAL) unit, which is in turn encapsulated into a packet. When H.264 is encoded with Data Partitioning enabled up to three NAL units per slice are produced. The three NAL units are termed partition A, B and C, respectively. Partition A contains the most important elements of the slice, including the slice header, macroblock types, quantisation parameters, prediction modes and the motion vectors. Partitions B and C contain the residual information for the intra and inter coded macroblocks, respectively. If partition A is lost then partitions B and C must be discarded. However if partition A is received then the quality is improved if partition B or C are also received.

For the transportation over an IP network the NAL units are firstly prefixed with the NAL unit type octet. The NAL unit type octet contains three fields: the Network Reference Indicator (NRI), the NAL unit type and a forbidden bit that must be one. The NRI is a two bit field which indicates if the content of the packet is used as a reference by other slices. If the NRI field is equal to 00 then the packet can be dropped without causing errors on other frames [11]. However if the value is not equal to 00 the packet should not be dropped. As any other NRI values are not defined we use 10 to represent IDR pictures and 01 to represent inter coded slices. The NAL unit type is a five bit field and contains details about the contents of the packet, as explained in [12]. Using the NAL unit type we are able to identify which packets contain partition A, B or C. The NAL unit and NAL until type octet are encapsulated within an RTP/UDP/IP packet for transporting over the network.

To enable the computation of the peak signal to noise ratio (PSNR) the decoder must produce complete frames. If data is missing the decoder uses a motion copy concealment scheme[13].

### III. Related Work

A number of previous works have investigated the assignment of H.264 data partitions to different wireless network access classes. However, the majority of them were completed through simulation. The authors of [4] presented the QoS architecture and simulated a H.264 video stream containing both IDR pictures and predicted slices. They show that by allocating partitions to access classes, in accordance with the packet’s importance, the loss percentages of the IDR and partition A packets can be reduced. This improvement comes at the cost of an increase in the number of partitions B and C lost. This paper limits their investigation to IDR and inter coded slices.

Some recent work conducted experimentally in [14] investigated the effect of varying the TxOP parameter on the video quality when using data partitioning. The TxOP parameter allows for a station to send a burst of data frames at once in a single contention attempt. This benefits streams, such as video, where the packets are dispatched in bursts. For example, IDR packets are large and can span a number of data frames. These could be transmitted in a single TxOP. In our work the standard TxOP parameters are used (see table I) and the allocation of packets into the wireless access classes (AC’s) is the main area for investigation, because changing the default parameter set dynamically would require significant modification to the standardised protocol.

The authors of [15] showed, through simulation, that the use of H.264 data partitioning provided a significant improvement (over 10 dB) over systems without data partitioning. The paper went on to compare data partitioning with the multiple description code and found that at bit error rates lower than approximately $4 \times 10^{-3}$ data partitioning performed better. However, when the bit error rate rose above this value then the multiple description code performed better. Our analysis will be focused on a data partitioning scheme. In addition, our previous work showed that video quality can be improved when the losses in partition A are reduced [16].
Currently differentiated QoS mechanisms such as DiffServ and EDCA only provide service differentiation on a class basis. A common problem in this situation is class hijacking, where all traffic is marked as the highest priority in order for users to gain unfair advantages over other users. However, the class hijacking approach reduces overall QoS, as there is no differentiation between competing traffic types. Our work aims to highlight the importance of correctly classifying packets, so the overall QoS for the network as a whole is increased.

IV. COMBINED STRATEGY

For improving the quality of the video received by a wireless client we implemented a cross-layer approach using a real-world testbed, in contrast to the previous work presented in [4] where the system is implemented through simulation. In the testbed an indication of the packets importance is provided to lower layers in the stack from the video streaming server. This is achieved through the use of the Differentiated Services Code Point (DSCP) field within the IP header. We can then set the DSCP field from our serving application and vary its value depending on the packet payload. Using the knowledge we have about the contents of the packet, which can be extracted from the NAL unit header, we are able to set the DSCP value such that when the packet is passed down to the data link layer of the stack, the DSCP field is then mapped to the access classes previously defined.

In the combined strategy, termed QoS Arch, the control information is assigned to AC_VO. IDR frames and Partition A of the predicted slices is assigned to AC_VI with partitions B and C assigned to AC_BE.

V. TESTBED SETUP

Using JM v13.2 [17] we encoded a 5 minute video sequence into a sequence of Real-time Transport Protocol (RTP) packets. A Common Intermediate Format (CIF) size video is encoded with a GOP of length 36 frames which is initially stored as a file containing a sequence of RTP packets. The number and size of the RTP packets are shown in Figure 1. We wrote a server which took the pre-coded RTP packets and placed them according to the details contained in the NAL[11] and RTP header.

![Fig. 1. Number and average size of the packets for each packet type](image)

We used a medium specification desktop PC running Debian Linux (2.6.19 kernel) equipped with a Fast Ethernet LAN and Atheros 5001X+ wireless card to create an access point (AP). The MadWifi 0.94 drivers were used in Master mode along with the bridging functions built into the kernel. CPU load was monitored during testing and the AP was not considered to be a bottleneck. The medium specification laptop clients were also Debian Linux based, using the same Atheros 5001X+ cards. The data transmission rate for the AP and clients were fixed at 54Mbps (64-QAM modulation) using the 802.11g PHY, disabling the auto fall back mechanism. RTS/CTS and specific enhancement features such as Turbo G and extended range were disabled. EDCA parameters on the access point were left at default as prescribed in Table I.

Underlying TCP traffic was generated using Iperf 2.02 [18] and placed in the AC_BE priority class. For TCP traffic we used TCP Reno combined with a 64k receiver window and a dummy data payload of 1460 bytes. The TCP traffic profile generated by Iperf is indicative of a bandwidth hungry protocol such as FTP or HTTP downloading.

The raw 802.11 packets were captured from the wireless channel using an AirPcap [19] wireless sniffer and tshark protocol analyser running on an additional PC. The raw packet captures were filtered in Wireshark to isolate the UDP/RTP video stream from the background TCP traffic. From this we were able to produce the MAC layer retransmission distribution by analysing the 802.11 frame sequence numbers. The same data was also used to extract the RTP packet numbers for calculating the number and type of the packets sent by the AP.

![Fig. 2. Testbed setup](image)

The testbed network topology is shown in Fig. 2, where three laptops were placed approximately 5M away from the access point, without direct line of sight, but with no major obstacles, such as walls, in the way. The laptops are used to represent a potential home wireless network, where video and other services may be used simultaneously.

Using the previously mentioned testbed we completed a series of experiments. Firstly, we streamed the video to a single client over a network using the legacy DCF protocol, where all the video and other traffic contended equally for the medium. Secondly, the video was assigned to just a single class using EDCA and streamed over the network. Finally, the packets were assigned across three EDCA access classes according to their importance, as shown in Table II.
TABLE II
ACCESS CLASS ASSIGNMENT

<table>
<thead>
<tr>
<th>Access Assignment</th>
<th>DCF</th>
<th>AC_BK</th>
<th>AC_BE</th>
<th>AC_VI</th>
<th>AC_VO</th>
<th>QoS arch</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC_BK</td>
<td>All Video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC_BE</td>
<td>All Video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC_VI</td>
<td>All Video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC_VO</td>
<td>All Video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCF</td>
<td>All Video</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE III
PERCENTAGE OF PACKETS TRANSMITTED BY THE ACCESS POINT

<table>
<thead>
<tr>
<th>Access Assignment</th>
<th>IDR packets</th>
<th>Partition A packets</th>
<th>Partition B packets</th>
<th>Partition C packets</th>
<th>Total packets</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC_VO</td>
<td>100.000%</td>
<td>99.972%</td>
<td>99.986%</td>
<td>99.971%</td>
<td>99.980%</td>
</tr>
<tr>
<td>AC_VI</td>
<td>99.502%</td>
<td>99.390%</td>
<td>98.905%</td>
<td>99.662%</td>
<td>99.000%</td>
</tr>
<tr>
<td>AC_BE</td>
<td>99.834%</td>
<td>99.948%</td>
<td>99.914%</td>
<td>99.947%</td>
<td>99.930%</td>
</tr>
<tr>
<td>AC_BK</td>
<td>99.834%</td>
<td>99.972%</td>
<td>99.894%</td>
<td>99.947%</td>
<td>99.950%</td>
</tr>
<tr>
<td>QoS arch</td>
<td>98.673%</td>
<td>99.127%</td>
<td>98.910%</td>
<td>99.102%</td>
<td>99.050%</td>
</tr>
<tr>
<td>DCF</td>
<td>98.673%</td>
<td>99.122%</td>
<td>98.905%</td>
<td>99.092%</td>
<td>99.040%</td>
</tr>
</tbody>
</table>

TABLE IV
PERCENTAGE OF PACKETS RECEIVED BY THE CLIENT

<table>
<thead>
<tr>
<th>Access Assignment</th>
<th>IDR packets</th>
<th>Partition A packets</th>
<th>Partition B packets</th>
<th>Partition C packets</th>
<th>Total packets</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC_VO</td>
<td>100.000%</td>
<td>99.972%</td>
<td>99.986%</td>
<td>99.971%</td>
<td>99.980%</td>
</tr>
<tr>
<td>AC_VI</td>
<td>100.000%</td>
<td>99.972%</td>
<td>99.986%</td>
<td>99.971%</td>
<td>99.980%</td>
</tr>
<tr>
<td>AC_BE</td>
<td>99.171%</td>
<td>99.944%</td>
<td>99.957%</td>
<td>99.710%</td>
<td>99.870%</td>
</tr>
<tr>
<td>AC_BK</td>
<td>99.171%</td>
<td>99.803%</td>
<td>99.726%</td>
<td>99.657%</td>
<td>98.890%</td>
</tr>
<tr>
<td>QoS arch</td>
<td>99.171%</td>
<td>99.122%</td>
<td>81.396%</td>
<td>87.312%</td>
<td>99.510%</td>
</tr>
<tr>
<td>DCF</td>
<td>98.673%</td>
<td>99.122%</td>
<td>98.905%</td>
<td>99.092%</td>
<td>99.040%</td>
</tr>
</tbody>
</table>

VI. RESULTS

Using Iperf for monitoring the throughput of the best effort traffic, we found that when the video was also sent in the AC_BE class the best throughput was obtained. When the video was allocated to the AC_VO the worst throughput was seen by Iperf.

In all the cases of AC_BK assignment the video decoder was unable to decode all of the frames within the sequence. The average decodable duration was 1.88 minutes. In all other scenarios the entire video was decoded. The reason for the video decoder to fail is when too much control information is not received then the decoder does not have enough information to operate.

Figure 3 shows that by replacing DCF with any access class other than AC_BK an improvement in the PSNR value of the video can be achieved. Of the individual classes, assigning the packets to the background class produces the biggest drop in picture quality. The reason for this is the larger AIFS and contention window being used in the AC_BK class. Table III highlights the issue of virtual contention at the access point, which displays the percentage of packets actually transmitted by the access point. It can be seen clearly that in the cases where all the packets are sent in the same access class (i.e. AC_BE and DCF) the percentage of video packets dropped by the access point is higher than in the other assignments.
VI. When video is streamed to the access point, there are fewer packet losses, and the video quality is better than when using QoS BE and DCF. The majority of packet losses in the wireless network occur at the access point, and they are not due to virtual contention at the access point. We also see that when the video is allocated to AC_VO or AC_VI, the video quality is better than when using the QoS arch assignment.

Table VI shows the percentage of packets received by the video client. It can be seen in some circumstances that a greater percentage of packets are lost due to virtual contention at the access point than in the transmission from the access point to the client. Figure 4 shows the jitter received by the video stream. The jitter values are within the range $\pm 2\text{ms}$. The smallest jitter is experienced when the video is streamed in the AC_VO class. Figure 5 shows that AC_VO requires fewer retransmissions than the AC_VI case, but, as we have already seen, it has the smallest jitter. The reason that AC_VO has both a smaller jitter and more retransmissions, compared to AC_VI, is because it has shorter contention time between successive attempts to access the medium.

VII. CONCLUSION

In our work, we set up a real IEEE 802.11e testbed for streaming video with class assignments. To our knowledge this is the first practical testbed combining 802.11e EDCA with H.264 data partitioning. Our testbed videos were streamed in each of the EDCA access classes as well as legacy DCF. On top of the video streams we also created best effort TCP traffic to congest the network. From our results we show that assigning video to either AC_VO or AC_VI provides a visible improvement to the received quality at the client, compared to using AC_BE or DCF.

In addition to the experiments for each access category we also tested the proposed QoS arch class assignment scheme. We found that although QoS arch can provide an improvement over assigning packets to the best effort class, it does not provide any improvement over assigning the video to AC_VO or AC_VI.

In our tests we have seen the effect of virtual contention on the packet stream. In the case of AC_BE and DCF the majority of packet losses occurred through virtual contention. For example with DCF the packet loss rate due to virtual contention is 0.05%, which is much higher than the 0.01% corrupted in the wireless network.

Our future work will extend the tests to involve a wider range of facilities in order to find out under which circumstances the QoS arch class assignment can prove to be beneficial. This will include the investigation of the effect of using different sets of hardware equipment with varied EDCA parameters and background traffic loads.

REFERENCES