Requirements for the Transmission of Streaming Video in Mobile Wireless Networks

Vasos Vassiliou, Pavlos Antoniou, Iraklis Giannakou, and Andreas Pitsillides*

Networks Research Group
Computer Science Department
University of Cyprus
vasosv@cs.ucy.ac.cy

Abstract. The ability to transmit video and support related real-time multimedia applications is considered important in mobile networks. Video streaming, video conferencing, online interactive gaming, and mobile TV are only a few of the applications expected to support the viability, and survival, of next generation mobile wireless networks. It is, therefore, significant to analyze the interaction of the particular media and applications. This paper presents the characteristics of mobile wireless networks and relates these characteristics to the requirements of video transmission. The relationship derived is based not only on the objective QoS metrics measured in the network, but also on the subjective quality measures obtained by video viewers at end hosts. Through this work we establish guidelines for the transmission of video based on the limits of mobile and wireless networks. We believe that the results help researchers and professionals in the fields of video production and encoding to create videos of high efficiency, in terms of resource utilization, and of high performance, in terms of end-user satisfaction.

1 Introduction

The basic factor behind the success of Third Generation mobile networks, like the Universal Mobile Telecommunications System(UMTS), is the availability of attractive, useful, and low cost services for the final user. Today, a very limited number of multimedia services for digital mobile communication networks exist, because of the limited abilities of user terminals, the low data transmission rates, and the relative cost. Recently, an increasing demand for digital services for the distribution stored video over the Internet is observed. With the spread of Third Generation mobile networks and the increased capabilities of mobile equipment with the ability of capture and playback video, an increase on the demand of these services is expected. Video has been an important media for communications and entertainment for many decades. The growth and popularity of the Internet in the mid-1990s motivated video communication over best-effort packet networks. Video over best-effort packet networks is complicated by a number of

* This work has been performed under the RPF project BINTEO and the UCY project ADAVIDEO

factors including unknown and time-varying bandwidth, delay, and losses, as well as many additional issues such as how to fairly share the network resources amongst many flows and how to efficiently perform one-to-many communication for popular content. Video communication over a dynamic environment, such as a mobile and wireless network is much more difficult than over a static channel, since the bandwidth, delay, and loss are not known in advance and are unbounded.

When the streaming path involves both wired and wireless links, some additional challenges evolve. The first challenge involves the much longer packet delivery time with the addition of a wireless link. The long round-trip delay reduces the efficiency of a number of end-to-end error control mechanisms. The second challenge is the difficulty in inferring network conditions from end-to-end measurements. In high-speed wired networks, packet corruption is so rare that packet loss is a good indication of network congestion, the proper reaction of which is congestion control. In wireless networks, however, packet losses may occur due to corruption in the packet. In the future, we will have access to a variety of mobile terminals with a wide range of display sizes and capabilities. In addition, different radio-access networks will make multiple maximum-access link speeds available. Because of the physical characteristics of cellular radio networks, the quality and, thus, the data rate of an ongoing connection will also vary, contributing to the heterogeneity problem. A related problem is how to efficiently deliver streamed multimedia content over various radio-access networks with different transmission conditions. This is achievable only if the media transport protocols incorporate the specific characteristics of wireless links.

This paper intends to give an understanding of the transmission of video over mobile wireless networks. Adopting the transmission of MPEG4-encoded video streams over wireless network environments, we investigate the types of errors that can be observed, using objective video quality metrics such as PSNR. Furthermore, we provide subjective video quality estimation based on the evaluation of decoded video streams by informed viewers.

The paper is organized as follows. Section 2 provides an overview of the characteristics of the most common mobile and wireless networks. Section 3 provides background information on the objective and subjective quality evaluation methods used in this paper. Section 4 describes the video characteristics, the setup, and the scenarios used to evaluate the transmission of streaming video in a wireless network. Section 5 presents the results of the objective and subjective evaluations. The paper ends with a last section on conclusions.

2 Characteristics of Mobile and Wireless Networks

2.1 Cellular Wireless Networks

Second Generation (2G) Cellular Networks. The main aim in the design of the 2G systems was the maximization of the system capacity, measured as the number of users per spectrum per unit area. 2G makes heavy use of digital technology through the use of digital vocoders, Forward Error Correction

(FEC), and high level digital modulation to improve voice quality, security and call reliability. The GSM technology has been a very stable, widely accepted and probably the most popular standard for mobile communication. The major drawback of GSM with respect to data and video is that GSM-enabled systems do not support high data rates. GSM supports only low rates for data services (up to 9.6 Kbps) and Short Message Services (SMS), thus, it is unable to handle complex data such as video. In addition, the GSM networks are not compatible with the current TCP/IP and other common networks because of differences in network hardware, software and protocols.

2.5G Cellular Networks (GPRS). The General Packet Radio Service (GPRS) is a standard developed by the European Telecommunications Standards Institute (ETSI) on packet data in GSM systems. GPRS is designed to provide a high data rate packet-switched bearer service in a GSM network. GPRS has a number of important benefits with respect to data and video. The most important are: (a) that it uses the same core infrastructure for different air interfaces, (b) it operates on an integrated telephony and Internet infrastructure, (c) it is always on, reducing the time spent in setting up and tearing down connections, (d) it is designed to support bursty applications, such as e-mail, telemetry, broadcast services and web browsing, and (e) it supports high-speed data services with rates up to 384Kbps.

3G Mobile Networks. 3G Systems are intended to provide a global mobility with wide range of services including telephony, paging, messaging, Internet and broadband data. UMTS offers teleservices and bearer services, which provide the capability for information transfer between access points. It is possible to negotiate the characteristics of a bearer service at session or connection establishment and renegotiate them during the session or connection. Bearer services have different QoS parameters for maximum transfer delay, delay variation and bit error rate. UMTS network services have different QoS classes for four types of traffic: Conversational class (voice, video telephony, video gaming) , Streaming class (multimedia, video on demand, webcast), Interactive class (web browsing, network gaming, database access), Background class (email, SMS, downloading).

Offered data rate targets are: 144 Kbps for satellite and rural outdoor, 384 Kbps for urban outdoor, and 2048 Kbps for indoor and low range outdoor. These are the maximum theoretical values in each environment for downlink speeds. The actual data rates may vary from 32Kbps, for a single voice channel, to 768 Kbps in urban low speed connections depending always on the class of service supported.

2.2 IEEE 802.11

Wireless local area networks (WLANs) based on the IEEE 802.11 standard are a significant and viable alternative to wireless connectivity. The standard has currently three variations widely deployed. The 802.11b operates on the 2.4GHz band and has a maximum theoretical data rate of 11Mbps, but operates also on 1, 2 and 5Mbps. The 802.11a and g operate on the 5GHz and 2.4GHz bands respectively and both have a maximum theoretical data rate of 54Mbps. Using different modulation schemes they can also operate on the lower scales of 6, 10, 12, 18, 36, and 48 Mbps.

Based on CSMA/CA, a common resource sharing MAC protocol, 802.11 also adheres to the characteristic that the data rate allocated to each user is inversely proportional to the number of users in the local network. Therefore, the practical data rates are usually lower than those mentioned above.

3 Video Quality Assessment Schemes

3.1 Objective QoS Measures

In an optimal case, the quality of video is monitored during transmission. According to measurements, adjustment of parameters and possible retransmission of the data is carried out. Objective quality assessment methods of digital video can be classified into three categories. In the first category, the quality is evaluated by comparing the decoded video sequence to the original. The objectivity of this method is owed to the fact that there is no human interaction; the original video sequence and the impaired one are fed to a computer algorithm that calculates the distortion between the two. The second category contains methods that compare features calculated from the original and the decoded video sequences. The methods of the third category make observations only on decoded video and estimate the quality using only that information. The Video Quality Experts Group (VQEG) calls these groups the full, the reduced and the no reference methods [1]. Traditional signal distortion measures use an error signal to determine the quality of a system. The error signal is the absolute difference between the original and processed signal. The traditional quality metrics are the Root Mean Square Error (RMSE), the Signal-to-Noise Ratio(SNR), and the Peak Signal-to-Noise Ratio (PSNR) in dB. In this work we employ a Full reference method and use the PSNR as the objective quality metric.

3.2 Subjective QoS Measures

There are numerous metrics used to express the objective quality of an image or video, which cannot, however, characterize fully the response and end satisfaction of the viewer. Perceived measure of the quality of a video is done through the human "grading" of streams which helps collect and utilize the general user view. There is a number of perceived quality of service measurement techniques. Most of them are explained in [2]. The following are the most popular: a) DSIS (Double Stimulus Impairment Scale) b) DSCQS (Double Stimulus Continuous Quality Scale) c) SCAJ (Stimulus Comparison Adjectival Categorical Judgement) d) SAMVIQ (Subjective Assessment Method for Video Quality evaluation)

In this work we have used the SAMVIQ [3] method. SAMVIQ is based on random playout of the test files. The individual viewer can start and stop the evaluation process as he wishes and is allowed to determine his own pace for performing grading, modifying grades, repeating playout when needed, etc. With the SAMVIQ method, quality evaluation is carried out scene after scene including an explicit reference, a hidden reference and various algorithms (codecs). As a result, SAMVIQ offers higher reliability, i.e. smaller standard deviations. A major advantage of this subjective evaluation scheme is in the way video sequences are presented to the viewer. In SAMVIQ video sequences are shown in multi-stimulus form, so that the user can choose the order of tests and correct their votes, as appropriate. As the viewers can directly compare the impaired sequences among themselves and against the reference, they can grade them accordingly. Thus, viewers are generally able to discriminate the different quality levels better with SAMVIQ than with the other methods. In addition, in this method there is only one viewer at a time, which alleviates a "group effect".

4 Evaluation Setup and Scenarios

4.1 Topology.

The evaluation topology consists of one Video Streaming Server, two backbone routers and video clients of variable types and connectivity methods (fixed, mobile, wired, wireless) as shown in Fig. 1. The video streaming server is attached to the first backbone router with a link which has 10Mbps bandwidth and 10ms propagation delay. These values remain constant during all scenarios. This router is connected to a second router using a link with unspecified and variable bandwidth, propagation delay, and packet loss. The different parameter values used to characterize this variable link are shown in Table 1. Using this topology, we conducted several experiments for two different sample sequences and with fixed-wired clients, fixed-wireless clients and mobile-wireless clients.

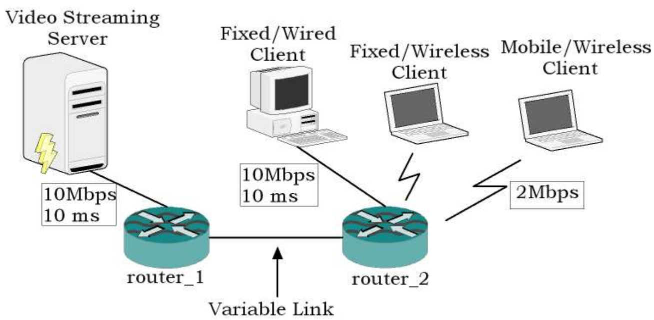

Fig. 1. Video Stream Evaluation Setup.

4.2 Variable Test Parameters.

The choice of the parameters used in the video quality evaluations (Table 1) was based on the typical characteristics of mobile and wireless networks, as these are described in Section 2. For example, the Link Bandwidth can be considered as either the last hop access link BW or the available BW to the user. The values chosen can represent typical wired home access rates (modem, ISDN, xDSL) or different bearer rates for UMTS.

Table 1. Variable Parameters

Video Stream Bit Rate	*Link Bandwidth*	*Propagation Delay*	*Packet Loss*
64 Kbps	64 Kbps	10 ms	
128 Kbps	100 Kbps	50 ms	10^{-5}
256 Kbps	256 Kbps	100 ms	10^{-3}
512 Kbps	512 Kbps	200 ms	
768 Kbps	1 Mbps	400 ms	

4.3 Test sequences.

The test sequences used in this work were the sample sequences Foreman and Claire. The sequences were chosen because of their different characteristics. The first is a stream with a fair amount of movement and change of background, whereas the second is a more static sequence. The characteristics of these sequences are shown in Table 2. The sample sequences were encoded in MPEG4 format with a free software tool called FFMPEG encoder [4]. The two sequences have temporal resolution 30 frames per second, and GoP (Group of Pictures) pattern IBBPBBPBBPBB. Each sequence was encoded at the rates shown in Table 1. The video stream bit rate[1] varies from 64Kbps to 768Kbps. This rate is the average produced by the encoder. Since the encoding of the sample video sequences is based on MPEG4, individual frames have variable sizes.

Table 2. Video Characteristics

Trace	*Resolution*	*Total Frames*	*# I Frames*	*# P Frames*	*# B Frames*
Foreman.yuv	176x144	400	34	100	266
Claire.yuv	176x144	494	42	124	328

4.4 Data Collection

All the aforementioned experiments were conducted with an open source network simulator tool NS2 [5]. Based on the open source framework called EvalVid [6] we were able to collect all the necessary information needed for the objective

[1] The terms video stream bit rate and video encoding rate are used interchangeably in this paper

video quality evaluation like PSNR values, frames lost, packet end to end delay and packet jitter. Some new functionalities were implemented in NS2 from [7] in order to support EvalVid. The whole data collection procedure and PSNR evaluation is illustrated in Fig. 2.

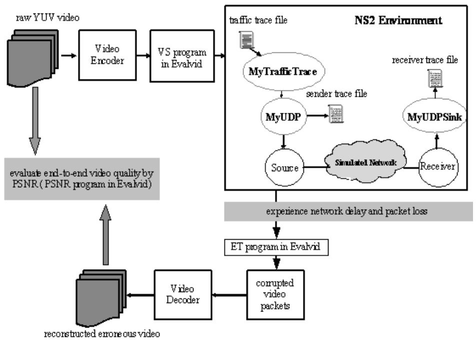

Fig. 2. PSRN calculation through evalvid.

5 Results

In this section we analyze results obtained from the above scenario evaluations. Given the very large number of produced streams, we chose to present and analyze only one scenario. The results presented are for the following case: single user, single video stream, No background traffic, Foreman test sequence, mobile and wireless terminal. All other parameters are variable as shown in Table 1. To identify if and how the different parameters affect the objective value of PSNR we compare them in pairs.

5.1 Link Bandwidth and Propagation Delay

The effect of propagation delay and link bandwidth on the PSNR while keeping the encoding rate steady at 64Kbps and 256Kbps is presented in Fig. 3. These graphs show that the objective values remain relatively constant with the change in either variable, with a slight general increase for high link BW values and counter-intuitively in high delay values as well. There is also an overall upward shift by 1dB when the encoding rate is increased from 64Kbps to 256Kbps. The PSNR is extremely low in the case where the encoding rate is higher than the link BW, as it is evident by Fig. 3b.

5.2 Video Encoding Rate and Propagation Delay

The effect of propagation delay and video encoding rate on the PSNR when keeping the link BW constant at 500Kbps and 1Mbps is presented in Fig. 4. The results show that for the 1M case the results are similar to those of Fig. 3.

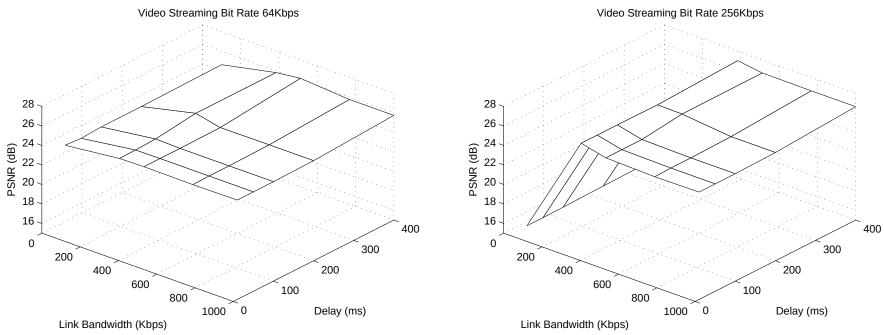

Fig. 3. Mean PSNR values vs Link Bandwidth and Delay (a) 64K Video Encoding Rate, (b) 256K Video Encoding Rate

For the 500K case we observe that the PSNR remains at the same levels with respect to delay, but is significantly reduced when the video encoding rate is at 512Kbps and 768Kbps with the PSNR of the latter being the worst at around 15dB. This leads us to believe that there is a stronger relationship between link BW and encoding rate, than between the link BW and the propagation delay.

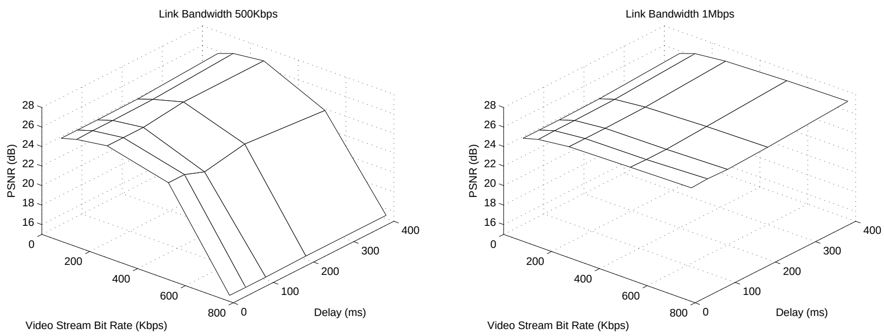

Fig. 4. Mean PSNR values vs Encoding Rate and Delay (a) 500K link BW, (b) 1M link BW

5.3 Link Bandwidth and Video Encoding Rate

Fig. 5 contains the most notable results. More specifically, for both values of delay considered (10ms and 400ms) the PSNR drops dramatically when the encoding rate is higher than the link bandwidth. This is somewhat intuitive if we consider that in those instances the packet losses of the video stream are very big, and approaching 100%, which in turn means that the PSNR is low as well.

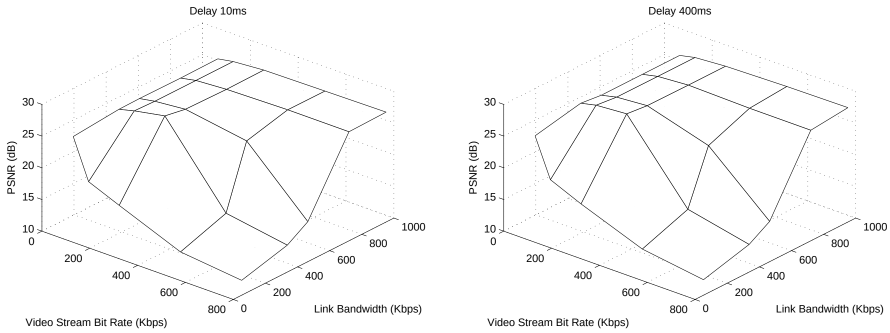

Fig. 5. Mean PSNR values vs Encoding Rate and Link Bandwidth (a) Delay 10ms, (b) Delay 400ms

5.4 Packet Loss

The packet loss rates used in the evaluation produced the same results both for the objective and subjective measures. The results in this paper refer to the 10^{-3} packet loss case.

5.5 Evaluation of Perceived Quality of Service

The set of video streams that were recorded on the receiving site of the evaluation setup was used as input to the PQoS evaluation method explained in Section 3.2. We used the software tool called "MSU Perceptual Video Quality tool" [8] which is a tool for subjective video quality evaluation implementing SAMVIQ. The score grades in this method range from 0 to 100. The videos were evaluated by a group of 20 students at the University of Cyprus. Table 3 presents the relationship between the average value of the students' subjective grading and the objective value obtained through EvalVid. It is clear from the table that

Table 3. Relationship of PSNR with MOS

PSNR (dB)	*MOS*	P-QoS Category
> 27.2	81-100	1 Excellent
26.9 - 27.2	61-80	2 Good
26.1 - 26.9	41-60	3 Fair
16.2 - 26.1	21-40	4 Poor
< 16.2	0-20	5 Bad

the range of PSNR values corresponding to each category is not similar and do not have a liner relationship with the MOS. The video streams which scored high had also an extremely high PSNR. The *Good* and *Fair* categories have also a small range of PSNR values (0.3dB and 0.8dB respectively) whereas the low categories get the bulk of the scores. This phenomenon illustrates clearly how inappropriate is PSNR to evaluate the actual QoS as perceived by the user.

6 Conclusions

In this paper we described the characteristics of mobile wireless networks and related these characteristics to the requirements of video transmission. The tests and simulations analyzed in this paper were designed to correlate objective video quality metrics with subjective video quality. Standard objective metrics such as PSNR were taken into consideration in order to evaluate objective quality. Many factors (bit-rate, link BW, propagation delay) had to be considered to specify effective subjective tests. A novel methodology called SAMVIQ was used for subjective evaluations. This method can be efficiently used for the evaluations of video sequences in both clear and error-prone environments. This set of values, when correlated with the conditions affecting PSNR help us reach some conclusions. Due to space limitations we could not include all additional metrics values for the resultant packet loss, delay, and jitter. From the results of the examined scenario we can conclude that the video quality, as this is reflected through PSNR values, depends on the percentage of lost frames as well as the end-to-and delay. The higher the percentage of lost frames the lower the PSNR values and hence the video objective quality.

Apart from the percentage of lost packets, jitter is an important factor that influences the video quality particularly if a video decoder does not provide buffering operation. Moreover jitter is influenced to a large extent by the network condition i.e. congestion conditions that may prevail in the network. Observing some of the presented graphs concerning PSNR we realize that the end to end delay does not play an essential role in the objective video quality. However, the end to end delay is a critical factor for real-time services and may influence the subjective video quality.

The experimental results show that the higher the video bit-rate the higher the QoS in terms of objective and subjective video quality evaluation measures. Of course the QoS depends primarily on the link bandwidth. As shown in Fig. 3(a) and Fig. 3(b) the best quality in terms of PSNR as well as user-perceived quality is achieved when the encoding rate is less than or equal to the link BW or available BW. Video sequences encoded at 256Kbps and transmitted over 500Kbps and 1Mbps link have almost the same mean PSNR value and viewers perceive the same quality. Needless to say that the most prevalent objective video quality metric does not correlate directly with viewer's perceived quality. Nevertheless the higher the PSNR values the higher the viewer perceived quality.

Through this work we establish guidelines for the transmission of video based on the limits of mobile and wireless networks. We believe that the results help researchers and professionals in the fields of video production and encoding to create videos of high efficiency, in terms of resource utilization, and of high performance, in terms of end-user satisfaction.

References

1. Rohaly M. et al. (2000) "Video Quality Experts Group: Current Results and Future Directions", In: SPIE Visual Communications and Image Processing, Perth, Australia, June 21-23, 2000, Vol. 4067, p.742-753.
2. ITU-R BT.500-11 "Methodology for the subjective assessment of the quality of television pictures"
3. F. Kozamernik, V. Steinmann, P. Sunna, E. Wyckens, "SAMVIQ: A New EBU Methodology for Video Quality Evaluations in Multimedia," SMPTE Motion Imaging Journal, April 2005, pp. 152-160.
4. FFMPEG Multimedia System.
   `http://ffmpeg.sourceforge.net/index.php.`
5. Network Simulator 2
   `http://www.isi.edu/nsnam/ns/`
6. J. Klaue, B. Rathke and A. Wolish, "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, pp. 255-272, Urbana, Illinois, USA, September 2003
7. Chih-Heng Ke, Cheng-Han Lin, Ce-Kuen Shieh, Wen-Shyang Hwang, A Novel Realistic Simulation Tool for Video Transmission over Wireless Network, The IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC2006), June 5-7, 2006, Taichung, Taiwan. Code obtained from:
   `http://140.116.72.80/~smallko/ns2/Evalvid_in_NS2.htm`
8. MSU Perceptual Video Quality tool
   `http://compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html`