2007-09-01

Video Frame Differentiation for Streamed Multimedia over Heavilty Loaded IEEE 802.11e WLAN using TXOP

Nicola Cranley
Technological University Dublin, nicola.cranley@tudublin.ie

Mark Davis
Technological University Dublin, mark.davis@tudublin.ie

Recommended Citation

Cranley, N. & Davis, M. (2007) Video frame differentiation for streamed multimedia over heavilty loaded IEEE 802.11e WLAN using TXOP. Proceedings of the 18th. *Annual IEEE International Symposioum on Personal, Indoor and Mobile Radio Communication (PIMRC),* Athens, Greece, 3-7 September, 2007.

The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)

VIDEO FRAME DIFFERENTIATION FOR STREAMED MULTIMEDIA OVER HEAVILY LOADED IEEE 802.11E WLAN USING TXOP

Nicola Cranley
CNRI, Focas Institute,
DIT Kevin Street, Dublin 8, Ireland

Mark Davis
CNRI, Focas Institute,
DIT Kevin Street, Dublin 8, Ireland

ABSTRACT

In this paper we perform an experimental investigation of using video frame differentiation in conjunction with the TXOP facility to enhance the transmission of parallel multimedia streaming sessions in IEEE 802.11e. The delay constraints associated with the audio and video streams that comprise a multimedia session pose the greatest challenge since real-time multimedia is particularly sensitive to delay as the packets require a strict bounded end-to-end delay. Video streaming applications are considered to be bursty. This burstiness is due to the frame rate of video, the intrinsic hierarchical structure of the constituent video frame types, and the different compression ratios for the different video frame types. The TXOP facility is particularly suited to efficiently deal with this burstiness since it can be used to reserve bandwidth for the duration of the packet burst associated with a packetised video frame. Through experimental investigation, we show that there is a significant performance improvement for video streaming applications under heavily loaded conditions by differentiating between the constituent video frame types. The results show that video frame differentiation reduces the mean loss rate by 12% and increases the mean PSNR by 13.1dB.

I. INTRODUCTION

There are many performance-related issues associated with the delivery of time-sensitive multimedia content using current IEEE 802.11 WLAN standards. Among the most significant are low delivery rates, high error rates, contention between stations for access to the medium, back-off mechanisms, collisions, signal attenuation with distance, signal interference, etc. For real-time multimedia applications packet loss and packets dropped due to excessive delay are the primary factors affecting the user-perceived quality. Real-time multimedia is particularly sensitive to delay as it has a strict bounded end-to-end delay constraint. Every multimedia packet must arrive at the client before its playout time with enough time to decode and display the contents of the packet. For video streams the delay incurred in transmitting the entire video frame from the sender to the client is of particular importance. The loss rates incurred due to packets being delayed past their playout time is heavily dependent on the delay constraint imposed on the video stream. Video streaming applications typically impose an upper limit on the tolerable packet loss. Specifically, the packet loss ratio is required to be kept below a threshold to achieve acceptable visual quality. Although WLAN networks allow for packet retransmissions in the event of an unsuccessful transmission attempt, the retransmitted packet must arrive before its playout time or within a specified delay constraint. If the packet arrives too late for its playout time, the packet is useless and effectively lost.

In IEEE 802.11b WLANs, the AP is a critical component that determines the performance of the network since it carries all of the downlink transmissions to wireless clients and is usually where congestion is most likely to occur. The AP can become saturated due to a heavy downlink load which results in packets being dropped from its transmission buffer and this manifests itself as bursty losses and increased delays [1]. Such losses and delays have a serious impact on multimedia streaming applications. This situation however need no longer apply following the approval of the IEEE 802.11e QoS MAC Enhancement standard which allows for up to four different transmit queues with different access priorities [2], allowing the AP to provide differentiated service to different applications and enable to meet their target QoS requirements. The IEEE 802.11e standard also defines a transmission opportunity (TXOP) as the interval of time during which a particular QSTA has the right to initiate transmissions without having to re-contend for access. During an EDCA TXOP, a QSTA is allowed to transmit multiple MPDUs from the same Access Category (AC) with a SIFS time gap between an ACK and the subsequent frame transmission [3]. The duration of the TXOP is determined by the value of the TXOP limit parameter. This TXOP mechanism is particularly suited to bursty traffic. In previous work we have shown how the *TXOPLimit* parameter should be dimensioned to improve the end-to-end delivery of video streaming applications by using a statistical analysis of the encoding characteristics of the video stream [4].

In this paper we experimentally investigate the performance of parallel multimedia streaming applications under heavily loaded conditions in conjunction with using the *TXOPLimit* parameter. We demonstrate that there is a significant performance improvement for all ACs by differentiating between the individual constituent I, P, and B video frame types. We show through experimental investigation that the mean loss rate is reduced by 12% and the mean PSNR is increased by 13.1dB.

II. EXPERIMENTAL TEST BED

To investigate the use of the 802.11e TXOP mechanism for video frame transmission, the video server was set up on the wired network and streamed video to a wireless client via the AP (Figure 1). The AP used was the Cisco Aironet 1200 using the firmware version IOS 12.3(8)JA which allowed us to access the 802.11e/WME capability of the device [5]. The AP was configured with a QoS policy where the Differentiated Services Code Point (DSCP) values in the IP

1-4244-1144-0/07/$25.00 ©2007 IEEE.

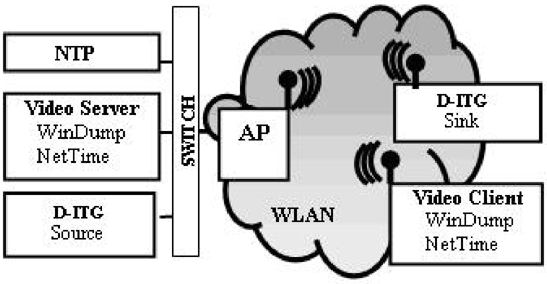


Figure 1 Experimental Test Bed

header are used to apply a particular Class of Service (CoS) to the incoming packets. Each CoS is then mapped to a particular AC where the CWmin, CWmax, AIFSN, and TXOP limit parameters can be configured. In the experiments reported here only the TXOP limit parameter is varied and the parameters CWmin, CWmax, and AIFSN were fixed with the original IEEE 802.11b settings.

The video streaming server consists of a modified version of RTPSender [6]. RTPSender reads from an encoded video file and identifies the different video frame types i.e. I, P, or B frames. The frame type indicator is used to set the IP DSCP value of the packets for this video frame. By modifying the IP DSCP value of video packets for the different frame types the AP can identify the different video frame types and assign them to the appropriate AC so that they can receive differentiated service as defined by the AP QoS policy.

Both the MultiMedia (MM) client and server used the packet monitoring tool WinDump [7] to log all packets transmitted and received and the clocks of both the client and server are synchronised before each test using NetTime [8]. However, in spite of the initial clock synchronisation, there was a noticeable clock skew observed in the delay measurements and this was subsequently removed using Paxson's algorithm as described in [9]. The delay measured here is the difference between the time at which the packet was received at the link-layer of the client and the time it was transmitted at the link-layer of the sender. The background traffic was generated using Distributed Internet Traffic Generator (D-ITG) [10]. The background traffic load had an exponentially distributed inter-packet time with a mean offered load of 6Mbps and an exponentially distributed packet size with a mean packet size of 1024B. The background traffic was transmitted from a wired source station via the AP to a wireless sink station.

III. MULTIMEDIA STREAM ANALYSIS

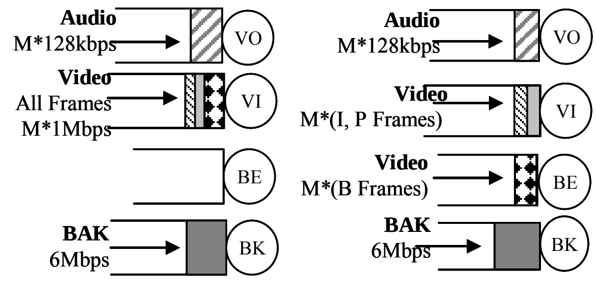
In the experiments reported here, the audio and video content was encoded using the commercially available X4Live MPEG-4 encoder from Dicas. In MPEG-4 the audio and video streams are transmitted separately through their own RTP/RTCP port pair. In this paper five different video content clips of approximately 10 minutes duration with different levels of spatial and temporal complexity were used during the experiments. DH is an extract from the film '*Die Hard*', DS is an extract from the film '*Don't Say a Word*', EL is an extract from the animation film '*The Road to Eldorado*', FM is an extract from the film '*Family Man*', and finally JR is an extract from the film '*Jurassic Park*'.

The audio track was encoded as MPEG-4 Advanced Audio Codec (AAC), 48kHz, and 128kbps CBR. The audio streams have the following characteristics: mean bit rate (130.93±15.27)kbps; mean sample size (341±40)B; maximum sample size 667B; minimum sample size 52B; Peak-to-Mean Ratio (PMR) of 1.96. The video track was encoded as MPEG-4 ASP (i.e. I, P, and B frames) with a frame rate of 24fps, a specified refresh rate of 10 (i.e. an I-frame every 10 frames), Group Of Picture (GOP) sequence (i.e. IPBBPBBPBB resulting in 3 I-frames, 6 P-frames, and 15 B-frames per second), CIF resolution and a target bitrate of 1Mbps using 2-pass encoding. In the experiments reported here the hint track MTU is 1024B for all video content types. The hint track tells the server how to optimally packetise a specific amount of media data. The hint track MTU setting means that the packet size will not exceed in the MTU size. Table 1 shows characteristics of each of the different video streams that were used in the experiments and the average over all content types. It can be seen that the combined load of the I and P-frames is less than the load of the B-frames only.

IV. EXPERIMENTAL DESIGN

In this work we shall investigate the benefits of using video frame differentiation for video streaming applications in conjunction with using the TXOP facility under heavily loaded conditions. The *TXOPLimit* parameter is an integer value in the range (0,255) and gives the duration of the TXOP interval in units of 32µs. If the calculated TXOP duration requested is not a factor of 32µs, that value is rounded up to the next higher integer that is a factor of 32µs. The maximum allowable *TXOPLimit* is 8160µs with a default value of 3008µs [3]. When there are no more packets to be sent during the TXOP interval and the channel becomes idle again, the 802.11 Hybrid Controller (HC) may sense the channel and reclaim the channel after a duration of PIFS (PCF Inter-Frame Space) after the TXOP. The *TXOPLimit* parameter is suited to bursty applications such as video as it allows for a number of packets to be transmitted during the allocated TXOP interval. In previous work [4] we have shown how the *TXOPLimit* parameter can be dimensioned by using the distribution of the encoding characteristics of the video stream. The *TXOPLimit* parameter is set to a value that relates to the mean video frame size, the number of packets required to transmit the video frame, and the time taken to transmit these packets [11]. The *TXOPLimit* parameter has shown to provide a significant improvement in the end-to-end delay to transmit video frames over WLAN. However, under heavily loaded conditions there may be significant loss due to buffer overflow at the AP and loss due to excessive delays.

Video frames are encoded in a sequence called the Group of Pictures (GOP) sequence consisting of I, P, and B frames. There is an inter-frame dependency between the different frame types. I frames are encoded autonomously, while P frames are encoded in reference to the most recent I or P frame. B frames are encoded using the closest previous and future I or P frames. The loss of an I frame results in the loss



M = Number of parallel multimedia streams

(a) (b)

Fig.2.(a) All frames are transmitted through the VI AC queue; (b) With frame differentiation I and P frames are transmitted through the VI AC queue, B frames are transmitted through the BE AC queue.

of the entire GOP and has a large impact on the video quality since the subsequent P and B frames require the I frame in order to be decoded. In contrast, the loss of a B frame that has no dependent frames has a small impact on the video quality. This frame priority information must be taken into account when dropping frames to minimize the effect of losses on the received video quality. In Table 1 we showed that the load of the B frames was greater than the combined load of the I and P frames. Video frame differentiation is achieved through the use of the DiffServ Code Point (DSCP) value in the IP header to indicate the video frame importance and drop preference. By differentiating between the constituent I, P, and B video frame types the AP can ensure higher throughput and better QoS for the video streaming applications under heavily congested conditions.

The 802.11e standard defines four AC queues into which different traffic streams can be directed: Voice (VO), Video (VI), Best-Effort (BE), and Background (BK). The AC queues were configured with IEEE 802.11b settings for CWmin, CWmax, and AIFSN while the value for *TXOPLimit* parameter is varied. In all experiments the audio streams are transmitted through the VO AC queue and a background traffic load of 6Mbps is transmitted through the BK AC queue. In this work we experimentally investigate two key scenarios as shown in Figure 2. Fig 2(a) where all video frames regardless of frame type are transmitted through the VI AC denoted as “All Frames”. Fig 2(b) where there is frame differentiation the I and P frames are transmitted through the VI AC and the B frames are transmitted through the BE AC denoted as “Differentiated Frames”.

For both scenarios the *TXOPLimit* parameter value is varied in the range: a value of 0 where the TXOP facility is not used; $\left(\overline{N}-\sigma\right)$ uses the mean minus one standard deviation number of packets required to transmit the video frames; $\left(\overline{N}\right)$ where the mean number of packets required to transmit the video frames; and $\left(\overline{N}+\sigma\right)$ uses the mean plus one standard deviation of packets required to transmit the video frames to dimension the *TXOPLimit* parameter.

V. RESULTS

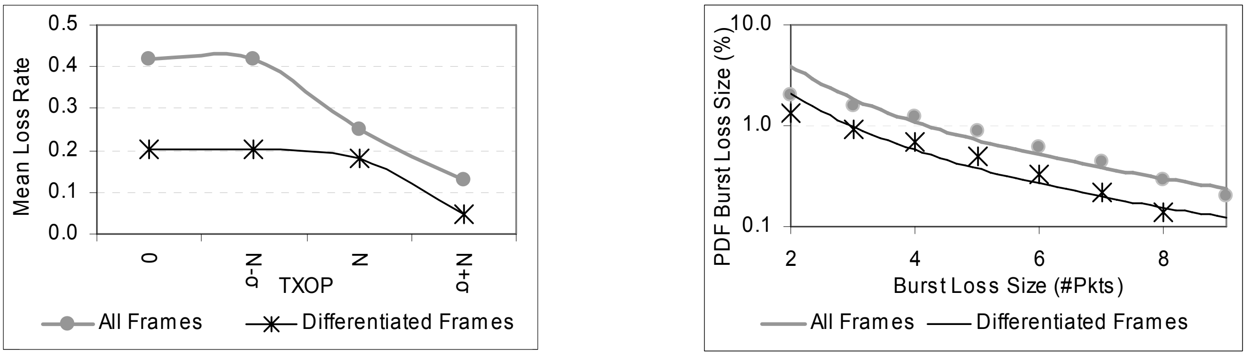
In the experiments there were 5 parallel multimedia streams with a background traffic load of 6Mbps. In our experiments the configurations “All Frames” and “Differentiated Frames” are compared. In our analysis we consider the Application Layer performance metrics: the Mean Loss Rate (MLR); the Playable Frame Rate (PFR); and the mean PSNR averaged over the 5 video streams.

Multimedia streaming applications have strict end-to-end playout delay constraint. This delay includes the transmission, queuing, and playout delay at the client device. Packets that arrive exceeding this playout delay constraint are effectively lost. The MLR corresponds to packets that have failed to be successfully received as well as those packets that have been dropped as a result of exceeding the delay constraint. The delay constraint associated with the video streaming applications used in this work is 500ms. In a WLAN environment, the bursty behaviour of video traffic has been shown to result in a sawtooth-like delay characteristic [12]. This sawtooth delay characteristic has significant implications for multimedia streaming applications and results in a bursty packet loss. Consider a frame is transmitted as a burst of n packets that are transmitted sequentially from the AP transmission buffer to the wireless client. If the $(n-k)$th packet experiences a queuing delay that causes it to exceed the delay constraint associated with the video frame resulting in it being effectively lost at the client, then the packets $(k \le n)$ packets will also be lost due to exceeding the delay constraint. Figure 3(a) show the MLR for both test scenarios and increasing *TXOPLimit* parameter value.

As the *TXOPLimit* parameter value is increased the MLR is also reduced. This is due to the fact that usage of the TXOP is

Table 1: Characteristics of the Video Content

	DH	**JR**	**EL**	**FM**	**DS**	**Mean Per Stream**
Mean Bitrate (kbps)	1633.0	980.0	1373.0	735.0	572.0	1058.6
PMR	35.4	27.9	40.0	35.4	39.4	35.6
Load I-frames (kbps)	239.0	161.0	404.0	120.0	115.0	207.8
Load P-frames (kbps)	407.0	315.0	457.0	202.0	170.0	310.2
Load B-frames (kbps)	987.0	504.0	512.0	413.0	287.0	540.6
Mean Num Pkts/I-frame	6.8	6.4	13.9	10.4	8.8	9.2
Mean Num Pkts/P-frame	2.4	2.1	1 4.7	3.5	2.9	3.1
Mean Num Pkts/B-frame	0.9	0.8	1.8	1.3	1.1	1.2



(a) (b)

Figure 3 (a) Mean Loss Rate for "All Frames" and "Differentiated Frames" with TXOP (b) PDF of Packet Burst Loss Event for "All Frames" and "Differentiated Frames"

not wasteful since when the AC queue has won a TXOP and has no more packets to send during the TXOP interval, the HC senses the channel as idle and reclaims the channel after a duration of PIFS after the TXOP. Under heavily loaded conditions with multiple video streams there is an increased likelihood that the AC queue will make use of the full duration of the TXOP interval to transmit the enqueued video packets. In this way the TXOP facility reduces the buffer occupancy which in turn reduces the likelihood of buffer flow. The PDF of the size of burst loss events consisting of 2 or more consecutively lost packets with a power trend line fitted through the data for both scenarios is shown in Figure 3(b). It can be seen that by differentiating between frames provides a significantly reduced in the MLR and reduces mean burst loss length. Although frame differentiation incurs a greater number of single packet loss events, error concealment algorithms can mask smaller packet losses. It has been shown that the burst loss length has a significant effect on the resulting distortion of the received video quality. In [13] the authors showed that as the burst length increased, the measured total distortion was much greater than the sum of the distortions for an equal number of individual losses. Regardless of *TXOPLimit* parameter value used, "Differentiated Frames" results in a MLR of 14% while "All Frames" results in a MLR of 26%. Packet loss events can be observed as distortion in the rendered video stream. The video streams were recorded and decoded using FFmpeg [14]. FFmpeg uses multiple-pass error concealment strategies to decode distorted video streams. The Peak-to Signal Noise Ratio (PSNR) was measured for each of the video streams and averaged over the 5 video streams as shown in Fig. 4. The scenario "All Frames" has a mean PSNR of 0dB for small values of the *TXOPLimit* parameter. This is due to the high loss rates incurred by the streams. Typically, video frames that are too distorted to be decoded are dropped by the decoder on the client device. Regardless of the *TXOPLimit* parameter value, the mean PSNR is 16.4dB (±15.8dB) when using a single AC queue for all video frames. In contrast by using frame differentiation the mean PSNR is 29.5dB (±6.26dB). Moreover video frame differentiation results in a lower standard deviation in PSNR which indicates the stability and continuity of the playout quality.

VI. CONCLUSIONS

In this paper, we have experimentally investigated the benefits of differentiating between the constituent I, P, and B frame types in conjunction with the TXOP facility for streamed multimedia over IEEE 802.11e WLAN under heavily loaded conditions. The TXOP facility is well-matched to efficiently deal with the inherent bursty characteristic that is associated with video streaming applications. The TXOP facility can be used to reserve bandwidth for the period of time required to transmit the video frame.

We showed that the bandwidth requirement of B-frames is greater than that of the I and P frames combined due to the higher frequency of B frames. By differentiating between the constituent frame types we can reduce the likelihood of packets relating to I or P frames being lost since these frames have a higher priority and a greater impact on the end-user QoS over B frames. Video frame differentiation is achieved through the use of the DiffServ Code Point (DSCP) value in the IP header to indicate the video frame importance and drop preference and to direct the packets relating to the I, P, and B frames to the appropriate Access Category.

Through experimental investigation, we show that there is a significant performance improvement for video streaming applications under heavily loaded conditions by differentiating between the constituent video frame types. The results show that video frame differentiation reduces the mean

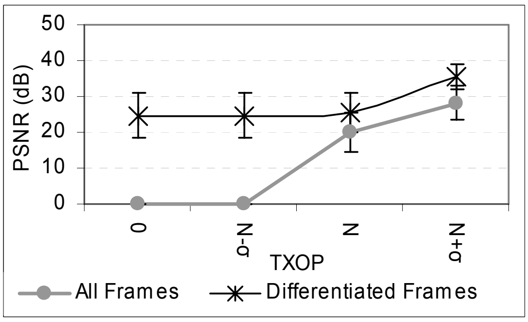


Figure 4 Mean PSNR for "All Frames" and "Differentiated Frames" with TXOP

loss rate by 12% and increases the mean PSNR by 13.1dB.

Further research is being conducted to provide prioritized differentiated service to the video streams and increase the number of parallel multimedia sessions that can be supported through an appropriate tuning of the AIFSN, CWmin, CWmax settings in conjunction with the TXOPLimit parameter.

VII. ACKNOWLEDGEMENT

The support of the Science Foundation Ireland, grant 03/IN3/I396, under the National Development Plan is gratefully acknowledged.

REFERENCES

---

[1] N. Cranley, M. Davis, "The Effects of Background Traffic on the End-to-End Delay for Video Streaming Applications over IEEE 802.11b WLAN Networks", IEEE Personal Indoor Mobile Radio Communications (PIMRC) 2006, Helsinki, Finland, September 2006

[2] Q. Ni, "Performance Analysis and Enhancements for IEEE 802.11e Wireless Networks", IEEE Network, Vol. 19, No. 4, July/August 2005, pp. 21-27.

[3] IEEE Std 802.11e, 2005 Edition, IEEE Standards for Local and Metropolitan Area Networks: Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements

[4] N. Cranley, M. Davis, "An Experimental Investigation of Parallel Multimedia Streams over IEEE 802.11e WLAN Networks using TXOP", To appear in Proc. IEEE ICC, June 25th 2007, Glasgow, Scotland.

[5] Cisco Aironet 1200, http://www.cisco.com/en/US/products/hw/wireless/ps430/index.html

[6] RTPTools, http://www.cs.columbia.edu/IRT/software/rtptools/

[7] WinDump, http://windump.polito.it/

[8] NetTime, http://nettime.sourceforge.net/

[9] S. B. Moon, P. Skelly, D. Towsley, "Estimation and Removal of Clock Skew from Network Delay Measurements", in Proc. of IEEE InfoComm'99, March 1999

[10] Distributed Internet Traffic Generator (D-ITG), http://www.grid.unina.it/software/ITG/download.php

[11] B-S Kim, S-W Kim, Y. Fang, T.F. Wong, "Two Step Multipolling MAC Protocol for Wireless LANs", IEEE JSAC in Communications, Vol. 23, No. 6, June 2005

[12] N. Cranley, M. Davis, "Delay Analysis of Unicast Video Streaming over WLAN", 2nd IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2006, Montreal, Canada, June 2006

[13] Yi J. Liang, J. G. Apostolopoulos, B. Girod, "Analysis of Packet Loss for Compressed Video: Does Burst-Length Matter?", Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-2003, vol. 5, pp. 684-7, Hong Kong, Apr. 2003.

[14] FFmpeg, http://ffmpeg.mplayerhq.hu/