Performance Evaluation of Audio-Video Telephony in WiMAX Networks

Kyungtae Kim and Young-June Choi
Dep. Computer Engineering, Ajou University, Suwon, 443-749, Korea
E-mail: kim.kyungtae@gmail.com, choiyj@ajou.ac.kr

***Abstract*—We consider efficient techniques for delivering audio-video telephony/conferencing service (AVS) traffic over WiMAX networks. Issues that need to be considered in error-prone and variable delay wireless networks specifically for real-time AVS traffic include delay, jitter and the possible lack of synchronization between audio and video traffic. First, we consider two different schemes for WiMAX connection management that allocate the same connection ID (CID) or two different CIDs for audio and video connections, which we refer to as integrated-CID scheme and separated-CID scheme, respectively. The integrated-CID scheme is amenable to connection management with reduced overheads, but causes cell-edge users to suffer longer voice delay variance. Second, we consider two different audio codecs, with low and high source rates. Generally, a codec with a high source rate gives more satisfaction to users, but we show that this does not hold for cell-edge users because it breaks the synchronization from video frames due to the longer reception delay. Finally, we investigate appropriate strategies for codec selection for both cell-edge and cell-interior, i.e., selecting a combination of high-rate audio/low-rate video and low-rate audio/high-rate video. We conduct simulation experiments to test the proposed scenarios over mobile WiMAX systems.**

I. Introduction

In recent years, Audio-Video Telephony/Conferencing Service (AVS) has become a promising application with the potential for conducting productive meetings without the need for travel [1]. In addition, mobile operators are motivated to look for revenue from enterprise deployments through the use of 4G technologies such as WiMAX; specifically, enterprise WiMAX femtocell deployments are being considered as an option for such deployments [2]. This motivates us to study the problem of AVS delivery over enterprise WiMAX networks that will be characterized by lower throughput (than WiFi which uses a much larger bandwidth) and potentially higher error rates.

The perceived quality of audio and video is affected by delay, jitter, and packet loss. In addition to these factors, AVS is also sensitive to the synchronization between audio and video; the level of synchronization is more susceptible to delay than packet loss. In wireless systems (such as WiMAX) with more packet losses and higher delay due to their low bandwidth and high channel error rates, the quality of AVS may be further degraded [3]. In this paper, we investigate various scenarios for AVS in mobile WiMAX systems which has inherent support in the standard for supporting multimedia traffic.

Mobile WiMAX systems define five service classes that include Unsolicited Grant Service (UGS), Real Time Variable Rate (RT-VR) service, and Extended RT-VR (ERT-VR) service [4]. UGS provides a fixed bandwidth for voice services and RT-VR provides variable bandwidth for video streaming services. ERT-VR mixes both UGS and RT-VR, so we can naturally use ERT-VR class of service for AVS. However, depending on the location of a user within a WiMAX cell, wireless users may see different transmission rates. When the transmission rate is low due to bad channel conditions, UGS may not guarantee the best quality for AVS. In this case, audio and video streams may have to be separated, such that each stream is transported over its own connection, where the quality of each stream is controlled separately to ensure a better AVS quality.

In this paper, we propose an adaptive connection management strategy to support AVS. The adaptive strategy relies on two different schemes to allocate audio and video connections. In one scheme, an AVS stream is managed by one connection through the allocation of one connection ID (CID), so audio and video streams are together delivered within the same connection which uses the ERT-VR class of service. We refer to this scheme as the *Integrated-CID scheme*. In another scheme, each stream is delivered through two separate connections, one which uses the UGS class of service and the other which uses the RT-VR classes of service. We refer to this scheme as the *Separated-CID scheme*. Through simulations, we show that cell-edge users achieve better performance with the Separated-CID scheme while the Integrated-CID scheme is more appropriate for the interior users in the cell.

We further test the above scenarios with two different audio codecs: G.729 (8 kbps) and G.711 (64 kbps). While G.711 in general guarantees a better audio quality thanks to its higher sampling rate, as we show, this is not always the case in WiMAX systems. We show that cell-edge users achieve better performances with G.729, because it reduces the gap between audio and video reception time, thereby achieving better synchronization. In general, cell-edge users suffer longer delay and larger delay variance in receiving the audio frame with G.711 and this degrades the overall QoS of AVS. We also compare the above scenarios with different video rates to find an appropriate codec option for different classes of users (cell-edge vs. cell-interior); in this regard, we investigate two options, namely G.729 audio + high video bandwidth and G.711 audio + low video bandwidth.

There have been some previous studies on the delivery of video conference services over various wireless systems

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

[3], [5]–[7]. The most relevant to our work is [7] where the performance of video conferencing traffic over mobile WiMAX systems was considered. This study was limited to evaluating the performance of different codecs and did not consider the general problem of connection management or codec selection. In our work, we consider both connection management and codec selection as parameters that can be appropriately selected to optimize AVS performance over WiMAX.

The remainder of this paper is organized as follows. Section II describes the characteristics of audio, video and AVS traffic. Section III presents our proposed scheme and also summarizes test scenarios that use different audio codecs and video bandwidths. Section IV shows our experimentation results and Section V concludes.

II. Audio-Video Telephony/Conferencing Systems

In this section, we describe the characteristics of audio and video traffic and further discuss the characteristics of AVS, i.e., integrated audio and video telephony/conferencing traffic.

A. Voice communication

VoIP marks a transformation from traditional circuit-switched voice to a more efficient packet-switched transport. To provide an acceptable voice quality in enterprise wireless networks similar to that in wired networks, a VoIP system should choose an appropriate codec. Mobile WiMAX and 3GPP long-term evolution (LTE) promise to be the first wireless access standards fully capable of supporting VoIP. Although VoIP has low-bandwidth requirements, it will still be a challenge to support bi-directional VoIP service with low delay in IP-based networks.

1) Speech codec: Most public switched telephone networks (PSTNs) operate with speech sampled at 8 kHz according to ITU-T G.711. This encodes speech at 64 kbps and introduces little audible distortion for most types of signals. In a number of applications, however, a much lower bit rate is desirable either because capacity is limited, i.e. wireless/mobile environment, or because the number of admitted calls need to be maximized. A voice encoder should be chosen to fit the particularities of the transport network (loss and delay). One of the popular voice encoders is G.729, which encodes at 8 kbit/s and uses 10 ms or 20 ms frames. In this paper, the applicability of two vocoders, namely G.729 (lower bit-rate) and G.711 (higher-bit rate), is considered and the performance of AVS using these two vocoders is studied.

2) VoIP quality assessment – E-Model: To measure the quality of a VoIP call, we use a metric called E-Model which takes into account mouth to ear delay, loss rate, and the type of the encoder [9]. The quality is defined by the *R-score* (a scalar measure shown below that ranges from 0 (poor) to 100 (excellent)), which for medium quality should be more than 70.

$$\begin{aligned} R &= 94.2 - 0.024d \\ &- 0.11(d - 177.3)H(d - 177.3) \\ &- 11 - 40\log(1 + 10e) \end{aligned} \tag{1}$$

where:

- $d = 25 + d_{jitter\_buffer} + d_{network}$ is the total ear to mouth delay comprising 25 ms vocoder delay, delay in the de-jitter buffer, and network delay.
- $e = e_{network} + (1 - e_{network})e_{jitter}$ is the total loss including network and jitter losses.
- The Heaviside function $H(x) = 1$, if $x > 0$; 0 otherwise.
- The parameters used are specific to the G.729a encoder with uniformly distributed loss.

B. Video communication

Video communication is starting to be a focus of investigation within 4G wireless networks such as mobile WiMAX and LTE. Most errors that occur in video transmission are caused by its large bandwidth requirements. Despite mechanisms used at the PHY and MAC layers to alleviate channel errors (such as Automatic Repeat Request (ARQ) mechanisms and intelligent scheduling), other effects such as a long delay may disturb the perceived quality for a video service. Although the channel capacity exceeds the required bit rate, channel errors can severely degrade the performance of the system and thus compression schemes and bit rates must be appropriately chosen in order to match the channel characteristics while maximizing the video quality.

Since raw video requires high transmission bit rates, video compression is usually employed to achieve transmission efficiency. MPEG-4 and H.263/264 are the two video compression standards used by most mobile networks and their compression technology produces good quality video at bit rates that are suitable for mobile and wireless transmission [8].

A peak signal-to-noise ratio (PSNR) is the most widespread objective metric used to assess the application-level QoS of video transmissions [10]. A PSNR measures the error between the reconstructed image and the original one frame-by-frame. In this paper, we use video delay as a QoS metric, rather than a PSNR, because in our simulations, video delays are found to be crucial to AVS quality whereas the impact of packet losses is negligible.

C. AVS

Audio/video synchronization means that visual lip movements of a speaker must match the sound of the spoken words. If the audio and video displayed at the receiving endpoint are not in sync, the misalignment between audio and video can degrade the perceived quality. Audio often arrives ahead of video in AVS systems since audio requires less time to deliver and process while the latencies involved in processing and sending video frames are greater than the latencies for audio. The conventional approach to synchronizing audio and video is to delay the media that arrives early so that the audio and video latencies are matched; however, the time required to deliver video can exceed the maximum perceived audio latency that is acceptable in a conversation. The latency deliberately introduced for synchronization might violate the QoS requirement of interactive communication. Thus, in some AVS systems, instantaneous audio delivery is supported rather

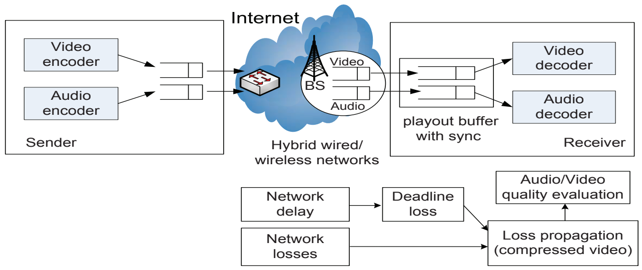

Fig. 1. Wireless networks impact on the AVS quality due to bandwidth limitation, which causes deadline losses of the video packets.

than trying to maintain lip synchronization by delaying audio to synchronize with video.

The upper part in Fig. 1 shows a simplified transmission path across a network. IP packets are generated by the audio/video encoders and forwarded through the network. The network introduces delay due to buffering, processing delay and transmission delay. Additionally, packets may be lost due to buffer overflows or link-layer errors. On the client side, the packets are stored in the playout buffer to resolve jitter and synchronization problems, and played back at their playout time.

The lower part of Fig. 1 illustrates the impact of data transmission on the video quality. In a hybrid wired/wireless network, network components, such as routers and base station, introduce network loss and delay. If the application has certain delay constraints, a longer network delay can cause so-called deadline losses for audio and video flows, especially when the video data arrives at the client after its anticipated playout time. Network losses and deadline losses both directly lead to video data losses, which results in further degradation due to video frame dependencies in case of compressed video such as MPEG.

III. Resource Allocation Strategy

We now consider several resource allocation strategies for mobile WiMAX systems that affect the QoS of AVS; these include the type of data delivery services, the audio codec used, and the video bit rate.

A. Type of data delivery services

In mobile WiMAX systems, five types (classes) of data delivery services are defined as in Table I. Typically, voice and video connections are classified as UGS and RT-VR, respectively, while data connections are classified as a Non-Real-Time Variable Rate (NRT-VR) or a Best Effort (BE) type. UGS dedicates a fixed bandwidth for each voice connection, so it is suitable for voice traffic with a fixed rate. RT-VR uses real-time polling service (rtPS) for uplink, so it is suitable for video traffic with variable rate. When AVS traffic is generated, we can naturally consider a strategy that assigns audio and video connections to UGS and RT-VR, respectively, thus creating two CIDs. We refer to this as the *separated-CID scheme*. However, this strategy may not always be appropriate for AVS service, because it may be possible for audio and video to

TABLE I
Type of data delivery services in mobile WiMAX systems

Symbol	Meaning	Uplink scheduling
UGS	Unsolicited Grant Service	dedicated bandwidth
RT-VR	Real-Time Variable Rate	frequent polling (rtPS)
NRT-VR	Non-Real-Time Variable Rate	less frequent polling (nrtPS)
BE	Best Efforts Service	Request/Transmission
ERT-VR	Extended RT-VR	UGS + rtPS

be delivered on the same connection while supporting the required data rate for each traffic.

We consider another scheme that assigns one CID for voice and video of AVS traffic. As seen in Table I, mobile WiMAX systems define an ERT-VR type where real-time traffic is based on UGS and rtPS together for uplink. This fits AVS traffic, although it may not keep the stringent requirement of UGS. We refer to this as *integrated-CID scheme*. This enables easy management for an AVS session while reducing the signal overhead caused by managing connections separately. Our scenarios are summarized as follows.

- Test scenario a1: separated-CID
- Test scenario a2: integrated-CID

B. Audio codec

Since the bandwidth required to send audio traffic is determined by the audio codec, it will affect the overall QoS of AVS. We consider two audio codecs, G.729 and G.711, with rates 8 kbps and 64 kbps respectively. Compared to G.711, G.729 consumes less bandwidth, so users may be less satisfied with the voice quality of G.729. G.711 will thus be a better choice in general, but the requirement for AVS is that voice traffic should be delivered together with video traffic in synchronization. Specifically, cell-edge users who may not be able to support higher bandwidths may benefit by the use of lower bandwidth codec, G.729, which in essence would enable them to synchronize audio and video traffic better. In such a case, G.729 would be better for cell-edge users, whereas G.711 may be applicable for interior users. Our test scenarios to explore this effect are as follows.

- Test scenario b1: G.729 (8 kbps) for all users
- Test scenario b2: G.711 (64 kbps) for all users

C. Video bit rate

Similarly, we can vary the video bit rate. As the video bandwidth increases, a user will be more satisfied. Ultimately, we would like to find a better combination between an audio codec and a video bit rate, when audio and video flows share a given bandwidth. To simplify this problem, we test the following scenarios. Different video rates are chosen with rate differences chosen to be 50 kbps to compensate for audio vocoder difference of around 50 kbps between 64 kbps and 8 kbps. In our experimentation, the maximum allowable video delay is 500 msec and the default video rate is set as 250 kbps.

- Test scenario c1: 8 kbps (audio) + 250 kbps (video)
- Test scenario c2: 64 kbps (audio) + 200 kbps (video)

IV. Experimentation

A. Simulation testbed

We use the *ns*-2 with WiMAX extensions provided by the WiMAX Forum. An audio module of *ns*-2 is programmed to create audio flows for G.711 and G.729 vocoders. On the receiver side, the sending time, receiving time, packet length, sequence number, and time stamp are traced to obtain the network loss, delay and jitter for each flow. The information is fed into the audio quality assessment module which produces R-score according to the voice quality assessment formula in Eq. 1.

Similar to the audio processing procedure, the effect of streaming video over mobile WiMAX is captured in a log file which is generated by *ns*-2 at the sender and receiver. The raw video clip in the YUV video format is fed to an H.263 video encoder which in turn generates an encoded video stream. In this test, an open-source H.263 encoder and decoder (i.e. ffmpeg) are used. The encoded video stream is read by the video clip transmitter provided by EvalVid [11] to generate a trace file, which contains information such as frame type, generated time, and frame size etc., for each video frame. The trace file is then input into the sender module in the *ns*-2 simulator to produce video streams to be sent over the emulated mobile WiMAX link. The log files at the sender and receiver are used by the trace program by EvalVid to generate possibly corrupted video files as a result of the transmission over the wireless link. The corrupted video file is used by PSNR and Mean Opinion Score (MOS) programs to evaluate end-to-end video quality.

For the WiMAX simulation environment, orthogonal frequency division multiple access (OFDMA) with 10 MHz is used. The effect of different service flows on QoS parameters such as packet loss, average jitter and average delay is measured to calculate users' perceived quality. In our simulation, packet loss and average jitter of video traffic are negligible, so we present only the video delay. An audio-video clip of 10 seconds is used in all scenarios and QoS metric is obtained to compare the performance of each scenario. A simple round-robin scheduler is exploited to allocate slots. Two different modulation and coding schemes (MCS), MCS-6 (64QAM-2/3) and MCS-2 (QPSK-1/2), are used for cell-interior and cell-edge users, respectively.

B. Separated-CID vs. Integrated-CID

1) Cell-interior users: We now present the results for cell-interior users who experience good channel quality. In case of Separated-CID (scenario a1) where voice has higher priority than video, audio traffic shows stable quality while the number of video flows increases until all available resources are used up. Fig. 2 shows that the number of voice flows with minimum R-score of 70 is 11 with stable network delay around 18 msec to 35 msec, and small network and jitter loss. However, the delay for video flows increases abruptly as the number of users increases, especially after 10 users have been accepted as shown in Fig. 3. The lower bandwidth vocoder, G.729

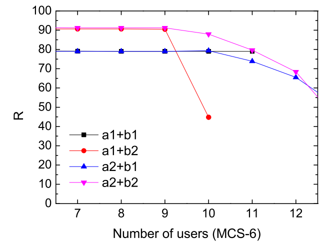

Fig. 2. Performance of voice traffic for cell-interior users.

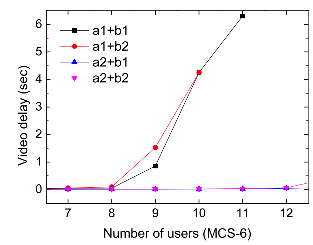

Fig. 3. Performance of video traffic for cell-interior users.

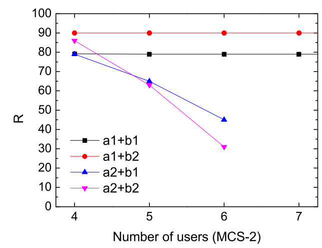

Fig. 4. Performance of voice traffic for cell-edge users.

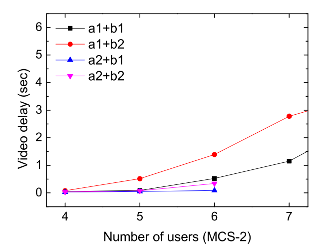

Fig. 5. Performance of video traffic for cell-edge users.

(scenario b1), is able to admit more users while providing acceptable voice quality when compared to G.711 (scenario b2). G.711 provides higher R-score (91) than G.729 (79) but at the cost of more bandwidth consumption (64 kbps). To accommodate more users for AVS, G.729 vocoder is recommended.

Fig. 3 shows that video delay is not acceptable with the Separated-CID scheme (scenario a1) but it remains low and acceptable with the Integrated-CID scheme (scenario a2). Therefore, synchronization is well ensured in the Integrated-ID scheme. Through the Integrated-CID scheme, we can reduce the number of CIDs, because MAC signalling overhead is also reduced and thus more users can be supported. In summary, the integrated-CID scheme achieves audio/video synchronization for cell-interior users with reduced MAC overhead while admitting more users.

2) Cell-edge users: Users who are farther from the base station will require a more robust MCS to guarantee the same level of bit-error-rate (BER) over a wireless link. Also they require more slots to achieve the same throughput as cell-interior users. This causes the number of admitted cell-edge users to be small. Thus, a maximum of 6 users for the Separated-CID scheme (scenario a2) is observed in Figs. 4 and 5. As these users are supported with a small data rate with a robust MCS, for the Integrated-CID scheme, audio packets with a smaller size, in most cases, will have to wait for longer time at the queue due to greater video delay. H.263 codec encodes frames with variable size [8]; e.g., 12,408 bytes for a first frame (I frame) and 111 bytes for a second frame (B frame) within the video clip used in our experimentation. This variable size of the video frame causes greater delay variance and higher jitter loss of the audio flow at the receiver. Due to

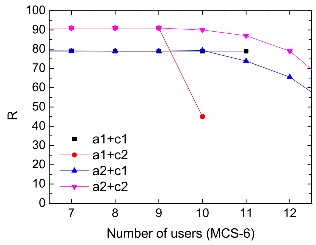

Fig. 6. Performance of voice traffic for various codecs.

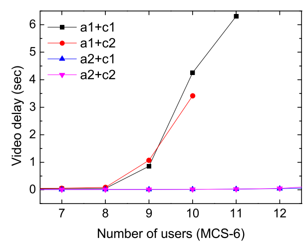

Fig. 7. Performance of video traffic for various codecs.

TABLE II
PERFORMANCES FOR CELL-EDGE USERS

Scheme	No. of users	R-score	Video delay (msec)
a1+c1	4	79	53
a1+c1	5	79	86
a1+c1	6	79	425
a1+c2	4	91	64
a1+c2	5	91	155
a1+c2	6	91	865
a2+c1	4	79	28
a2+c1	5	65	51
a2+c1	6	43	85
a2+c2	4	87	33
a2+c2	5	81	126
a2+c2	6	57	634

the jitter loss, the number of acceptable cell-edge users is just 4 in the case of scenario a2, as shown in Fig. 4. On the other hand, the Separated-CID scheme (scenario a1) admits more cell-edge users with good voice quality, because the audio flows are treated with higher priority without causing higher delay variance. But these users may suffer a little more video delay as seen in Fig. 5.

C. Combination of different vocoder and video rate

We now consider the combination of two different audio codecs and two different video rates with low and high sampling rates. Figs. 6 and 7 shows the audio and video qualities for cell-interior users. The overall performance is very close to Figs. 2 and 3 in that the Integrated-CID scheme outperforms the Separated-CID scheme. That is, b1 and c1, and b2 and c2 show similar performance. This implies that the voice rate determines the overall performance rather than the video rate, when voice and video flows share a given bandwidth. For cell-interior users, a2+c2 combination provides the best audio quality as well as the best video quality.

Table II summarizes the audio and video performances for cell-edge users. As shown in Figs. 4 and 5, the Separated-CID scheme (scenario a1) maintains both acceptable R-score and video delay. When a high-rate video codec is used at the cost of low-rate audio codec (scenario c1), video delay is small and R-score is low. Thus, there is a tradeoff between audio quality and video quality when they share a given bandwidth. To admit as many cell-edge users as possible, the combination of Separated-CID, lower-rate vocoder and higher-rate video (a1+c1) is recommended.

V. CONCLUSION

In this paper, we investigated a solution for allocating CIDs for AVS traffic in mobile WiMAX systems and proposed two schemes, namely the Integrated-CID and Separated-CID schemes. Through simulations, we found that each is a good choice for different classes of users, namely cell-interior users and cell-edge users, respectively. In addition, we studied the effect of different audio and video sampling rates on the performance of AVS for these two classes of mobile WiAMX users. For future work, we will extend our study to include various MCS types and users with different mobility patterns.

ACKNOWLEDGEMENT

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-(C1090-1221-0011)).

REFERENCES

[1] "How should en anterprise move toward Unified Communications?", http://www.avaya.com/usa/resource/assets/whitepapers/moving towards uc (uc3858).pdf (oneline), Avaya, 2009

[2] D.-Y. Kwak, J.-S. Lee, Y. Oh, and S.-C. Lee, "Development of WiBro (Mobile WiMAX) Femtocell and Related Technical Issues," in Proc. *Global Telecommunications Conference, 2008*. pp.1-5, Nov. 30 2008.

[3] D. Chattopadhyay, A. Sinha, T. Chattopadhyay, and A. Pal, "Adaptive rate control for H.264 based video conferencing over a low bandwidth wired and wireless channel," in Proc. *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting 2009*. pp.1-6, May 13-15, 2009.

[4] IEEE 802.16e-2005, "*Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands"*, Feb. 2006.

[5] D.B. Khedher, R.H. Glitho, and R. Dssouli, "Media handling aspects of multimedia conferencing in broadband wireless ad hoc networks," *IEEE Network*, vol. 20, no. 2, pp. 42-49, Mar. 2006.

[6] G.-M. Su; Z. Han, M. Wu, and K.J.R. Liu, "Joint uplink and downlink optimization for video conferencing over wireless LAN," in Proc. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2005*, May 2005.

[7] B. Kaarthick, V.J. Yeshwenth, N. Nagarajan, and Rajeev, "Performnance analysis of Video Conferencing and Multimedia application Services over WiMAX," in Proc. *IEEE International Advance Computing Conference (IACC) 2009*, pp. 1109-1113, Mar. 2009.

[8] Iain E.G.Richardson, *H.264 and MPEG-4 Video Compression*, ISBN 0-470-84837-5.

[9] R.G. Cole and J. Rosenbluth, "Voice over IP performance monitoring", *Computer Communication Review*, vol. 31, pp. 9-24, Apr. 2001.

[10] N. Thomos, N.V. Boulgouris, and M.G. Strintzis, "Optimized transmission of JPEG2000 streams over wireless channels," *IEEE Transactions on Image Processing*, vol. 15, no. 1. pp. 54-67, Jan. 2006.

[11] J. Klaue, B. Rathke, and A. Wolisz, "EvalVid - A framework for video transmission and quality evaluation," in Proc. *the 13th International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*, pp. 255-272, 2003.