ANALYSIS OF DISTORTION DUE TO PACKET LOSS IN STREAMING VIDEO TRANSMISSION OVER WIRELESS COMMUNICATION LINKS

Lai U Choi[1], *Michel T. Ivrlač*[2], *Eckehard Steinbach*[1], *and* *Josef A. Nossek*[2]

[1]Institute of Communication Networks
Media Technology Group
Munich University of Technology
email: {laiuchoi, eckehard.steinbach}@tum.de

[2]Institute for Circuit Theory and Signal Processing
Wireless Communications
Munich University of Technology
email: {ivrlac, josef.a.nossek}@tum.de

ABSTRACT

In this paper[1], we provide an accurate and fully analytical model for the distortion due to lost frames in wireless video transmission. Our analysis combines the properties of the video sequence and those of the wireless transmission link. Built on the *long-term average* properties of the video source, we first develop a model for the average distortion due to the loss of frames, and then derive a mathematical model for analysis of transmission errors occurring during wireless transmission. Experimental results show a surprisingly accurate behaviour of the proposed model.

1. INTRODUCTION

Optimization of end-to-end quality of compressed video transmission has recently gained attention in the research community, both for the wireline Internet and for its wireless extensions [1, 2, 3, 4]. Due to its conceptual simplicity and mathematical convenience, it is common to associate video quality with the average pixel-by-pixel distortion of the video frames. This distortion results from both the effects of lossy compression introduced by source encoding (source distortion), and the lossy transmission channel (loss distortion). Accurate modelling of the different types of distortion is the foundation to successful end-to-end quality optimization.

Obtaining good models is always a challanging task, especially for the loss distortion, as it depends both on the properties of the video encoding and on the transmission system used to transfer the compressed video. Previously proposed analysis of loss distortion tend to model packet losses by superposition of independent single frame losses (e.g. [1, 5]). This gives however only reliable results when single losses are spaced sufficiently far apart. While the impact of channel memory has been recognized (e.g. [6, 7]), there is still lack of theoretically founded models for the loss distortion in wireless systems.

In this paper, we provide an accurate and fully analytical model for the distortion due to lost frames in wireless systems. The proposed model is built on *long-term average* properties of the video source. Since the instantaneous distortion of a group of pictures (GOP) due to the loss of frames is random in nature and changing from GOP to GOP, it is difficult to model. On the other hand, the long-term distortion is deterministic and allows for simple, yet highly accurate modelling, as will be demonstrated. Our analysis combines the properties of the video sequence (e.g., source rate, display frame rate, and frame size) and those of the wireless transmission link (e.g., packet error probability, packet size, and forward error control coding). In particular, we take into account the time-varying fading nature of the wireless channel and develop a relationship between the properties of the wireless link and the model parameters of a two-state Markov process (See Section 5). In order to provide a complete analysis of the expected distortion, we first develop a model for the average distortion due to the loss of frames (See Section 4). Then, we derive a mathematical model for analysis of transmission errors occuring during wireless transmission (See Section 5). Moreover, we provide a way to simplify the model (See Section 6). We validate the proposed model with the experimental results, which are shown to be surprisingly accurate (See Section 7). This also reveals the potential of our applied long-term average concept. The application of this proposed model can be, for instance, finding an optimum trade-off between source and channel coding. Another application may be providing an objective function to multi-user resource allocation in the so-called cross-layer optimization [3, 4], in conjunction with adaquate models for the source distortion. The long-term average characteristics of the proposed model is well suited for such applications.

2. ENCODING STRUCTURE & ERROR CONCEALMENT

In the following, we will focus on predictive source encoding in IPP$\cdots$P - structure, i.e. each group of pictures (GOP) consists of one intra-frame (I-frame) followed by $(F-1)$ inter-frames (P-frames). Calling T_{GOP} the duration of the GOP, we have a frame rate of F/T_{GOP} frames per second (fps). This encoding structure is prone to error propagation due to inter-frame dependencies introduced by the predictive encoding. A typical measure to mitigate the effect of frame loss is error concealment. In the simplest, yet commonly used way, an incorrectly decoded frame and all subsequent frames in the GOP are replaced by the most recently correctly received frame. In this paper, we assume this type of error concealment, which is referred to as *previous frame error concealment.*

3. LONG-TERM AVERAGE CONCEPT

The distortion analysis developed in this paper is based on the concept of long-term averaging. That is, distortion is defined as the *average distortion* over *multiple* GOPs, during which the video sequence can be assumed to be more or less stationary. The instantanous distortion $\tilde{D}_i$ of a GOP due to the loss of a particular frame, say the i-th frame, is random in nature from GOP to GOP and difficult to model. On the other hand, its long-term average value $D_i = \mathrm{E}[\tilde{D}_i]$ is deterministic and allows for simple, yet accurate modelling, as will be shown in the remaining of the paper. Moreover, we will make use of the *average* size $\mathrm{E}[S_i]$ (in bits) of a frame, say the i-th frame, instead of its instantaneous size S_i:

$$\mathrm{E}[S_i] = \alpha_i R_{\mathrm{s}} T_{\mathrm{GOP}}, \tag{1}$$

[1] This work is partially supported by an Alexander von Humboldt (AvH) Research Fellowship.

0-7803-9134-9/05/$20.00 ©2005 IEEE

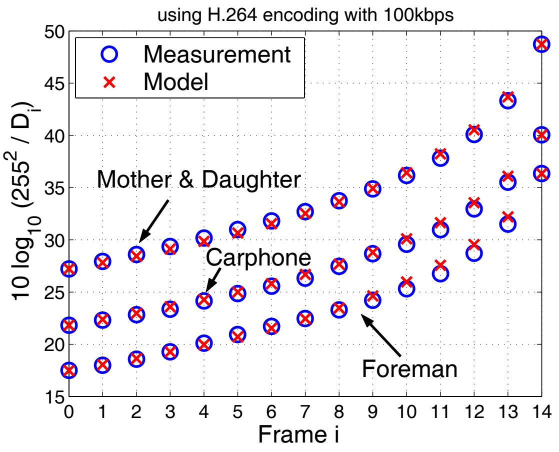

Fig. 1. Average distortion due to lost frame at position i.

where R_s is the average source rate in bits per second (bps), while the factor $\alpha_i \in]0;1]$ is the *average relative size* of the i-th frame. Note that $R_\mathrm{s}T_\mathrm{GOP}$ is the *average* number of bits in a GOP and

$$\sum_{i=0}^{F-1} \alpha_i = 1. \tag{2}$$

Due to previous frame error concealment and the IPP...P-structure, the average distortion D_i introduced by a lost frame depends only on the number $i \in \{0, 1, \ldots, F-1\}$ of the *first* frame which cannot be decoded correctly in a GOP. Calling P_i the probability that this happens, the loss distortion D_L is defined as:

$$D_\mathrm{L} = \sum_{i=0}^{F-1} D_i \cdot P_i, \tag{3}$$

which is the *expected distortion* due to frame losses concealed by previous frame error concealment. In the following, we will develop analytical models for both D_i and P_i, hence providing complete analysis of the distortion due to frame loss. Note that D_i is related to the video sequence and the applied error concealment technique, while P_i is determined by the characteristics of the transmission system. In this paper, we focus on streaming video transmission over a wireless fading link.

4. DISTORTION MODEL

Suppose the i-th frame is the first lost frame in a GOP, then the i-th frame and all its successors in the GOP are replaced by the $(i-1)$-th frame. The average distortion is then given by

$$D_i = \frac{1}{F}\sum_{m=i}^{F-1} d_{m,i-1} \; = \frac{1}{F}\sum_{n=1}^{F-i} d_n, \tag{4}$$

where $d_{m,i-1}$ is the average MSE between the m-th and $(i-1)$-th frame. Assuming stationarity, we have $d_{m,i-1} = d_{m-i+1} = d_n$, i.e. the average MSE does not depend on the position i of the frame loss, but only on the difference $n = m - i + 1$. Due to temporal proximity, the average error d_1 will be the smallest. With larger time separation, the error will increase, i.e. $(d_{n+1} - d_n) > 0$. It turns out that by setting $(d_{n+1} - d_n)$ to be a positive constant, say Δd, a close correspondence to empirical data is achieved. Hence,

$$d_n = d + (n-1)\Delta d, \quad \text{for} \;\; n \in \{1, 2, \ldots, F-i\}, \tag{5}$$

where d and Δd are constants, is a suitable model. By substituting (5) into (4), we obtain

$$D_i = \frac{1}{2F} \cdot (F-i)\,(2d + (F-1-i)\Delta d)\,. \tag{6}$$

The maximum distortion $D_\mathrm{max} = D_0$ is achieved when the first frame is lost and the distortion is minimum $D_\mathrm{min} = D_{F-1}$ in case the last frame is lost. Thus, we can rewrite (6) as:

$$D_i = (F-i) \cdot \frac{F \cdot i \cdot D_\mathrm{min} + (F-i-1) \cdot D_\mathrm{max}}{(F-1)F}\,. \tag{7}$$

The values D_max and D_min depend on the video sequence and have to be determined by measurement. Figure 1 shows the average distortion D_i obtained from three video test sequences encoded by a H.264 compliant source encoder ($F = 15$ frames per GOP). The circles indicate measured data, while the crosses mark the model prediction from (7). As can be seen from the Figure, the model gives a fairly accurate estimation of the measured D_i. It also turns out from our experiment (not shown here for brevity) that D_max depends only slightly on the source rate.

5. TRANSMISSION ERROR ANALYSIS

Let us now develop a mathematical model for the event probabilities P_i, that the i-th frame in a GOP is the first to be decoded wrongly. The P_i depend on the transmission system, which in our case, consists of the wireless channel, signal-processing and forward error control (FEC) coding. The frames are transmitted over the wireless channel in data packets of size L bits. The data packets get encoded by a FEC code in order to protect them from the influence of noise. At the other end of the wireless link, the FEC decoder outputs a sequence of decoded packets. The resulting packet error probability PEP depends on the FEC code, the transmit and receive signal processing and on the wireless channel itself. We can simplify the analysis by assuming that the FEC code yields correct results as long as the channel quality is above a certain threshold and returns wrongly decoded packets otherwise. Due to mobility, the channel quality depends on time, and can be modelled by a block fading random process. The channel quality is thereby modelled as being *constant* for the so-called *decorrelation time* T_dec and then changing abruptly to an independent random value. The decorrelation time depends on the mobility of the wireless terminal which leads to a Doppler spectrum [8]. Its shape depends on the spatial distribution of the obstacles (scatterers) in the proximity of the wireless terminal and its relative velocity with respect to the scatterers. A useful simplification which is frequently used in theoretical analysis of wireless radio propagation, results if one assumes that the scatterers are aligned at the perimeter of a circle around the wireless terminal. This results in the so-called Jakes Doppler spectrum [9]. The decorrelation time T_dec then relates to velocity v and wavelength λ as:

$$T_\mathrm{dec} = 0.4 \cdot \lambda / v. \tag{8}$$

After the time T_dec, the channel quality takes on a new random value, which falls below the threshold for the FEC decoder with probability PEP. The probability of an error burst for exactly j decorrelation times is therefore given by $\mathrm{PEP}^{j-1}(1-\mathrm{PEP})$. Hence, the average number N_B of packets in an error burst is

$$N_\mathrm{B} = \frac{T_\mathrm{dec}}{T_\mathrm{P}} \cdot \sum_{j=0}^{\infty} j \cdot \mathrm{PEP}^{j-1}(1-\mathrm{PEP}) \tag{9}$$

$$= \frac{T_\mathrm{dec}}{T_\mathrm{P} \cdot (1-\mathrm{PEP})}, \tag{10}$$

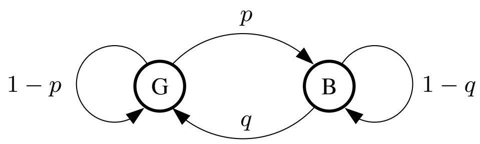

Fig. 2. Markov model for packet loss.

where $T_{\mathrm{P}} = L/R_{\mathrm{s}}$ is the time between two packet transmissions. In order to know if a particular packet can be decoded correctly, we make use of the popular two-state Markov chain depicted in Figure 2. The state is updated prior to each packet transmission. When the model is in state "G", the packet is decoded correctly, otherwise incorrectly. The probability of having an error burst of exactly j wrongly decoded packets is given be $(1-q)^{j-1}q$, and therefore the average number of wrongly decoded packets in an error burst results in $N_{\mathrm{B}} = 1/q$. Comparing to (10) we find

$$q = \frac{L}{R_{\mathrm{s}}T_{\mathrm{dec}}}\left(1 - \mathrm{PEP}\right), \tag{11}$$

where we have used $T_{\mathrm{P}} = L/R_{\mathrm{s}}$. In order to determine the transition probability p of the Markov chain, we look at the ergodic probability P_{B} of being in the state "B":

$$P_{\mathrm{B}} = \frac{p}{p+q}. \tag{12}$$

Since $P_{\mathrm{B}} = \mathrm{PEP}$, it follows from (11) and (12) that

$$p = \frac{L}{R_{\mathrm{s}}T_{\mathrm{dec}}}\mathrm{PEP}. \tag{13}$$

With (1), the average number of packets b_i in frame i is given by:

$$b_i = \alpha_i R_s T_{\mathrm{GOP}}/L. \tag{14}$$

The number of all packets of all frames in a GOP up to and including the i-th frame is then given by

$$B_i = \sum_{n=0}^{i} b_n = \frac{R_s T_{\mathrm{GOP}}}{L}\sum_{n=0}^{i}\alpha_n = \frac{R_s T_{\mathrm{GOP}}}{L}\beta_i, \tag{15}$$

where we have introduced

$$\beta_i = \sum_{n=0}^{i}\alpha_i, \tag{16}$$

for notational convenience. Now, we can calculate the probabilities P_i. Let us first restrict i to be ≥ 1, i.e. the first frame (I-frame) shall be decoded correctly. Hence, the first packet of the first frame has to be correct, which happens with probability $(1-\mathrm{PEP})$. The following $(B_{i-1}-1)$ packets have to be correct, too. This happens with probability $(1-p)^{-1+B_{i-1}}$. Finally, at least one packet of the i-th frame must be lost. Since the i-th frame would have been completely correct with probability $(1-p)^{b_i}$, at least one packet is lost with probability $1-(1-p)^{b_i}$. This leads for $i \geq 1$ to the following expression for P_i

$$P_i = (1-\mathrm{PEP})\cdot(1-p)^{-1+B_{i-1}}\cdot\left(1-(1-p)^{b_i}\right). \tag{17}$$

The probability P_0 of having the first error in the group occuring in the first frame is given by:

$$P_0 = 1-(1-\mathrm{PEP})(1-p)^{-1+b_0}. \tag{18}$$

To this end, (17) and (18) constitute an analytical solution for P_i, for all $i \in \{0, 1, \ldots, F-1\}$. In order to obtain more insight, we proceed as follows. Using $B_i = B_{i-1} + b_i$, we obtain from (17) for $i \geq 1$:

$$P_i = (1-\mathrm{PEP})\cdot\left[(1-p)^{-1+B_{i-1}} - (1-p)^{-1+B_i}\right]. \tag{19}$$

Because of

$$\frac{\left(a^{-1+B_{i-1}} - a^{-1+B_i}\right) - \left(a^{B_{i-1}} - a^{B_i}\right)}{a^{-1+B_{i-1}} - a^{-1+B_i}} = 1-a, \tag{20}$$

where we set $a = 1-p$, we can neglect the -1 terms in the exponents of (19), provided that $1-a = p \ll 1$. From (13), this requests that the packets are not too large:

$$L \ll R_{\mathrm{s}}T_{\mathrm{dec}}/\mathrm{PEP}. \tag{21}$$

As in practical applications the packet error probability is usually small, it usually suffices to have

$$L \leq R_{\mathrm{s}}T_{\mathrm{dec}}. \tag{22}$$

With the definition of p from (13), we can write:

$$(1-p)^{B_i} = \left(1 - \frac{\mathrm{PEP}\cdot\beta_i T_{\mathrm{GOP}}/T_{\mathrm{dec}}}{B_i}\right)^{B_i}. \tag{23}$$

Because of the identity $\lim\limits_{x\to\infty}(1-y/x)^x = \exp(-y)$, we have an approximation of (23):

$$(1-p)^{-1+B_i} \approx \exp\left(-\gamma_i\cdot\mathrm{PEP}\right), \tag{24}$$

where the coefficient γ_i is defined as

$$\gamma_i = \beta_i\frac{T_{\mathrm{GOP}}}{T_{\mathrm{dec}}}. \tag{25}$$

In order to check the validity of the approximation (24), notice that for $y > 0$

$$x \geq \frac{y^2}{2\epsilon} \quad\Leftrightarrow\quad \frac{\exp(-y) - (1-y/x)^x}{\exp(-y)} \leq \epsilon, \tag{26}$$

for any $\epsilon \in]0;1[$. Setting $x = B_i$ and $y = \mathrm{PEP}\cdot\beta_i T_{\mathrm{GOP}}/T_{\mathrm{dec}}$, we see with the help of (15) that approximating $(1-y/x)^x$ with $\exp(-y)$ produces a relative error no larger than ϵ, as long as

$$\mathrm{PEP} \leq \sqrt{\frac{2\epsilon}{L}\cdot\frac{R_{\mathrm{s}}T_{\mathrm{dec}}^2}{T_{\mathrm{GOP}}}}. \tag{27}$$

Using the approximation (24) in (19), we obtain for $i \geq 1$:

$$P_i = (1-\mathrm{PEP})\left[\exp(-\gamma_{i-1}\cdot\mathrm{PEP}) - \exp(-\gamma_i\cdot\mathrm{PEP})\right]. \tag{28}$$

Since $b_0 = B_0$, we can directly apply (24) on (18), so that

$$P_0 = 1-(1-\mathrm{PEP})\cdot\exp(-\gamma_0\cdot\mathrm{PEP}). \tag{29}$$

The equations (28) and (29) constitute a simplified analytical model for the packet loss probabilities, which remains accurate as long as (21) and (27) are obeyed. From (28) and (29), it is interesting to note that the P_i are independent of the source rate R_{s} and the packet size L. They rather depend on the characateristics of the video source (via β_i and T_{GOP}) and on the characteristics of

Table 1. Estimated and measured average MSE

$T_{\rm dec} = 55{\rm ms}$			
PEP →	0.1	0.01	0.001
FM	521 (577)	67.7 (69.2)	6.96 (8.07)
MD	60.9 (76.2)	8.14 (6.93)	0.84 (0.97)

$T_{\rm dec} = 11{\rm ms}$			
PEP →	0.02	0.005	0.001
FM	463 (496)	141 (134)	29.8 (33.7)
MD	55.4 (54.3)	17.2 (17.2)	3.66 (3.56)

the transmission system (via PEP and $T_{\rm dec}$). Together with the distortion D_i due to the loss of the i-th frame from (7), this completes the analytical model for the average distortion given in (3). Validation with measured data obtained from H.264/AVC encoded video test sequences will be presented in Section 7, which reveals the excellent match of the proposed model to the measured results. Besides its accuracy, another advantage of this model is its low and predictable computational complexity.

6. SIMPLIFICATION

The proposed distortion model specified in (3), (7), (28) and (29) is completely analytical. We can nevertheless introduce some simplifications, which essentially do not change its accuracy but make the model better accessible. For example, P-frames ($i \geq 1$) have on average almost the same size. Therefore all α_i with $i \geq 1$ can be defined to be equal. Let the I-frame be on average A times larger than the P-frames. We can write

$$\alpha_i = \frac{1}{F + A - 1} \cdot \begin{cases} A & \text{for } i = 0 \\ 1 & \text{else} \end{cases} . \qquad (30)$$

With (25) and (16), it therefore follows:

$$\gamma_i = \frac{T_{\rm GOP}}{T_{\rm dec}} \cdot \frac{A + i}{F + A - 1}. \qquad (31)$$

When we substitute (31) into (28) and (29), we arrive at a simplified solution. One advantage of this simplification is the reduction of the number of parameters, since all γ_i are derived from the single constant A via (31).

7. EXPERIMENTAL VALIDATION

The proposed model for the loss distortion from (3), (7), (28), (29), and (31) is validated by comparison to measured data with videos generated from H.264/AVC encoded test sequences (”Foreman” (FM) and ”Mother & Daughter” (MD)). We have $F = 15$ frames per GOP with $T_{\rm GOP} = 0.5{\rm sec}$ and the following parameters:

	$D_{\rm min}$	$D_{\rm max}$	A	$R_{\rm s}$
FM	15	1175	6.07	130 kbps
MD	0.87	123	12.3	83 kbps

The test sequences are repeated 100 times to ensure adaquate duration (20 minutes) in order to obtain accurate averaging results. We choose two decorrelation times $T_{\rm dec} \in \{55, 11\}$ ms which correspond to a velocity of $v \in \{3, 20\}$ km/h, at 2GHz carrier frequency, respectively. Table 1 shows sample results from the model with simplifications for selected values of PEP. The measured loss distortion is displayed in parenthesis, which is obtained by transmitting the encoded video with packet size of $L = 432$ bits through the wireless link. Close observation of the results shows that the model predictions deviate less than 1 dB from the measured values of distortion. This demonstrates the potential of the proposed long-term average model.

8. CONCLUSION

A fully analytical model for the distortion due to lost frames in a wireless video transmission system has been proposed. This model builds on the long-term average properties of the video source. Based on the developed model for the average distortion due to the loss of frames, a mathematical model for analysis of transmission errors is derived. The analytical distortion model is shown to be highly accurate by experimental validation. The model may well serve as part of an objective function for cross-layer optimization.

9. ACKNOWLEDGEMENT

The authors like to express their sincere thanks to the Alexander von Humboldt (AvH) foundation for kindly supporting this research.

10. REFERENCES

[1] K. Stuhlmüller, N. Färber, M. Link, and B. Girod, “Analysis of Video Transmission over Lossy Channels”, *IEEE Journal on Selected Areas in Communications*, vol. 18, no 6, pp. 1012-1032, June 2000.

[2] Z. He, and C. W. Chen, “End-to-End Quality Analysis and Modeling for Video Streaming over IP Network”, *Proc. IEEE International Conference on Multimedia and Expo*, ICME 2002.

[3] L. U. Choi, W. Kellerer, and E. Steinbach, “Cross-Layer Optimization for Wireless Multi-User Video Streaming”, *Proc. IEEE International Conference on Image Processing*, ICIP 2004, Singapor, Oct. 2004.

[4] M. T. Ivrlač, and J. A. Nossek, “Cross Layer Design - An Equivalence Class Approach”, *Proc. IEEE International Symposium on Signals, Systems, and Electronics*, ISSSE-04, Linz, Austria, 2004.

[5] I. M. Kim, and H. M. Kim, “A New Resource Allocation Scheme based on a PSNR Criterion for Wireless Video Transmission to Stationary Receivers of Gaussian Channels”, *IEEE Trans. Wireless Commun.*, vol. 1, no 3, pp. 393-401, July 2002.

[6] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, ”Analysis of Packet Loss For Compressed Video: Does Burst-Length Matter?”, *Proc. IEEE Int. Conf. Accoust., Speech, and Signal Proc.*, ICASSP 2003.

[7] J. G. Apostolopoulos, W. Tan, S. J. Wee, and G. W. Wornell, ”Modelling Path Diversity for Multiple Description Video Communication”, *Proc. IEEE Int. Conf. Accoust., Speech, and Signal Proc.*, ICASSP’02, May 2002.

[8] M. J. Gans, ”A Power Spectral Theory of Propagation in the Mobile Radio Environment”, *IEEE Transactions on Vehicular Technology*, Vol. VT-21, pp. 27-38, February 1972.

[9] R. H. Clarke, ”A Statistical Theory of Mobile-radio Reception”, *Bell Systems Technical Journal*, Vol. 47, pp. 957-1000, 1968.