Optimal settings of maximum transfer unit (MTU) for efficient wireless video communications

C.K. Kodikara, S.T. Worrall and A.M. Kondoz

Abstract: Following its success over wired networks, the deployment of the Internet Protocol (IP) over wireless networks is an increasingly popular topic. Protocol optimisations implemented at this level only consider the traditional Internet applications. Path maximum transfer unit (MTU) discovery algorithms are designed to avoid packet fragmentation along the transmission path and hence improve protocol efficiency. However, the design criteria do not take into account the potentially significant influences on the performance over the wireless links. A theoretical analysis is presented of the effect of MTU size upon the performance of wireless video. The theoretical calculation is based on a distortion model, which accurately estimates the video quality in terms of average frame peak signal-to-noise ratio (PSNR). The theoretical analysis are validated for MPEG-4 coded video transmission over a simulated UMTS network.

1 Introduction

Owing to high operational flexibility and system efficiency, packet-switched communication has become a key technology in third-generation wireless communication systems. In addition to capacity enhancement, the use of packet switching over the radio interface provides a combined transport mechanism for heterogeneous services such as video, voice and data.

Extensive research activity has been presented in the area of optimisation of IP and its underlying network protocols (e.g. TCP, UDP). IP allows network nodes to fragment packets that are too large to be forwarded. Such fragmentation not only results in poor network performance but can also lead to total communication failure in some circumstances [1]. However, datagrams that are much smaller than the allowed MTU size waste network resources and result in suboptimal system throughput. IP path MTU discovery [2] is used to obtain the optimal maximum packet size for a given connection while avoiding datagram fragmentation along the transmission path. Design criteria mainly follow the response time requirement, where the interactive response time is limited to 100–200 ms [3].

Third-generation wireless systems, such as CDMA2000 and UMTS, use IP as the transport mechanism for packet switched based communication [4]. Much research has been conducted into modifications of IP and the underlying protocols, such as RTP/RTCP and UDP, for wireless applications [5–8]. Most of this research focuses on the reliable transmission of IP over error-prone wireless networks. However, little attention is shown regarding the higher loss probability of the wireless link on the design and optimisation of IP path MTU discovery algorithms [9].

IEE Proceedings online no. 20045067

doi:10.1049/ip-com:20045067

Paper first received 27th May 2004 and in revised form 23rd February 2005

The authors are with the Centre for Communications Systems Research (CCSR), University of Surrey, Guildford GU2 7XH, UK

E-mail: c.kodikara@eim.surrey.ac.uk

This paper models the performance of fully error resilience enabled MPEG-4 video based on end-to-end distortion estimation. Quantisation distortion, concealment distortion (temporal and spatial) and distortion caused by error propagation over predictive frames are considered in the distortion calculation. Using the developed model, the effect of packet size on the performance of wireless video applications is investigated and the optimal datagram size for given channel conditions is derived.

2 Video coding algorithm

Although, motion prediction in video codes generally achieves high compression efficiency, it is more sensitive to channel errors as it promotes error propagation. Carefully designed error resilience techniques have been incorporated into the MPEG-4 compression algorithm to make the video codec more resilient to channel degradations [10]. The main error resilience techniques include video packet resynchronisation code words, data partitioning, reversible variable length codes and header extension codes.

The data format used in MPEG-4 is shown in Fig. 1. Each frame (video object plane, VOP) is split into video packets (VPs), which we formed by grouping an integer number of coded consecutive macro blocks (MBs) (16×16 pixels blocks). Synchronisation markers are used to isolate VPs from each other, hence increasing the amount of recovered information in the presence of bit stream corruption. The minimum video packet size is fixed by the encoder settings. Following the concept of data partitioning, data within a video packet are further divided into two main parts. The motion-related information for all MBs contained in a given VP is placed in the first partition and the relevant discrete cosine transform (DCT) data are placed in the second partition.

The most important information that the decoder needs to decode the compressed video data is placed in the VOP header. A header extension code (HEC) reduces the sensitivity of header information to channel errors by repeating important header information in selected video packets. The repetition is controlled by toggling the one-bit HEC field in each VP header.

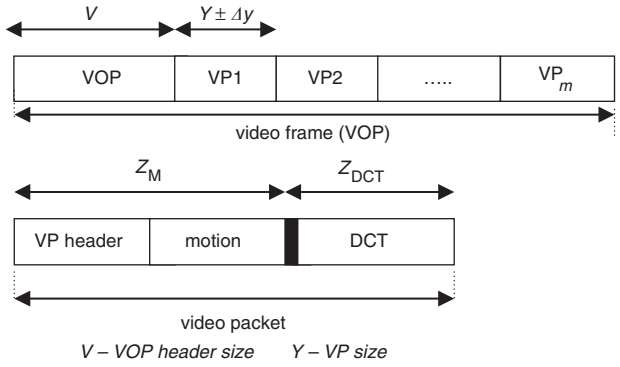

Fig. 1 *MPEG-4 data format*

Video packetisation should be designed so as to enhance the error resiliency resulting from the error resilience tools. Following the MPEG-4 video packetisation rules listed in RFC 3016 [6], information data belonging to different VOPs are segmented into different IP packets. One video frame is split into K IP packets. The VOP header data are repeated N times using the HEC. Assume $N <= K$ and no more than one header extension code is present in an IP packet. The number of video packets (M) encapsulated into one IP packet is calculated in such a way that the resulting transport packet is not larger than the selected MTU size, L (see Fig. 2).

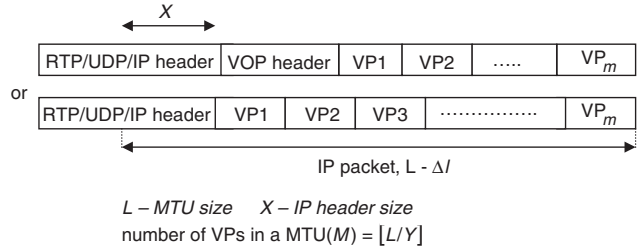

Fig. 2 *IP packetisation of MPEG-4*

3 Video performance

Video performance is often shown as a combination of quantisation distortion and channel distortion [11–13]. Quantisation distortion is mainly a function of video source rate and scene content and is unrecoverable at the decoder. There are three types of channel distortion; spatial concealment, temporal concealment and error propagation.

If the video frame configuration information or video packet header data are lost during transmission, the decoding process is impossible and the data belonging to that frame or video packet has to be discarded at the decoder. However, the error concealment tools implemented at the decoder replace this discarded data with error-concealed data from the neighbouring packets. The distortion resulting from this process is called spatial-concealment distortion as only spatial error concealment is involved in the process. Spatial-concealment distortion can further be divided into two parts considering the presence or absence of network layer header corruption in packet video networks. In the presence of network layer header corruption, the information data encapsulated in that packet should be discarded. This causes the loss of a large amount of data, hence accurate concealment is difficult, and the result is severe spatial-concealment distortion. If spatial-concealment is caused by the corruption of the video packet header or motion information in each individual packet, then concealment is relatively easy and reasonable concealment may be possible depending on the correct reception of neighbouring packets.

On the other hand, if the header and motion information is received correctly but DCT information is corrupted, then the decoder only discards the corrupted DCT data and replaces them with the corresponding concealed data from the previous frame. The distortion resulting from motion-compensated error concealment is called temporal-concealment distortion.

Errors can propagate in two ways, either in the temporal domain or in the spatial domain. Frame-to-frame error propagation through motion prediction and temporal concealment is called temporal domain error propagation. The propagation of errors from neighbouring video packets via spatial concealment is referred to as spatial domain error propagation.

Taking a video packet as the base unit, the expected average frame quality in terms of peak signal-to-noise ratio (PSNR) can be written as

$$\begin{aligned} \tilde{E}(Q_f) &= 10 \cdot \log\left(g/\tilde{E}(D_f)\right) \\ \tilde{E}(D_f) &= I \cdot \tilde{E}(D_{pv}) \end{aligned} \tag{1}$$

where $\tilde{E}(Q_f)$ and $\tilde{E}(D_f)$ are the expected average frame quality and the expected average frame distortion and $\tilde{E}(D_{pv})$ denotes the expected average distortion of a video packet. The average number of video packets in a frame is denoted by I, while g is a constant defined by the dimensions of the frame. Now, $\tilde{E}(D_{pv})$ can be written as

$$\begin{aligned} \tilde{E}(D_{pv}) = {} & \tilde{E}(D_{Q,pv}) + \tilde{\rho}_{u,pv}\tilde{E}(D_{s\_con,pv}) \\ & + \tilde{\rho}_{d,pv}\tilde{E}(D_{t\_con,pv}) + \tilde{\rho}_{um,pv}\tilde{E}(D_{sm\_con.pv}) \\ & + \tilde{f}_{tp} + \tilde{f}_{sp} \end{aligned} \tag{2}$$

$\tilde{E}(D_{Q,pv})$, $\tilde{E}(D_{s\_con,pv})$, $\tilde{E}(D_{sm\_con,pv})$ and $\tilde{E}(D_{t\_con,pv})$ represent expected average quantisation distortion, spatial-concealment distortion due to video packet corruption, spatial-concealment distortion due to IP packet corruption and temporal-concealment distortion, respectively. The probability of receiving an undecodable video packet is denoted by $\tilde{\rho}_{u,pv}$. This includes the probability of corruption of a VOP header, a VP header or the motion data. The probability of receiving a decodable video packet, but with errors, where the DCT data is corrupted but other information is received correctly, is denoted by $\tilde{\rho}_{d,pv}$. The probability of network layer packet header corruption is given by $\tilde{\rho}_{um,pv}$, while $\tilde{f}_{tp}$ and $\tilde{f}_{sp}$ represent the distortion caused by the propagation of errors in the temporal domain and spatial domain, respectively.

4 Problem formulation

To achieve better video quality in bandwidth-constrained error-prone packet networks, it is necessary to maximise the source bit rate, and at the same time minimise the effects of packet loss. A larger transport packet reduces the packet overhead resulting in higher source throughput. However, accommodating a large amount of information in one packet results in a large amount of information loss at one time if the packet is lost due to packet header corruption. Therefore, the optimal packet size is a function of packet overhead and packet loss ratios and it greatly depends on the operating environment.

The goal is to find the optimal IP packet size (MTU size), which maximises received video quality given a particular

mobile channel. The channel is characterised by the probability of channel bit errors and the channel bandwidth. The optimisation problem can formally be written as

$$\underset{\{MTU\ size\}}{\text{minimise}}\ \tilde{E}(D_f) = I \cdot \tilde{E}(D_{pv}) \tag{3}$$

$$\text{subject to } OH_{MTU} + R_{source} = R_{channel}$$

where OH_{MTU}, R_{source} and $R_{channel}$ denote overhead rate due to IP packetisation, source rate and channel bit rate, respectively.

4.1 Probability calculation

If the probability of receiving an IP header with errors is $\tilde{\rho}_{IP}$, the probability of receiving a VOP header with errors is $\tilde{\rho}_{VOP}$, the probability of receiving a video packet header and motion information with errors is $\tilde{\rho}_M$ and the probability of finding an error in the DCT part is $\tilde{\chi}$, then

$$\tilde{\rho}_{u,pv} = 1 - (1 - \tilde{\rho}_{IP}).(1 - \tilde{\rho}_{VOP}).(1 - \tilde{\rho}_M) \tag{4}$$

$$\tilde{\rho}_{d,pv} = (1 - \tilde{\rho}_{IP}).(1 - \tilde{\rho}_{VOP}).(1 - \tilde{\rho}_M).\tilde{\chi} \tag{5}$$

$$\tilde{\rho}_{um,pv} = \tilde{\rho}_{IP} \tag{6}$$

For a given probability of channel bit error ρ_b, it can be shown that

$$\tilde{\rho}_{IP} = \sum_{i=1}^{X} (1 - \rho_b)^{i-1}\rho_b = 1 - (1 - \rho_b)^X \tag{7}$$

where X is the IP header size. Similarly

$$\tilde{\rho}_{VOP} = (1 - (1 - \rho_b)^V)^N \tag{8}$$

The number of times the VOP header is repeated in the video packet is denoted by N. The size of the VOP header is given by V.

$$\tilde{\rho}_M = 1 - (1 - \rho_b)^{Z_M} \tag{9}$$

$$\tilde{\chi} = \sum_{i=1}^{Z_{DCT}} \binom{Z_{DCT}}{i} (1 - \rho_b)^{Z_{DCT}-i}\rho_b^i = 1 - (1 - \rho_b)^{Z_{DCT}} \tag{10}$$

where Z_M and Z_{DCT} are the average size of the first and second partitions of the video packets. For fixed IP header sizes (assume no IP header compression) $\tilde{\rho}_{IP}$ is clearly independent of the MTU size. Similarly $\tilde{\rho}_{VOP}$ only depends on the size of the VOP header part and the occurrence of a header extension within the video packet. $\tilde{\rho}_M$ and $\tilde{\chi}$ are slightly affected by the MTU size as it might change the encoder quantisation step size due to the change of source bit rate, hence Z_{DCT} and Z_M. However, for fixed video packet sizes (as used in the experiments discussed in this paper), the probability terms $\tilde{\rho}_M$ and $\tilde{\chi}$ can be considered to be independent of MTU size.

4.2 Distortion modelling

4.2.1 Quantisation distortion: Let $\varphi_{k,i,j}(u, v)$ and $\tilde{\varphi}_{k,i,j}(u, v)$ be the original and reconstructed luminance values, respectively, of the (u, v) pixel of the kth macro block (MB) of the ith video packet in the jth video frame. Video MB reconstruction follows the inverse of the encoding process, as shown in Fig. 3. The quantisation distortion of a video packet can now be defined as

$$E(D_{Q,pv}^{i,j}) = \sum_{i=1}^{\eta}\sum_{v=1}^{16}\sum_{u=1}^{16} (\varphi_{k,i,j}(u, v) - \tilde{\varphi}_{k,i,j}(u, v))^2 \tag{11}$$

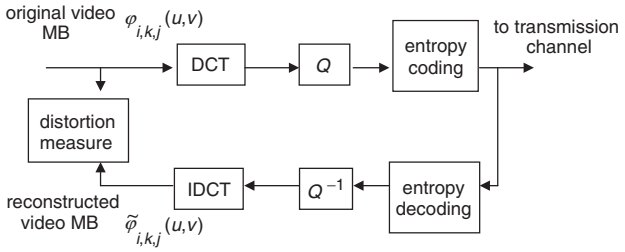

Fig. 3 *Quantisation distortion calculation*

where η is the total number of MBs in the ith video packet. Then, the average quantisation distortion is given by

$$\tilde{E}(D_{Q,pv}) = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{I^j}\sum_{i=1}^{I^j} E\left(D_{Q,pv}^{i,j}\right) \tag{12}$$

where J and I^j are the total number of video frame and the total number of video packets in the jth frame, respectively.

Experimentally evaluated quantisation distortion values for two specified channel bit rates are shown in Fig. 4. $\tilde{E}(D_{Q,pv})$ decreases with the increase of MTU size. This is due to the IP packet overhead, which increases with decreasing MTU size. The overhead reduces the available throughput for source data in the bandwidth-limited channel and, hence, increases the quantisation distortion.

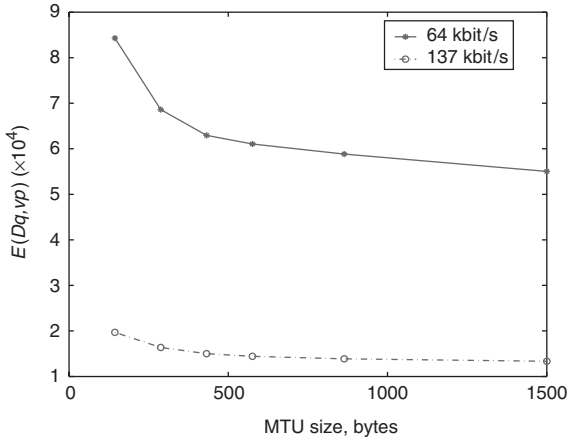

Fig. 4 *Variation of* $\tilde{E}(D_{Q,pv})$

4.2.2 Concealment distortion: Concealment distortions are computed in a similar manner to quantisaton distortion. As shown in Fig. 5, the transmitted video data belonging to each MB is corrupted using the noise generator located at the encoder. Corrupted data is replaced by the concealed data, and data belonging to the original

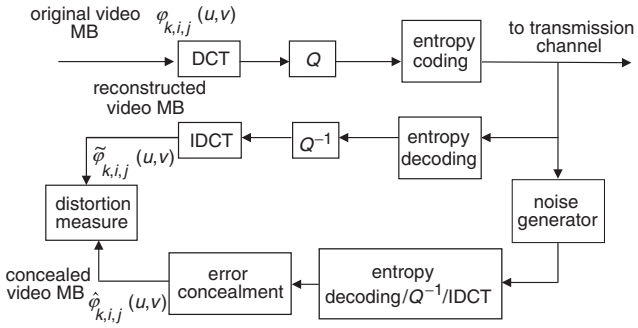

Fig. 5 *Concealment distortion calculation*

and concealed MBs is compared. Thus, concealment distortion of the video packet can be written as

$$E\left(D^{i,j}_{con,\,pv}\right) = \sum_{i=1}^{\eta}\sum_{v=1}^{16}\sum_{u=1}^{16}\left(\tilde{\varphi}_{k,i,j}(u,\,v) - \hat{\varphi}_{k,i,j}(u,\,v)\right)^2 \tag{13}$$

and the average concealment distortion is given by

$$\tilde{E}\left(D_{con,\,pv}\right) = \frac{1}{J}\sum_{j=1}^{J}\frac{1}{I^j}\sum_{i=1}^{I^j}E\left(D^{i,j}_{con,\,pv}\right) \tag{14}$$

where $\hat{\varphi}_{i,k,j}(u,\,v)$ indicates the luminance values of the $(u,\,v)$ pixel of the kth MB after the concealment.

In the calculation of $\tilde{E}(D_{s\_con,\,pv})$, only the spatial-concealment algorithms are used to generate the concealed data. However, only the temporal concealment algorithm is used to conceal the corrupted data for the $\tilde{E}(D_{t\_con,\,pv})$ calculation. The calculations of $\tilde{E}(D_{s\_con,\,pv})$ and $\tilde{E}(D_{t\_con,\,pv})$ assume the correct reception of neighbouring video packets, while the calculation of $\tilde{E}(D_{sm\_con,\,pv})$ relies on the correct reception of neighbouring MTU packets.

Figure 6 illustrates the calculation of concealment distortion graphically. For example, to calculate $\tilde{E}(D_{s\_con,\,pv})$, data belonging to each MB in a video frame is corrupted using the noise generator. A spatial-concealment algorithm is used to conceal each corrupted MB if the data belonging to surrounding MBs is received correctly. This provides the maximum concealment distortion of a MB, if it is corrupted due to the channel errors. This process is repeated for all the MBs in the video frame. Note that the total spatial-concealment distortion due to the channel errors is $\tilde{\rho}_{u,\,pv}\tilde{E}(D_{s\_con,\,pv})$ as given in (2). $\tilde{\rho}_{u,\,pv}$ is the probability of receiving an undecodable video packet. Thus, $\tilde{\rho}_{u,\,pv}$ captures the effect of bit errors on video stream in a practical environment. Moreover, the distortion due to the error propagation (temporal and spatial) is calculated separately as describe in Section 4.3. In this way, $\tilde{E}(D_{t\_con,\,pv})$, $\tilde{E}(D_{s\_con,\,pv})$ and $\tilde{E}(D_{sm\_con,\,pv})$ are calculated for different MTU sizes and are plotted in Figs. 7–9.

Figures 7–9 illustrate the effect of MTU size upon the expected distortion types; $\tilde{E}(D_{t\_con,\,pv})$, $\tilde{E}(D_{s\_con,\,pv})$ and $\tilde{E}(D_{sm\_con,\,pv})$. $\tilde{E}(D_{s\_con,\,pv})$ and $\tilde{E}(D_{t\_con,\,pv})$ show more or less constant values with varying MTU sizes. This is because both terms assume the correct reception of the IP header in the calculation. Hence, the introduction of IP packetisation

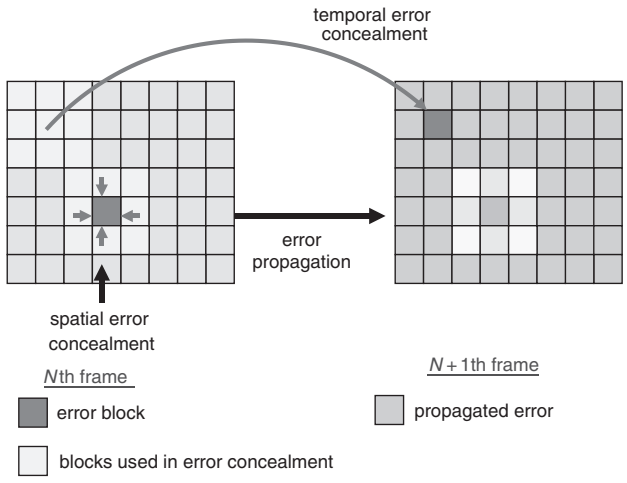

Fig. 6 *Graphical illustration of concealment-distortion calculation*

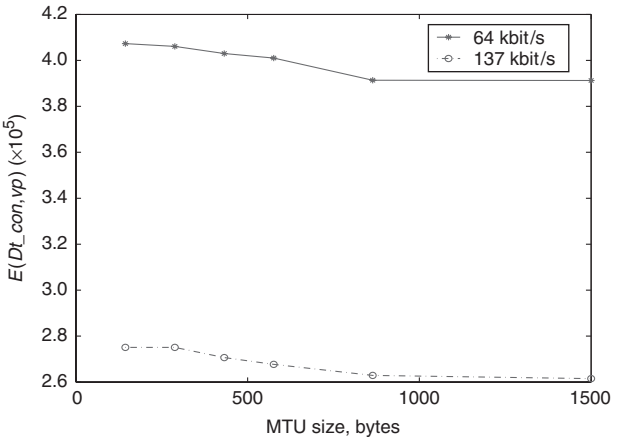

Fig. 7 *Variation of* $\tilde{E}(D_{t\_con,\,pv})$

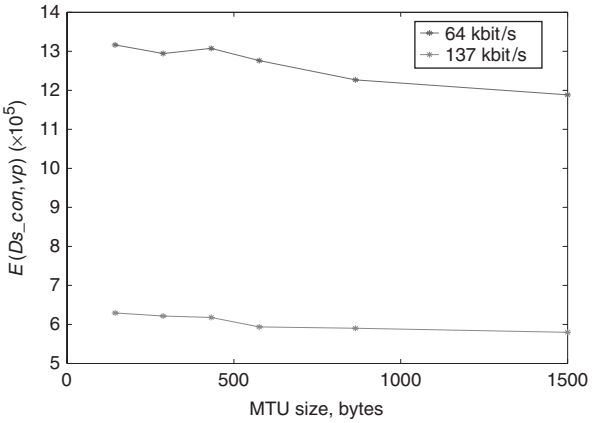

Fig. 8 *Variation of* $\tilde{E}(D_{s\_con,\,pv})$

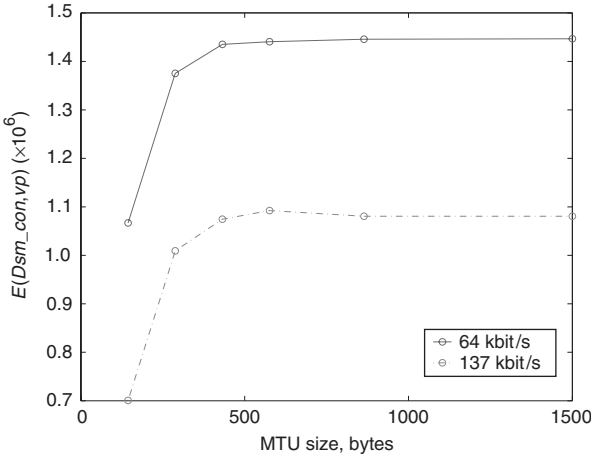

Fig. 9 *Variation of* $\tilde{E}(D_{sm\_con,\,pv})$

does not affect the concealment performance. However, $\tilde{E}(D_{sm\_con,\,pv})$, where the corruption of IP headers is taken into account, is greatly affected by the IP packetisation. Figure 8 shows a dramatic increase in $\tilde{E}(D_{sm\_con,\,pv})$ with increasing MTU size for relatively low MTU sizes. However, it reaches an asymptotical maximum value when MTU size gets closer to the VOP size. As VOPs are segmented into separate IP packets, a significant number of IP packets are much smaller than the specified MTU size. This reduces the effectiveness of MTU size upon $\tilde{E}(D_{sm\_con,\,pv})$ at large MTU sizes.

4.3 Propagation loss modelling

The expected distortion calculation assumes error-free neighbouring packets and reference video frames. The correlation between video packets in the same frame and adjacent video frames is quantified by the spatial and temporal error propagation terms in (2).

The propagation of corrupted information through predictive coding is represented by $\tilde{f}_{tp}$. An adaptive intra refresh algorithm [14], which uses a selected number of intra-coded MBs in a video frame, is used to prevent the error propagation. Therefore, $\tilde{f}_{tp}$ is clearly a function of intra refresh frequency and is calculated as

$$\tilde{f}_{tp} = P_{tp} \cdot \left[\tilde{\rho}_{u,\,pv}\tilde{E}(D_{s\_con,\,pv}) + \tilde{\rho}_{d,\,pv}\tilde{E}(D_{t\_con,\,pv}) + \tilde{\rho}_{um,\,pv}\tilde{E}(D_{sm\_con.pv}) + \tilde{f}_{sp}\right] \quad (15)$$

where

$$P_{tp} = f_{refresh} \cdot \left(1 - (1 - \rho_b)^F\right) \quad (16)$$

and $f_{refresh}$ is the ratio between the number of intra-coded MBs and the total number of MBs in a video frame. F is the average frame size.

Similarly, the average spatial-error propagation $\tilde{f}_{sp}$ depends on the number of corrupted video packets in a frame and only propagates through the spatial-concealment process. $\tilde{f}_{sp}$ is modelled as

$$\tilde{f}_{sp} = P_{sp} \cdot \left[\tilde{\rho}_{u,\,pv}\tilde{E}(D_{s\_con,\,pv}) + \tilde{\rho}_{um,\,pv}\tilde{E}(D_{sm\_con.pv})\right] \quad (17)$$

where

$$P_{sp} = N_{vp} \cdot \left(1 - (1 - \rho_b)^{\vartheta_{vp}}\right) \quad (18)$$

and $N_{vp}$and ϑ_{vp} represent the average number of video packets in a frame and the average size of a video packet, respectively.

5 Experimental results

Transmission of the standard 'Suzie' sequence over a simulated UMTS channel is considered in the experiments. A UMTS physical link layer simulator is developed using the Signal Processing WorkSystem (SPW) software simulation tools developed by Cadence Design System Inc. The simulator includes all the radio configurations, channel structures, channel coding/decoding, spreading/despreading, modulation parameters, transmission modelling and their corresponding data rates according to the UMTS specifications. The transmitted signal is subjected to a multipath fast fading environment, where the power-delay profiles are specified in [15]. The multipath-induced intersymbol interference is implicit in the developed chip-level simulator. A detailed discussion of the channel simulations can be found in [16]. Using the developed simulator, error characteristics of the transmission channel are simulated for a range of channel condition and for different physical layer configurations.

The Vehicular A propagation condition and down link transmission is considered in the experiment discussed in this paper. The mobile speed is set to 50 km/h. Spreading factors of 32 and 16 are used in the physical channel configuration. Source data is protected with the use of 1/3 rate convolutional code at the physical layer. The experimentally evaluated channel block error rates (BLER) over the Vehicular A environment are listed in Table 1.

MPEG-4 encoded video is transmitted over the UMTS simulated channel using both circuit-switched and packet-switched connections. The received video quality is measured in terms of average frame PSNR and the results are plotted in Fig. 10. For packet-switched connections, an MTU size of 576 bytes is used. The corresponding video source rate is calculated to be 64 kbit/s and 58.75 kbit/s for circuit-switched and packet-switched radio bearer configurations, respectively. Video performance for the same video sequence over a given channel condition is estimated using the developed distortion estimation model as explained in Sections 3 and 4.

Table 1: Channel BLER for Vehicular A environment

E_b/N_o	SF 32	SF 16	E_b/N_o	SF 32	SF 16
12	0.0000	0.0006	7	0.0460	0.0600
11	0.0003	0.0010	6	0.1371	0.1500
10	0.0010	0.0020	5	0.3173	0.3500
9	0.0030	0.0110	4	0.5335	0.5500
8	0.0130	0.0240	3	0.7828	0.7900

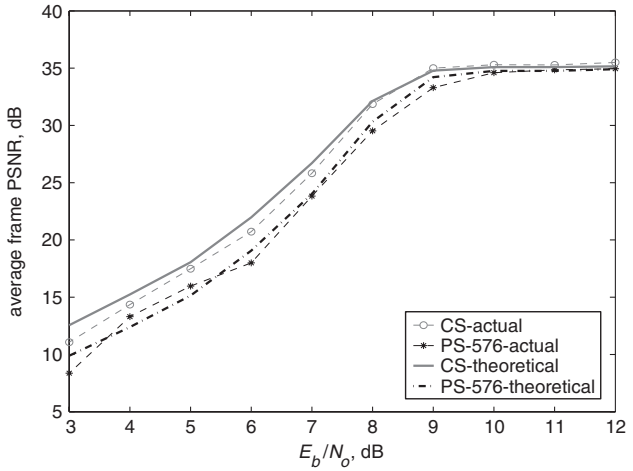

Fig. 10 *Video performance comparison*

Let ρ_{bl} and X_{bl} be the probability of channel block error and the number of data bits contained in a radio block. Then

$$\rho_{bl} = \sum_{i=1}^{X_{bl}} (1 - \rho_b)^{i-1}\rho_b = 1 - (1 - \rho_b)^{X_{bl}} \quad (19)$$

$$\text{and } \rho_b = 1 - \sqrt[X_{bl}]{(1 - \rho_{bl})}$$

Using (19), the probability of channel bit error ρ_b is computed from the experimentally evaluated channel BLER values ρ_b, which are shown in Table 1. This calculation captures the effect of bursty channel errors on the performance. The calculated ρ_b is used to estimate the video performance (1)–(18). The performance calculation for circuit-switched connections uses the same distortion model with $\tilde{\rho}_{IP} = 0$. Figure 10 shows the comparison of the estimated performance and experimentally evaluated video performance over the error-prone channel. The Figure clearly illustrates the close match between the experimental and theoretical results. Furthermore, the Figure shows that the video quality loss resulting from the use of packet-switched connections instead of circuit-switched connections ranges from 1–3 dB.

Using the developed distortion model, the optimal MTU sizes for the transmission of video over the UMTS

Vehicular A propagation environment with varying channel conditions are calculated and the results are shown in Fig. 11. The use of spreading factor 32 and 16 in the radio bearer configuration provides the total available channel bit rates of 64 kbit/s and 137 kbit/s, respectively, [16]. The performance shown in Fig. 11 can be divided into three main regions based on the channel conditions experienced. Small packet sizes show better performance for poor channel conditions, while large packet sizes are preferable with good channel conditions. Packet size variation from small to large is visible with moderate channel conditions.

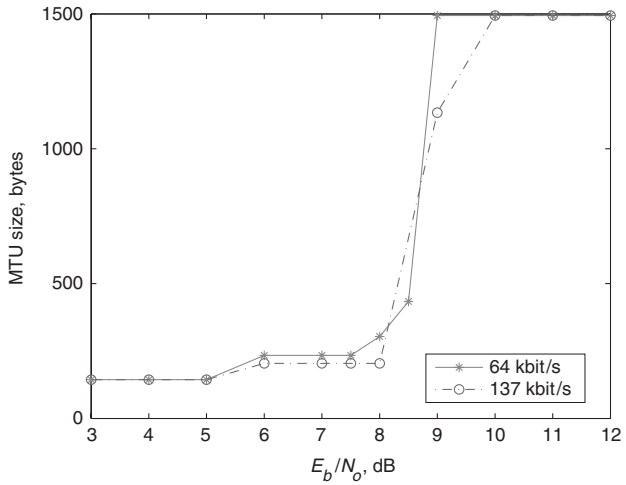

Fig. 11 *Optimal video packet size*

This behaviour can further be explained based on the expected distortion values, which are described in Section 4. Grouping the distortion terms, (2) can be rewritten as

$$\tilde{E}(D_{pv}) = \lambda_1 + \lambda_2 + \lambda_3 \tag{20}$$

where

$$\begin{aligned}\lambda_1 &= \tilde{E}(D_{Q,\,pv}) \\ \lambda_2 &= \tilde{E}(D_{s\_con,\,pv}) \cdot \tilde{\rho}_{u,\,pv}(1 + P_{tp} + P_{sp} + P_{sp} \cdot P_{tp}) \\ &\quad + \tilde{E}(D_{t\_con,\,pv}) \cdot \tilde{\rho}_{d,\,pv}(1 + P_{tp}) \\ \lambda_3 &= \tilde{E}(D_{sm\_con,\,pv}) \cdot \tilde{\rho}_{um,\,pv}(1 + P_{tp} + P_{sp} + P_{sp} \cdot P_{tp})\end{aligned} \tag{21}$$

The calculated distortion factors λ_1, λ_2 and λ_3, for 576 bytes MTU size, and with varying channel conditions are listed in Table 2. The last column in the Table shows the dominating distortion factor (DDF) for a given channel condition.

For channels where E_b/N_o is less than 6 dB, λ_3 dominates the others. λ_3 is proportional to the $\tilde{E}(D_{sm\_con,\,pv})$. As can be seen in Fig. 9, $\tilde{E}(D_{sm\_con,\,pv})$ increases with increasing of MTU size, hence a small MTU size is expected to have a better performance for poor channel conditions (channel E_b/N_o less than 6 dB). For good channel conditions, λ_1, which is the quantisation distortion, significantly dominates λ_2 and λ_3. The quantisation distortion decreases with increasing packet size. Therefore, large packets are preferable for good channel conditions. However, for moderate channel conditions where E_b/N_o varies in the range 6–8 dB, λ_2 becomes significant compared to the others. According to (20), λ_2 is proportional to $\tilde{E}(D_{s\_con,\,pv})$ and $\tilde{E}(D_{t\_con,\,pv})$, which are independent of MTU size. Therefore, video performance is expected to be independent of the packet size for moderate channel conditions. Also, it shows a sharp rise in the optimal MTU size when the channel condition changes from moderate to good. This is due to the nonlinearity associated with the distortion terms. In order to illustrate the nonlinearity of the distortion, the gradient of the theoretical PSNR performance curve for 576 bytes MTU size, with varying channel conditions, is plotted in Fig. 12. As can be seen from the Figure, the gradient value falls dramatically (i.e. the total distortion falls dramatically) at channel conditions of 8–9 dB. This behaviour causes the sharp rise in optimal MTU size shown in Fig. 11.

These theoretical analysis are verified by the experimental results shown in Fig. 13. A performance improvement of 2–3 dB of average frame PSNR can be achieved by using

Table 2: Expected distortion calculation for packet video

E_b/N_o	λ_1	λ_2	λ_3	DDF
12	61044	0	0	λ_1
10	61044	992.6	1091.6	λ_1
9	61044	5197.2	5385.2	λ_1
8	61044	77602.1	56880.5	λ_2
7	61044	346674	341978.1	λ_2
6	61044	1157580	1155251	λ_2
5	61044	2702683	3025136	λ_3
4	61044	4752671	6085720	λ_3
3	61044	7332380	12002987	λ_3

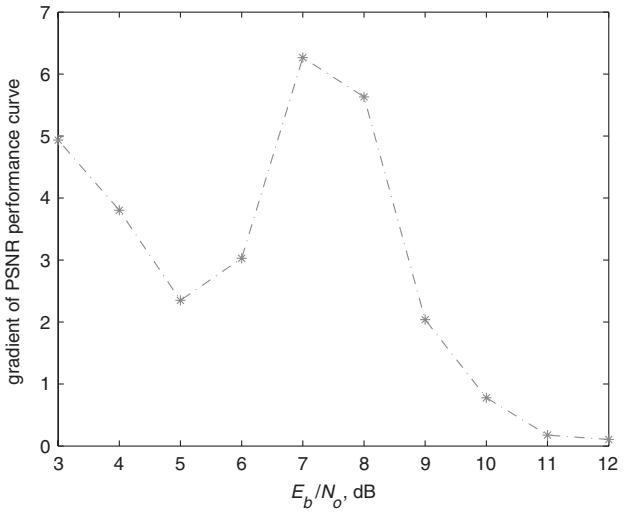

Fig. 12 *Gradient of PSNR performance curve*

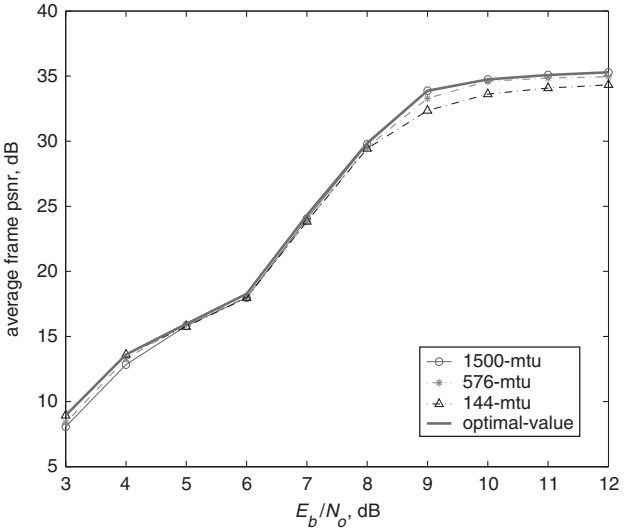

Fig. 13 *Video performance over packet networks*

optimal packetisation for good channels, while an improvement of 1–2 dB is possible for poor channel conditions. The performance is independent of the packet size with moderate channel conditions. However, experimental results show that if the average frame PSNR value is lower than 20 dB, the resulting video output is unacceptable in terms of perceptual quality. Taking 20 dB as a marginal value, it can be concluded that the use of the largest possible MTU size provides the optimal results for video communications over wireless networks.

6 Application to UMTS

The proposed optimal setting of MTU for real-time video transmission over a UMTS network is performed in a number of steps. In the first step, the instantaneous channel condition is predicted based on channel quality measurement or channel information fed back through a feedback channel. The second step involves calculation of expected distortion. The expected distortion calculation can be performed either dynamically or statistically.

In the dynamic distortion calculation, the expected distortion terms $\tilde{E}(D_{Q,pv})$, $\tilde{E}(D_{t\_con,pv})$, $\tilde{E}(D_{s\_con,pv})$ and $\tilde{E}(D_{sm\_con,pv})$ are calculated for a set of MTU sizes for the current video frame at the encoder. Note that different MTU sizes permit different source rates for operation over the same channel bandwidth due to the overhead added by the packet header information.

As described in Section 4, this calculation requires implementation of the decoder at the encoder, including the decoder error concealment algorithms. However, in practice, the decoder performs the inverse operation of the encoder. Therefore, it is possible to optimise the encoder–decoder combination. This reduces the complexity required for the proposed expected distortion calculation. In the statistical case, pre-evaluated performance curves for distortion $\tilde{E}(D_{Q,pv})$, $\tilde{E}(D_{t\_con,pv})$, $\tilde{E}(D_{s\_con,pv})$ and $\tilde{E}(D_{sm\_con,pv})$ are used. After the evaluation of $\tilde{E}(D_{Q,pv})$, $\tilde{E}(D_{t\_con,pv})$, $\tilde{E}(D_{s\_con,pv})$ and $\tilde{E}(D_{sm\_con,pv})$, the total expected distortion is computed using an equation for the predicted channel bit error rate performance. Finally, the optimal MTU size for the current transmission is derived as explained in Section 4.

The algorithm proposed is evaluated for video transmission over a single dedicated channel in real-time conversational video applications. The applicability of the algorithm for video streaming applications [17], is currently being investigated, where a number of video users share the same radio channel. Furthermore, the proposed optimal MTU setting algorithm can be used in conjunction with other video performance enhancement algorithms such as unequal error protection (UEP) [18], unequal power allocation (UPA) [19, 20] or link adaptation algorithms [21, 22]. The integration of the proposed MTU setting algorithms with UEP, UPA and link adaptation is currently under investigation.

7 Conclusions

The effect of MTU sizes on the performance of video transmission over IP-based wireless networks has been analysed. The video performance over packet switched networks degrades in terms of perceptual quality because of the information loss caused by corrupted packet header and the throughput loss due to the packet header overhead. Larger packet size results in less throughput loss. However, a large amount of information data will be lost in the presence of packet header corruption and a smaller packet size increases the overhead rates. In this paper, an optimal packet size estimation method is designed to minimise the distortion seen in received video when operating over error-prone bandwidth limited channels. The developed method is tested for MPEG-4 video transmission over simulated UMTS networks. Experimental results show that a quality improvement of 1–3 dB of average frame PSNR can be achieved with the use of optimal MTU sizes in packet-switched wireless networks.

8 References

1 Kent, C.A., and Mogul, J.C.: 'Fragmentation considered harmful'. Proc. ACM SIGCOMM'87 Workshop on Frontiers in Computer Communications, Aug. 1987, pp. 314–329
2 Mogul, J., Deering, S.: 'RFC-1063: Path MTU discovery' Nov. 1990
3 Jacobson, V.: 'RFC-1144: Compressing TCP/IP headers for low speed serial links' Feb. 1990
4 3GPP TS 26.235: '3rd Generation Partnership Project, Technical Specification Group Services and System Aspects, Packet switched conversational multimedia applications, Default codecs (Release 6)' 3GPP TS 26.235 v 6.0.0 (2003–06)
5 Stockhammer, T., Wiegand, T. *et al.*: 'Video coding & transport layer techniques for H.264/AVC based transmission over packet lossy networks'. Int. Conf. on Image Processing, ICIP 2003, Sept. 2003, Vol. 3, pp. 481–484
6 Kikuchi, Y., and Nomura, T.: 'RFC 3016: RTP payload format for MPEG-4 audio/visual streams' Nov. 2000
7 Worrall, S., Sadka, A., and Kondoz, A.: 'Optimal packetisation of MPEG4 over mobile networks', *IEE Proc., Commun.*, 2001, **148**, (4), pp. 197–201
8 Etoh, M., and Yoshimura, T.: 'Advances in wireless video delivery', *Proc. IEEE*, 2005, **93**, (1), pp. 111–122
9 Hong, B., and Nosratinia, A.: 'Analysis of packet header effects in rate allocation for packet video'. Int. Conf. on Image Processing, ICIP 2002, Sept. 2002, Vol. 2, pp. 177–180
10 Talluri, R.: 'Error-resilient video coding in the ISO MPEG-4 standard', *IEEE Commun. Mag.*, 1998, pp. 112–119
11 Kim, I.M., and Kim, H.M.: 'An optimal power management scheme for wireless video service in CDMA systems', *IEEE Trans. Wirel. Commun.*, 2003, **2**, (1), pp. 81–91
12 Cheng, L., Zhang, W., and Li Chen: 'Rate-distortion optimized unequal loss protection for FGS compressed video', *IEEE Trans. Broadcast.*, 2004, **50**, (2), pp. 126–131
13 Wei, J., Soong, B.H., and Li, Z.G.: 'A new rate-distortion model for video transmission using multiple logarithmic functions', *IEEE Signal Process. Lett.*, 2004, **11**, (8), pp. 694–697
14 ISO/IEC JTC 1/SC 29/WG 11. 'Information technology—generic coding of audio-visual objects—part 2: Visual' ISO/IEC 14496-2 2001, July 2001
15 3GPP TR 101 112, 1998. 'Selection procedures for the choice of radio transmission technologies of the UMTS' v.3.2.0
16 Kodikara, C., Worrall, S., Fabri, S., and Kondoz, A.: 'Performance evaluation of MPEG-4 video telephony over UMTS'. Proc. IEE 3G 2003, UK, June 2003, pp. 73–77
17 Ahmed, T., Mehaoua, A., Boutaba, R., and Iraqi, Y.: 'Adaptive packet video streaming over IP networks: a cross-layer approach', *IEEE J. Sel. Areas Commun.*, 2005, **23**, (2), pp. 385–401
18 Bystrom, M., and Stockhammer, T.: 'Dependent source and channel rate allocation for video transmission', *IEEE Trans. Wirel. Commun.*, 2004, **3**, (1), pp. 258–268
19 Kodikara, C., Worrall, S., and Kondoz, A.M.: 'Energy efficient video telephony over UMTS'. IEEE 59th Vehicular Technology Conf., 2004. VTC 2004-Spring, 17–19 May 2004, Vol. 5, pp. 2719–2723
20 Costa, C.E., Eisenberg, Y., Zhai, F., and Katsaggelos, A.K.: 'Energy efficient wireless transmission of MPEG-4 fine granular scalable video'. IEEE Int. Conf. on Commun., 2004, 20–24 June 2004, Vol. 5, pp. 3096–3100
21 Kodikara, C., Fabri, S.N., and Kondoz, A.M.: 'Link adaptation for real-time video communications in E-GPRS networks', *IEE Proc., Commun.*, 2004, **151**, (5), pp. 438–444
22 Zhang, Q., Zhu, W., and Zhang, Y.-Q.: 'Channel-adaptive resource allocation for scalable video transmission over 3G wireless network', *IEEE Trans. Circuits Syst. Video Technol.*, 2004, **14**, (8), pp. 1049–1063