GOP Length Effect Analysis on H.264/AVC Video Streaming Transmission Quality over LTE Network

Ulil S. Zulpratita

Abstract—H.264/AVC provides an interface for flexible, bandwidth-optimized transmission of broadcast video streaming in Long Term Evolution (LTE) network. The video quality and compression ratio depends on Group of Pictures (GOP) structure. GOP structure also affects the distortion sensitivity of the video stream due to packet losses. In this paper, the effect of the GOP length to the video quality is investigated. The simulation results show that by increasing the GOP length, the video quality also increases. But after reaching the highest PSNR (Peak Signal to Noise Ratio) or MOS (Mean Opinion Score) value, it starts to decrease. In the first increasing period, the efficient coding plays major role, so higher of GOP lengths lead to a better quality. After the optimal of GOP settings, when the video quality is the highest, the error propagation effect becomes more significant. Hence, using higher Intra frame (I-frame) intervals, the error will be spreading because of significant distortion. Each type of transmitted video has different optimal GOP, depend on its characteristic. For four types of reference videos that are used in the simulation, the optimal GOP is reached at $5 \leq$ GOP length ≤ 8.

Keywords—Error propagation, GOP length, H.264/AVC, LTE.

I. Introduction

THE combination of increased data rates in Long Term Evolution (LTE) technology and emergence of advanced video coding standard H.264 allow mobile operators to offer multimedia-based services with a high quality of experience to end users [1]. LTE technology allows a significant higher capacity at a lower cost per bit, leading to improve commercial viability of video services. However, it will not be able to eliminate other phenomena that may lead to a distortion of the signal. Moreover, streaming video is sensitive to packet losses. Therefore, the transmission of delay sensitive video streams over lossy wireless links needs a special attention [2].

In the paper, we examines several Group of Pictures (GOP) lengths on H.264/AVC video streaming transmission over the LTE network, and analyzes the optimal GOP length, which provides the best video quality. The aim is to find the optimal distance between the I-frame, in order to maximize the coding efficiency and minimize the quality distortion due to error propagation.

To make the analyses, we implement a simulation tool. As the result of the measurements, it able to recommend the adequate GOP length for the given LTE link parameters, in order to achieve the best received video quality.

Author is with the Informatics Engineering Department, Widyatama University, Jl. Cikutra 204A, Bandung, Jawa Barat, Indonesia 40124 (phone: +6222-7275855; e-mail: ulil.zulpratita@widyatama.ac.id).

II. Related Works

In previous works, the error propagation due to packet losses is analyzed in the GOP pattern. Maugey et al. [3] proposed a theoretical model for the error propagation phenomenon generated by a frame loss in a distributed video coding framework.

Lin et al. [4] analyzed the effect of wireless link characteristics on the video quality and found that burst packet losses on the video delivered quality is less than distributed packet losses in the same packet loss rate.

In paper [5], the number of B-frames between two reference frames is investigated. According to the result, the number of following B frames should be from 1 to 4, while in [6] the conclusion is that the number should be varied from 0 to 2. Paper [7] studies the impact of the choice of GOP by evaluating the effects of GOP on both static MPEG videos and on MPEG videos streaming over a lossy network. The result consistently suggests two guidelines. First, the number of B-frames between two reference frames should be close to 2. Second, the number of P frames should be 5 or fewer.

Huszak and Imre [8] analyze the distortion due to error propagation and the MPEG–4 coding efficiency together. Based on the result, it proposes video coding guidelines for videos transmitted over lossy wireless links. The measurement result shows that the coding efficiency is more beneficial than the distortion caused by the error propagation. Hence, the GOP length should be increased to achieve higher streamed video quality improvement at the receiver.

III. Research Method

In the research, it uses two types of simulator. LTE System Level Simulator [9], [10]. is used to simulate the LTE network and LTEVidSim presents a complete framework and tool-set for evaluation of the video transmitted quality over LTE network.

A. LTE System Level Simulator (LTE-SLS)

The LTE system level simulator is published by Institute of Communications and Radio-Frequency Engineering Vienna University of Technology, Austria under a non-commercial academic use license. The simulator supplements an already freely-available LTE link-level simulator. The combination allows for detail simulation of both the physical layer procedures to analyze link-level related issues and system-level simulations where the physical layer is abstracted from link level results and network performance is investigated. In system-level simulations, the physical layer is abstracted by simplified models that capture its essential characteristics with high accuracy and simultaneously low complexity [9], [10].

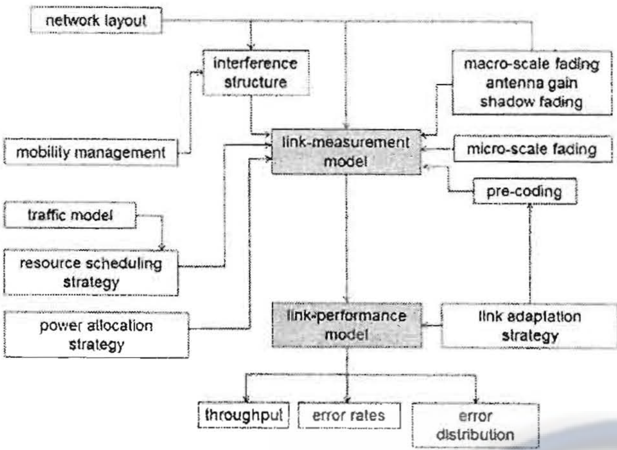


Fig 1. Block diagram of the LTE-SLS [9], [10]

Fig 1 depicts a schematic block diagram of the LTE-SLS. Similarly to other system-level simulators, the core part consists of: (i) a link measurement model and (ii) a link performance model.

B. LTEVidSim

The LTEVidSim is an LTE video transmission simulator. It is a complete framework and tool-set for an evaluation of the quality of video transmitted over a simulated LTE network. It supports a subjective video quality evaluation of the received video based on the frame-by-frame PSNR (Peak Signal to Noise Ratio) calculation. The basic idea of designing the LTEVidSim is referred to the EvalVid tool which is developed by J. Klaue et al. [11]. While the EvalVid is developed by using NS-2, the LTEVidSim is developed by using MATLAB

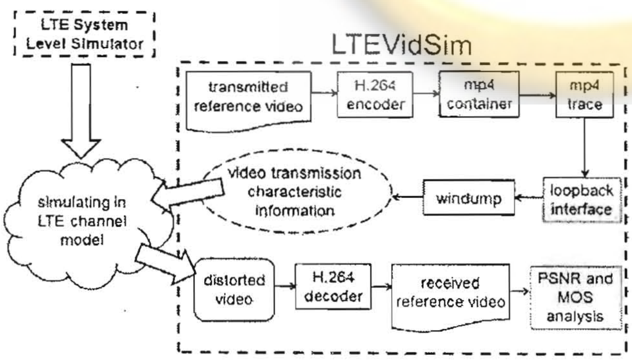


Fig 2. LTEVidSim Block Diagram

The components of LTEVidSim are:

1. **Reference Video** is a video source that needed. Raw uncoded video files are stored in the YUV format, since this is the preferred input format of many available video encoders.
2. **H.264/AVC Codecs** consist of H.264/AVC encoder and H.264/AVC decoder.
3. **Video Sender (VS)** consists of mp4 container and mp4 trace. The mp4 container contains information of every frame in the video file, while mp4 trace contains hint tracks to divide the data into packet-sized for transport over RTP (Real-time Transport Protocol) and then transmit them to a specified destination host.
4. **Loopback Interface** is a virtual interface that is always stay-up even if the outbound interface is down. Any packet transmitted over the virtual loopback interface is immediately received by the self-same interface. In the LTEVidSim, the loopback interface is used to packet sniffing purpose.
5. **Windump** is a packet sniffer that runs under the command line. From the evaluate traces, there are three types of information that can be determined such as:
   a. At sender side: transmission time-stamp and payload size of every packet sent.
   b. At receiver side: receiving time-stamp and payload size of every packet received.
   c. Type of RTP packet information which is packetized by mp4 trace.

Hence, it shows the characteristic of video transmission. The packet is regarded as loss if it arriving too late at the receiver. A maximum tolerable delay for the simulation is 72ms. The value is obtained from coherence time, T_c calculation using Rappaport's rule of thumb [12].

$$T_c = \frac{1}{f_d}\sqrt{\frac{9}{16\pi}} = \frac{0.423}{f_d} \quad (1)$$

Where f_d is a maximum Doppler frequency.

$$f_d = f_o \frac{v}{c} \quad (2)$$

6. **MOS and PSNR Analysis**. There are basically two approaches to measure the digital video quality, namely subjective quality measure and objective quality measure. Subjective quality metrics always grasp crucial factors which are the impression of the user when watching the video. The human quality impression is usually given in a scale from 5 (best) to 1 (worst). This scale is called Mean Opinion Score (MOS) [11], [13]. While the objective quality is measured by taking the average PSNR (Peak-Signal-to-Noise-Ratio) over all the decoded frames. PSNR computes the maximum possible signal energy to noise energy. PSNR measures the difference between the reconstructed video file and the original video file. As long as the video content and the codec type are not changed, PSNR is a valid quality measure [14], [15].

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \quad (3)$$

Mean Square Error (MSE) is the cumulative square between compressed and the original image.

TABLE I
PSNR TO MOS CONVERSION [11]

PSNR (dB)	MOS
>37	5 (Excellent)
31-36.9	4 (Good)
25-30.9	3 (Fair)
20-24.9	2 (Poor)
<19.9	1 (Bad)

C. *Simulation Procedures*

1. LTE system level simulator which is used in this research is 1.3 r427 versions. TABLE II contains the configuration parameters for LTE system level simulator.
2. The LTE system level simulation file result is saved in results folder. The file is used as an LTE channel model for simulating each video simulation scenario at LTEVidSim.
3. Next, the LTEVidSim simulation is run using parameter setting as shown in TABLE III.

Human visual system is more sensitive to brightness level (luminance) rather than colors (chrominance). This is the reason why the chrominance component needs smaller resolution than the luminance. The sub sampling of 4:2:0 means that each of 2x2 luminance pixel, there is 1 chrominance component, where the vertical and horizontal component resolution is half of luminance. The simulation scenarios for the research are shown in TABLE IV.

TABLE II
CONFIGURATION PARAMETERS FOR LTE SYSTEM LEVEL SIMULATOR

Parameter	Value
Frequency	2.1 GHz
System bandwidth	10 MHz
Resources Blocks (RBs)	50 (1 RB = 180 kHz)
Transmission mode	OLSM (Open Loop Spatial Multiplexing)
nTX x nRX (number of transmit and receive antennas)	2x2
Simulation length	12000 TTIs (Transmission Time Intervals, 1 TTI = 1ms)
Latency time scale	25 TTIs [16]
Inter eNodeB distance	500 m
eNodeB (enhanced-NodeB) rings	2
Macroscopic path loss model settings environment	urban
Minimum coupling loss	70 dB [17]
eNodeB TX power	46 dBm [17]
UEs (User Equipments) position	Homogeneous. UEs are located in target sector only, 8 UEs/sector
UE speed (assuming users are pedestrians)	3 km/h = 0.83 m/s
Scheduler	Max Min
Uplink delay	3 TTIs

TABLE III
CONFIGURATION PARAMETERS FOR LTEVIDSIM

Parameter	Value
Reference Video	a. flower garden b. football c. silent d. coastguard
Number of frames	300 frames/sequence
Video resolution (width x height)	176 x 144 pixels, QCIF (Quarter Common Intermediate Format)
Encoding parameters	a. Video bit rate = 0 (auto) b. Frame rate = 30 fps
Video transmission parameter	a. Maximum Tolerable Unit (MTU) = 450 byte b. Maximum tolerable delay = 72 ms c. Stabilization time = 500 ms
Subsampling scheme	4:2:0

TABLE IV
SIMULATION SCENARIOS

Reference Video	H. 264 profile	GOP length
a. flower garden b. football c. silent d. coastguard	Baseline, without B (bi-predictive) frame	From N=2 to N=35, N is the distance between the I-frame.

The LTEVidSim simulation is done by using Monte Carlo iteration method. This method is a probabilistic quantitative analysis method. The Monte Carlo Simulation is used to answer "what if" questions comprehensively. The iteration number which is applied in the simulation is 100 iterations. With 8 UEs, it results 800 data per each simulation scenario.

IV. SIMULATION RESULTS AND ANALYSIS

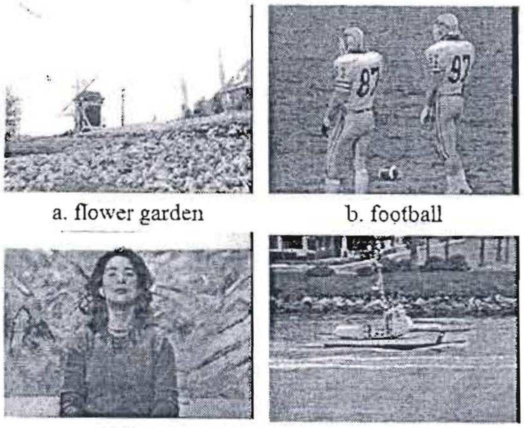
In the research, it shows the impact of program's content on picture quality. The tests involved a wide array of test picture sequences, where each designed to stress encoders in different ways. There are four types of reference video that be used, which are flower garden, football, silent, and coastguard. Each video has different characteristic.

The first video of flower garden contains a moderate motion or a panning of the camera. There are a lot of details information for the background landscapes and colours.

The second video, which is football, contains period of rapid motion, large object and a very complex background.

The third video namely silent contains a little motion and it is very easy for the compression engine. It shows a talking gesture woman. With a little difference between frames, more bits can be used for detail.

The fourth video is a coastguard where contains a moderate motion or a panning of the camera. It has less detail information than flower garden video.



a. flower garden b. football

c. silent d. coastguard

Fig 3. The four video test sequences which are simulated

Below are simulation results for each scenario.

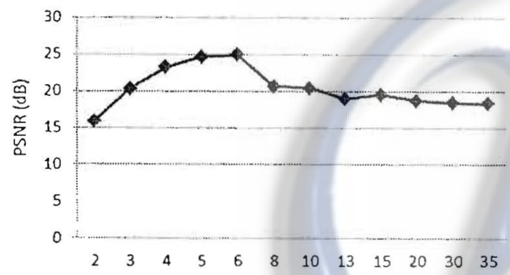


Fig 4. GOP length effect on PSNR of flower garden video

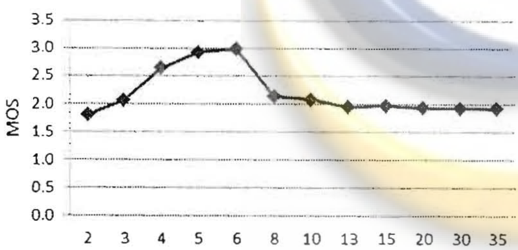


Fig 5. GOP length effect on MOS of flower garden video

Fig 4 and Fig 5 show that by increasing the GOP length the video quality first increases, but after reaching the highest PSNR and MOS value, it starts to decrease. In the first increasing period the efficient coding plays major role, so higher GOP lengths leads to better quality. After the optimal GOP (GOP length = 6) settings, when the video quality is the highest, the error propagation effect becomes more significant. Hence, using higher I frame intervals, the error spreading will cause significant distortion.

The similar phenomena result is also happened for other types of reference videos. For football, silent, and coastguard videos, the optimal GOP lengths are reached at $5 \leq$ GOP length ≤ 8.

Fig 6 and Fig 7 are the combination of four reference videos simulation graphic results. Fig 6 and Fig 7 show the impact of video sequence characteristics on picture quality. The silent video has the best video quality than others, while the football video has the worst. For the silent video, with a little difference between frames, more bits can be used for detail. There is a little difference between the original and the compressed picture.

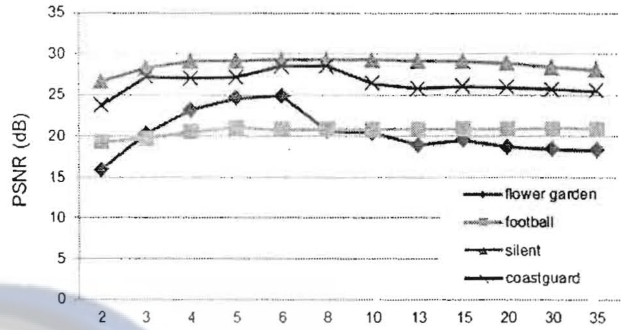


Fig 6. GOP length effect on PSNR value based on video sequence characteristic

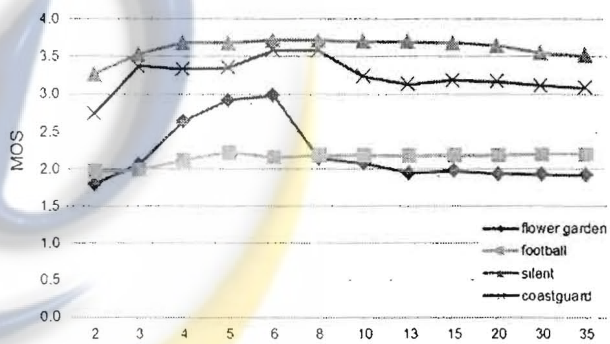


Fig 7. GOP length effect on MOS value based on video sequence characteristic

The obvious results of this test show that a compression rate does not solve all needs. In terms of costs, H.264/AVC allows significant savings on storage costs by choosing compression settings based on the material you are encoding. Sports footage of football video requires significantly higher bit rates than an interview show. Clips with lots of chrominance, typical for commercials, require a higher bit rate to maintain quality.

V. Conclusion

The choice of GOP structure affects static video properties such as frame and file size. The structure also impacts the streaming video quality. The successful decoding ability of the compressed video stream with inter–frame dependencies depends heavily on the receipt of reference frames (I and P frames). While the loss of packets in a frame can degrade the video quality, the more problematic situation is the propagation of errors to dependent frames. By increasing of the packet loss rate, the quality of the decoded frames becomes too poor for viewing. The GOP structure defines the frame type layout of the video therefore the error propagation extent depends heavily on the structure. The error will propagate till the next reference frame.

The optimal GOP length in order to maximize the coding efficiency and minimize the quality distortion due to error propagation is different based on video sequence characteristic which is transmitted over the LTE network. For sports footage video which contains periods of rapid motion, large objects, and a very complex background, it is very difficult to compress. For the same bit rate, this type of video has a worse quality video rather than an interview show. Also clips with lots of chrominance, typical of many commercials, require a higher bit rate to maintain the quality.

REFERENCES

[1] P. McDonagh, C. Vallati, A. Pande, P. Mohapatra, P. Perry, E. Mingozzi, "Investigation of scalable video delivery using H.264 SVC on an LTE network", *Wireless Personal Multimedia Communications (WPMC)*, 2011 14th International Symposium, Brest, France, 3-7 Oct. 2011.

[2] A.D.N. Gordo, "Evaluating video quality degradation from data packet losses in an LTE environment", Unpublished master's thesis, Lulea University of Technology, Sweden, November 2009.

[3] T. Maugey, T. André, B. Pesquet–Popescu, J. Farah, "Analysis of error propagation due to frame losses in a distributed video coding system", *EUSIPCO 2008*, Lausanne, Switzerland, Aug. 2008.

[4] Cheng-Han Lin, Chih-Heng Ke, Ce-Kuen Shieh, N. K. Chilamkurti, "The packet loss effect on MPEG video transmission in wireless networks", *AINA*, Vienna, Austria, 18-20 April 2006.

[5] A. Dumitrasand B. G. Haskell, "I/P/B frame type decision by collinearity of displacements", in *Proc. of ICIP 2004*, Singapore, Oct. 2004.

[6] Y. Yokoyama, "Adaptive GOP structure selection for real-time MPEG-2 video encoding", in *Proc. of ICIP 2000*, Vancouver, Canada, Sept. 2000.

[7] Huahui Wu, M. Claypool, R. E. Kinicki, "Guidelines for selecting practical MPEG group of pictures", *EuroIMSA 2006*, Innsbruck, Austria, 13-15 February 2006

[8] Á. Huszák, S.r Imre, "Analysing GOP structure and packet loss effects on error propagation in MPEG–4 video streams", in *Proc. of the 4th ISCCSP 2010*, Lymassol, Cyprus, 3-5 March 2010

[9] J.C. Ikuno, M. Wrulich, M. Rupp, *System level simulation of LTE networks,* Institute of Communications and Radio-Frequency Engineering Vienna University of Technology, Austria, 2010.(http://www.nt.tuwien.ac.at/ltesimulator)

[10] J. C. Ikuno, M. Wrulich, M. Rupp, "System level simulation of LTE networks," *in Proc. 2010 IEEE 71st Vehicular Technology Conference*, Taipei, Taiwan, May 2010

[11] J. Klaue, B. Tathke, and A. Wolisz, "Evalvid – A framework for video transmission and quality evaluation", In Proc. Of the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, Urbana, Illinois, USA, 2003, pp. 255-272.

[12] T.S. Rappaport, "Wireless Communications: Principles and Practice", Prentice Hall, Upper Saddle River, NJ, 1996.

[13] J. Korhonen, J. You, "improving objective video quality assessment with content analysis", In *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM'10)*, Scottsdale, AZ, USA, Jan 2010.

[14] ITU-T Rec. J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference", *International Telecommunication Union*, Geneva, 2008.

[15] Q. Hyunh-Thu, and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electronics Letters*, 44(13), pp. 800-801, June 2008.

[16] P. Viswanath, D. Tse R. Laroia, "Opportunistic beam forming using dumb antennas", *IEEE Transactions on Information Theory*, vol. 48(6), June, 2002.

[17] Technical Specification Group RAN, "E-UTRA: LTE RF system scenarios," 3GPP, Tech. Rep. TS 36.942, 2008-2009.