Measurement and Analysis of Video Streaming Performance in Live UMTS Networks

Ralf Weber[#1], *Mauricio Guerra*[*2], *Salil Sawhney*[*3],
Leonid Golovanevsky[*4], *Ming Kang*[*5]

[#]QUALCOMM CDMA Technologies GmbH. Engineering Services Group (ESG)
Nordostpark 98, 90411 Nuremberg, Germany
[1]ralf.weber@qualcomm.com

[*]QUALCOMM Incorporated. Engineering Services Group (ESG)
5775 Morehouse Drive, San Diego CA 92121, USA
[2]mguerra@qualcomm.com, [3]ssawhney@qualcomm.com, [4]leonidg@qualcomm.com, [5]mkang@qualcomm.com

Abstract — **This paper discusses an approach to assess Video Streaming (VS) performance over live UMTS networks. The suggested methodology can be used to evaluate the service performance of Release 99 (Rel'99) or HSDPA network implementations. In this paper, measurement and analysis results of VS performance in a deployed Rel'99 UMTS network are presented. The measurements were carried out through field tests in an unloaded network cluster. RTP statistics and video quality metrics are shown for single-user VS tests in both stationary and mobile environments. The results indicate that RTP performance metrics are a good means to correlate RF conditions and protocol observations with terminal and network related behaviors, whereas the suggested set of video quality metrics can be used to evaluate the corresponding user perception.**

***Keywords:** UMTS, Release 99, HSDPA, Multimedia, Video Streaming, Network Performance*

1 Introduction

Currently, Video Streaming (VS) is emerging as a new differentiating service in networks based on Universal Mobile Telecommunication Systems (UMTS) [1]. This type of service can be offered over either UMTS Release 99 (Rel'99) or its evolution based on High Speed Downlink Packet Access (HSDPA). While Rel'99 can guarantee dedicated resources in terms of bandwidth through a fixed assignment of either interactive or streaming bearers, the data rate of HSDPA can vary depending on radio conditions and the number of users if a given Quality-of-Service (QoS) cannot be supported by the network.

Although quality assessments of video codec performance have been an active research area for more than one decade (See e.g., [2-17] and references therein), most of the previous work was mainly focused on the impairments introduced by source coding and compression algorithms [19]. On the other hand, transmission errors in forms of packet delay, jitter (packet inter-arrival variations), and packet loss can have major impact on the end user experience. This is particularly true in a cellular network environment where the channel condition can vary dramatically due to fading and other network effects such as handover. Understanding the impact of these transmission impairments on the overall quality seen by end users is important for a successful deployment of VS services over cellular networks such as UMTS.

The recent study by Heijenk and Niemegeers [19] addressed this topic with the help of a network simulator to compare video quality. Herein, a comparison was done between radio link control (RLC) acknowledged mode (AM) and unacknowledged mode (UM) in terms of average peak signal to noise ratio (PSNR) (with a simple mapping to subjective mean opinion score, MOS), jitter, and video frame error rate. In this paper, we investigate VS performance in a live UMTS network in order to examine effects of different network issues (such as RF condition change, data rate switching etc.) on the overall quality.

Performance considerations for VS assessment over mobile radio networks should include the following aspects:

- Packet loss, jitter, delay, audio-video synchronization and occupancy of media buffers that mitigate delay/jitter problems for audio and video components of a clip
- Encoded clip can have variable instantaneous data rates
- User perception of real-time application like VS is more sensitive to transmission errors as compared to other non-real time applications like FTP, HTTP, etc. (given the use of unreliable transport protocol like UDP, coupled with tighter delay requirements)
- Resulting video metrics are clip dependent, making the selection of clip content important
- Correlation of RF related problems to video impairments is not straight forward due to layered link layer retransmission schemes inherent in mobile communication systems

The remainder of this paper is organized as follows: Section II is a brief overview of some basic concepts and definitions used in this paper for assessing the quality of VS.

Section III describes the overall assessment process. Section IV describes the test scenarios and Section V presents the performance results. Section VI summarizes the paper with some concluding remarks.

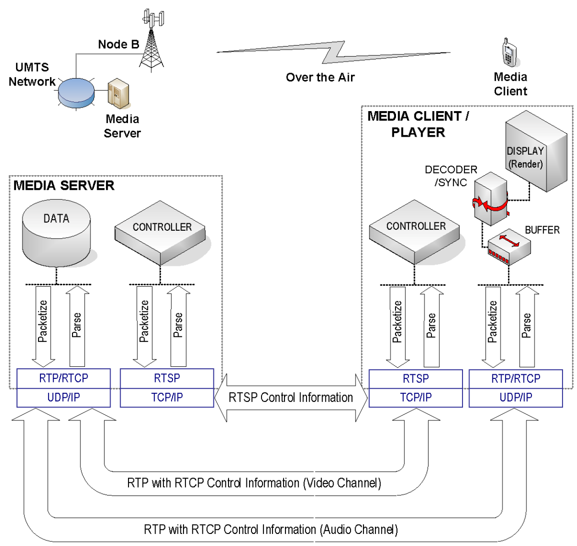

Figure 1 Typical VS Protocol Architecture

2 Video Streaming Assessment Concept

2.1. General considerations and protocol architecture

In this work, we are focusing on the performance of streaming MPEG-4 encoded videos over UMTS Rel'99 networks in the packet switched (PS) domain. We assume that one interactive Radio Access Bearer (RAB) class [20] of either 64/128/384kbps is used for content delivery. The typical protocol architecture of Figure 1 is considered. Control information is sent over Real Time Streaming Protocol (RTSP) over TCP/IP to allow clients to request, setup, play, pause, record and teardown VS sessions. Media data stream and receiver feedback statistics are delivered using Real time Transport Protocol/Real time Transport Control Protocol (RTP/RTCP) [21-24] over UDP/IP. The RLC protocol is configured in Acknowledged "No-Discard" Mode [25]. A target downlink (DL) transport channel Block Error Rate (BLER) of 1% is adjusted and a media de-jitter buffer of 2.5s is employed in the application layer to offset the inter-arrival variations.

2.2. Quality metrics

We consider two types of quality metrics:

1) RTP statistics and application layer metrics including the RTP throughputs (total, video, and audio), RTP video/audio jitter, video/audio buffer occupancy, numbers of lost video/audio packets, etc. These RTP metrics provide insight into the impact of network conditions on the resulting trends in audio and video quality.

2) Perceptual video quality metrics (see [1-16]) including average PSNR [19], Structural Similarity [16] (SSIM, which is based on the difference in the mean and variance between the reference and impaired video clips as well as the correlation coefficients between the two clips), PIQE (Psycho-visually-based Image Quality Evaluator) Similarity [6], and PIQE Blockiness [6]. SSIM and PIQE Similarity are full reference estimators. While SSIM focuses on luminance, contrast and structural elements, PIQE Similarity focuses on remaining edge correspondence between original and degraded video frames. PIQE Blockiness primarily estimates the blockiness artifacts in a video frame due to block-based video compression that can result from transmission errors and packet losses. To assess the PSNR degradations introduced by the radio transmission, we define the PSNR calculation based on the encoded MPEG4 clip (not the raw, un-compressed source) originally transmitted and the impaired clip received by the user equipment (UE) client. To distinguish this metric from the commonly defined PSNR [19] between the raw source clip and encoded clip, we denote it here as E-PSNR.

3 Video Assessment Process

The applied principal video assessment process is depicted in Figure 2.

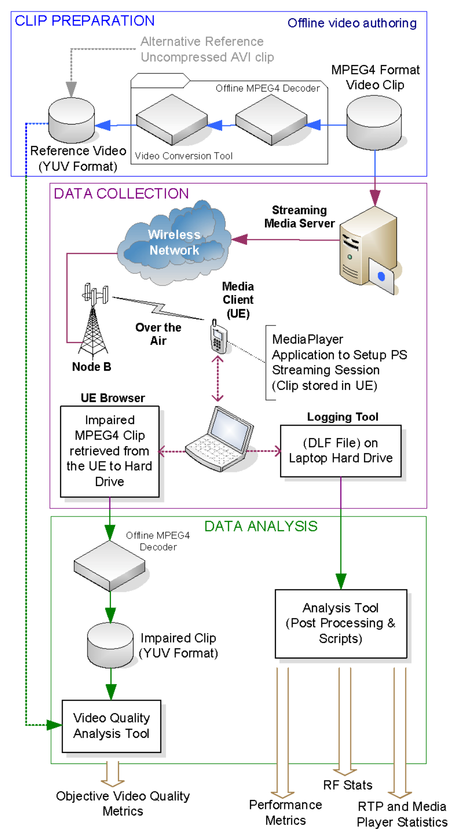

Figure 2 Applied VS Assessment Process

Independent of the assessment methodology, it is important to mention that the encoding rate, the display size (resolution) as well as the clip content (e.g. high motion and scene complexity) can impact VS performance. Furthermore, audio and video settings should be defined consistently between the media server and the clients. For this, setup and configuration (parameters) need to be specified and the selected encoder implementation has to meet pre-defined quality requirements.

4 Test Scenarios

The following tests were executed:

1) Two test scenarios were used to assess single-user performance in stationary environments:
   - Mid Cell Scenario – allows an evaluation of streaming performance in such common RF environments
   - Far Cell Scenario – allows an evaluation of streaming performance in RF environments that could be seen by users in weak or marginal RF conditions (cell-edge scenario).
2) To assess single-user performance in mobility, a selected metric drive route was selected, which allows an evaluation of streaming performance in mobile environments with both good and weak RF conditions.

Table 1 summarizes the RF conditions at selected mid cell and far cell locations. In both cases, the average Active Set Size was one.

Table 1 RF Environment in Unloaded Test Locations

RF Metrics (Average)	Mid-Cell Location	Far-Cell Location
Combined Ec/No [dB]	-3.4	-9.7
Rx Power [dBm]	-83.3	-100.7
Tx Power [dBm]	-17.6	5.7

The cluster of cells had minimal intra-cell interference (unloaded cell) and inter-cell interference (small cluster).

The selected metric route intentionally included mixed RF environments with near, mid and far cell conditions. Statistics were obtained over a whole metric route, including streaming of 3 identical clips in sequence. Each clip consisted 4503 frames and had a duration of 5 minutes.

The selected clips had an audio coding rate of 24 kbps (AAC format) and a video coding rate of 80 kbps (MPEG-4 format) resulting in a nominal total encoding rate of 104 kbps. No bit rate adaptation was employed. The average speed along the metric route was 34 km/h.

5 Performance Results

Both RTP layer statistics and video quality metrics were considered to assess VS performance. Table 2 and Table 3 summarize the obtained RTP statistics and perceptual video quality metrics, respectively, indicating trends for perceived VS performance of the respective scenarios.

Table 2 RTP and Buffer Performance Comparison

Average Values	Performance Metrics for 104kbps Clip	Mid-Cell Location	Far-Cell Location	Mobility
Misc Metrics	SDU Throughput [kbps]	96.2	89.7	92.4
	Audio/Video Sync Offset [ms]	-30.2	-29.3	-30.1
Video Metrics	RTP Throughput [kbps]	62.6	61.9	60.5
	Total # RTP Packets Dropped	13 (0.4%)	188 (4.3%)	438 (3.3%)
	Jitter [ms]	26.7	33.6	39.2
	Differential RTP Packet Delay [ms]	677.6	1460.7	-58.0
	Buffer Occupancy [s]	4.8	4.9	4.7
Audio Metrics	RTP Throughput [kbps]	25.0	24.3	23.8
	Total # RTP Packets Dropped	34 (1.0%)	427 (7.0%)	1017 (5.5%)
	Jitter [ms]	21.6	30.7	30.9
	Total # Re-buffering Events	0	2	4
	Re-buffering Events Duration [s]	-	13.9	7.4
	Differential RTP Packet Delay [ms]	868.1	1587.9	17.9
	Buffer Occupancy [s]	4.4	4.4	4.2

Table 3 Perceptual Video Quality (104 Kbps Clip)

Average Performance Metrics	Mid-Cell Location	Far-Cell Location	Mobility
Total # Dropped Frames	13 (~0%)	170 (3.8%)	409 (3.0%)
PIQE Blockiness	0.0	0.03	0.02
PIQE Similarity	1.0	0.95	0.96
Structural Similarity	1.0	0.96	0.98

RTP statistics show a clear downward trend with degraded RF environment going from mid-cell to far-cell and mobility scenarios. Since all re-buffering events were triggered by reduced audio buffer occupancy, there were no re-buffering events listed as part of the video metrics.

As in stationary cases, significant impairments were observed in the mobility scenario, when the assigned RAB rate was lower than the total clip encoding rate (down-switch to 64kbps). There have been correlations noticed between audio re-buffering events and RAB changes.

From the RF environment shown in Figure 3, it is apparent that just before the down-switch from 384 kbps RAB to 64 kbps RAB, the Ec/No and RSSI decrease significantly while UE Tx power increases. At the same time, the sudden increase in DL BLER indicates that the

Node-B seem to hit the maximum DCH power threshold for this RAB and thus is unable to transmit more power during this weak RF environment. As a result hereof, the RNC switches the UE to a 64 kbps RAB.

Figure 4 shows the impact of the RAB down-switch on the trend in audio and video quality metrics. The frame render rate goes to zero during the re-buffering period while audio and video buffers are being replenished.

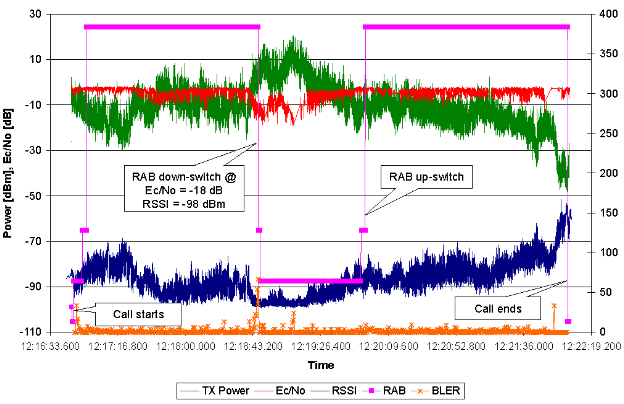

Figure 3 RF Conditions and RAB assignment

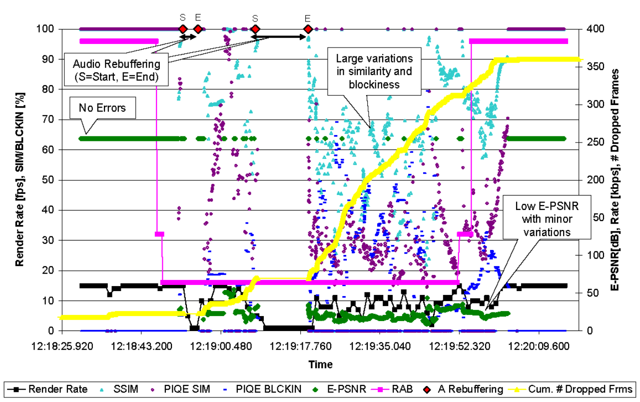

Figure 4 RAB down-switch impacts on Audio and Video Quality

As soon as a 64 kbps RAB is assigned, the frame render rate is always lower than the expected 15 fps. This is due to the fact that frames can only be rendered as fast as they are received, because the restricted 64 kbps air interface rate is not sufficient to accommodate the nominal audio plus video clip encoding rate of 104 kbps.

The E-PSNR values shown in Figure 4 experience a significant reduction during the 64 kbps RAB interval, while before and after the down-switch no degradations (E-PSNR value set to 255 dB) are observed.

According to Figure 4, PIQE Similarity and SSIM have values equal to or close to 1 before and after the RAB down-switch, indicating an error free situation. However, during the RAB down-switch, they also show significant degradations (PIQE Similarity more than SSIM). Also PIQE Blockiness indicates no errors (values equal to or close to 0) before and after the RAB down-switch, while significant degradations occur during the 64 kbps RAB period. In contrast to E-PSNR, blockiness and similarity metrics undergo high variations within this interval, which makes it difficult to correlate the degradations with the performance perception during this time.

Figure 5 depicts the RAB down-switch impact on RTP quality metrics. A number of audio and video RTP packets are lost continuously during the 64 kbps RAB assignment. Given that all RABs are configured here as RLC-AM "No Discard" Mode, the interface between RNC and UE can be considered reliable. Hence, the RTP packet loss is assumed to occur over the RNC-SGSN-GGSN server interface.

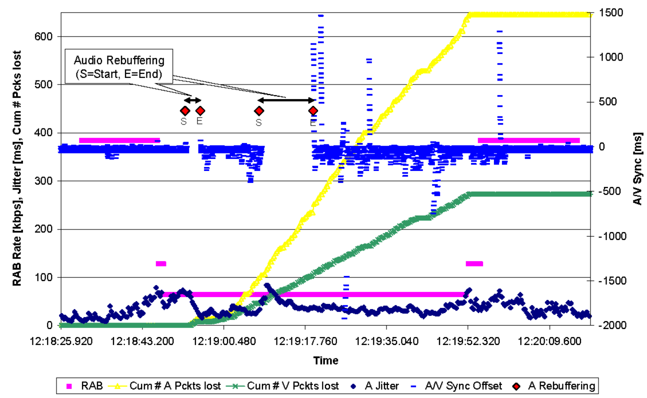

Figure 5 RAB down-switch impacts on RTP Quality Metrics

As can be seen from Figure 5, RTP audio packet jitter is low and has limited variance during the 64 kbps RAB. However, jitter spikes can be observed just around the time of the down-switch occasions from 384 kbps to 64 kbps RAB and stabilizes afterwards. High values of RTP audio packet jitter are also apparent during the time intervals when re-buffering is in progress. Further spikes are also seen after the completion of re-buffering.

The synchronization offset between audio and video RTP packets (A/V Sync Offset) shows spikes not only around the time RTP packet losses are clustered, but also just before a RAB down-switch occurs, as well as before and after re-buffering events (see Figure 5).

Dropped frames are a result of consecutive loss of RTP packets (see Figure 4 and Figure 5). This causes the user to see some intermittent freeze frames. The same effect happens during re-buffering periods when rendering is stopped completely.

Each audio re-buffering event freezes video rendering to ensure that audio buffer fills up to a specified level so that a continuous synchronous playback of audio and video is possible afterwards. However, re-buffering always results in a strong negative impact on user perception.

As shown above, higher number of video impairments were observed when the network down-switches the interactive RAB to 64kbps. This usually resulted in a higher number of re-buffering events owing to a depletion of the media buffer as the buffer outflow rate (playback rate) exceeded the buffer inflow rate. As expected, the lowest number of dropped frames was observed at the mid cell location where the best overall RF conditions were present.

To mitigate the impact of RAB down-switching, several approaches could be considered. One of them would be to implement bit rate adaptation techniques driven either by the media server or the media client. Dynamic streaming rate adaptations can improve streaming performance by reducing the effect of RAB down-switching. Another approach would be to employ clips with a nominal coding rate lower than the 64 kbps RAB. While using the latter approach, the tradeoff would be that this could decrease the quality of the clip, while the number of supported VS users (cell capacity) can increase. To accommodate higher clip rates at the expense of reduced coverage, the DCH power thresholds for the 128 kbps RAB could be optimized to disallow or minimize switching to the 64 kbps RAB. It is up to the operators to decide which technology (Rel'99 or HSDPA) and which strategy to follow in order to provide adequate video quality without sacrificing cell capacity.

6. Conclusions

This paper demonstrated a practical approach to assess video streaming performance over mobile radio networks. Real-time measurement results obtained in a deployed Rel'99 UMTS network were shown. The same assessment concept can also be used for VS services over HSDPA.

RTP statistics and video quality metrics were employed to study single-user VS performance of a 104 kbps encoded clip in stationary mid/far-cell as well as in mobility conditions. Attention was paid to interactions of underlying protocol layers and mechanisms.

It turned out that the video and audio performance was impacted mainly when the networked down-switched the radio bearer to 64 kbps and thus below the effective clip rate. This resulted in a drastic increase of re-buffering events leading to perceived visual impairments. The latter were reflected in reduced E-PSNR, higher blockiness as well as lower similarity. While the resulting E-PSNR showed only marginal variations during the impaired time intervals, all other video metrics had much higher variations and did not indicate a clear quality trend. As long as the bearer rate was 128 kbps or higher, no visual impairments were observed with a Rel'99 RAB target downlink transport channel BLER of about 1%. As a result, a few dropped/freeze frames were always randomly distributed across the 5 minutes clip with minimal effect on the user perceived quality.

It could be shown that the applied assessment concept allows a combined investigation of video quality and network related issues including radio conditions and RTP protocol aspects. The achieved correlation of user perceived impairments with network related events can be used to improve and optimize UMTS systems. The applied approach might need to be revised for VS assessments when a bit rate adaptation concept is implemented.

References

[1] 3GPP Specifications, www.3gpp.org

[2] P. Marziliano et. al, "A no-reference perceptual blur metric", ICIP'02.

[3] Y. Yoshida et al, "Parameter estimation of uniform image blur using DCT", IEICE Trans. Fundamentals, pp . 1154 – 1157, July, 1993.

[4] X. Marichal et. al, "Blur determination in the compressed domain using DCT informtation", ICIP'99

[5] J. Caviedes et. al, "No-reference sharpness metric based on local edge Kurtosis", ICIP'02.

[6] A. Leontaris and A.R. Reibman, "Comparison of blocking and blurring metrics for video compression", ICASSP, 2005

[7] R. W. Chan and P. B. Goldsmith, "A psychovisually-based image quality evaluator for JPEG images", ICSMC'00

[8] J. Yang et al, "Noise estimation for blocking artifacts reduction in DCT coded images", IEEE Trans. CSVT, pp. 1116-1134, Oct. 2000

[9] S. Minami et. al., " An optimization approach for removing blocking effects in transform coding", IEEE Trans. on CSVT, pp. 74-82, Apr. 1995

[10] W. Gao et al, "A de-blocking algorithm and a blockiness metric for highly compressed images", IEEE Trans. On CSVT, pp. 1150—1159, Dec. 2002

[11] H. R. Wu et. al., "A generalized block-edge impairment metric for video coding", IEEE SP Let., Nov. 1997

[12] S. Suthaharan, "Perceptual quality metric for digital video coding", IEE EL. Let., Mar. 2003

[13] T. Vlachos, "Detection of blocking artifacts in compressed video", IEE El. Let., Jun. 2000.

[14] S. Liu et. al, "Efficient DCT-domain blind measurement and reduction of blocking artifacts", IEEE Trans. On CSVT, pp. 1139—1149, Dec. 2002

[15] Z. Wang et al, " Blind measurement of blocking artifacts in compressed video", ICIP'00.

[16] Z. Wang et al, "Image quality assessment: From error visibility to structural similarity", IEEE Trans. Image Processing, pp. 600—612, Apr. 2004

[17] A.B. Watson, "Toward a perceptual video quality metric", Human Vision, Visual Processing, and Digital Display VIII, 3299, pp. 139--147, 1998

[18] R. Weber, "Multimedia Performance Assessments in Deployed UMTS Networks", IEEE Proceedings of ICMSC 2005 Conference, Montreal, Canada, Aug. 14 – 17, 2005

[19] A. Lo, G. Heijenk, I. Niemegeers, "Performance Evaluation of MPEG-4 Video Streaming over UMTS Networks using an Integrated Tool Environment", Proc. of 2005 Int. Symp. on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'05), Philadelphia, PA, USA, July 24-28, 2005.

[20] A.N. Netravali and B.G. Haskell, Digital Pictures: Representation, Compression, and Standards (2nd Ed), Plenum Press, NY, 1995.

[21] 3GPP Technical Specification: Group Services and System Aspects: QoS Concept and Architecture (3G TS 23.107 version 3.0.0)

[22] RFC 3551, Standard 65, RTP Profile for Audio and Video Conferences with Minimal Control

[23] RFC 3550, Standard 64, RTP : A Transport Protocol for Real-Time Applications

[24] RFC 1890, Obsolete, RTP Profile for Audio and Video Conferences with Minimal Control

[25] RFC 1889, Obsolete, RTP : A Transport Protocol for Real-Time Applications, 3GPP Technical Specification: Radio Link Control (RLC) protocol specification (3G TS 25.322)