2005-01-01

Performance Analysis of Network-level QoS with Encoding Configuration for Unicast Video Streaming over IEEE 802.11 WLAN Networks

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

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

Recommended Citation

Cranley, N. & Davis, M. (2005) Performance analysis of network-level QoS with encoding configuration for unicast video streaming over IEEE 802.11 WLAN networks. IEEE WirelessCom 2005, *International Conference on Wireless Networks,* Communications and Mobile Computing, Maui, Hawaii, June, 2005.

Performance Analysis of Network-level QoS with Encoding Configurations for Unicast Video Streaming over IEEE 802.11 WLAN Networks

Nicola Cranley
Communications Network Research Institute,
School of Electronic and Communications Engineering,
Dublin Institute of Technology,
FOCAS Institute,
Dublin 8, Ireland
Nikki.Cranley@CNRI.DIT.ie

Mark Davis
Communications Network Research Institute,
School of Electronic and Communications Engineering,
Dublin Institute of Technology,
FOCAS Institute,
Dublin 8, Ireland
Mark.Davis@DIT.ie

Abstract

Video streaming has a large impact on the resource requirements of the WLAN. However, there are many variables involved in video streaming, such as the video content being streamed, how the video is encoded and how it is sent. This makes the role of radio resource management extremely difficult. In this paper we investigate the effect that video encoding configurations has on the network resource requirements for unicast video streaming in a WLAN environment. We compare the network resource requirements of several content types encoded at various encoding configurations with varying I-frame frequencies, target encoding bit rates and hint track settings. We present two key findings: We show that by halving the hint track MTU values, the access requirements of the WLAN are increased by 20%. Furthermore, we show how the I-frame frequency of the encoded file relates to the resource requirements of the WLAN.

1. Introduction

Streaming multimedia over wireless networks is becoming an increasingly important service. A content provider is unlikely to have the resources to provide real-time adaptive encoding for each unicast request and as such reserves these resources for “live” multicast sessions only. Typically, pre-encoded content is transmitted by unicast streams where the client chooses the connection that most closely matches their requirements. For such unicast sessions, the adaptive streaming server can employ several techniques to adapt the pre-encoded content to match the clients’ resources. In such adaptive streaming systems, two techniques that are most commonly used are frame dropping [1] and stream switching [2].

In this paper we evaluate the effect that video encoding configurations and parameters have on the network resource requirements for unicast streaming of pre-encoded content over WLAN networks. There are a large and diverse number of variables that must be taken into consideration for unicast video streaming each of which has an impact on the resource requirements video stream on the WLAN. Such variables include:

- The actual content and complexity of the content being streamed which in turn affects the efficiency of the encoder to compress the stream.
- The compression scheme being used, that is, different compression schemes have differing levels of efficiency.
- The encoding configuration. There could be any number of possible encoding configurations possible such as the frame rate, the I-frame rate, the quantization parameter, the target bit rate (if any) supplied and target stream type i.e. VBR, CBR or near CBR.
- If the file to be streamed is .MP4 or .3gp, then a hint track must be prepared that indicates to the server how the content should be streamed.
- The streaming server being used, the rate control adaptation algorithm being used, and the methods of bit rate adaptation used by the server.

Given the large number of variables required to analyse video, in this paper we have focused on investigating the effect the encoding configuration has on the resource requirements of the WLAN, or more specifically, the effects the hint track setting, the I-frame rate, and target encoding bit rate variations have on the resource requirements in the WLAN.

This paper is structured as follows. Section one gives a brief discussion of MPEG-4 encoding, MP4 files and the importance of hint tracks. Hint tracks are required to stream MP4 and .3gp multimedia files as it tells the server how to packetise and transmit the visual elementary stream. The following section provides an analysis of the video content used during the experiments and demonstrates the burstiness and variability of the video streams used. The test bed used for the experiments and the WLAN probe used to measure the resource requirements of the WLAN are described briefly. The next section describes the experiments conducted. We show the impact on the resource utilisation of different hint track settings when transmitting the same video elementary stream. Then we discuss the importance of I-frames and investigate the network resource usage with varying I-frame frequencies. Finally, we present some conclusions and directions for future work.

2. MPEG-4

MPEG-4 dramatically advances audio and video compression, enabling the distribution of content and services from low bandwidths to high-definition quality across broadcast, broadband, wireless and packaged media [3]. In the MPEG-4 standard, there are a number of profiles, which determine the capabilities of the player to play out encoded content. The purpose of these profiles is that a codec only needs to implement a subset of the MPEG-4 standard whilst maintaining inter-working with other MPEG-4 devices built to the same profiles. The most widely used MPEG-4 visual profiles are the MPEG-4 Simple Profile (SP) and the MPEG-4 Advanced Simple Profile (ASP) and are part of the non-scalable subset of visual profiles. The main difference between MPEG-4 SP and ASP is that SP contains only I and P-frames whereas ASP contains I, P and B-frames. MP4 files comprise a hierarchy of data structures called atoms [4]. A parent atom is of type *moov* and contains the following child atoms: *mvhd* (the movie header) and a series of *trak* atoms (the media tracks and hint tracks). A *trak* represents a single independent data stream and an MP4 file may contain any number of video, audio, hint, Binary Format for Scenes (BIFS) or Object Descriptor (OD) tracks.

2.1. Hint Tracks for Streaming

Within an MP4 file, each video and audio track must have its own associated hint track. Hint tracks are used to support streaming by a server and indicate how the server should packetise the data. As with MP4 streaming, .3gp files use the "hint track" mechanism for streaming the content, although in .3gp files the BIFS and OD tracks are optional and can be ignored.

Streaming media requires that the media be sent to the client as quickly as possible with strict delay requirements. Hint tracks allow a server to stream media files without requiring the server to understand media types, codecs, or packing. Each track in a media file is sent as a separate stream, and the instructions for packetising each stream is contained in a corresponding hint track [5]. Each sample in a hint track tells the server how to optimally packetise a specific amount of media data. The hint track sample contains any data needed to build a packet header of the correct type, and also contains a pointer to the block of media data that belongs in the packet. For each media track to be streamed there must be at least one hint track. It is possible to create multiple hint tracks for any track, each optimised for streaming over different networks. Hint tracks have the same structure as media tracks and are atoms of type *trak*. Hint samples are protocol specific by specifying the protocol to be used and providing the necessary parameters for the server. The *stsd* child atom contains transport-related information about the hint track samples. It specifies the data format (currently only RTP data format is defined), the RTP timescale, the maximum packet size in bytes (MTU) and additional information such as the random offsets to add to the stored RTP timestamps and sequence number.

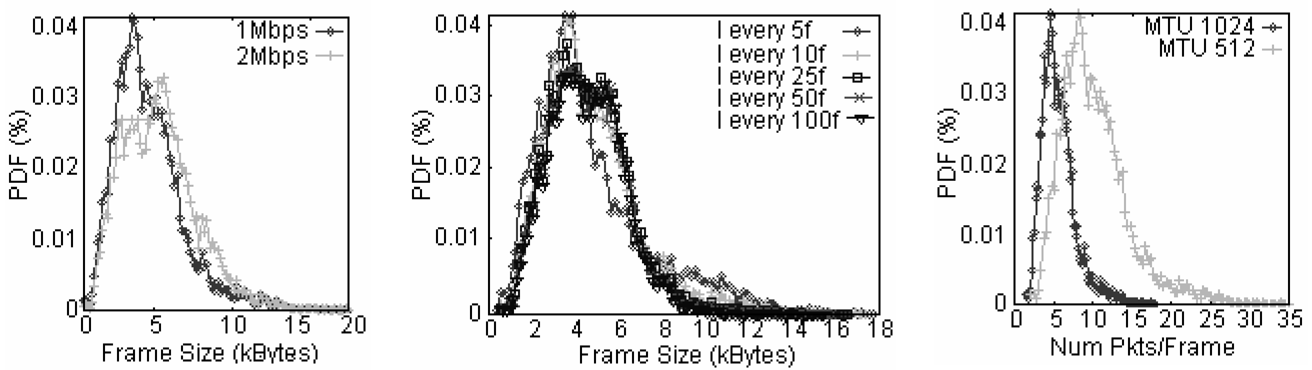
Hint track settings are required for streaming MP4 and .3gp multimedia files. However, given that in general most video-frames are quite large and so at most one video frame can be packetised into a single 1024B packet, hint tracks are especially important for audio streaming since multiple audio samples can be packetised into one packet.

2.2. Video Analysis

In the experiments reported here, the video content was encoded using the commercially available X4Live MPEG-4 encoder from Dicas. This video content, JR, is a 5 minute extract from the film 'Jurassic Park' with a CIF display size. Table 1 shows the encoding configuration for this content type encoded as MPEG-4 SP in 7 different ways by adjusting the I-frame frequency (from 1 I-frame every 5 frames to 1 I-frame every 100 frames) and adjusting the target CBR bit rate (from 1Mbps to 2Mbps) using 2-pass encoding. Although a target bit rate is specified, it is not always possible for an encoder to achieve this rate. Columns 5-7 of Table 1 show the peak to mean ratio overall frames, I-frames and P-frames respectively. Figures 1(a) and (b) show the probability distribution function (PDF) of the frames sizes for each of the encoding configurations. Figure 1(c) shows the PDF of the number of packets required to send each frame for encoding configuration, JR1 with hint track MTU 512B and 1024B.

Table 1: JR Content Type at Different Resolutions

Clip	Bit Rate	I-Freq	Peak (B)	F	I	P
JR1	1Mbps	10	17299	3.57	1.92	3.02
JR2	1.5Mbps	10	17299	3.15	1.92	2.60
JR3	2Mbps	10	17299	3.15	1.92	2.60
JR4	1Mbps	5	17635	3.59	1.98	3.15
JR5	1Mbps	25	16403	3.47	1.81	2.92
JR6	1Mbps	50	15715	3.36	1.75	2.91
JR7	1Mbps	100	15363	3.30	1.70	2.89



(a) (b) (c)

Figure 1(a): PDF of frame sizes for JR encoded at VBR 1Mbps and 2Mbps (b) PDF of frame sizes with varying I-frame frequencies (c) PDF of number of packets required to send each frame with different hint track MTU settings

3. Experimental Test Bed

To evaluate unicast video streaming a video server was set up on the wired network and streamed to wireless clients via the Access Point (Figure 2) under lightly loaded conditions where there are no other wireless stations contending for access to the medium. Under these conditions, it is possible to isolate and study the resource requirements of a high quality video streaming session. There are two open-source streaming servers available, Helix from Real [6] and Darwin Streaming Server (DSS) from Apple [7-8]. There have been several papers that have evaluated the performance of the Helix streaming system [9]. In this paper, we have chosen DSS to be the streaming server for our experiments. Although, our future work will investigate the behavioural and performance-related differences between streaming servers with differing adaptation algorithms. DSS is an open-source, standards-based streaming server that is compliant to MPEG-4 standard profiles, ISMA streaming standards and all IETF protocols. The DSS streaming server system is a client-server architecture where both client and server consist of the RTP/UDP/IP stack with RTCP/UDP/IP to relay feedback messages between the client and server. The client can be any QuickTime Player or any player that is capable of playing out ISMA compliant MPEG-4 or .3pg content. The client connects to the server via RTSP to establish a unicast video streaming session. In the experiments here, the client used a 3 second pre-buffering delay. This buffering delay minimized any the effects of any quality degradation due to delay and/or loss. This was necessary to ensure that the server did not use any quality or transmission rate adaptation as a result of RTCP feedback messages from the client.

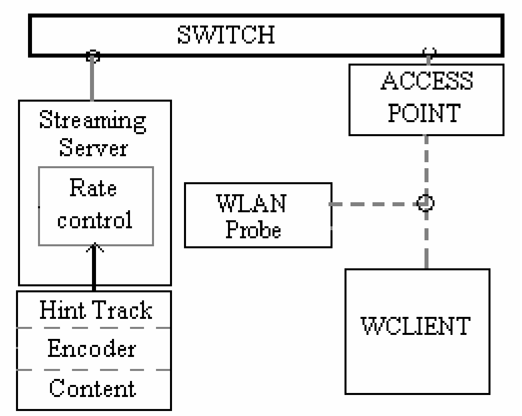
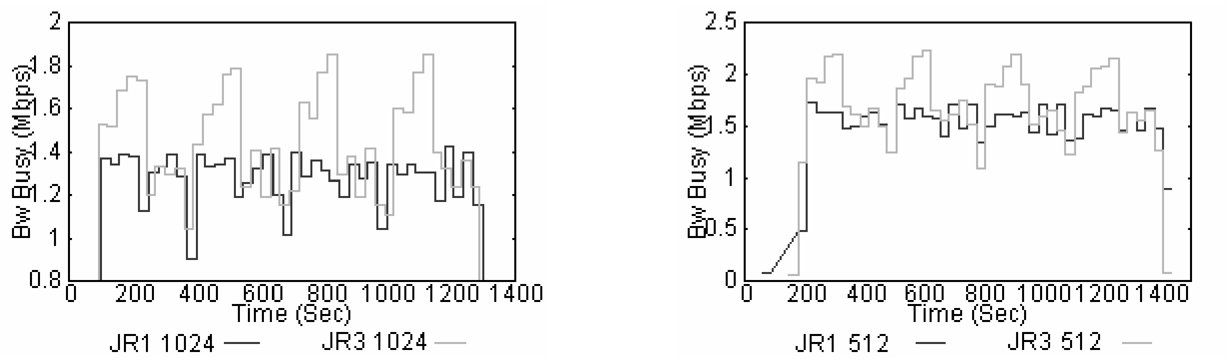


Figure 2: Experimental test bed

At the wireless side, a WLAN resource monitoring application reported in [10] was used to measure the resource utilisation of the video streams. This application non-intrusively monitors and records the busy and idle intervals on the wireless medium and by analysing the temporal characteristics of these intervals infers the resource usage on a per-STA basis. The WLAN resource utilisation is characterised in terms of MAC bandwidth components that are derived from the line rate of the WLAN, i.e.11Mbps. Specifically, three MAC bandwidth components are defined: A load bandwidth (BW_{LOAD}) associated with the transport of the traffic stream and is related to the throughput, an access bandwidth requirement (BW_{ACCESS}) that represents the "cost" of accessing the wireless medium, and a free bandwidth (BW_{FREE}) that gives a measure of the likely QoS. An access efficiency may be defined as the ratio of the BW_{LOAD} to the BW_{ACCESS} and gives an indication of how efficiently a STA accesses the medium. This technique has been shown to be particularly effective in characterising WLAN resource utilisation in a manner that is both compact and intuitive.



(a) **(b)**

Figure 3(a) Variations in 'Busy Bandwidth', BW_{BUSY} over time for clip JR1 encoded at 1Mbps and JR3 encoded at 2Mbps and hint track MTU 1024B (b) Hint track MTU 512B

3.1. Resource Usage Variations with Hint Track Settings

The test duration was approximately 20min for all tests and WLAN probe measurements were taken every second. Figures 3(a) and (b) show variations in busy bandwidth with time averaged over periods of 30sec for different video files encoded with a target bit rate of 2Mbps and 1Mbps and with ea hint track setting of MTU 1024B and 512B. The video files used in these tests have a duration of 5min and were played in 4 loops over the test duration. These loops can be seen as the periodic repeated patterns in the busy bandwidth measured by the probe. It can be seen that the encoding configurations have a clear impact of the bandwidth variations. As expected, the video encoded at 2Mbps has greater bit rate requirement than that encoded at 1Mbps, however the bit rate variations of the 2Mbps file are much greater than that of the 1Mbps file.

In addition, it can also be seen that the hint track setting has an impact on the busy bandwidth. The smaller the MTU packet size set in the hint track, the greater the number of packets required to send each frame, resulting in not only a greater packet header overhead, but also a greater bandwidth access requirement. The busy bandwidth usage for the different encoding configurations and hint track settings can be seen in Tables 2 and 3. Interestingly, regardless of the encoding configuration used, using a hint track setting of 1024B MTU reduces the busy bandwidth requirement by at least 20%. It can be seen that the access efficiency can be doubled by using the larger hint track setting. Similarly, by using a hint track MTU of 512B, the bandwidth required to access the medium is doubled, since there are approximately twice the amount of packets required to send the same video frame.

Table 2: Overall WLAN Characteristics

Clip	Ratio	MTU 1024	MTU 512
	BW_{BUSY}	BW_{BUSY}(Mbps)	BW_{BUSY}(Mbps)
JR1	0.79	1.19	1.49
JR2	0.84	1.37	1.62
JR3	0.83	1.37	1.65
JR4	0.81	1.21	1.49
JR5	0.80	1.16	1.44
JR6	0.81	1.15	1.41
JR7	0.83	1.15	1.38

Table 3: Mean Resource Usage at the AP

Clip	Hint MTU	Access Efficiency	BW_{ACCESS} (Mbps)	BW_{LOAD} (Mbps)
JR1	1024	2.16	0.55	1.19
JR2	1024	2.16	0.63	1.36
JR3	1024	2.17	0.63	1.37
JR4	1024	2.16	0.56	1.21
JR5	1024	2.15	0.54	1.16
JR6	1024	2.17	0.53	1.15
JR7	1024	2.15	0.53	1.14
JR1	512	1.28	1.16	1.48
JR2	512	1.28	1.27	1.62
JR3	512	1.28	1.29	1.65
JR4	512	1.28	1.16	1.48
JR5	512	1.27	1.13	1.44
JR6	512	1.27	1.11	1.41
JR7	512	1.27	1.08	1.37

3.2. Resource Usage Variations with I-Frame Frequency

As we have seen, the choice of encoding parameters has a serious impact on the bandwidth requirements of the WLAN. In this section, we shall analyze the effect that the I-frame frequency has on the bandwidth requirements. There is an inter-frame dependency between the different frame types, i.e. I-frames are individually "decodable", whilst P-frames are predictively encoded from the pervious I-frame and as such require the pervious I-frame

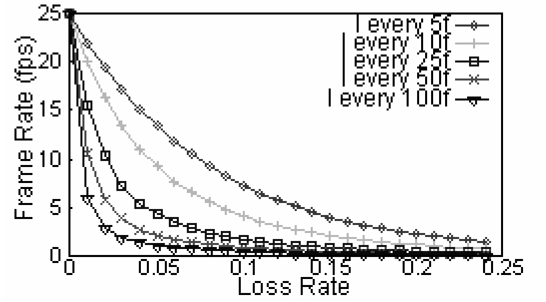


Figure 4. The effect of I-frame frequency on "decodability" in the presence of packet loss

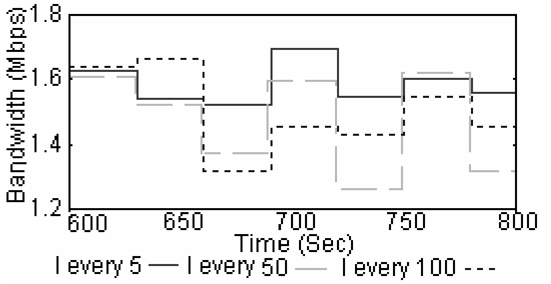


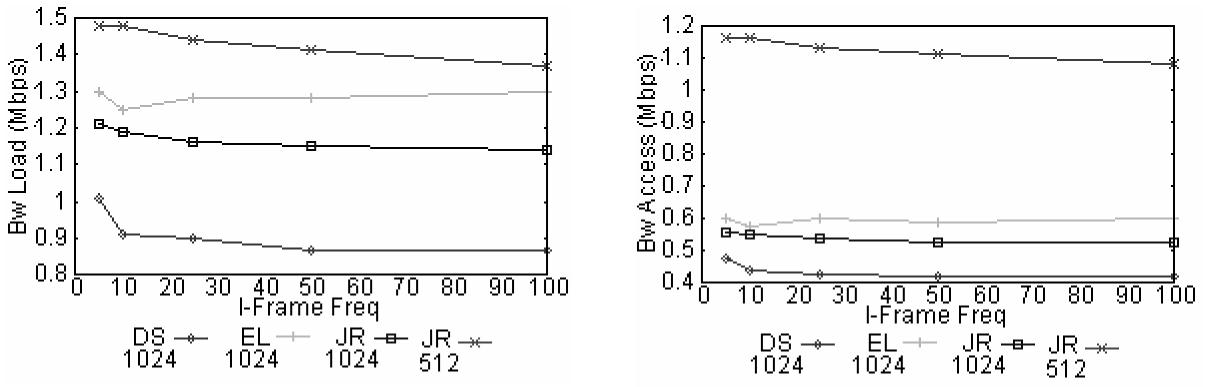
Figure 5 Close-up of variations in 'Busy Bandwidth' over time for clip JR encoded with different I-frame frequencies

to be correctly decoded in order to be correctly decoded itself. Similarly, B-frames are bi-directionally encoded and as such depend on the previous and subsequent I or P-frames. For example, if data is lost pertaining to a P-frame, the lost data will cause an error in the decoded frame and this error will propagate throughout the frames until the next I-frame is received to refresh the frame and the error. Using the analysis described in [11], it can be seen how the I-frame frequency significantly improves the ability of the decoder to play out the received stream in the presence of lost packets when there are no error concealment strategies being employed. The analysis was performed on content JR1 with a hint track setting of MTU 1024B using a mean of 8 packets to send an I-frame and 4 packets required to send a P-frame. From Figure 4, it can be seen that increasing the I-frame frequency significantly improves the "decodability" of the subsequent frames in the presence of loss. However, frequent I-frames can cause large periodic spikes in the bit rate. It is a common misconception that video frames follow a size relationship where I>P>B. If, for example, there is a very low I-frame frequency and there is a high level of scene activity and scene changes and cuts within the content, then much more information is required to encode the P or B-frame, although some encoders support automatic generation of I-frames when a scene cut is detected. However, the frequency of scene cuts is entirely content-dependent. During the analysis of the video files used in the experiments here for example, in video files JR6 and JR7 the largest frame size corresponded to a P-frame. Figure 5 shows variations in the busy bandwidth averaged over 30second intervals over a period of time for the content, JR, encoded at differing I-frame frequencies. It can be seen that having more frequent I-frames requires only a slightly more bandwidth than infrequent I-frames. This is important given the importance of frequent I-frames on the "decodability" of the video in the presence of loss on the network.

Table 4: AP data for different content types

Clip	I-Freq	Access Efficiency	BW_{ACCESS} (Mbps)	BW_{LOAD} (Mbps)
EL1	10	2.15	0.58	1.25
EL2	10	2.19	0.72	1.58
EL3	10	2.18	0.68	1.48
EL4	5	2.17	0.60	1.30
EL5	25	2.13	0.60	1.28
EL6	50	2.17	0.59	1.28
EL7	100	2.17	0.60	1.30
DS1	10	2.07	0.44	0.91
DS2	10	2.11	0.44	0.93
DS3	10	2.12	0.43	0.91
DS4	5	2.10	0.48	1.01
DS5	25	2.09	0.43	0.90
DS6	50	2.02	0.42	0.87
DS7	100	2.07	0.42	0.87

Figures 6(a) and 6(b) show how the access bandwidth and load bandwidth are only slightly reduced with I-frame frequency, indicating there is very little overall bandwidth gain by reducing the I-frame frequency. However, more importantly, this does affect the playout of the received video if there are packets lost in the network. To validate this result, two different video content types were tested. The video sequence DS corresponds to a 5 minute extract from the movie 'Don't Say a Word' and EL corresponds to a 5 minutes extract from the animated movie 'The Road to Eldorado'. Animated videos are very challenging for encoders since animations generally consist of line art and as such have greater spatial complexity and detail. Both of these video sequences were encoded using exactly the same encoding methodology and configurations as JR. It can be seen, that in all cases, regardless of the content type, there is very little bandwidth gain by reducing the I-frame frequency. The different content types have different mean bandwidth requirements since despite being encoded using the same configurations, different content types result in varying levels of encoding efficiency which in turn affects the encoded bit rate and bit rate variations. The summarised results of these tests for the content types, EL and DS are shown in Table 4.



(a) (b)

Figure 6(a) Load bandwidth with I-frame Frequency for several content types (b) Load bandwidth with I-frame Frequency for several content types

4. Conclusions

In this paper, we have demonstrated the effect the encoding configuration has on the resource requirements of WLAN networks. In particular, we have shown how hint track settings impact on the access bandwidth requirement. The results demonstrate that by using a hint track setting of 1024B rather than 512B, the access efficiency is doubled and the required access bandwidth is reduced by 20%. This indicates that ideally video packets should be as large as possible. However the trade-off is that if packets are lost, more bandwidth is required to retransmit the lost packets and it makes the task of error concealment at the receiver more difficult since more data is lost. We have discussed the importance of frequent I-frames for decoding the video stream in the presence of loss. However, when there is a high level of scene activity, there will be large frequent spikes in the encoded bit rate regardless of the encoding frame type. We have shown that the bandwidth requirements are only slightly reduced with I-frame frequency, demonstrating that there is very little advantage in reducing the I-frame frequency. We have seen that the animated content types have different requirements to other content types. The contents characteristics affect the encoding efficiency and resulting encoded bit rate variations. Future work is planned to investigate the impact different content types have on resource requirements despite having the same encoding configuration. In addition, the application-level QoS and user-perceived quality will be assessed.

References

[1] C. Krasic, J. Walpole, W-C. Feng, 2003, "Quality Adaptive Media Streaming by Priority Drop", Proc. NOSSDAV'03, Monterey, CA, 1-4 June 2003.

[2] G. Conklin, G. Greenbaum, et al., "Video Coding for Streaming Media delivery on the Internet", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, No. 3, March 2001

[3] R. Koenen, MPEG-4 WG11, “MPEG-4 Overview” Version 21, ISO-IEC JTC1/SC29/WG 11: N4668 Coding of Moving Pictures and Audio, MPEG

[4] Apple Computer, “QuickTime File Format”, http://developer.apple.com/techpubs/quicktime/qtdevdocs/PDF/QTFileFormat.pdf

[5] “Hint Track Format”, http://developer.apple.com/documentation/QuickTime/REF/Streaming.29.htm#pgfId=19901

[6] Helix Streaming Server, https://www.helixcommunity.org/

[7] “Darwin Streaming Server”, http://developer.apple.com/darwin/projects/streaming/

[8] “QuickTime Streaming”, http://developer.apple.com/documentation/QuickTime/RM/Streaming/StreamingClient/StreamingClient.pdf

[9] T. Kuang, C Williamson, “RealMedia Streaming Performance on an IEEE 802.11b Wireless LAN”, Proc. Of IASTED Wireless and Optical Communications (WOC) Conference, Banff, AB, Canada, pp.306-311, July 2001

[10] M. Davis, “A Wireless Traffic Probe for Radio Resource Management and QoS Provisioning in IEEE 802.11 WLANs”, ACM MSWiM Conference 2004 in Venice, Italy, October, 2004

[11] N. Feamster, H. Balakrishnan, “Packet loss recovery for Streaming Video”, Proc. of 12th International Packet Video Workshop, April 2002