

EFFECTS OF CELL LOSS ON MPEG VIDEO: ANALYTICAL MODELING AND EMPIRICAL VALIDATION

Christopher W Snyder, Uttam K Sarkar, and Dilip Sarkar

Department of Computer Science, University of Miami
POBox 248154, Coral Gables, FL 33124

Emails: snyder@mail.cs.miami.edu, uttam@cs.miami.edu, sarkar@cs.miami.edu

ABSTRACT

MPEG-encoded variable bit rate (VBR) video traffic is expected to incur cell losses during its transmission over a broadband network. These cell losses introduce distortion into the received video. Frame drop rate and Root Mean Square Error (RMSE) can be used to objectively measure this distortion. These two metrics are useful in identifying the *threshold* cell loss rate (CLR) beyond which the quality of received video is unacceptable to viewers. Thus, they are useful in identifying CLRs to be used in evaluating VBR video traffic models. A novel analytical model is developed for predicting both frame drop rate and RMSE as a function of (i) uniform cell loss rate and (ii) MPEG encoding parameters. The model is validated by an empirical study using an actual MPEG video.

1. INTRODUCTION

MPEG encoded *variable bit rate* (VBR) video traffic is expected to dominate the bandwidth of broadband networks [CR99]. Accurate models of VBR video traffic load in a broadband channel are necessary for prediction of performance of any proposed (and/or designed) network during its operation, and significant research efforts have been dedicated to develop these traffic models [HTL92, Hey97, WCJ95, CR99, RK98, SRS01].

A measure of quality of these traffic models is the closeness of cell loss ratios (CLR) between the original video traffic trace and model-generated video traffic trace when those are independently transmitted over two identical communication channels. VBR video frame size patterns are so complicated no model-generated video traffic can be expected to show a CLR identical to that of original traffic. As the transmission rate (the capacity of the channel) and/or the buffer size increases the CLR naturally decreases for both original and any realistic model-generated traffic. At very low transmission rates CLRs would be very high and those could be quite close for both original and modelled traffic traces. However, such high CLRs would be of little use in practice as the quality of received video will be too poor to be acceptable. Consequently, comparisons at high CLRs do not speak much about the quality of a model. As the transmission rate is gradually increased, the difference of CLRs between original and model-generated videos also increases and gradually becomes significant irrespective of the sophistication of any model. In order to study the effectiveness of traffic models it is, therefore, necessary to identify CLRs to be used for evaluating the models. This requires identifying CLRs beyond which the quality of received video sustains perceptually annoying degradations.

Two major issues in user-oriented QoS analysis of MPEG video are (i) encoding bit rate and (ii) CLR during transmission in the network. Standard video quality metrics used for these analyses include \hat{S} [WJP⁺93], Peak Signal to Noise Ratio (PSNR), MPQM[Lam96], MQUANT[VFH99] and Root Mean Square Error (RMSE). Since our study concentrates on the effect of CLR in developing metrics for evaluating video traffic models we focus on RMSE measure.

In Section 2 we develop an analytical model for estimation of frame drop rate and RMSE at a given CLR when cell losses are uniformly distributed over the bitstream. The framework of experimental study is elaborated in Section 3. Section 4 explains the details of how the experiments have been carried out. Section 5 discusses and compares experimentally obtained and analytically modelled results. Section 6 draws concluding remarks.

2. AN ANALYTICAL MODEL

Our study sought to analyse and measure two effects of cell loss on an MPEG bitstream: frame loss and root mean square energy loss (RMSE). Frame loss occurs when the decoder is unable to decode a frame. That frame is entirely lost and will not be included in the output sequence. Root mean square energy loss, defined later in Equation 5, quantifies the effect of pixel-level mismatches.

2.1. Frame Drop Rate

Let N_{cells} be the total number of cells needed to transmit the bitstream, and P_{header} be the probability that any given cell contains an I, P, or B-frame header. Then P_{header} is defined by $P_{header} =$

$$\frac{N_I * CPH_{FH(I)} + N_P * CPH_{FH(P)} + N_B * CPH_{FH(B)}}{N_{cells}}, \quad (1)$$

where $CPH_{FH(I)}$, $CPH_{FH(P)}$, and $CPH_{FH(B)}$ are the average number of cells needed to transmit one I, P, or B-frame header, and N_I , N_P , and N_B are the number of I, P, and B-frames, respectively.

If the header size (HS) is small relative to the cell size (CS), specifically, if $1 < HS \leq CS$, the header will require either one or two cells, depending on the relative positions of the cell and the header (see Fig. 1). Thus, the expected number of cells needed to transmit a given header can be defined as $CPH =$

$$2 * \frac{2 * (HS - 1)}{(HS - 1) + (CS - 1)} + 1 * \left(1 - \frac{2 * (HS - 1)}{(HS - 1) + (CS - 1)}\right) \quad (2)$$

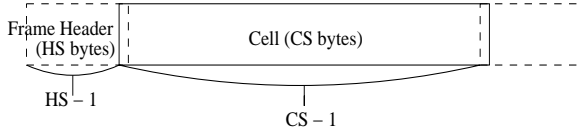


Figure 1: Relationship of Cell Size and MPEG Frame Header Size

Since the loss of a part of a frame header forces the frame to be dropped, the expected frame drop rate is given by

$$FL_{expected}(N_{cells}, CLR, P_{header}) = N_{cells} * CLR * P_{header}. \quad (3)$$

It is possible that $FL_{expected}$, the expected number of frames lost as defined in equation 3, may exceed 1. When this happens we say that the probability of losing at least one frame is 1. In other words, the probability of losing at least one frame is given by

$$P_{(lose \geq 1)} = \min(FL_{expected}(N_{cells}, CLR, P_{header}), 1). \quad (4)$$

For a specific MPEG video both N_{cells} and P_{header} are constants making $FL_{expected}$ and $P_{(lose \geq 1)}$ functions of CLR alone. The frame drop rate is obtained by dividing the expected number of frames lost by the total number of frames.

2.2. Root Mean Square Energy Loss

Root mean square energy loss (RMSE) is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - P'_i)^2}{n}}, \quad (5)$$

where P_i is the value of the i^{th} byte of the undistorted video, P'_i is the value of the i^{th} byte in the distorted video, and n is the total number of bytes compared.

Based on this definition of RMSE we define the expected RMSE as

$$RMSE_{expected} = \sqrt{\frac{DBC_{total} * SD_{avg}}{N_{decode}}}, \quad (6)$$

where SD_{avg} is the average square difference between distorted and unchanged decoded byte values, DBC_{total} is the total number of corrupted decoded bytes, and N_{decode} is the number of bytes in the decoded video. DBC_{total} can be further broken down to

$$DBC_{total} = DBC_{FH} + DBC_{SH} + DBC_{MBH} + DBC_{MBD},$$

where DBC_{FH} , DBC_{SH} , DBC_{MBH} , and DBC_{MBD} are the numbers of decoded bytes corrupted as a result of frame header, slice header, macroblock header, and macroblock data byte corruption, respectively. Due to space constraints we omit the complete expressions for the DBC terms. The details can be found in [SSS02]. For simplicity, we view each macroblock structure as consisting of a header part and a data part; we do not attempt to further break down the macroblock data into motion vectors and blocks. Because of the dependencies between frames in an MPEG video sequence, corruption of a structure in one frame will propagate to the corresponding structures in all dependent frames. In the next Section we describe the experimental setup to validate the proposed analytical model.

3. EXPERIMENTAL SETUP

3.1. High-Level Description of the System

Figure 2 shows the setup for our empirical study. Loss of cells during the transmission of an MPEG-encoded bitstream was simulated (see next section) to produce a distorted copy of the bitstream. Both the distorted copy of bitstream and the original bitstream were decoded separately, producing a sequence of distorted frames, a sequence of unchanged frames, and header offset reports for both bitstreams. The frame sequences were then passed together through a pixel-level analyzer to measure the differences between the distorted frames and the unchanged frames.

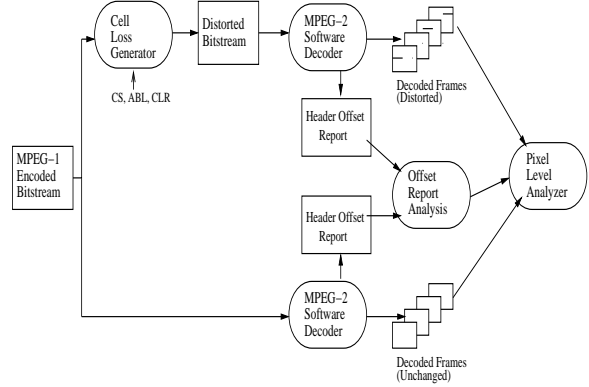


Figure 2: Experimental Setup

3.2. Cell Loss Generator

A two-state Markovian model known as the Gilbert model [RGS91], where states 0 and 1 correspond to incorrect and correct cell reception, was used to simulate cell loss on the video bitstream (see Fig. 3). The model was controlled by the cell size (CS), the expected cell loss rate (CLR), and the average error burst length (ABL). CLR is defined as $\frac{P}{P+Q}$, where P is the number of cells lost and Q the number of cells received. A lost cell was represented by setting the bytes of the cell to all zeros.

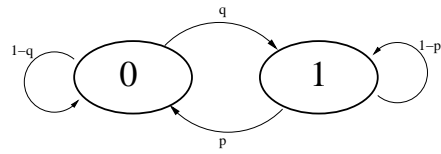


Figure 3: Gilbert Model

4. EMPIRICAL STUDY

A sample video of 194 seconds (30 frames/sec) depicting sea lion foraging habits [lab] and featuring several different indoor and outdoor scenes, some with minimal motion and some with fast motion, was compressed in MPEG-1 format [tmp]. The compressed video had a (15, 3) GOP structure and a data rate of 1 Mbits/sec. A complete specification of the video is available in [SSS02].

4.1. Cell Loss Parameters

The cell loss simulator was implemented as a Java program and the Random class included in the Java API was used to generate uniformly distributed random numbers. Our implementation used a non-bursty (ABL = 1) ATM cell with 48 bytes payload. Multiple repetitions of the experiment were performed at various cell loss ratios (CLRs), ranging from 10^{-2} to 10^{-6} ; 20 repetitions were performed at selected CLRs ranging from $1.0 * 10^{-6}$ to $3.3 * 10^{-6}$; 15 repetitions at CLRs from $1.0 * 10^{-5}$ to $2.5 * 10^{-3}$; and 10 repetitions at CLRs from $3.7 * 10^{-3}$ to $1.0 * 10^{-2}$.

4.2. Frame Loss

The MPEG Software Simulation Group's MPEG-2 software decoder (version 1.2, [Gro]) decoded the video, producing a sequence of picture (frame) files of raw 24-bit RGB pixel values in PPM (Portable Pixmap) format. This decoder also produced a header offset report for each bitstream. By comparing the header offset report for the unchanged bitstream with the header offset report for the distorted bitstream, we were able to identify all frame headers that were lost or corrupted beyond recognition in the distorted bitstream. Thus, by knowing which picture headers were lost we could determine both number and type of frames lost.

4.3. Root Mean Square Energy Loss

A byte-level comparison program was used to compare each decoded frame of the unchanged sequence with its corresponding decoded frame from the distorted sequence. The root mean square energy loss per byte was calculated using Equation 5 for RMSE. Whenever a frame was missing from the distorted sequence due to frame loss, the program compensated by comparing the unchanged frame corresponding to the missing frame to the most recent intact frame in the distorted frame sequence (see Fig. 4). This method simulates the behavior of a video player that simply redisplay the last intact frame whenever it encountered any missing frame.

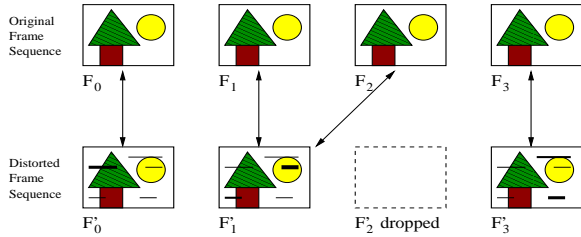


Figure 4: Mechanism for Comparing Dropped Frames

5. RESULTS AND DISCUSSIONS

In this section we compare and contrast analytical results with experimental observations.

5.1. Frame Drop Rate

After the substitution of all video-specific data, we get

$$FL_{expected} = CLR * 7508.52.$$

Figure 5 compares the expected and observed frame losses. The analytical curve gives slightly higher values than the empirical data. This overestimation may be due to the way in which the decoder drops frames. The decoder will drop a frame only if the frame header is missing or corrupted to the point that it cannot be recognized as a frame header, which will only happen if the frame start code is dropped. In our calculation of the expected number of frames lost (see Eqn. 3), we base the average number of cells required to transmit a frame header (see Eqn. 2) on the total frame header size. However, the frame start code, the critical structure in determining whether or not the frame is dropped, is only a part of the whole frame header. Thus, the average number of cells required to transmit the entire frame header is slightly larger than the number required to transmit the critical section of the frame header. This leads to a slight overestimate of P_{header} (see Eqn. 1), and in turn to an overestimate of $FL_{expected}$ (see Eqn. 3).

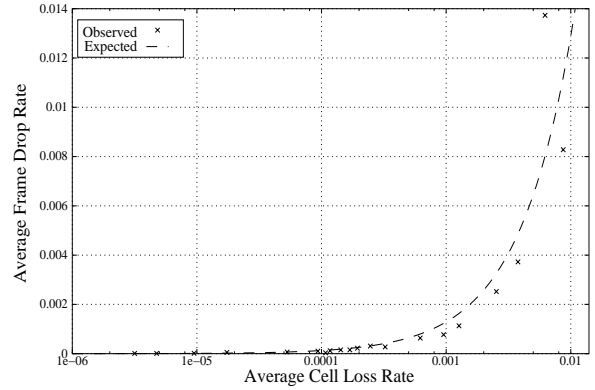


Figure 5: Average Frame Drop Rate

5.2. Probability of Losing At Least One Frame

Figure 6 shows plots of the expected and observed probabilities of losing at least one frame. Like the expected number of frames dropped on which it is based, the expected probability appears to slightly overestimate the observed probability. This difference is believed to be partly due to the reasons given in Section 5.1 and partly due to the limited number of experimental trials performed at each CLR (see 4.1) shown as a scatter diagram. The difference is expected to be much narrower if a large number of trials is performed at each CLR.

5.3. Root Mean Square Energy Loss

Due to space limitation we omit details, but taking the average squared byte value difference as

$$SD_{avg} = \left(\frac{255 - 0}{2} \right)^2 = 127.5^2,$$

and calculating DBC_{total} from the video specific data, we obtain,

$$RMSE_{expected} = \sqrt{CLR} * (1.027 * 10^3).$$

The observed RMSE loss curve follows the analytical curve quite closely for CLR values less than approximately $5 * 10^{-3}$. Beyond that point, the observed and expected values diverge more

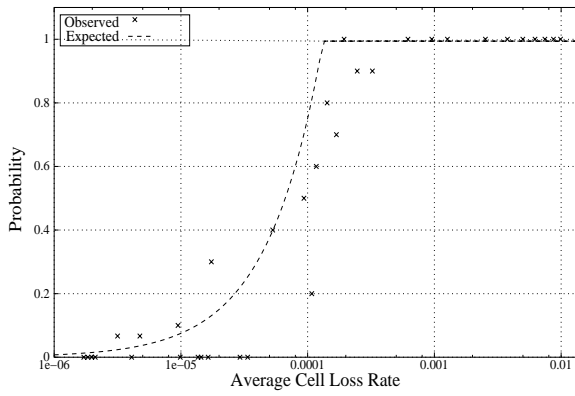


Figure 6: Probability of Losing At Least One Frame

and more as the slope of the observed curve decreases. This divergence may be due to our method for handling dropped frames. When a frame is dropped from the distorted sequence, the most recent intact frame in the distorted sequence is used as a substitute for it. Because of the high temporal redundancy between adjacent video frames, most bytes of the substitute frame will be very similar to the bytes of the original frame to which they are compared.

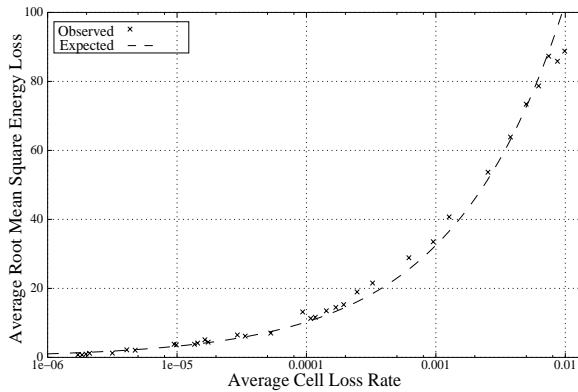


Figure 7: Root Mean Square Energy Loss

6. CONCLUDING REMARKS

An analytical model for predicting frame drop rate and RMSE as functions of uniform cell loss rate and MPEG encoding parameters has been proposed. Empirical observations with one video suggest that our analytical model is quite accurate. The observed frame drop rate and probability of dropping at least one frame is close to, although slightly lower than, the predicted value. The observed RMSE is very close to the expected up to CLR of approximately 5×10^{-3} , beyond which the analytically predicted RMSE is increasingly higher than the observed. This result is natural because the analytical model does not take dropped frame replacement into account. However, the high CLR at which our model differs from the observed values are unlikely to be used for video delivery. Currently, we are refining this model to make it more accurate.

7. REFERENCES

- [CR99] K. Chandra and A. R. Reibman, *Modeling one- and two-layer variable bit rate video*, IEEE/ACM Trans. Networking **7** (1999), no. 3, 398 – 413.
- [Gro] MPEG Software Simulation Group, *MPEG-2 encoder / decoder, version 1.2, July 19, 1996*, url= <http://www.mpeg.org/MSSG/>.
- [Hey97] D. P. Heyman, *The GBAR source model for VBR videoconferences*, IEEE/ACM Trans. Networking **5** (1997), no. 4, 554 – 560.
- [HTL92] D. P. Heyman, A. Tabatabai, and T. V. Lakshman, *Statistical analysis and simulation study of video teleconferencing traffic in ATM*, IEEE Transactions on Circuits Syst. Video Technol. **2** (1992), 49–59.
- [lab] *The laboratory for applied biotelemetry and biotechnology at Texas A and M university galveston website*, http://www.tamug.tamu.edu/lab/Projects/OptFor/Video_clips.htm.
- [Lam96] V. D. B. Lambercht, *Perceptual models and architectures for video coding applications*, Ph.D. thesis, Swiss Federal Institute of Technology, Lausanne, 1996.
- [RGS91] J.W. Roberts, J. Guibert, and A. Simonian, *Network performance considerations in the design of a VBR codec*, 1991, pp. 77–82.
- [RK98] Q. Ren and H. Kobayashi, *Diffusion approximation modeling for markov modulated bursty traffic and its applications to bandwidth allocation in ATM networks*, Journal on Selected Areas in Comm **16** (1998), no. 5, 679–691.
- [SRS01] U. K. Sarkar, S. Ramakrishnan, and D. Sarkar, *Modeling full-length video using markov-modulated gamma-based framework*, Proceedings of IEEE Globecom 2001, Multimedia QoS/Video Applications, 2001, pp. 1 – 5.
- [SSS02] C.W. Snyder, U.K. Sarkar, and D. Sarkar, *Study of ATM network cell loss on quality of MPEG video*, Tech. Report CSC-Jan02-01, Department of Computer Science, University of Miami, Coral Gables, FL 33124, 2002.
- [tmp] *Tsunami MPEG encoder (TMPGEnc)*, http://www.tmpgenc.com/e_main.html.
- [VFH99] O. Verscheure, P. Frossard, and M. Hamdi, *User-oriented QoS analysis in MPEG-2 video delivery*, Real Time Imaging **5** (1999), 305 – 314.
- [WCJ95] J.-L. Wu, Y.-W. Chen, and K.-C. Jiang, *Two models for variable bit rate MPEG sources*, IEICE Trans. Commun. **E78-B** (1995), 737–745.
- [WJP⁺93] A. Webster, C. Jones, M. Pinson, S. Voran, and S. Wolf, *An objective video quality assessment system based on human perception*, SPIE - Human Vision, Visual Processing and Digital Display, vol. 1913, 1993, pp. 15 – 26.