University of Cyprus
Department of Computer Science
Networks Research Laboratory

Adaptive Methods for the Transmission of Video Streams in Wireless Networks

Deliverable 1.1

Objective Quality of Service Estimation Methods

Project Title	:	Adaptive Methods for the Transmission of Video Streams in Wireless Networks
Deliverable Type	:	Intermediate WP Report (M06)

Deliverable Number	:	D1.1
Title of Deliverable	:	Objective Quality of Service Estimation Methods
Work Package Title	:	Content Adaptation Techniques
Work Package Number	:	WP1
Internal Document Number	:	
Contractual Delivery Date	:	
Actual Delivery Date	:	
Author(s)	:	Pavlos Antoniou Vasos Vassiliou Andreas Pitsillides
Email(s)	:	paul.antoniou@cs.ucy.ac.cy vasosv@cs.ucy.ac.cy Andreas.Pitsillides@ucy.ac.cy

Abstract

Over the last years, emphasis has been put on developing methods and techniques for evaluating the perceived quality of digital video content. These methods are mainly categorized into two classes: The objective and subjective ones. In this deliverable we investigate perceptual objective quality assessment schemes as well as subjective quality assessment schemes. In order to provide some kind of correlation between these two schemes, we conducted some experiments using the open source network simulator NS2. Furthermore we used a video subjective assessment method (SAMVIQ) proposed in [1] for mapping user-perceived quality to actual QoS in video stream applications.

Keywords: objective quality, subjective quality, Subjective Assessment Method for Video Quality evaluation (SAMVIQ).

Table of Contents

List of Figures

List of Tables

1. Introduction

Streaming audio and video over the Internet is becoming more popular. This rapid expansion underlies a new challenge for maintaining the quality of service of each stream. On the other hand new mobile systems are envisioned to offer wireless services to a wide variety of mobile terminals ranging from cellular phones and Personal Digital Assistants (PDAs) to laptops. These mobile terminals are referred to as heterogeneous terminals. Heterogeneous terminals have various processing power, memory, display and data rate capabilities. So, decoding data rate and resolution of content must be adapted to the networking environment and display device (terminal). This quality is necessary when transmitting rich media over heterogeneous networks, as well as for applications where the aforementioned terminals are not capable of displaying the full resolution or full quality images.

During this project, we gave emphasis especially on MPEG4 video streams as this emerging ISO/IEC standard can be used in a wide range of networking environments (allows for streaming of very low bit rate content over all types of networks), provides scalable video encoding and makes provisions for streaming in error-prone environments. The object scalability that has been introduced in MPEG4 can be used in order to adapt video quality to network state and parameters. This solves the problems of heterogeneity of terminals and redundancy of data.

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 video client-side decoded erroneous video streams. In the final stage we propose some kind of correlation between objective quality and subjective quality schemes.

2. An Overview of Digital Video

First, we give a brief overview of digital video. Let us start with an analogue video signal generated by an analogue video camera. The analogue video signal consists of a sequence of video frames. The video frames are generated at a fixed frame rate (30 frames/s in the National Television Standards Committee, NTSC, format). For each video frame, the video camera scans the frame line by line (with 455 lines in NTSC). To obtain a digital video signal the analogue video signal is passed to a digitizer. The digitizer samples and quantizes the analogue video signal. Each sample corresponds to a picture element (pel). The most common digital frame formats are Common Intermediate Format (CIF) with 352 × 288 pels (i.e., 352 pels in the horizontal direction and 288 pels in the vertical direction), Source Intermediate Format (SIF) with 352 × 240 pels, and Quarter CIF (QCIF) with 176 × 144 pels. In all three frame formats, each video frame is divided into three components. These are the luminance component (Y), and the two chrominance components: hue (U) and intensity (saturation) (V). Since the human eye is less sensitive to the colour information than to the luminance information, the chrominance components are sampled at a lower resolution. Typically, each chrominance component is sampled at half the resolution of the luminance component in both the horizontal and vertical directions. (This is referred to as 4:1:1 chroma subsampling.) In the QCIF frame format, for instance, there are 176 × 144 luminance samples, 88 × 72 hue samples, and 88 × 72 intensity samples in each video frame, when 4:1:1 chroma subsampling is used. Finally, each sample is quantized; typically, 8 bits are used per sample. As an aside we note that the YUV video format was introduced to make color TV signals backward compatible with black-and-white TV sets, which can only display the luminance (brightness) components. Computer monitors, on the other hand, typically use the RGB video format, which contains red, green, and blue components for each pel. MPEG4 encoding standard employ discrete cosine transform (DCT) [17] to reduce the spatial redundancy in the individual video frames. Each video frame is divided into macroblocks (MBs). An MB consists of 16 × 16 samples of the luminance component and the corresponding 8 × 8 samples of the two chrominance components. The 16 × 16 samples of the luminance component are divided into four blocks of 8 × 8 samples each. The DCT is applied to each of the six blocks (i.e., four luminance blocks and two chrominance blocks) in the MB. For each block the resulting DCT coefficients are quantized using an 8 × 8 quantization matrix, which contains the quantization step size for each DCT coefficient. The quantization matrix is obtained by multiplying a base matrix by a quantization parameter. This quantization parameter is typically used to tune the video encoding. A larger quantization parameter results in coarser quantization, which in turn results in lower quality as well as smaller size (in bits) of the encoded video frame. The quantized DCT coefficients are finally variablelength-coded for a more compact representation. Both, MPEG–4 and H.263 employ predictive encoding to reduce the temporal redundancy, that is, the temporal correlation between successive video frames. A given MB is either intracoded (i.e., without reference to another frame) or interceded (i.e., with reference to a preceding or succeeding frame). To intercode a given MB, a motion search is conducted to find the best matching 16 × 16 sample area in the preceding (or succeeding) frame. The difference between the MB and the best matching area is DCT coded, quantized, variable-length-coded, and then transmitted along with a motion vector to the matching area.

3. MPEG4 Standard

3.1 General

MPEG-4 (ISO14496) is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group). The first version of the MPEG-4 standard was finalized in October 1998 and became an international standard at the beginning of 1999. Although defined as one standard, MPEG-4 is actually a set of compression/decompression formats and streaming technologies that address the need for distributing rich interactive media over narrow and broadband networks. The communication revolution triggered by the Internet, the advent of wireless devices and the promise of the Next Generation Internet (broadband Internet) underscores the importance of an international standard that defines a universal way of transmitting rich media. To this end, MPEG-4 aims to pave the way toward a uniform, high-quality streaming standard that would replace the many proprietary streaming technologies in use today.

MPEG-4 has been designed to address the following issues:

Interoperability. The standard is not specific to any one platform but is designed for all platforms.

Transport Independence. MPEG-4 leaves the choice of transport mechanism up to the service provider. This allows MPEG-4 to be used in a wide range of networking environments. During the next deliverables we will refer to some adaptive methods for the transmission of video streams mainly in wireless networks.

Compression and Transmission of Rich Media. MPEG-4 has been designed for the low and mid bit-rate compression and transmission of rich media streams.
Interactivity. MPEG-4 allows content authors and viewers to influence how they interact with a stream.
Scalability. MPEG-4 allows for flexibility in the way multimedia streams are decoded (scalable video encoding). Decoding bit rate and resolution of content is adapted to the networking environment and display device. This quality is necessary when transmitting rich media over heterogeneous networks, as well as for applications where the receiver is not capable of displaying the full resolution or full quality images. The object scalability that has been introduced in MPEG4 can be used in order to adapt video quality to network state and parameters. This solves the problems of heterogeneity of receivers and redundancy of data.
Profiles. MPEG-4 offers different technology profiles for different applications. In this way, service providers need not use the entire set of technologies, but only the sub-set that suits their applications needs.

MPEG-4 aims to achieve its objectives by applying certain principles to the way data is represented. MPEG-4 relates to the components that comprise a multimedia scene as media objects. For example, a sound track, animation, video or images are all individual media objects. Media objects can be grouped together to form compound objects. These are the building blocks of multimedia scenes. But these media objects are only one part of an MPEG-4 stream. Additional information that governs how the objects are rendered on the screen and how they are transmitted over networks is also needed. For these purposes, MPEG-4 streams include Stream Description information and Coding information. The Screen Description information describes the relation between the media objects and how they are presented. The Coding information describes how the media objects are linked to the resources that are transmitting the media objects.

MPEG-4 has several characteristics that make it the ideal standard for streaming rich media over the Internet.
a) For the narrowband Internet, applications can use content compressed at rates as low as 24 Kbit/s. For the broadband Internet, applications can use the same content encoded at higher bit rates.
b) The interactive nature of MPEG-4 means that MPEG-4 content can be used in advanced multimedia applications. Using some feedback techniques (we will be referred to them at a forthcoming deliverable) between a video streaming server and a corresponding client, a video stream can be adapted to the network parameters.
c) Because MPEG-4 allows for scalability, the same content can be streamed to different devices over heterogeneous networks.

Moreover, the MPEG-4 standard allows for streaming of very low bit rate content over all types of networks. In addition, MPEG-4 makes provisions for streaming in error-prone environments. These qualities are crucial when streaming rich content to wireless devices.

3.2 An Overview of MPEG4 Video Compression

In this section we provide a brief overview of MPEG–4 video coding. MPEG–4 provides very efficient video coding covering the range from the very low bit rates of wireless communication to bit rates and quality levels beyond high-definition television

(HDTV). In contrast to the framebased video coding of MPEG–1 and H.263, MPEG–4 is objectbased. Each scene is composed of video objects (VOs) that are coded individually. (If scene segmentation is not available or useful, e.g., in very simple wireless video communication, the standard defines the entire scene as one VO.) Each VO may have several scalability layers (i.e., one base layer and one or several enhancement layers), which are referred to as video object layers (VOLs) in MPEG-4 terminology. Each VOL in turn consists of an ordered sequence of snapshots in time, referred to as video object planes (VOPs). For each VOP the encoder processes the shape, motion, and texture characteristics. The shape information is encoded by bounding the VO with a rectangular box and then dividing the bounding box into MBs. Each MB is classified as lying:

• Inside the object
• On the object's border
• Outside the object (but inside the bounding box)

The border MBs are then shape coded. The texture coding is done on a per-block basis similar to the "frame-based standards (e.g., MPEG–1 and H.263). In an intracoded (I) VOP the absolute texture values in each MB are DCT coded. The DCT coefficients are then quantized and variable-length lengthcoded. In forward predicted (P) VOPs each MB is predicted from the closest match in the preceding I (or P) VOP using motion vectors. In bidirectionally predicted (B) VOPs each MB is predicted from the preceding I (or P) VOP and the succeeding P (or I) VOP. The prediction errors are DCT coded, quantized, and variable-length-coded. The I, P, and B VOPs are arranged in a periodic pattern referred to as a group of pictures (GoP). A typical GoP structure is IBBPBBPBBPBB. For the transmission the shape, motion, and texture information is multiplexed at the MB level; that is, for a given MB the shape information is transmitted first, then the motion information, and then the texture information, then the shape information of the next MB, and so on. To combat the frequent transmission errors typical in wireless communication, MPEG–4 provides a number of error resilience and error concealment features. We will refer to these techniques in a forthcoming deliverable.

4. Video quality assessment schemes

In recent years, video communications services such as video-streaming services and video-conferencing services have been among the most promising of multimedia communications applications. To realize high-quality video services that are comfortable to view, we need to design appropriate networks and applications, and then monitor the quality of service. To prime criterion for the video quality is subjective quality, the users' perceptions of service quality. This can be measured through subjective quality assessment. However, while subjective quality assessment is most reliable method, it is time-consuming and expensive. In particular, it is most impracticable as a method for quality monitoring in real-time. We thus need objective methods that are solely based on physical measurement but produce results comparable with those of subjective testing.

4.1 Objective quality assessment methods

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 (in this the decoded compressed video) are fed to a computer algorithm that calculates the distortion between the two as shown below.

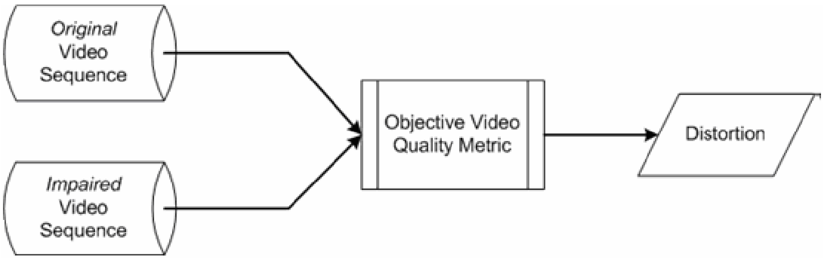


Figure 1: Comparing the decoded video sequence to the original

This model was initially applied to images and later extended to video sequences by simply applying the image quality metric in every frame of the video sequence.

Root-Mean-Square Error (RMSE)

The simplest method that compiles to this model is the Root-Mean-Square Error. It calculates the "difference" between two images. It can be applied to digital video by averaging the results for each frame.

For an MxN image, RMSE can be calculated as:

$$RMSE = \sqrt{\frac{1}{M \times N} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[f'(m,n) - f(m,n) \right]^2}$$

where f is the original image and f' the impaired.

Signal-to-Noise Ratio (SNR)

SNR is the simple mathematical advancement of RMSE and can be calculated as:

$$SNR = \frac{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[f'(m,n) \right]^2}{\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left[f'(m,n) - f(m,n) \right]^2}$$

Peak-Signal-to-Noise Ratio (PSNR)

PSNR is yet another extension based on RMSE and can be calculated as follows:

$$PSNR = 20 \times \log_{10} \left(\frac{255}{RMSE} \right)$$

PSNR is defined as the ratio between signal and RMS noise observed between the original (reference) video and the impaired (processed) video. The advantage of PSNR is that it is very easy to compute. However, PSNR does not match well to the characteristics of human visual system [2]. It does not take account human vision and thus cannot be a reliable predictor of perceived visual quality. Human observers will perceive different kinds of distortion in digital video, e.g. jerkiness (perceptual measure

of motion that does not look smooth), blockiness, blurriness and noise. These cannot be measured by PSNR.
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 [3].
Clearly, the more information about the original sequence that is needed, the harder the task is to use the method at the receiving end. The full reference methods are only possible for out-of-service situations when using predefined test video sequences. The reduced reference methods can be used if reference features can be sent to the receiver using very little bandwidth for comparison with the features calculated from the received sequence. 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 aforementioned Root Mean Square Error (RMSE) and PSNR. These metrics are only effective when error is additive, not correlated with the signal as is the case with digital video.
Objective quality assessment methods try to achieve a high correlation with subjective video quality assessment without losing the advantages of that objective quality assessment has to offer.
The ANSI objective video quality standards T1.801.03-1996 [4] as well as the metrics developed by Institute for Telecommunication Sciences (ITS) [5] capture the relationship between the measurable video quality parameters and perceptual quality distortion (blurring, tiling, noise, etc.).

4.2 Subjective quality assessment methods

Subjective video quality is a subjective characteristic of video quality. It is concerned with how video is perceived by a viewer and designates his or her opinion on a particular video sequence. Subjective measures are extracted from marks given by people, and can help to evaluate users' opinion.
There is enormous amount of ways how you can show video sequences to expert and to record his or her opinion, and a few of them have been standardized. Most of them are thoroughly described in ITU-R BT.500-11 “Methodology for the subjective assessment of the quality of television pictures” and ITU-R BT.700 as shown below.

ITU-R BT.500 Recommendation and SAMVIQ (ITU-R BT.700)

	Methodology					
Parameter	**DSIS**	**DSCQS**	**SS**	**SSCQE**	**SDSCE**	**SAMVIQ**
Explicit reference	yes	no[1]	no	no	yes	yes (uncompressed)
Hidden reference	no	yes[1]	no	no	no	yes
High anchor	no	yes[1]	no	no	no	no (hidden ref)
Low anchor	no	yes[1]	no	no	no	yes
Scale	5 grades	bad -excellent (continuous quality scale)	5 grades	bad - excellent (continuous quality scale)	bad - excellent (continuous quality scale)	bad - excellent (continuous quality scale)
Sequence length	10 s	10 s	10 s	≥5 min	10 s	10 s
Picture format	all	all	all	all	all	all
2 simultaneous stimulus	no	no	no	no	yes	no
Presentation of test material	variant I: once variant II: twice in succession	twice in succession (double stimulus)	once	once	once	as often as user likes (multi stimuli)
Videos per trial	2	2	1	1	2	max 10[2]
Voting	only test sequence	only test sequence and reference	only test sequence	only test sequence	difference between the test sequence and ref. simultaneously shown	test sequence and reference
Possibility to change the vote before proceeding	no	no	no	no	no	yes
Continuous quality evaluation	no	no	no	yes (moving slider in a continuous way)	yes (moving slider in a continuous way)	no
Minimum accepted votes	15	15	15	15	15	15
Rejection criteria		yes, but not stable				yes
Observers per display	1 to many	1 to many	1 to many	1 to many	1 to many	1
Display	all (mainly TV)	all (mainly TV, DLP)	all (mainly TV)	all (mainly TV)	all (mainly TV)	all (mainly PC, PDA)
Quality results	relative, depending on reference quality	relative, depending on compared sequence	relative	relative	relative, depending on ref quality	absolute measure of video quality
Standard	ITU-R BT.500-11	ITU-R BT.500-11	ITU-R BT.500-11	ITU-R BT.500-11	ITU-R BT.500-11	ITU-R SG6 WP 6Q (new: ITU-R BT.700)

[1] not mandatory (could be any test sequence), [2] different bit rates in one trial to avoid contextual effects

DSIS: Double Stimulus Impairment Scale (≈ Degradation Category Rating DCR in ITU-T P.910)
DSCQS: Double Stimulus Continuous Quality Scale
SS: Single Stimulus (≈ Absolute Category Rating ACR in ITU-T P.910)
SSCQE: Single Stimulus Continuous Quality Evaluation
SDSCE: Simultaneous Double Stimulus for Continuous Evaluation
SAMVIQ: Subjective Assessment Methodology for VIdeo Quality

Figure 2: ITU-R BT.500 Recommendation and SAMVIQ (ITU-R BT.700)

Some of these methods mentioned in Figure 2 are described in detail below [6].

4.2.1 Double Stimulus Continuous Quality Scale (DSCQS) Method

In the DSCQS method, each trial consists of a pair of stimuli: one stimulus is the reference, and the other is the test. The test stimulus is usually the reference after undergoing some type of processing. The two stimuli are each presented twice in a trial, in alternating fashion, with the order of the two randomly chosen for each trial.

Test subjects are not informed of the ordering of the test and reference stimuli, and they rate each stimulus by marking a continuous quality scale. Thus, two ratings are made for each trial in the DSCQS method: one for the reference and the other for the test condition. An example of the rating scales for one DSCQS trial is given in Figure 3. Occasionally in a DSCQS test, both stimuli presented in a trial are the reference stimulus. Such trials are used to detect erratic test subject behaviour.

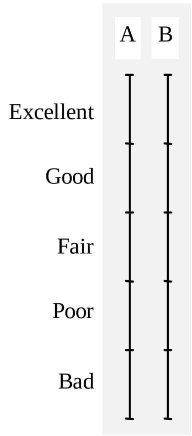


Figure 3: Rating scales for a trial with the DSCQS method.

The DSCQS method is typically applied for evaluations where the quality difference between the test and reference sequences is not too large.

4.2.2 Double Stimulus Impairment Scale (DSIS) Method

As in the DSCQS method, each trial consists of a pair of stimuli: the reference and the test. However, in the DSIS method, the two stimuli are always presented in the same order: the reference is always first, followed by the test.

In the DSIS method, test subjects compare the two stimuli in a trial and rate the impairment of the test stimulus with respect to the reference, using a five-level degradation scale. Thus, only one vote is made for each DSIS trial. The rating scale used for one DSIS trial in the MPEG-4 tests is shown in Figure 4.

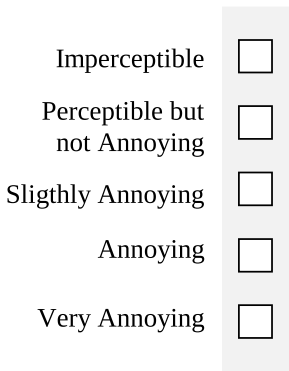


Figure 4: Rating scales for a trial with the DSIS method.

The DSIS method is typically applied for evaluating the annoyance of video impairments; so it is primarily suited for evaluating the performance of systems that introduce clearly visible impairments. For this reason it was applied for evaluating the performance of proposals at low and medium bit rates.

Since the evaluation in the DSIS method is made with respect to a reference, it is important that the quality level of the test stimulus and the corresponding reference are not too different, or the usefulness of the reference could be greatly reduced.

4.2.3 Single Stimulus (SS) Method

In the SS method, only one stimulus is presented in each trial. The test subject rates each stimulus, typically using a five-level quality scale (i.e. Excellent, Good, Fair, Poor, Bad).

The SS method is appropriate when references are not available. An example of the rating scales for one SS trial is given in Figure 5.

EXCELLENT	☐
GOOD	☐
FAIR	☐
POOR	☐
BAD	☐

Figure 5: Rating scales for a trial with the SS method.

Since explicit references are not used in SS methods, context dependency (the effect of previously seen trials on the rating given to the current trial) is stronger than for test methods which use an explicit reference. To compensate for this effect, each SS test was performed twice, with two different trial presentation orders.

4.2.4 Subjective Assessment Method for Video Quality Evaluation (SAMVIQ)

There are multiple differences between BT.500 (all the aforementioned methods are included in BT.500) and SAMVIQ. A major difference 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.

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). There is no continuous sequential presentation of the sequences as in the DSCQS method, where the viewer can make errors of judgement due to a lack of concentration. As a result, SAMVIQ offers higher reliability, i.e. smaller standard deviations. Viewers are generally able to discriminate the different quality levels better with SAMVIQ than with BT.500.

In SAMVIQ there is only one viewer at a time, which alleviates a "group effect". A hidden reference is mandatory. The explicit reference is uncompressed which allows the viewer to determine near-absolute measure of video quality. The SAMVIQ method provides an overall quality score for relatively short multimedia sequences. The duration of a sequence is typically in the range of 10 to 15 seconds in order to give the subject sufficient time to formulate a stable grading. The content of a sequence has to be homogeneous. A large quality range is required to stabilize the viewers' quality scores; otherwise, when the quality range is reduced, viewers try to discriminate among the quality of the sequences even if the differences are not perceptible. Therefore, the reliability of results decreases, as the quality of the codecs tested is similar.An example of the rating scales for one SS trial is given in Figure 6.

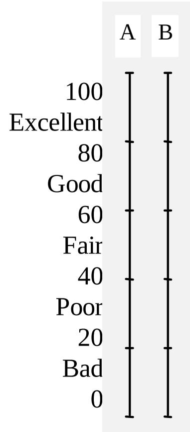


Figure 6: Rating scales for a trial with the SAMVIQ method.

The viewers are asked to assess the overall picture quality of each presentation by inserting a slider mark on a vertical scale. The scales provide a continuous rating system to avoid quantizing errors, but they are divided into five equal lengths which correspond to the normal ITU–R BT.500 five point quality scale. The associated terms categorizing the different levels are the same as those normally used; but here they are included for general guidance. The grading scale is continuous and is divided in five equal portions, as follows:
Excellent (80 to 100)
Good (60 to 80)
Fair (40 to 60)
Poor (20 to 40)
Bad (00 to 20)
The lowest quality perceived should be scored "0" (bottom of the scale) and the highest quality should be marked "100" (top of the scale).

5. Experiments

5.1 Test Sequences

The specification of the test sequences *Foreman* and *Claire* are in table 1. The two test sequences were encoded in MPEG4 format with a free software tool called FFMPEG encoder [7]. The *Foreman* sequence shows a man speaking and showing with his hands a house under construction. The *Claire* sequence shows a woman speaking. There is no moving background in the second sequence.

samples	*Spatial Resolution*	*# of frames*	*Bitrate (Kbps)*	*# of I frames*	*# of P frames*	*# of B frames*	*Mean PSNR (dB)*	*Stdv PSNR (dB)*
foreman	176x144	400	64	39	95	266	30.87	2.29
			256	34	100	266	35.91	1.84
			768	34	100	266	40.53	1.36
claire	176x144	494	64	42	124	328	41.85	1.33
			256	42	124	328	45.50	0.63
			768	42	124	328	45.50	0.63

Table 1: Technical data on the encoded sequences.

The two sequences have temporal resolution 30 frames per second, and GoP (Group of Pictures) pattern IBBPBBPBBPBB. Each sequence was encoded at low, medium and high quality video streams which correspond to 64Kbps, 256Kbps and 768Kbps samples respectively.
As can be seen from Table 1, as far as the PSNR values are concerned, the *Cliare* sequence has higher mean value as well as lower standard deviation for all bitrates.

5.2 Topologies – Scenarios

The simulations in NS2 have been carried out for the topology as shown in Figure 7.

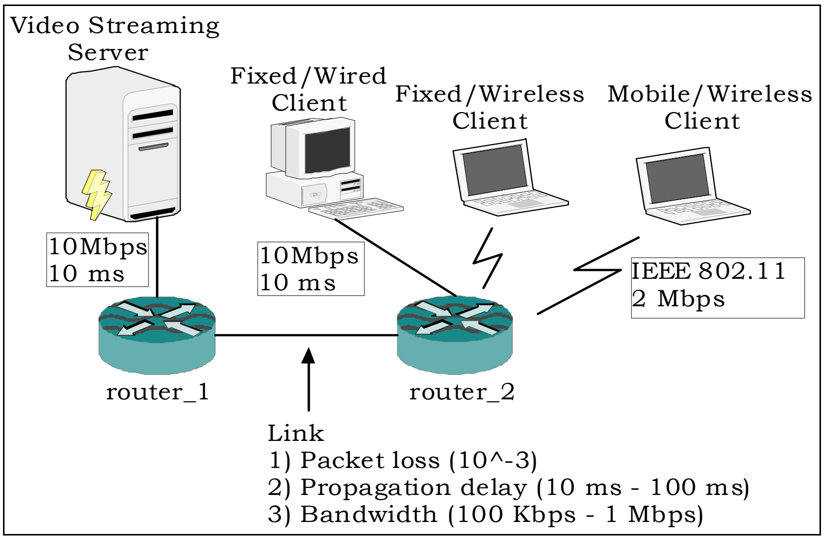


Figure 7: Topology used in the first scenario.

This topology consists of a Video Streaming Server, 2 backbone routers and some (fixed/mobile, wired/wireless) video clients. Video Streaming Server is attached to the first backbone router with a link which has 10Mbps bandwidth and 10ms propagation delay. These values will remain unchanged during all scenarios. This router is connected to a second router using a link with unspecified bandwidth and propagation delay. The bandwidth will vary from 100Kbps up to 1Mbps. Also propagation delay will vary from 10ms up to 100ms. The packet loss probability will be constant 0.001.
Using this topology, we conducted different scenarios with fixed wired clients, fixed wireless clients and mobile wireless clients.
An additional topology is illustrated in Figure 8.

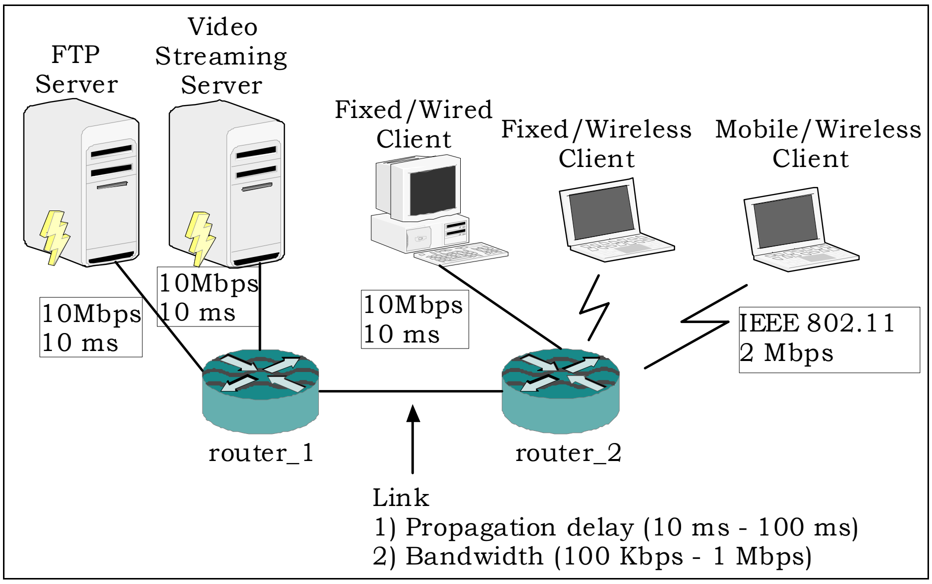


Figure 8: Topology used in the second scenario.

In addition to the previous topology we present another topology used in the second scenario which is almost the same as the previous one but incorporates an extra node that acts as an FTP Server. This server – which is attached to the first router – is used to provide some kind of background bursty traffic.

5.3 Data Collection

All the aforementioned experiments were conducted with an open source network simulator tool NS2.

Based on the open source framework called EvalVid [8] we were able to collect all the necessary information needed for the objective video quality evaluation like PSNR values, frames lost, packet end to end delay and packet jitter.

Some new functionalities were implemented in NS2 from [9] in order to support EvalVid. The whole data collection procedure and PSNR evaluation is illustrated below.

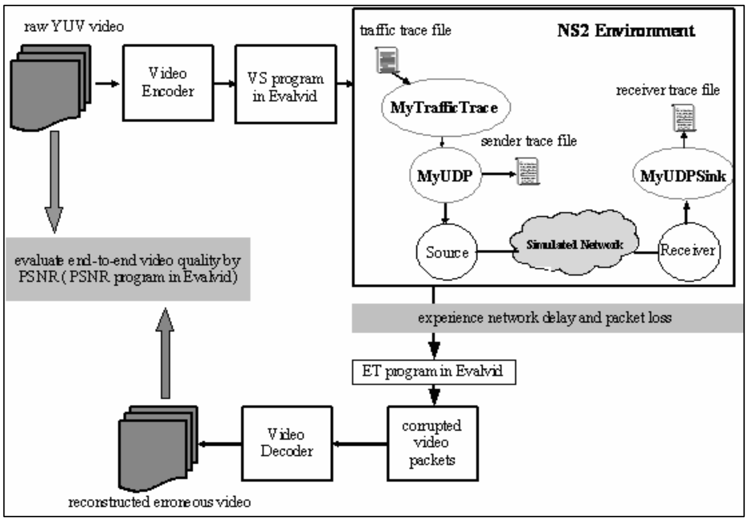


Figure 9: Data collection and PSNR evaluation using EvalVid.

6. Results

6.1 First Scenario

In this first scenario we used the two sequences (*Foreman* and *Claire*) coded in MPEG4 with bit-rates 64Kbps, 256Kbps and 768Kbps so as to we have streams of different qualities. Also we had variable bandwidth (100Kbps, 500Kbps, 1Mbps) as well as variable propagation delay (10ms, 50ms, 100ms) on the link that connects the two routers. We considered also packet loss probability equal to 10^{-3} on the same link. All the variables of this scenario are presented in following table.

Bit-rate	*Bandwidth*	*Propagation delay*
64 Kbps	100 Kbps	10 ms
256 Kbps	500 Kbps	50 ms
768 Kbps	1 Mbps	100 ms

Table 2: Variable values of the first scenario.

6.1.1 64Kbps

Figure 10 is a sample from a set of initial experiments where the test sequence is *Foreman*. In this script, the bandwidth of the link in question varies from 100Kbps up to 1Mbps and the propagation delay varies from 10ms to 100ms. This scenario involves a wired fixed client which is attached on the second router. Client streams video encoded at 64Kbps from the video streaming server which is attached to the first router. The first graph illustrates the packet end to end delay.

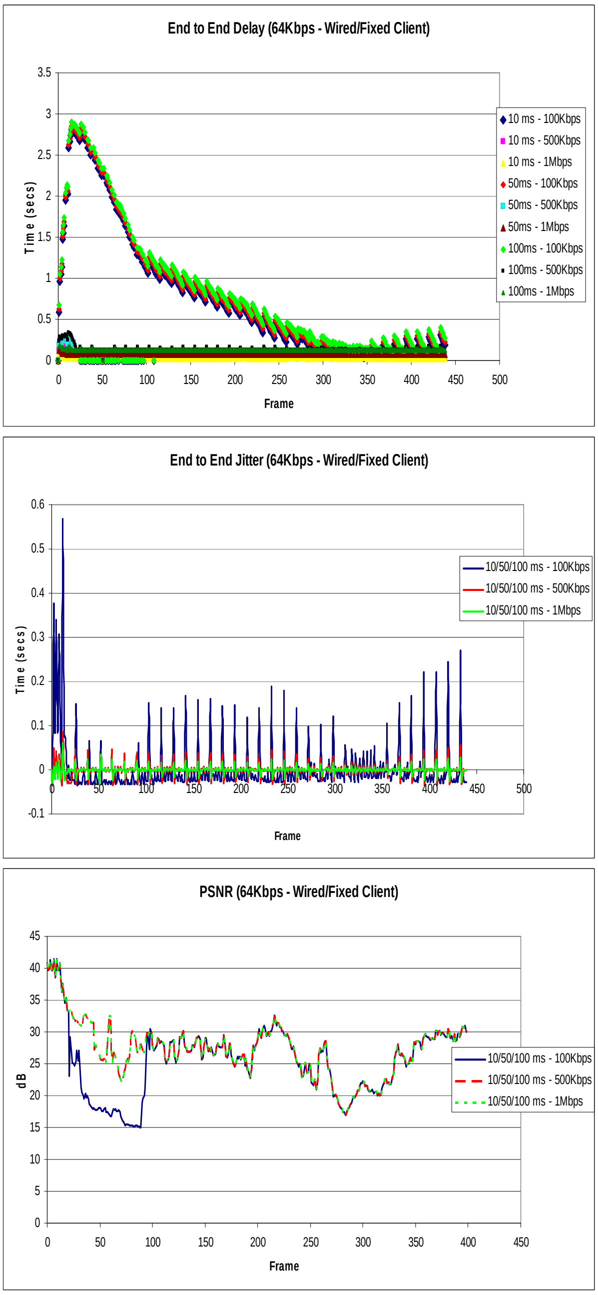


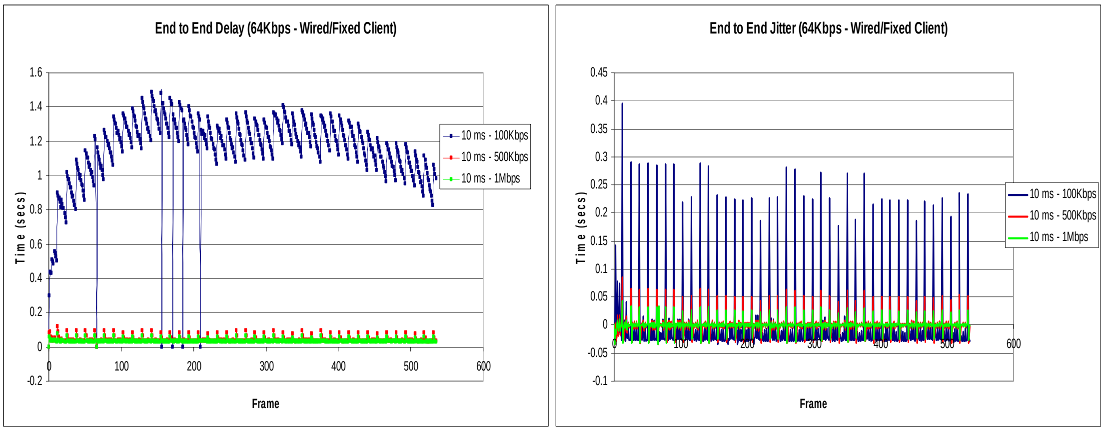
Figure 10: End to End Delay, End to End Jitter and PSNR respectively. Scenario of a fixed wired client who streams 64Kbps encoded video (Foreman).

As can be seen, there is an initial exponential increase of the delay concerning 100Kbps link bandwidth. This observation stands for all values of propagation delay. However, there is a slight difference between the different scenarios of propagation delay. The higher the propagation delay, the higher the packet end to end delay and this was absolutely expected. This exponential rise of the per packet end to end delay is due to fact that at the beginning of the encoded (MPEG4) video sequence there are some large frames (intracoded (I) variable-length VOPs). These frames are inserted in the first router's queue before they are able to be transmitted because the link between the 2 routers can transfer only 100Kbps. The packets that wait for a long time in this queue have greater end to end delay. Some of the packets are lost because the queue is full. It can be seen that this issue does not apply to the other cases (500Kbps and 1Mbps) because the link in question is capable of transferring larger amount of data and the queue in the first router will never overflow.

The second graph depicts the packet end to end jitter. The first observation is that jitter is independent of the propagation delay. Long-range density fluctuations of unexpected magnitude are observed in case we have bandwidth of 100Kbps because of the fact that each individual packet of the stream faces different end to end delay. Again this observation does not apply to the other cases (500Kbps and 1Mbps).

The third graph shows the PSNR values. Once again we see that these values are independent of the propagation delay. Furthermore for 500Kbps and 1Mbps link bandwidth the corresponding values range from 20dB to 35dB. That means that we have accepted video quality whereas in case of 100Kbps there was a deep decrease concerning the initial values of PSNR due to high packet loss (queue overflow). Once one or more packets are lost, the video quality considerably deteriorate in the video with smaller PSNR values.

The following Figure 11 corresponds to the same scenario that involves the second video sequence *Claire*.



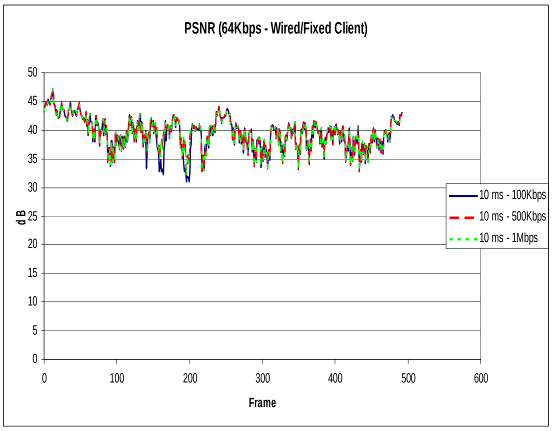


Figure 11: End to End Delay, End to End Jitter and PSNR respectively. Scenario of a fixed wired client who streams 64Kbps encoded video(Claire).

The results demonstrate that there are some differences between this and the previous scenarios. First of all the end to end delay is longer than before as far as 100Kbps link bandwidth is concerned. The main reason for this attitude is that at the beginning of this encoded video sequence there are no large frames as before. Instead, there are larger intracoded VOPs through all the duration of the video. The variation of end to end delay is due to the fact that whenever an I frame is transmitted, all the other packets have to wait longer in the buffer.

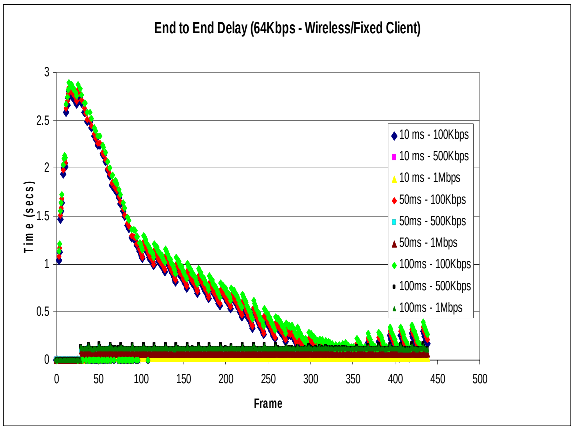
The second graph depicts the packet end to end jitter. The first observation is that jitter is independent of the propagation delay. Long-range density fluctuations of unexpected magnitude are observed in case we have bandwidth of 100Kbps because of the fact that each individual packet of the stream faces different end to end delay.

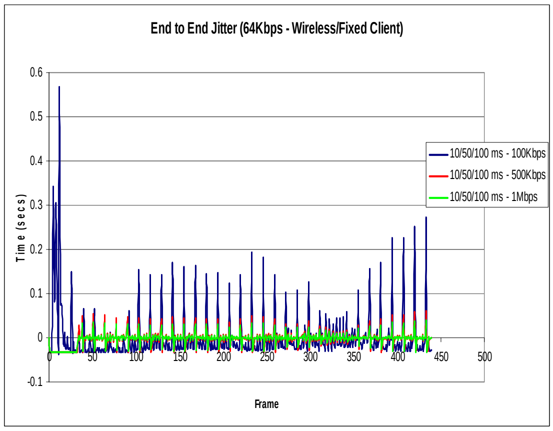
Finally the third graph indicates that in video sequences that contain less motion (like *Claire*), PSNR values are higher compared with video sequences that contain more motion (like *Foreman*).

Bandwidth	*Propagation delay*	*lost I frames percentage foreman/claire*	*lost P frames percentage foreman/claire*	*lost B frames percentage foreman/claire*
100Kbps	10ms	15.4% / 0.0%	9.5% / 1.6%	9.0% / 0.9%
	50ms	15.4%	9.5%	9.0%
	100ms	15.4%	9.5%	9.0%
500Kbps	10ms	2.6% / 0.0%	0.0% / 0.0%	0.0% / 0.3%
	50ms	2.6%	0.0%	0.0%
	100ms	2.6%	0.0%	0.0%
1Mbps	10ms	2.6% / 0.0%	0.0% / 0.0%	0.0% / 0.3%
	50ms	2.6%	0.0%	0.0%
	100ms	2.6%	0.0%	0.0%

Table 3: Percentage of lost I,P,B frames concerning scenarios with fixed wired client (64Kbps video stream).

Table 3 reveals that there is lower packet loss probability in scenarios that contain video sequences with less motion. Furthermore, we observe that the higher the link bandwidth, the lower the packet loss probability.
Except for these, it is obvious that intracoded VOPs (I frames) have higher loss probability than forward predicted VOPs (P frames) and bidirectionally predicted VOPs (B frames). This is because I frames have larger size compared with P and B frames.





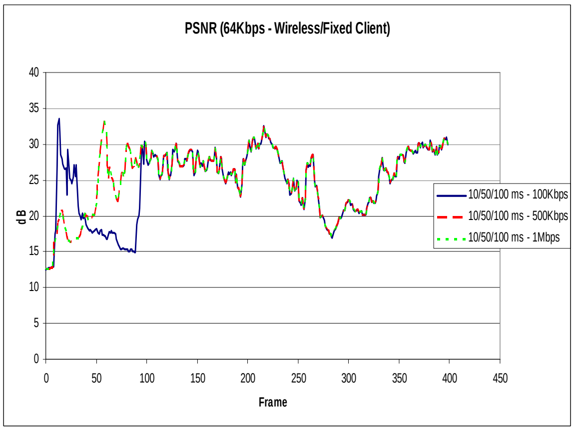


Figure 12: End to End Delay, End to End Jitter and PSNR respectively. Scenario of a fixed wireless client who streams 64Kbps encoded video (Foreman).

Figure 12 illustrates end to end delay, jitter and PSNR for a fixed wireless client who streams 64Kbps encoded video. As far as the end to end delay and jitter are concerned the results resemble that of wired fixed client scenarios. On the other hand the graph that depicts the PSNR values reveals that at the beginning of the video there is a sharp increase followed by a steep decrease in case we have 100Kbps link bandwidth whereas a step-wise variation occurred in case of 500s and 1Mbps. This is due to packet loss that occurs because of buffer overflow and can be seen in Table 4.
The graphs concerning *Claire* are similar with these of wired fixed client.

Bandwidth	*Propagation delay*	*lost I frames percentage foreman/claire*	*lost P frames percentage foreman/claire*	*lost B frames percentage foreman/claire*
100Kbps	10ms	17.9% / 4.8%	10.5% / 3.2%	9.0% / 3.4%
	50ms	17.9%	11.6%	8.6%
	100ms	17.9%	10.5%	9.0%
500Kbps	10ms	7.7% / 7.1%	8.4% / 5.6%	7.1% / 5.5%
	50ms	7.7%	7.4%	6.4%
	100ms	7.7%	7.4%	6.0%
1Mbps	10ms	7.7% / 7.1%	8.4% / 6.5%	6.8% / 5.5%
	50ms	7.7%	7.4%	6.4%
	100ms	7.7%	7.4%	6.0%

Table 4: Percentage of lost I,P,B frames concerning scenarios with fixed wireless client (64Kbps video stream).

It is obvious that the packet loss probabilities are higher compared with scenarios with fixed wired client.

The graphs shown above as well as packet loss probabilities concerning scenarios that involve one mobile wireless client are identical to those that involve one fixed wireless client, so they can be omitted.

From the observations of scenarios that involve streams of 64Kbps we can conclude that there are no essential differences between the three different cases of clients (wired/fixed, wireless/fixed, wireless/mobile) with regard to end to end delay and jitter. As far as PSNR values are concerned there are certain differences that are due to the increase of packet loss probability in case of wireless users. Also this is obvious from the percentages of lost I,P,B frames.

6.1.2 256Kbps

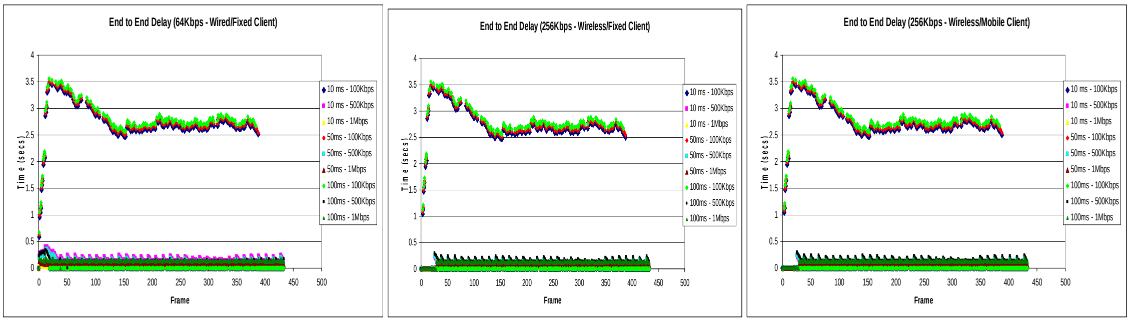


Figure 13: End to End Delay. Streaming of 256Kbps encoded video (Foreman).

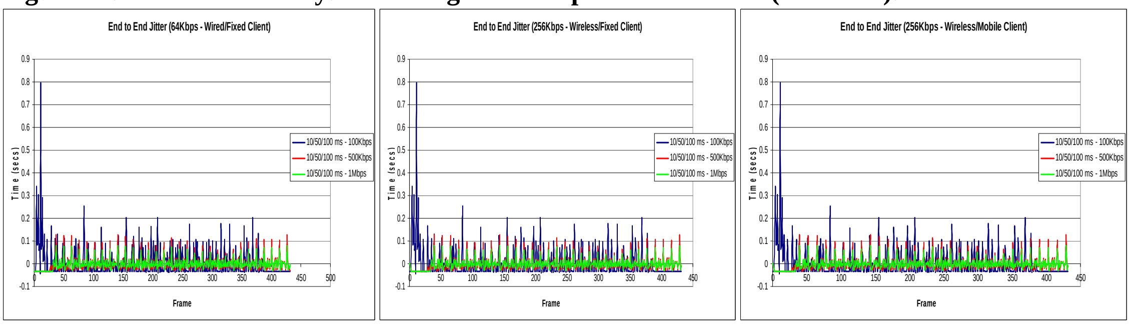


Figure 14: End to End Jitter. Streaming of 256Kbps encoded video (Foreman).

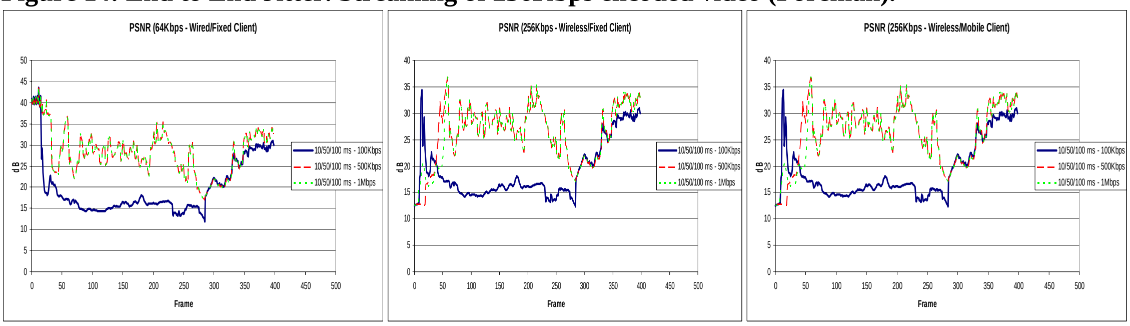


Figure 15: PSNR. Streaming of 256Kbps encoded video (Foreman).

In this section we compare the end to end delay, jitter and PSNR values for all scenarios. The first scenario involves a wired fixed client, the second one involves a wireless fixed client and the final involves a wireless mobile client. Each of them is attached on the second router. Client streams video (*Foreman*) encoded at 256Kbps from the video streaming server which is attached to the first router. Figure 13 illustrates the end to end delay and Figure 14 depicts the end to end jitter for every packet concerning the aforementioned scenarios. As can be seen the three graphs in each figure are identical.

However, there are some differences compared with the previous section which involves 64Kbps streams. Firstly, there are certain differences concerning end to end delay in case of bandwidth 100Kbps. At the beginning there is a steep increase of end to end delay as in 64Kbps scenarios but there is no gradual decrease as before. Instead of this, there is slight decrease of the delay but afterwards it remains constant around 2.5 seconds. This is because larger amount of data are transmitted and the buffer at the first router never empties. Therefore more packets have to wait in the queue. Secondly, we can see that the end to end delay in case of 500Kbps and 1Mbps is slightly longer compared with 64Kbps scenarios.

Figure 15 shows the PSNR values for all scenarios. We observe that the graphs for wireless clients are identical but they differ compared with wired fixed client. This is because in the latter case the packet loss probability is lower. This is obvious from the two tables that are presented below.

Bandwidth	*Propagation delay*	*lost I frames percentage foreman/claire*	*lost P frames percentage foreman/claire*	*lost B frames percentage foreman/claire*
100Kbps	10ms	94.1% / 85.7%	90.0% / 54.0%	47.4% / 34.5%
	50ms	94.1%	90.0%	47.4%
	100ms	94.1%	90.0%	47.4%
500Kbps	10ms	2.9 % / 0.0%	0.0 % / 0.8%	0.0 % / 0.0%
	50ms	2.9 %	0.0%	0.0 %
	100ms	2.9 %	0.0 %	0.0 %
1Mbps	10ms	2.9% / 0.0%	0.0 % / 0.8%	0.0 % / 0.0%
	50ms	2.9%	0.0 %	0.0 %
	100ms	2.9%	0.0 %	0.0 %

Table 5: Percentage of lost I,P,B frames concerning scenarios with fixed wireless client (256Kbps video stream).

Bandwidth	*Propagation delay*	*lost I frames percentage foreman/claire*	*lost P frames percentage foreman/claire*	*lost B frames percentage foreman/claire*
100Kbps	10ms	97.1% / 95.2%	91.0% / 43.5%	47.4% / 38.1%
	50ms	97.1%	91.0%	47.4%
	100ms	97.1%	91.0%	47.7%
500Kbps	10ms	8.8% / 7.1%	6.0% / 6.5%	5.6% / 5.5%
	50ms	8.8%	6.0%	5.3%
	100ms	8.8%	6.0%	5.3%
1Mbps	10ms	8.8% / 7.1%	8.0% / 6.5%	6.8% / 5.5%
	50ms	8.8%	7.0%	6.4%
	100ms	8.8%	7.0%	6.0%

Table 6: Percentage of lost I,P,B frames concerning scenarios with fixed and mobile wireless client (256Kbps video stream).

From the observations that involve streams of 256Kbps we can say that are in effect almost the same conclusions that we had in case of 64Kbps. The main difference in case of 256Kbps is that there is higher packet loss probability and hence reduction of received video quality in the scenarios that involves link bandwidth of 100Kbps. This is because the transmission rate exceeds the offered bandwidth. In scenarios which involve links of 500Kbps and 1Mbps there is no such problem and also the results are identical.

6.1.3 768Kbps

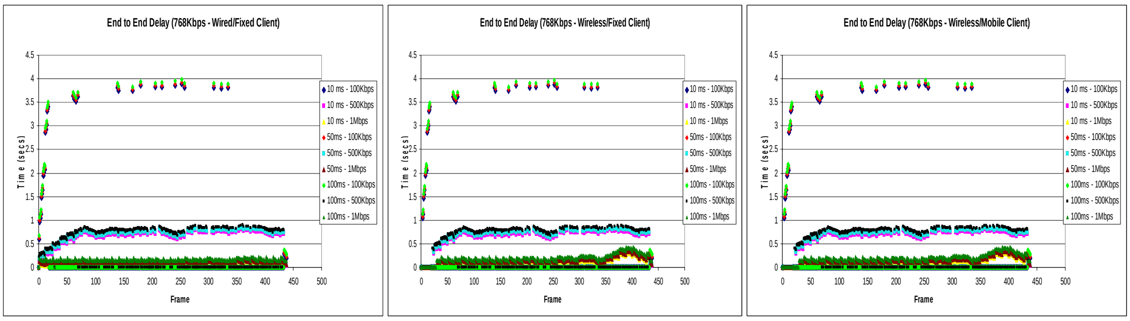


Figure 16: End to End Delay. Streaming of 768Kbps encoded video (Foreman).

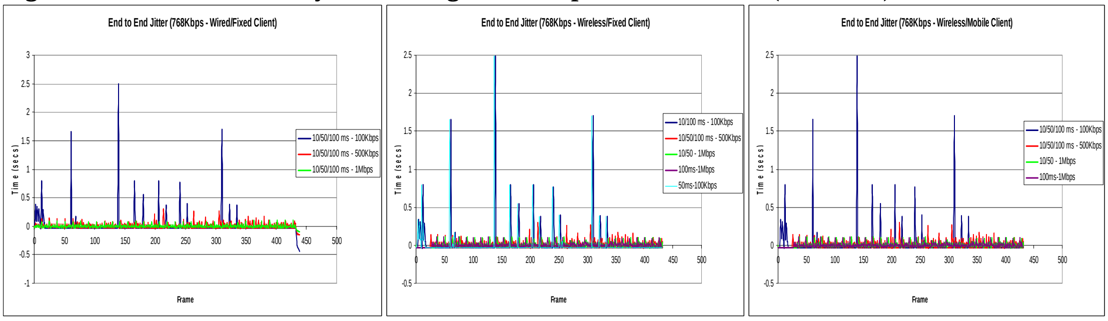


Figure 17: End to End Jitter. Streaming of 768Kbps encoded video (Foreman).

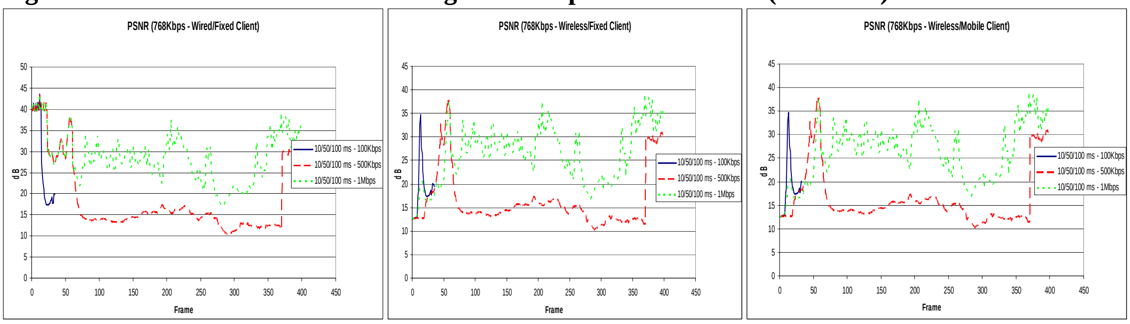


Figure 18: PSNR. Streaming of 768Kbps encoded video (Foreman).

In this section we compare the end to end delay, jitter and PSNR values for all scenarios. The first scenario involves a wired fixed client, the second one involves a wireless fixed client and the final involves a wireless mobile client. Each of them is attached on the second router. Client streams video (*Foreman*) encoded at 768Kbps from the video streaming server which is attached to the first router. Figure 16 illustrates the end to end delay and Figure 17 depicts the end to end jitter for every packet concerning the aforementioned scenarios. As can be seen the three graphs in each figure are identical.

However, there are some differences compared with the previous sections which involve 64Kbps and 256Kbps streams. Firstly, there are certain differences concerning end to end delay in case of bandwidth 100Kbps. At the beginning there is a steep increase of end to end delay as in 64Kbps and 256Kbps scenarios but there is no gradual decrease as in 64Kbps scenarios not even slight decrease of the delay as in 256Kbps scenarios afterwards. Moreover large amount of data is lost (see gaps). This is because larger amount of data are transmitted and the buffer overflows as the link is unable to support the stream rate. Finally only a small number of packets manages to be

transmitted over the link. Secondly, we can see that the end to end delay in case of 500Kbps and 1Mbps is much longer compared with 64Kbps and 256Kbps scenarios.
Figure 18 shows the PSNR values for all scenarios. We observe that the graphs for wireless clients are identical but they differ compared with wired fixed client. This is because in the latter case the packet loss probability is lower. Furthermore the EvalVid tool cannot calculate all the values of PSNR during the 100Kbps scenarios due to high percentage of lost packets. This is obvious from the two tables that are presented below.

Bandwidth	*Propagation delay*	*lost I frames percentage foreman/claire*	*lost P frames percentage foreman/claire*	*lost B frames percentage foreman/claire*
100Kbps	10ms	94.1% / 85.7%	96.0% / 54.0%	92.9% / 34.5%
	50ms	94.1%	96.0%	92.9%
	100ms	94.1%	96.0%	92.9%
500Kbps	10ms	82.4% / 0.0%	75.0% / 0.8%	20.3% / 0.0%
	50ms	82.4%	75.0%	20.3%
	100ms	82.4%	75.0%	20.3%
1Mbps	10ms	0.0% / 0.0%	1.0% / 0.8%	0.0% / 0.0%
	50ms	0.0%	1.0%	0.0%
	100ms	0.0%	1.0%	0.0%

Table 7: Percentage of lost I,P,B frames concerning scenarios with fixed wireless client (768Kbps video stream).

Bandwidth	*Propagation delay*	*lost I frames percentage foreman/claire*	*lost P frames percentage foreman/claire*	*lost B frames percentage foreman/claire*
100Kbps	10ms	97.1% / 95.2%	97.0% / 43.5%	92.9% / 38.1%
	50ms	97.1%	97.0%	92.9%
	100ms	97.1%	97.0%	92.9%
500Kbps	10ms	91.2% / 7.1%	80.0% / 6.5%	25.9% / 5.5%
	50ms	91.2%	80.0%	25.2%
	100ms	91.2%	80.0%	25.2%
1Mbps	10ms	8.8% / 7.1%	8.0% / 6.5%	7.5% / 5.5%
	50ms	8.8%	7.0%	7.1%
	100ms	8.8%	7.0%	6.0%

Table 8: Percentage of lost I,P,B frames concerning scenarios with fixed and mobile wireless client (768Kbps video stream).

From the observations that involve streams of 768Kbps we can conclude that are in effect almost the same conclusions that we had in case of 64Kbps and 256Kbps. The main difference in case of 768Kbps is that there is higher packet loss probability and hence reduction of received video quality in the scenarios that involves link bandwidth of 100Kbps and 500Kbps. This is because the transmission rate exceeds the offered bandwidth. In scenarios which involve link of 1Mbps there is no such problem and also the results are identical.
In conclusion we realized that user perceived quality is the same for a fixed wireless client and a mobile wireless client. In order to find out if this happens while having more than one client in the same network, we repeated some of the aforementioned scenarios. These scenarios are presented in the next section.

6.1.4 Scenarios with 2 clients

In order to observe if there are some essential differences (increase or reduction in end to end delay and jitter as well as PSNR) in cases where there are more than one client in the network, we conducted some simulations that involve (a) two mobile wireless clients, (b) a mobile and a fixed wireless client as well as (c) a mobile with a fixed wired client. Our main aim was to compare these scenarios with previous ones that involve fixed wireless client. During these scenarios we maintained constant bandwidth (1Mbps) and propagation delay (10ms) on the link that connects the two routers. Also we considered that the packet loss probability over the link was 10^{-3}.

Figures below illustrate the results taken by the aforementioned scenarios for 64Kbps and 256Kbps video. We can conclude that there is no differentiation in the metrics shown below with regard to the scenarios that involve video of 64Kbps but while having the video transmission rate increased (256Kbps), different phenomena is observed. The following graphs portray the end to end delay and PSNR in the wireless mobile client side when he coexists with other (mobile or fixed) clients.

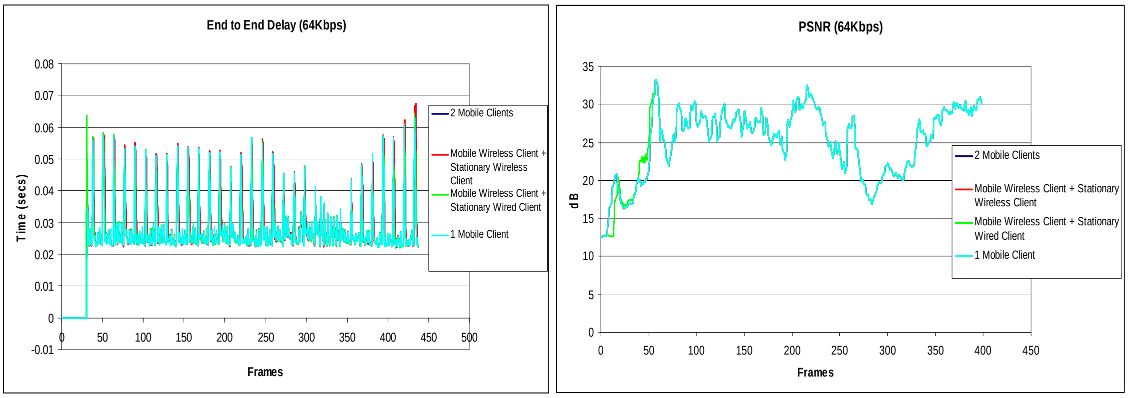


Figure 19: End to End Delay and PSNR in case of 64Kbps video while having more than one clients.

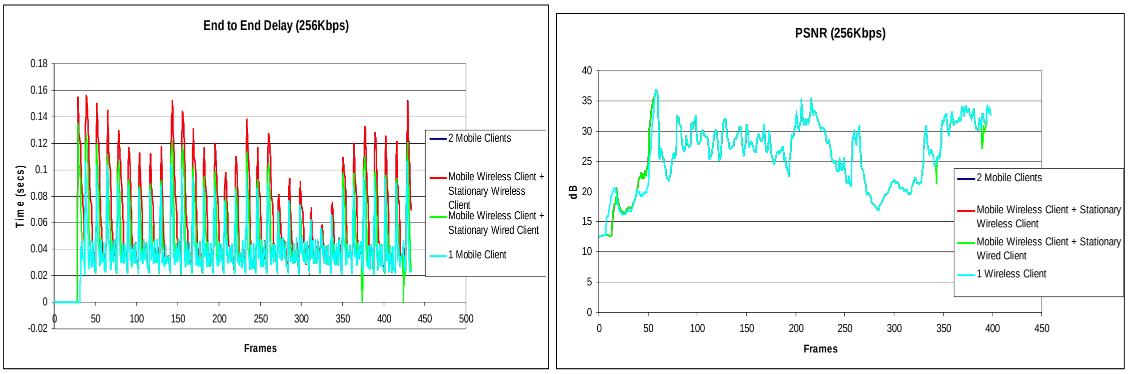


Figure 20: End to End Delay and PSNR in case of 256Kbps video while having more than one clients.

In scenarios that involve 2 mobile wireless clients as well as a mobile wireless client and a wireless fixed client, end to end delay is identical and longer than having a wireless client with a wired fixed client. This is because in the former situation packets are transmitted exclusively over the wireless channel so there is queuing delay in the second router/base station whereas in the latter situation packets follow different channels.

Furthermore, it is obvious that there are no fundamental differences between these scenarios as far as PSNR is concerned.

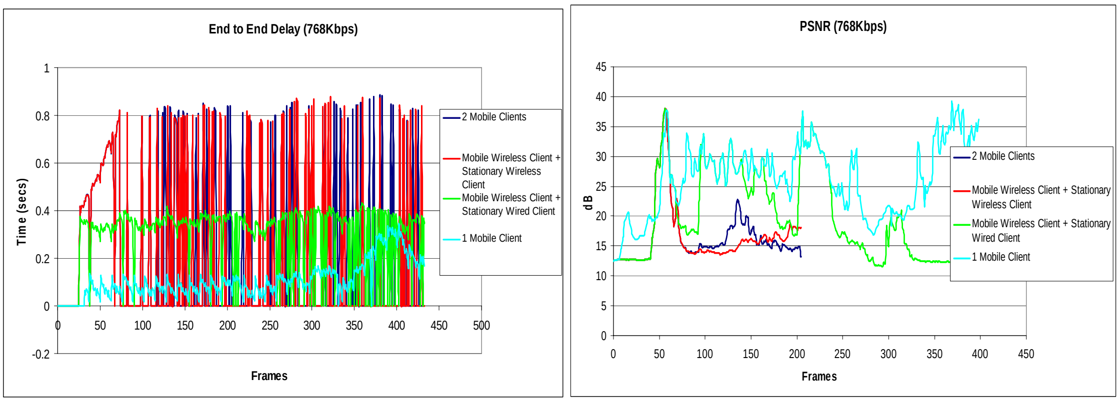


Figure 21: End to End Delay and PSNR in case of 768Kbps video while having more than one clients.

Figure 21 portrays end to end delay and PSNR in case of 768Kbps video. We observe important differences between the 4 different cases. Scenarios that involve 2 mobile wireless client as well as scenarios that involve a mobile with a fixed wireless clients, end to end delay is comparatively longer (0.8 secs.) and this is due the fact that the aggregated video transmission rate exceeds the link bandwidth so many packets have to wait in the queue for a long time (congestion occurs). Also large amount of packets is transmitted over the wireless channel while that does not happen in case where we have a mobile wireless user with a fixed wired user. In the latter case end to end delay is perceptible smaller (0.4 secs.) but longer compared with the corresponding delay in 256Kbps scenarios. In scenario that only a mobile wireless user exists, end to end delay is in acceptable levels (there is no congestion in the network).

As far as PSNR is concerned, we can see that in the first two scenarios the corresponding values range from13dB to 23dB. In scenarios that we have a (mobile) wireless client and a (fixed) wired client the corresponding values are perceptibly improved. In one's mobile wireless client scenario, PSNR implies acceptable video quality (25-40dB).

	Scenarios	*lost I frames percentage*	*lost P frames percentage*	*lost B frames percentage*
64Kbps	2 Mobile Wireless Clients	7.7%	8.4%	6.8%
	Mobile Wireless + Fixed Wireless	7.7%	8.4%	6.8%
	Mobile Wireless + Fixed Wired	7.7%	7.4%	6.8%
	1 Mobile Client	7.7%	8.4%	6.8%
256Kbps	2 Mobile Wireless Clients	8.8%	7.0%	6.0%
	Mobile Wireless + Fixed Wireless	8.8%	7.0%	6.0%
	Mobile Wireless + Fixed Wired	8.8%	8.0%	6.4%
	1 Mobile Client	8.8%	8.0%	6.8%
768Kbps	2 Mobile Wireless Clients	91.2%	82.0%	65.8%
	Mobile Wireless + Fixed Wireless	94.1%	82.0%	65.4%
	Mobile Wireless + Fixed Wired	61.8%	32.0%	6.8%
	1 Mobile Client	8.8%	8.0%	7.5%

Table 9: Percentage of lost I,P,B frames.

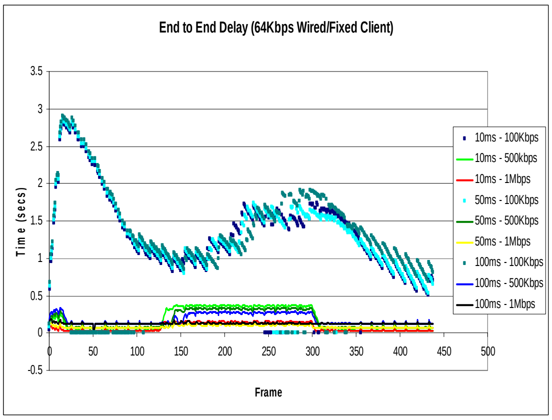
Table 9 presents the percentage of lost I,B,P frames for all the aforementioned scenarios in mobile wireless client's side. In 64Kbps and 256Kbps scenarios there is a relatively small number of lost frames compared with 768Kbps scenarios. This is due to the fact that the bandwidth of the link exceeds the aggregate video transmission rate so we have low packet drop rate (the queue never overflows). On the other hand in 768Kbps scenarios important differences emerge. Firstly the aggregate video transmission rate exceeds the offered link bandwidth (high packet drop rate) and secondly all the packets are transmitted through the wireless channel (in scenarios concerning 2 wireless clients).

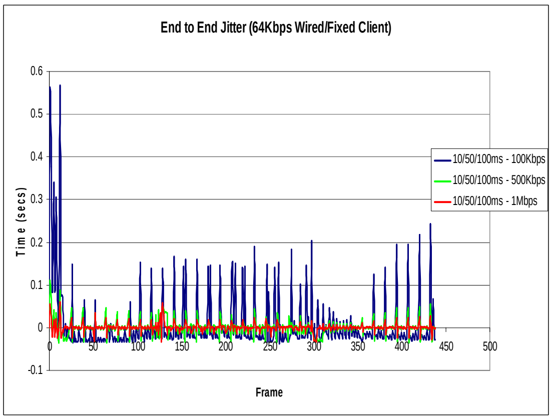
6.2 Second Scenario

In this second scenario we used only *Foreman* video sequence encoded at 64Kbps and 256Kbps. The main difference is that a new node has been added in our experimental network called FTP Server. Its main role is to provide some kind of bursty background traffic. In the following scenarios client streams video from Video Streaming Server and downloads a file from FTP Server at the same time. Video lasts for 14 seconds while FTP downloading starts from 4 second and stops at 10 second.
Needless to say that (a) FTP application runs over TCP Reno protocol and (b) there is no priority queuing policy in any of the 2 routers (droptail queue is used).

6.2.1 64Kbps

Figure 22 is a sample from a set of initial experiments where the test sequence is *Foreman*. In this script, the bandwidth of the link in question varies from 100Kbps up to 1Mbps and the propagation delay varies from 10ms to 100ms. This scenario involves a wired fixed client which is attached on the second router. Client streams video encoded at 64Kbps from the video streaming server and downloads a file at the same time from the FTP server .These 2 servers are attached to the first router. The first graph illustrates the packet end to end delay.





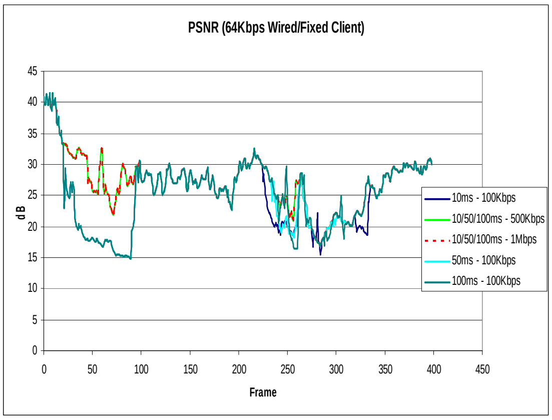


Figure 22: End to End Delay, End to End Jitter and PSNR respectively. Scenario of a fixed wired client (video streaming + FTP traffic).

Some similarities as well as differences can be seen in this scenario compared with the corresponding scenario shown in Figure 10. In scenarios that involve 100Kbps link bandwidth as far as end to end delay is concerned, we observe that after the first peak there is a gradual decrease of the delay followed by a slight increase. This second peak is due to the fact that FTP traffic coexists with video streaming traffic. This unexpected flow of packets cannot be served by the link and that's why many packets are inserted in the queue.

In scenarios having link bandwidth 500Kbps there is a small rise of end to end delay at the time interval that the two traffic patterns coexist. On the other hand in 1Mbps scenarios, end to end delay remains constant because the aggregated data stream does not exceed the offered bandwidth so there is no additional queuing delay.
There are also some differences between scenarios having different propagation delays and this is obvious in case of 500Kbps link bandwidth. In scenario that involves 10ms propagation delay we can see that some packets face longer end to end delay because RTT is smaller and the slow start procedure (during FTP traffic) increases the sending rate faster (senders open their congestion windows faster) than in scenarios with longer propagation delay.
Needless to say that the video quality deteriorates as this is testified by the PSNR.
The following graphs portray the client-side FTP traffic bandwidth. As can be seen TCP takes advantage of the available link bandwidth. Except for this, we observe that the longer the propagation delay the slower the reaction of the slow start procedure. In particular, it is shown that when multiple connections share a common bottleneck, those session with a smaller RTT are able to grab the available bandwidth at that link more quickly as it becomes free that is, open their congestion windows faster.

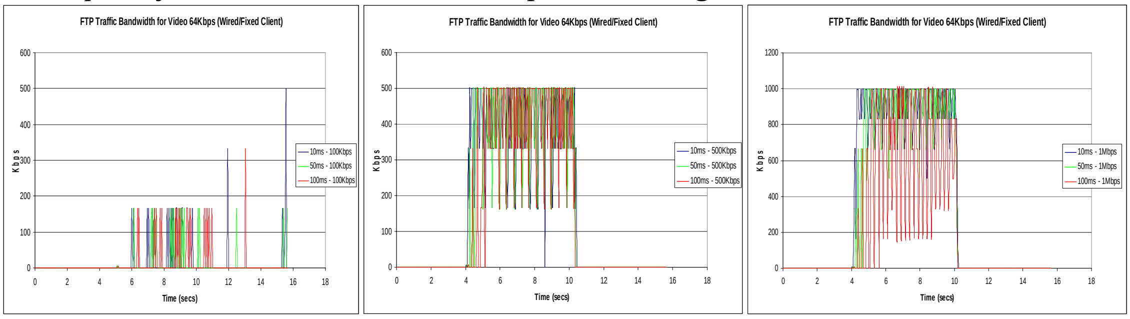
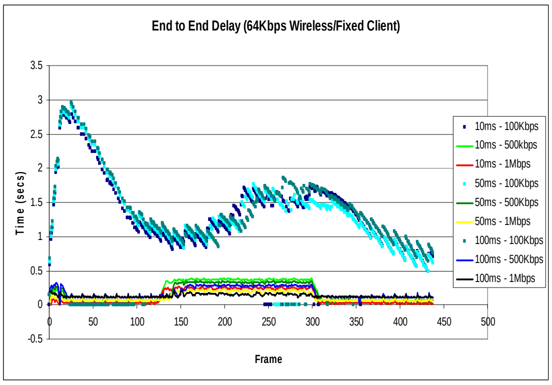


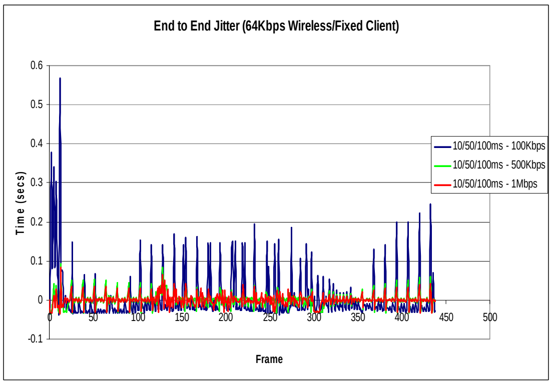
Figure 23: FTP Traffic Bandwidth for Video 64Kbps.

Bandwidth	*Propagation Delay*	*lost I frames percentage*	*lost P frames percentage*	*lost B frames percentage*
100Kbps	10ms	25.6%	11.6%	12.4%
	50ms	23.1%	10.5%	12.0%
	100ms	23.1%	11.6%	11.7%
500Kbps	10ms	2.6%	0.0%	0.0%
	50ms	2.6%	0.0%	0.0%
	100ms	2.6%	0.0%	0.0%
1Mbps	10ms	2.6%	0.0%	0.0%
	50ms	2.6%	0.0%	0.0%
	100ms	2.6%	0.0%	0.0%

Table 10: Percentage of lost I,P,B frames concerning scenarios with fixed wired client (64Kbps video stream + FTP traffic).

Table 10 shows that in scenarios which the link bandwidth is 100Kbps the percentages of lost frames are considerably high. Also these percentages are higher compared with the corresponding ones of Table 4 (only video streaming). This is an absolutely logical conclusion since now a larger amount of packets is being transmitted through the network. This phenomenon leads to congestion at the first router having many packets discarded due to queue overflow.





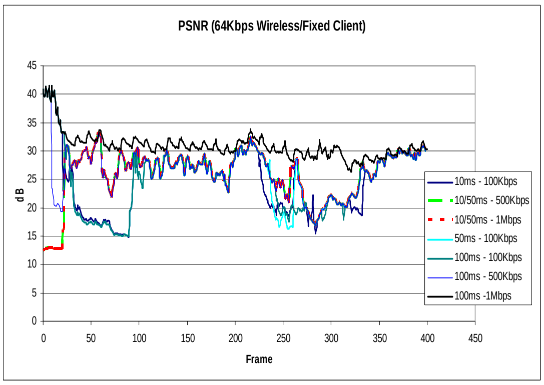


Figure 24: End to End Delay, End to End Jitter and PSNR respectively. Scenario of a fixed wireless client (video streaming + FTP traffic).

Figure 24 illustrates end to end delay, jitter and PSNR for a fixed wireless client. As far as the end to end delay and jitter are concerned the results resemble that of wired fixed client scenarios. On the other hand the graph that depicts the PSNR values reveals that better video quality is anticipated while having 1Mbps (link bandwidth) and 100ms (propagation delay). The smaller the link bandwidth, the lower the anticipated video quality. This can be observed in the Table shown below.

Bandwidth	*Propagation Delay*	*lost I frames percentage*	*lost P frames percentage*	*lost B frames percentage*
100Kbps	10ms	25.6%	11.6%	12.4%
	50ms	17.9%	13.7%	10.9%
	100ms	20.5%	11.6%	11.3%
500Kbps	10ms	5.1%	0.0%	0.0%
	50ms	5.1%	0.0%	0.4%
	100ms	2.6%	0.0%	0.0%
1Mbps	10ms	5.1%	1.1%	0.0%
	50ms	2.6%	0.0%	0.0%
	100ms	0.0%	0.0%	0.0%

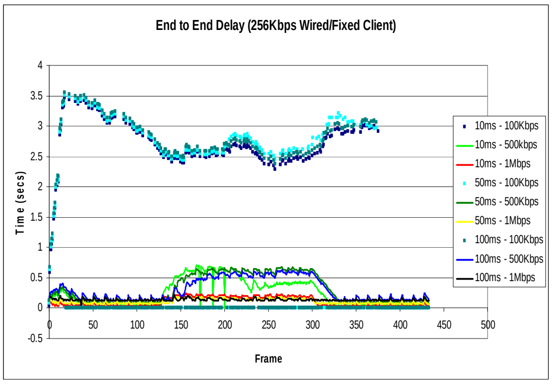
Table 11: Percentage of lost I,P,B frames concerning scenarios with fixed wireless client (64Kbps video stream + FTP traffic).

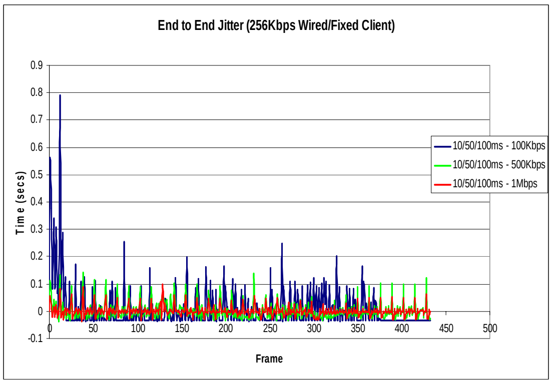
The graphs shown above as well as packet loss probabilities concerning scenarios that involve one mobile wireless client are the identical to those that involve one fixed wireless client, so they can be omitted.

We can conclude that there are no essential differences between the three different cases of clients (wired/fixed, wireless/fixed, wireless/mobile) with regard to end to end delay and jitter. As far as PSNR values are concerned there are certain differences that are due to the increase of packet loss probability in case of wireless users.

In conclusion we can say that the objective video quality deteriorates in relation to previous scenarios which involve only video streaming applications especially during the time intervals that FTP traffic and video traffic coexist.

6.2.2 256Kbps





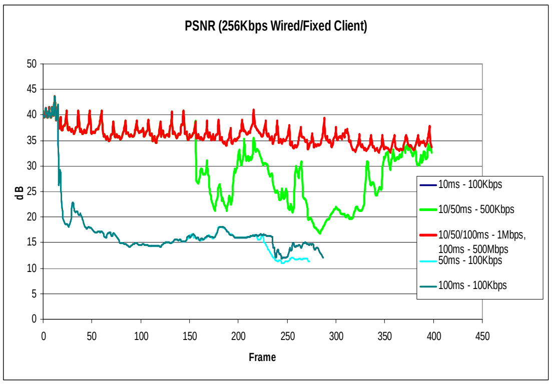


Figure 25: End to End Delay, End to End Jitter and PSNR respectively. Scenario of a fixed wired client (video streaming + FTP traffic).

In this section we present a set of graphs concerning video streams encoded at 256Kbps. In scenarios that involve 100Kbps link bandwidth we observe some fluctuations in the graph that portrays the end to end delay in contrast to the corresponding Figure 8. This phenomenon was explained in the previous paragraph 6.2.1. Also in scenarios where we have 500Kbps link an increase of the end to end delay during FTP traffic transmission occurred because of the burstiness of TCP flows.

From the graph that depicts the PSNR values we can see that video streams that are transmitted over 1Mbps link are considered to have better quality compared with video streams that are transmitted over 500Kbps as well as 100Kbps. In the latter situation video streams have very low PSNR values due to the high percentage of lost packets.

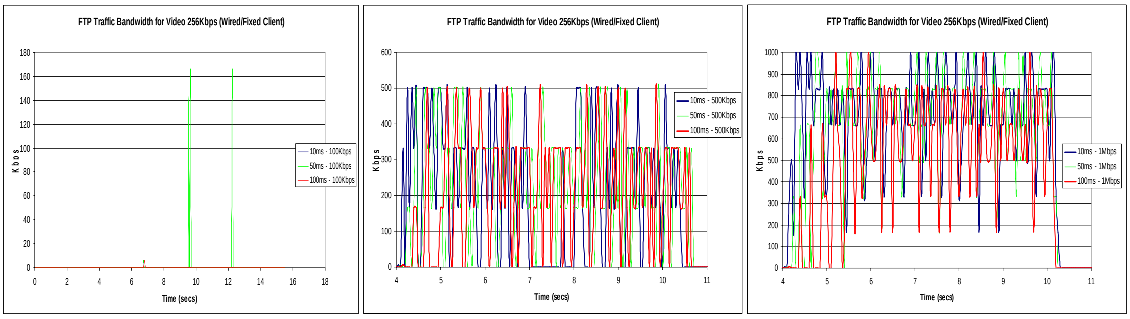


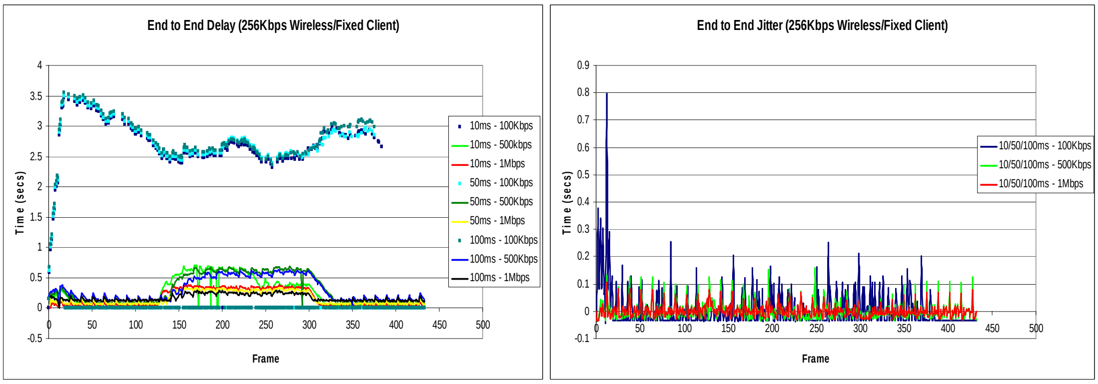
Figure 26: FTP Traffic Bandwidth for Video 256Kbps.

In case of 100Kbps link, FTP application was not be able to transmit adequate amount of data because of the inadequate link bandwidth. The larger the link bandwidth, the larger the amount of data being transmitted.

Bandwidth	*Propagation Delay*	*lost I frames percentage*	*lost P frames percentage*	*lost B frames percentage*
100Kbps	10ms	94.1%	89.0%	47.7%
	50ms	94.1%	86.0%	51.5%
	100ms	94.1%	89.0%	47.7%
500Kbps	10ms	0.0%	2.0%	0.4%

	50ms	0.0%	0.0%	0.0%
	100ms	0.0%	0.0%	0.0%
1Mbps	10ms	0.0%	0.0%	0.0%
	50ms	0.0%	0.0%	0.0%
	100ms	0.0%	0.0%	0.0%

Table 12: Percentage of lost I,P,B frames concerning scenarios with fixed wired client (256Kbps video stream + FTP traffic).



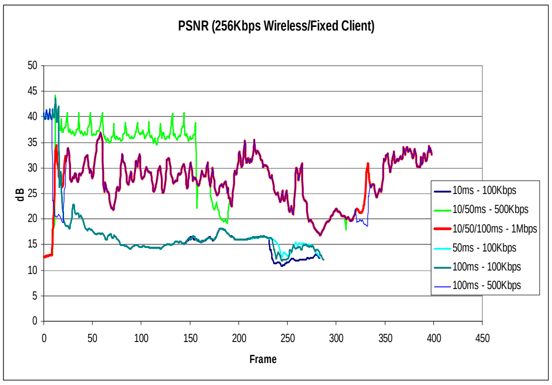


Figure 27: End to End Delay, End to End Jitter and PSNR respectively. Scenario of a fixed wireless client (video streaming + FTP traffic).

This is the last scenario that we simulated in NS2. It has to do with the transmission of a video stream encoded at 256Kbps. The wireless fixed client also downloads a file from the FTP server. The conclusions are similar to those concerning the previous scenario as far as the end to end delay and jitter are concerned. The main difference is that we have an increase in the end to end delay even when we have 1Mbps link during the time interval that the two traffic patterns coexist.

PSNR values are still small while having 100Kbps link whereas scenarios that involve 1Mbps link (for all values of propagation delay) and 500Kbps link with 100ms propagation delay, have similar higher PSNR values. PSNR values are improved in scenarios concerning 500Kbps link and 10ms/50ms propagation delay.

Bandwidth	*Propagation Delay*	*lost I frames percentage*	*lost P frames percentage*	*lost B frames percentage*
100Kbps	10ms	94.1%	87.0%	49.2%
	50ms	94.1%	88.0%	48.9%
	100ms	94.1%	89.0%	47.7%
500Kbps	10ms	5.9%	1.0%	1.5%
	50ms	5.9%	0.0%	0.4%
	100ms	5.9%	0.0%	0.0%
1Mbps	10ms	2.9%	1.0%	0.0%
	50ms	2.9%	0.0%	0.0%
	100ms	2.9%	0.0%	0.0%

Table 13: Percentage of lost I,P,B frames concerning scenarios with fixed wired client (256Kbps video stream + FTP traffic).

As shown in the above table, there is an increase in the number of lost packets compared with the previous scenario. This is probably due to the rapidly changing wireless channel.

7. Subjective Quality Evaluation Tests

The test material for the subjective tests consists of the video sequences received by the clients in the scenarios conducted in NS2. These malformed and erroneous sequences were gathered together in order to be evaluated by a group of users using the SAMVIQ method that mentioned before. The viewers were asked to assess the overall picture quality of each sequence by inserting a slider mark on a vertical scale.
We used a software tool called "MSU Perceptual Video Quality tool" [10] which is a tool for subjective video quality evaluation. An instance of this tool is shown at the figure below.

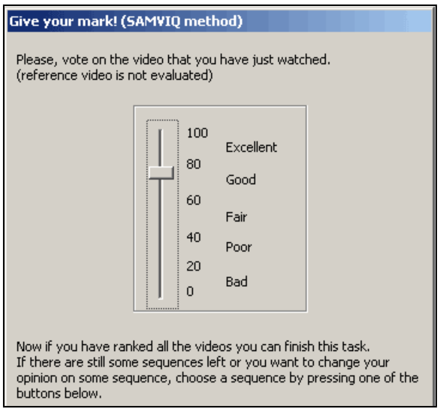


Figure 28: Instance of MSU tool used in subjective video quality evaluation.

We conducted a classroom test. There were 30 students in the class. Students were instructed to view more than 30 15-second clips collected from NS2 concerning *Foreman.* Students used the aforementioned tool in order to mark each individual clip. Some instances of these clips are depicted in figure below.

Figure 29: Foreman video clip instances.

8. Evaluations

From the results of the first 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. Also it is obvious that the percentage of lost packets is increased as the video transmission rate rises from 64Kbps up to 768Kbps.
Another remarkable observation is that I frames have higher possibility to be lost compared with B and P frames. This is due to the fact that they have larger size. It is well known that I frames (intra-coded frames) contain the information that results from encoding a still image, i.e., with no reference to any other image. I frames are points of reference and random access in the video stream and they can be decoded without the need for any other frames. That's why by loosing these frames the objective video quality (PSNR) deteriorates.
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. However, this conclusion cannot be extracted during these simulations.
We also observed that scenarios concerning a wireless fixed client and a wireless mobile client are identical while they slightly differ from scenarios concerning fixed wired client. In scenarios that involve two wireless clients (mobile-mobile or mobile-fixed) video quality (in a mobile wireless client side) is lower in relation to scenarios with only one mobile wireless client. In scenarios involving a mobile wireless client with a fixed wired client we observe that video quality (in mobile wireless client side) is better than having two wireless clients but worse than having only one mobile wireless client.
Besides we investigated two different widely used video sequences *Foreman* and *Claire*. We realize that static video streams such as *Claire* have smaller percentages of lost packets compared with video streams with mobility such as *Foreman*.
In the second scenario we provided background traffic in terms of FTP traffic. Simulations show that the objective video quality degrades due to the fact that the coexistence of a video stream and FTP traffic increases the percentages of lost packets.
Taking all these simulation results and subjective evaluation tests into consideration we tried to provide some kind of correlation between PSNR values and subjective evaluation test results. Table 14 presents mean PSNR values evaluated in NS2 for a given set of test sequences in relation to the viewers' mean grades. The following results apply for scenarios that do not involve FTP traffic.

Sequence	Mark (SAMVIQ)	Mean PSNR (dB)
FOREMAN\fixed_wireless\foreman256kbit_0.001prop_10ms_100kbit	11.3125	16.23631
FOREMAN\fixed_wireless\foreman256kbit_0.001prop_10ms_1Mbit	74.3125	25.95526
FOREMAN\fixed_wireless\foreman256kbit_0.001prop_10ms_500kbit	72.3125	26.11404
FOREMAN\fixed_wireless\foreman64kbit_0.001prop_10ms_100kbit	37.1875	24.50526
FOREMAN\fixed_wireless\foreman64kbit_0.001prop_10ms_1Mbit	51.75	25.15456
FOREMAN\fixed_wireless\foreman64kbit_0.001prop_10ms_500kbit	50.125	25.15456
FOREMAN\fixed_wireless\foreman768kbit_0.001prop_10ms_100kbit	15.3125	19.01528
FOREMAN\fixed_wireless\foreman768kbit_0.001prop_10ms_1Mbit	**87.4375**	**26.28639**
FOREMAN\fixed_wireless\foreman768kbit_0.001prop_10ms_500kbit	8.6875	16.25707
FOREMAN\mobile_wireless\foreman256kbit_0.001prop_10ms_100kbit	13.1875	16.23631
FOREMAN\mobile_wireless\foreman256kbit_0.001prop_10ms_1Mbit	70.75	25.95526

FOREMAN\mobile_wireless\foreman256kbit_0.001prop_10ms_500kbit	77	26.11404
FOREMAN\mobile_wireless\foreman64kbit_0.001prop_10ms_100kbit	40.375	24.50526
FOREMAN\mobile_wireless\foreman64kbit_0.001prop_10ms_1Mbit	52.5625	25.15456
FOREMAN\mobile_wireless\foreman64kbit_0.001prop_10ms_500kbit	49.625	25.15456
FOREMAN\mobile_wireless\foreman768kbit_0.001prop_10ms_100kbit	16.125	18.975
FOREMAN\mobile_wireless\foreman768kbit_0.001prop_10ms_1Mbit	77.8125	26.28639
FOREMAN\mobile_wireless\foreman768kbit_0.001prop_10ms_500kbit	7.4375	16.25707

Table 14: PSNR values compared with subjective evaluation test results (scenarios without FTP traffic).

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 Table 14 the video sequence encoded at 768Kbps and transmitted over 1Mbps link with 10ms propagation delay in scenario that involves one fixed wireless client considered to have the best quality in terms of PSNR as well as user-perceived quality.

The most remarkable observation is that video streams received by a fixed wireless client have the same mean PSNR values as those received by a mobile wireless client. On the other hand viewers perceive different QoS among these two kinds of scenarios. Their grades reflect that a fixed wireless user receives slightly higher overall QoS in relation to a mobile wireless user. 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.

Video sequences encoded at 256Kbps and transmitted over 500Kbps and 1Mbps link in scenario that involves one fixed wireless client have almost the same mean PSNR value and viewers perceive the same quality. In case of a mobile wireless client the corresponding sequences have the same mean PSNR values as in the latter case but viewers’ perception differs.

Table 15 summarizes the results shown in Table 14. In particular the table below presents the PSNR values in relation to SAMVIQ grading scale.

PSNR (dB)	*SAMVIQ*
> 26.2	81-100 (Excellent)
25 – 26.2	61-80 (Good)
24.5 – 25	41-60 (Fair)
19 – 24.5	21-40 (Poor)
< 19	0-20 (Bad)

Table 15: PSNR values (scenarios without FTP traffic) in relation to SAMVIQ grading scale.

Table 16 presents mean PSNR values evaluated in NS2 for a given set of test sequences in relation to the viewers’ mean grades. The following results apply for scenarios that do involve FTP traffic.

Sequence	Mark (SAMVIQ)	Mean PSNR (dB)
FOREMAN\fixed_wireless\foreman256kbit_0.001prop_10ms_100kbit	9.8125	17.07035
FOREMAN\fixed_wireless\foreman256kbit_0.001prop_10ms_1Mbit	76.3125	27.14709
FOREMAN\fixed_wireless\foreman256kbit_0.001prop_10ms_500kbit	59.25	29.75506
FOREMAN\fixed_wireless\foreman64kbit_0.001prop_10ms_100kbit	31.0625	24.88907
FOREMAN\fixed_wireless\foreman64kbit_0.001prop_10ms_500kbit	50	25.72604
FOREMAN\fixed_wireless\foreman768kbit_0.001prop_10ms_100kbit	13	24.47777

FOREMAN\fixed_wireless\foreman768kbit_0.001prop_10ms_1Mbit	51.1875	25.88662
FOREMAN\fixed_wireless\foreman768kbit_0.001prop_10ms_500kbit	8.3125	24.1105
FOREMAN\mobile_wireless\foreman256kbit_0.001prop_10ms_100kbit	10.4	17.07035
FOREMAN\mobile_wireless\foreman256kbit_0.001prop_10ms_1Mbit	76.4	27.14709
FOREMAN\mobile_wireless\foreman256kbit_0.001prop_10ms_500kbit	62.66667	29.75506
FOREMAN\mobile_wireless\foreman64kbit_0.001prop_10ms_100kbit	31.73333	24.88907
FOREMAN\mobile_wireless\foreman64kbit_0.001prop_10ms_500kbit	55.33333	25.72474
FOREMAN\mobile_wireless\foreman768kbit_0.001prop_10ms_100kbit	52.8	25.72604
FOREMAN\mobile_wireless\foreman768kbit_0.001prop_10ms_1Mbit	7.866667	24.47777
FOREMAN\mobile_wireless\foreman768kbit_0.001prop_10ms_500kbit	54.53333	26.31216

Table 16: PSNR values compared with subjective evaluation test results (scenarios with FTP traffic).

Table 16 that contains simulation and test results concerning scenarios with FTP traffic provides different results than Table 14. The last table reveals that mean grades which correspond to video sequences having high PSNR values are unexpectedly low compared with scenarios without FTP.

In contrary to the previous scenarios without FTP traffic, these results show that as the link bandwidth increases viewer perceived quality degrades. This is due to the burstiness of TCP flows.

All in all our studies show that PSNR as traditionally used objective measure is not adequate as perceptually meaningful measure.

9. Conclusion

The tests and simulations described in this paper were designed to correlate objective video quality with subjective video quality. Standard objective metrics such as PSNR were taken into consideration in order to evaluate objective quality.

Many factors (bit-rate, complexity of test sequence) had to be considered to specify fair and 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.

Our experimental results show that objective video quality metrics does not always provide and adequate knowledge about the actual QoS of video stream applications. In particular we proved that PSNR as traditionally used objective measure is not adequate as perceptually meaningful measure.

The focus of our future work is to investigate and evaluate error resilience techniques in video streaming over wireless networks as well as video layering techniques.

References

[1] F. Kozamernik, V. Steinmann, P. Sunna, E. Wyckens "SAMVIQ – A New EBU Methodology For Video Quality Evaluations in Multimedia".

[2] S. Winkler, "A perceptual distortion metric for digital colour images", Proc. of the 5th IEEE International Conference on Image Processing (ICIP '98), Chicago, IL, October 4-7, 1998.

[3] 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.

[4]. ANSI T1.801.03-1996, "American National Standard for telecommunications-digital transport of one-way video signals-parameters for objective performance assessment", American National Standards Institute.

[5] Stephen Wolf, and Margaret H. Pinson, “Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system”, Proc. of SPIE International Symposium on Voice, Video, and Data Communications, Boston, MA, September 11-22, 1999.
[6] T. Alpert, V. Baroncini, D. Choi, L. Contin, R. Koenen, F. Pereira, H. Peterson, “Subjective Evaluation of MPEG-4 Video Codec Proposals: Methodological Approach and Test Procedures”.
[7] http://ffmpeg.sourceforge.net/index.php.
[8] J. Klaue, B. Rathke and A. Wolish, “EvalVid – A Framework for Video Transmission and Quality Evaluation”.
[9] How to evaluate MPEG video transmission using the NS2 simulator – http://140.116.72.80/~smallko/ns2/Evalvid_in_NS2.htm
[10] MSU Perceptual Video Quality tool http://compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html.