Dependence video quality on NALU size

Alexander Chuykov, Felix Taubin
State University of Aerospace Instrumentation
Saint-Petersburg
alex@chkv.net

Abstract

This article review RTP packets transmission. Each packet contains compressed video information by H.264 AVC standard. Various types of errors in communication channels and ways how to decrease impact of them are considered. The Gilbert channel model is used. Results show, the first, how the video quality depends on the size of RTP packet; the second, how the bitstream size depends on the size of RTP packet. Future research directions are planned.

I. Introduction

The market of mobile devices grows up. There are lots of new devices with new features. They include mobile communication, mobile internet, photo- and video-cameras, GPS and etc. Software and customers services improve existing services and add new such as conference call, video call, social networks access, multimedia libraries and so on. Mobile devices have various connections now. They can be 2G, 3G, 4G and 802.11. Different kinds of the connection offer different quality of service. The quality identify (form video transmission point of view):

1) FPS (Frames Per Second) and low latency
2) Losses and fading block of the frame
3) Video quality (includes visual video quality)

So, the main problem is choosing between transfer rate and quality of service (described higher). It is easy to believe, that small packet size decreases lost probability but increase transfer rate and power consumption.

II. Codec Review

The H.264MPEG-4 AVC (Advanced Video Coding, [1]) or SVC (Scalable Video Coding) is offered for transfer video over mobile networks. Main features are:

1) A lot of profiles for video compression for various bitrates, FPS and resolutions
2) Slice support, various macroblock encoding: I, P, B (bidirectional prediction), SI and SP
3) $\frac{1}{4}$ pixel motion compensation for blocks sizes from 16x16 to 4x4 (few reference frames can be used)
4) Integer transformation for blocks
5) Adaptive non-linear deblocking filter to remove boundaries of blocks or/and macroblocks
6) Entropy encoding
7) Hierarchical bitstream

Modern mobile devices have enough speed and memory to use H.264 to encode and decode video information. The RTP or similar proprietary protocols are used for transfer video streams over IP networks. The H.264 compressed video puts to the RTP payload according rules in [2]. Note. The resolution of the frame (width and height) is not refer quality measure fully. The frame resolution is fixed for mobile devices. It depends on application and/or video codec options.

III. Features of Video Transferring over the Networks

The different kinds of noise influence are affected on wireless transfer. They can be interference, fading and it is no possible to predict them. Obvious, the video information transfer requires special facilities to noise protection [3]. For future review, we split the system of encoding, decoding, sending and receiving to three layers:

1) Physical layer. It equals physical layer OSI model
2) MAC layer. It equals data link layer (MAC sub-layer mostly) and network layer of the OSI model
3) Application layer. It equals all higher layers of the OSI model (application layer mostly)

Let see different kinds of noise which are exist in thee video transferring channels:

1) Fade and noise (like AWGN) on physical layer
2) Packet loosing is result of collision on MAC layer

So, let see how to protect video transferring on different layers:

1) Unequal FEC (Forward Error Correction, [4]) on physical layer
2) The optimal packet size can be choose on MAC layer
3) The application layer provides various types of protection. For example, data partition and slice usage

IV. Review of the H.264 Compression Video Interpretation

The H.264 encoder splits the frame to the slices. It uses the FMO (Flexible Macroblock Ordering, [5]) to do this. Following examples explain two trivial split of frame to slice(s). First, whole frame is one slice. Second, ROI (Region Of Interest) selects the foreground as slice #0; all other macroblocks are background as slice #1. Thus, slice is group of the macroblocks in space and time. The slice encodes independence from other slice. But enhanced slices could encode and transfer to increase video quality. The order of transfer can by any for slices. One slice will used for this article only. It describes whole frame.

The encoded slice consists of encoded macroblocks. This group of encoded macroblocks consists of few NAL (Network Abstraction Layer) units (NALU). They include headers, parameters and compressed texture. NALU includes NALU header with NALU type, payload length and non-empty payload data. Following items describe the NALUs transfer rules over RTP packets:

1) One NALU $\equiv$ one RTP packet, if full NALU size does not exceed limit of RTP packet payload size. The limit of RTP payload size sets by system of video transfer
2) Few NALUs $\equiv$ one RTP packet, if sum of all NALUs size does not exceed limit of RTP packet payload size
3) One NALU $\equiv$ few RTP packet, if NALU size exceeds limit of RTP packet payload size. But fragments from different NALU cannot merge to the one packet

Loss of the NALU leads slice decoding fail (partly of fully). Thus some macroblocks cannot decode (partly of fully) which slice describes. Obviously, decreasing of the packet size decrease impact of packet losses, but increases the channel usage.

Data partition means that all NALU groups roughly by importance:

1) Most important information macroblock type, quantization parameter, motion compensation vectors are type A
2) Intra macroblocks are type B
3) Intra macroblocks are type C

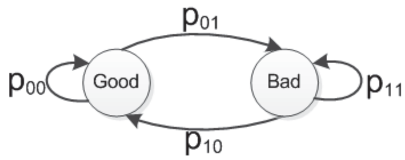

Fig. 1. Two-state Markov chain

This way optimizes the low layer protection (optimize the unequal FEC codes).

The optimal packets size should be choose on the basis of MAC layer and physical layer. So, we have two ways. The first one, small packets decreases impact packet lost and packet errors. On the other hand, small packets increase the channels usage (packets header) and receiver/transmitter loading (power consumption is most important). The second, large packet includes few NALUs. The channel usage decreases and receivertransmitter loading decreases too. But the packet losses and packet errors impact more on video quality.

V. Model Parameters

The PER (Packet Error Rate) and PLR (Packet Loss Rate) are very important parameters for packetized wireless transferring. Let see the relation between video quality and PERPLR. The packetized wireless network was simulated. The RTP packet payload size is introduced (like MTU). The simulation options include requirements above (packets fragmentation or merging). The bitrate was fixed [6]:

1) 192 Kbit 176×144 @ 30 fps
2) 768 Kbit 352×288 @ 30 fps
3) >1 Mbit 352×288 @ 30 fps

The RTP packet sizes are 600, 1000, 1400 and 2000 bytes. The 1400 bytes size approximates the limit IEEE 802.3 networks packet size. The 2000 bytes size approximates the limit IEEE 802.11 networks packet size respectively. The Gilbert model was chosen to simulate wireless network [7]. The model consists of two-state Markov chain (see Fig. 1) with states Good and Bad and transition matrix (1).

$$P = \begin{vmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{vmatrix} \tag{1}$$

This model is very popular for wireless channel simulation and well known. The transition matrix can be written via average burst bit error rate and average burst bit error length.

The average PLRs are calculated for various average burst bit error rate, the average burst bit error length equals to 50 bits and the message packet lengths.

VI. Simulation Results

The average PLRs are calculated. Average relative decoded frames number can be obtain by using these PLRs. Each RTP packet has maximum payload utilization but without changing order of NALUs. Each RTP packet has maximum NALUs as possible or one fragment. For example, the set of NALUs lengths are 100, 100, 100, 800 and 100 bytes. Set RTP payload limit equals 600 bytes. So RTP packets will have following sizes 300 (three NALUs), 600 (fragment 1), 200 (fragment 2) and 100 bytes. This example is not includes header and similar information but it is has quite precise to describe the package idea.

TABLE I
AVERAGE PLR

Average burst bit error	Message packet length, bytes	Average PLR
10^{-5}	600	0.0010
	1000	0.0016
	1400	0.0022
	2000	0.0032
10^{-4}	600	0.0097
	1000	0.0160
	1400	0.0222
	2000	0.0316
10^{-3}	600	0.0925
	1000	0.1488
	1400	0.2016
	2000	0.2748
10^{-2}	600	0.6234
	1000	0.8033
	1400	0.8970
	2000	0.9609

The Fig. 2 shows that packet size is not impact strongly on average number decoded frames for low average probaility (10^{-5} and 10^{-4}). For higher average probabilities (10^{-3} and 10^{-2}) the 1000 bytes length is optimal for RTP packet. The rate-distortions curves have similar view like average number of decode frames. The bitstream size modifications are shown following table. It depends on target encoder bitstream and packet size. The results are percentage relationship between the target bitstream and bitstream size on MAC layer.

TABLE II
BISTREAM ON MAC LAYER

Target encoder bitstream size	Packet size, bytes			
	600	1000	1400	2000
196 Kbit	2.43%	1.35%	1.11%	0.81%
768 Kbit	2.60%	1.74%	1.36%	1.09%
> 1 Mbit	2.48%	1.62%	1.23%	0.96%

Encoder creates NALUs without size limitaion (the 64 kbyte limit exists). Bit stream constists of various NALUs with different lengths. Lengths depend on target encoder bitstream. Higher bitrates make NALU with longer length. Fragmented NALU requires more bits then not fragmented NALU. Obviously, the encoder should control the NALU size to decrease average NALU length. In this article only one slice was used. In future researches two or more slices will be used.

VII. FUTURE WORKS

Future researches have following directions: research how slices depend on video quality (FMO and ASO usage); channel mode improvement (collisions and burst bit error length); increasing efficiency of encode and transfer.

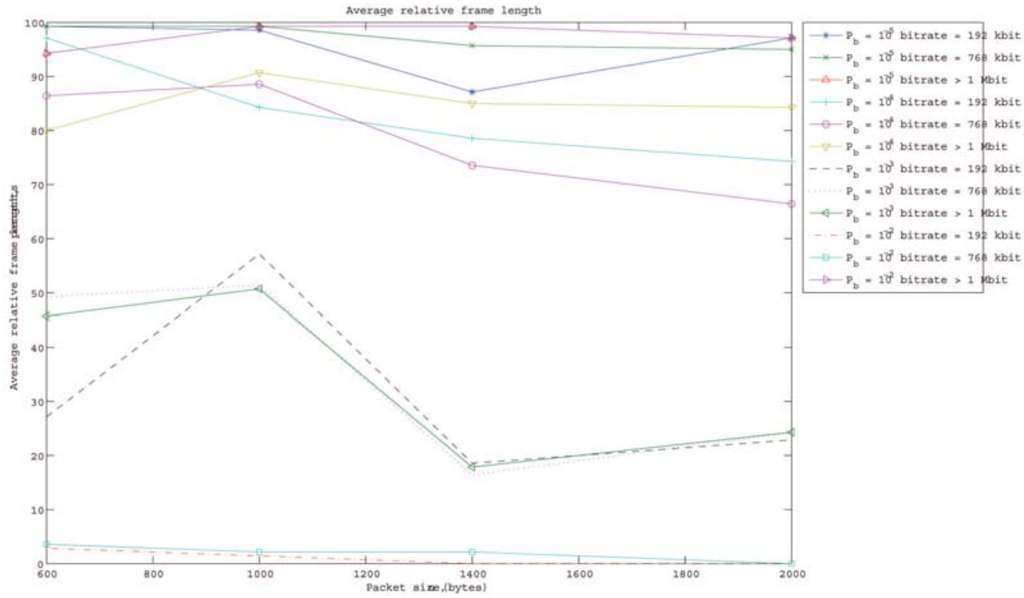

Fig. 2. Simulation results

REFERENCES

[1] I. T. Union, "ITU-T recommendation H.264. Advanced video coding for generic audiovisual services," International Telecommunication Union, Tech. Rep., 2009.

[2] S. Wenger, M. Hannuksela, T. Stockhammer, M. Westerlund, and D. Singer, "RTP Payload Format for H.264 Video," RFC 3984 (Proposed Standard), Internet Engineering Task Force, Feb. 2005. [Online]. Available: http://www.ietf.org/rfc/rfc3984.txt

[3] B. Sklar, *Digital communications. Fundamentals and Applications*. Prentice Hall PTR, 2004.

[4] J. Rosenberg and H. Schulzrinne, "An RTP Payload Format for Generic Forward Error Correction," RFC 2733 (Proposed Standard), Internet Engineering Task Force, Dec. 1999, obsoleted by RFC 5109. [Online]. Available: http://www.ietf.org/rfc/rfc2733.txt

[5] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression. Video coding for Next-generation Multimedia*. John Wiley & Sons, 2003.

[6] J. Keith, *Video Demystified*, 5th ed. Elsevier, 2007.

[7] H. Wang and N. Moayeri, "Finite state Markov channel — a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.