Packet-Switched H.264 Video Streaming Over WCDMA Networks

Carlos Murillo, MS.
Florida Atlantic University, Boca Raton, FL USA.
cmurill3@fau.edu

Cyril-Daniel Iskander, Ph.D.
Florida Atlantic University, Boca Raton, FL USA.
iskander@fau.edu

Abstract

A current research challenge consists of providing high quality real-time multimedia services to a mobile user over packet-switched wireless networks, in a time-varying interference-limited wireless channel. The new H.264/AVC digital video compression standard, combined with the RTP/UDP/IP protocol stack, render possible the efficient transport of video services over WCDMA wireless networks. This paper reviews the mechanisms which make possible streaming H.264/AVC video over WCDMA wireless networks, using RTP/UDP/IP as the protocol stack. In particular, it describes the packetization schemes and control mechanisms, and compares different error control schemes to obtain an aceptable end-to-end QoS. These components of a packet-switched streaming service are integrated into a software simulation model which is used to evaluate and predict the end-to-end H.264/AVC video quality in a next-generation WCDMA wireless network.

The results indicate that the RLC-ACK scheme, on WCDMA networks, can be flexibly configured to allow a trade-off between the required reliability and maximum delay allowed in the RLC layer. Moreover, the simulations show that the H.264/AVC codec provides a suitable video stream adapted to packet-switched format for wireless transmission. The paper is structured as follows. First, it provides an overview of the packet-switched video streaming transmission over the wireless channel. Second, the WCDMA system reference model and its counterpart modeled software are described. Third, the H.264/AVC software architecture functionality and the WCDMA link layer model are detailed. Finally, representative results, performance evaluation, and conclusions are presented.

Keywords

WCDMA, H.264/AVC, RTP/UDP/IP, RLC.

1.- Introduction

The main objective of video compression algorithms is to achieve high compression ratios with good quality of video presentation. In practice, this is accomplished by exploiting temporal, spatial, and statistical redundancy in video sequences. A relevant strategy, for example is adopted in the H.264/AVC standard [1]. It is known that there is a trade-off in this process between the fidelity of the approximation to the original image and the number of bits required to represent that image (coding distortion *versus* data rate). Compression performance can then be evaluated by considering the rate required to achieve a

specific quality level. Thus, compression techniques that exploit spatial and temporal redundancy will result in different rates for different frames at a given quality level. This means that the degree of redundancy and the rate for a given distortion may fluctuate widely from scene to scene; that is, scenes with high motion content will require more bits than stationary ones. This variable bit rate transmitted via the wireless medium is performed usually in packet- and circuit switched formats. Evaluating the performance of such video streaming transmissions over packet-switched networks is the motivation of this study and relevant simulations form the scope of the underlying research.

Efficient digital compression of video signals such as that obtained via the H.264/AVC standard [1] allow new wireless multimedia applications. Representative examples include the implementation of conversational video and video streaming. The purpose of compression is to achieve data rate reduction and efficient channel bandwidth utilization so as to lower the transmission cost. Packet video streaming is becoming a main issue in these multimedia applications because of the inherent variability of the frame rate. In addition, end-to-end delay, network congestion, and transmission losses are other issues that hamper the delivery of high quality of service (QoS) applications to end-users.
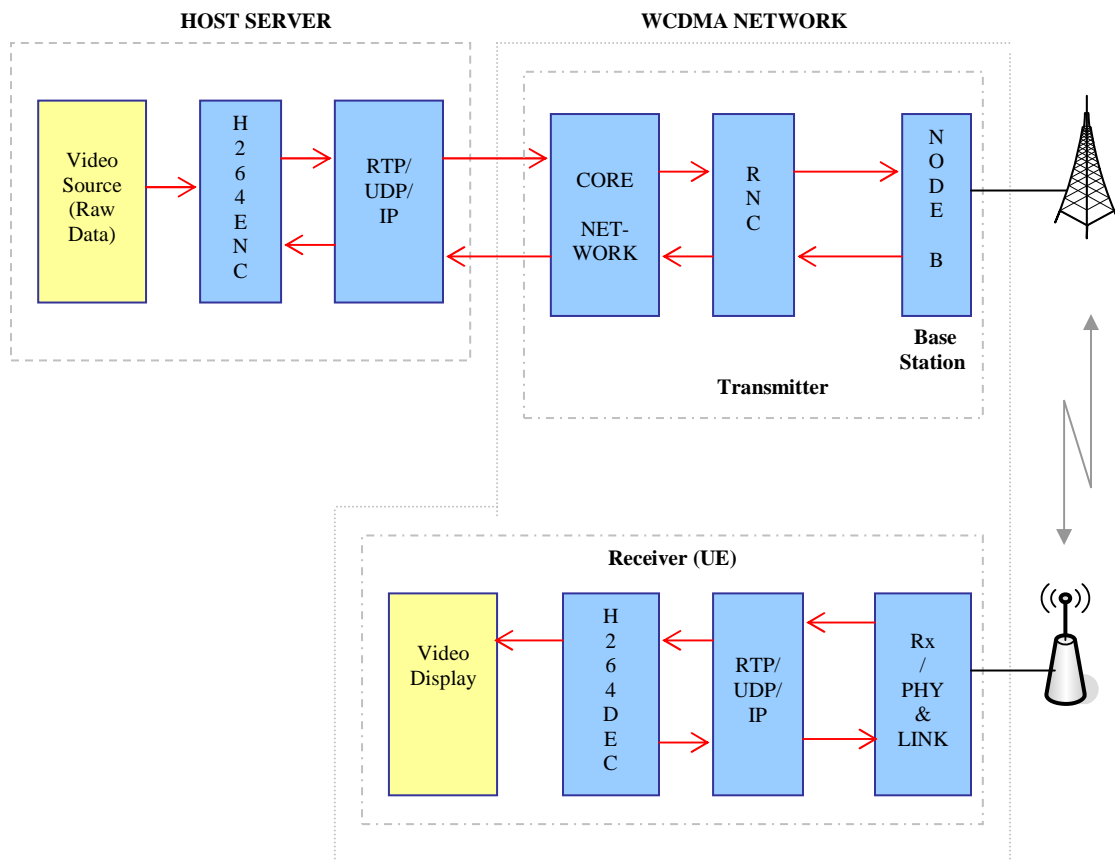

Figure 1: *System Reference Model*

Historically, Second Generation (2G) wireless systems were primarily targeted at the transmission of speech; and even though they can support extensive user mobility and coverage, they provide however only a limited bandwidth on the order of 10 Kbps on average [2]. Clearly, these bandwidths are inadequate for the transmission of multimedia applications like high bandwidth video. In contrast, Third Generation (3G) cellular systems such as IMT-2000 have been conceived for the delivery of multimedia services and applications [4]. Such 3G systems are based mainly on Code Division Multiple Access (CDMA) and are able to support applications with a variety of rate requirements. Wideband CDMA

(3GPP) and cdma2000 (3GPP2) are the main 3G cellular standards that offer multimedia services. Pertinent to 3G systems, the main improvements over 2G systems are more capacity, greater coverage area, and a high degree of services. In addition, 3G systems provide larger bandwidths and adaptive rate transmission to deliver video to mobile users [3].

It is known that the uncompressed video signal requires higher bandwidth for its transmission. This wide bandwidth requirement makes transmission difficult over a wireless channel for a user in an outdoor environment because of the limited channel capacity. Therefore, video compression techniques, such as the H.264/AVC standard, are adopted to make such video transmissions feasible over interference-limited wireless channels.

The H.264/AVC [1] codec is a new standard for video compression and a same code implementation is available in the public domain [6]. This codec and the WCDMA link layer models have been configured under diverse evaluation scenarios and simulation results obtained are presented in this paper. The simulation of the WCDMA system corresponds to the link layer as per the 3GPP specification [7] [8] [11] indicated for the transport of packet-switched data. The wireless system reference model under consideration refers to the WCDMA illustrated in Figure 1 and its corresponding software simulation model is presented in Figure 2.

The main objective of this study is to evaluate the end-to-end transmissions of PDU-frames, average delay, and QoS over a WCDMA (RNC transmitter to RNC receiver/end-user) system. The system is evaluated with special emphasis on the RLC layer (L2) functions, which provide additional tools for correcting errors that were not recovered by the error correction schemes at the physical layer. That is, the link-layer increases the channel reliability based on ARQ recovery mechanism by reducing the FER. Also, evaluated via the present model is the transport of end-to-end packet-video streaming in a downlink scenario. Simulations are done to ascertain the following: Minimum end-to-end QoS characteristics for a maximum BER and PLR (based on the application type to be transmitted), ARQ scheme implementation, and average delay.

In an ARQ scheme implementation, the receiver acknowledges (ACK) the correct data frame sequence and signals a negative acknowledgement (NACK) when erroneous frames and/or lost frames are received. The NACK messages are fed back to the transmitter to indicate retransmission of the respective frame; and, it will be sent back until the frame is received successfully based on a limited number of frames retransmissions and a predefined (acceptable) delay. The expected arrival time of a frame at the receiver is typically more than a round trip delay time and only after this time a NACK is triggered (Figure 2).

2. - System Reference Model

The system reference model is composed of a host server and a WCDMA network system as shown in Figure 1. The server can be located within the wire-line or can be part of the wireless network. Presently, the server is assumed to be the part of the WCDMA wireless system and the protocols of the interconnection to the core network at the physical layers are assumed specific sub-network protocols. Since the objective of the research is confined to the upper layers, such protocols are not reviewed here. As such, the main functions of the server in this case are the compression of raw video data into H.264 video streams, RTP/UDP/IP packetization, and transmission of packets to the wireless network.

The main functions of the core network include: adapting the incoming data from the wire-line network into the wireless environment protocol and transferring this data to the appropriate radio network controller (RNC) unit and to the base station (node B). The RNC controls the communication link status and performs erroneous frames retransmissions working at the higher layers either at the transmitter or at the receiver. The ARQ schemes configured here are basically part of the RNC functions at the transmitter

and receiver sides where the corresponding node B (base station) functionalities are seen as an agent providing data frames to the RNC.

The node B does the physical layer process transmitting an error robust RF signal to the wireless channel as well as providing robust signal detection at the receiver side. In this model, the physical layer is considered as an agent introducing errors to the transmitted frames. Thus, the link layer gets diverse bit error rates patterns added to the bit stream. In order to make feasible the support of next-generation services such as video streaming, the UMTS-WCDMA system is considered. It supports video streaming and a wide range of additional applications.
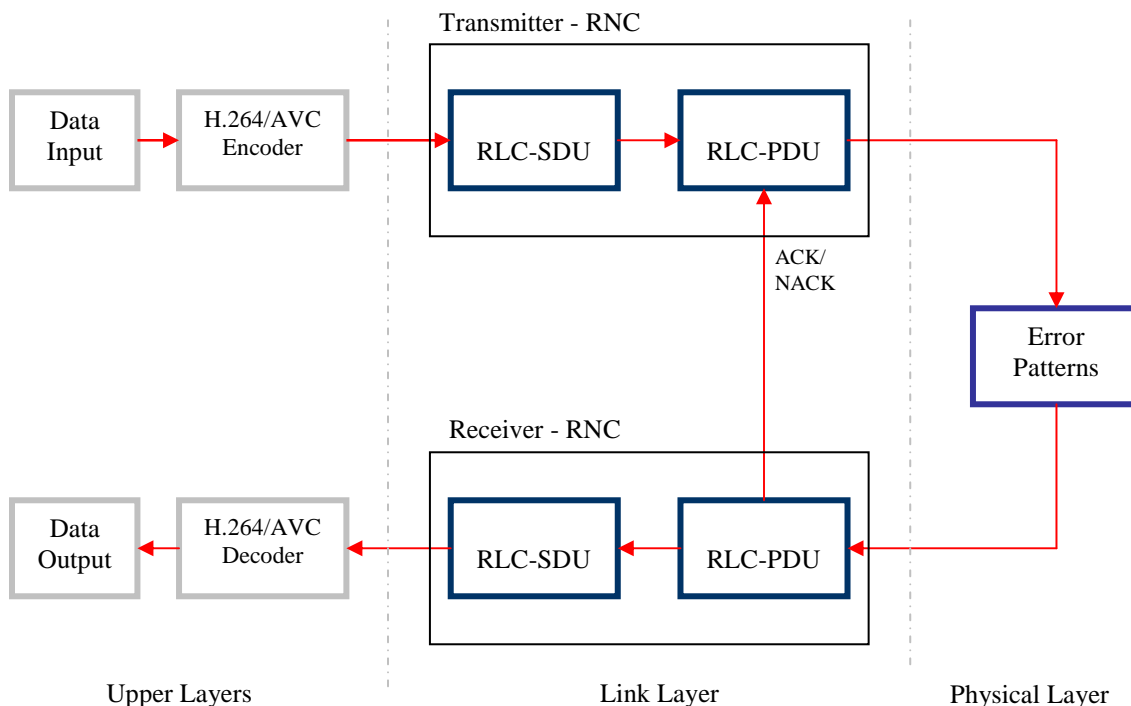

Figure 2: Simulation Model

Since the main interest here is the wireless channel (where its main components are located in the Radio Network Control and Node B), the core network is assumed for modeling purposes as a component not affecting end-to-end transmission quality. Thus its functionalities are not included in the software model. The interconnections to/from the core network to both the server and RNC/Node B are assumed to be through fiber optic, offering enough bandwidth, avoiding congestion delay, and introducing no errors, thus emulating an error-free channel for packet transmissions.

In short the current modeling and simulation environment refers to a video source plus a H.264/AVC video encoder (that reside in a host server) and a simplified WCDMA link-layer. Compression, packetization, and transmission of packets are assumed to be done by the host server. In the WCDMA wireless domain, the functionalities of the RNC-link level are characterized, mainly for the MAC and RLC protocols at the transmitter and receiver side, defined by the 3GPP [8]-[10].

3.- H.264/AVC Architecture Functionality

H.264/AVC consist of two conceptual layers: (i) Video Coding Layer (VCL), which is the lower coding layer defining a specific compression quality for its efficient video representation, and (ii) Network Adaptation Layer (NAL), which translates the compressed sequence of bits provided by the VCL layer into a format suitable for a specific network delivery or storage medium. This representation can be in a packet format to be transported over RTP/UDP/IP (Figure 3) by any kind of real-time wired or wireless networks or in a byte stream format to be delivered to any circuit-switched network. VCL and NAL are media aware; that is, they may know the properties and constraints of the underlying networks, such as the prevailing or expected packet loss rate, Maximum Transfer Unit (MTU) size, and transmission delay jitter. The VCL exploits this knowledge when it adjusts error resilience features, such as intra macro blocks rate and coded slice size.

A coded video sequence in H.264 consists of a sequence of coded pictures. A coded picture can represent either an entire frame or a single field. Generally, a video frame can be considered to contain two interleaved fields, a top field and a bottom field. A picture may be split into one or several slices. Slices are a sequence of macro-blocks which are processed generally in scan order. Slices are self-contained and can be decoded without the use of data from other slices. Each slice can be coded using different coding types such as (a) the "I slice" which is coded using intra-prediction; (b) the "P slice" using inter-prediction with at most one motion compensated prediction signal per prediction block; (c) the "B slice" using inter-prediction with at most two motion compensated prediction signals per prediction block; (d) "SP slice" where the P slice can be switched between different pre-coded pictures; and (e) the "SI slice" where the I slice is switched in this case. Detailed information on the H.264/AVC standard is given in [1].
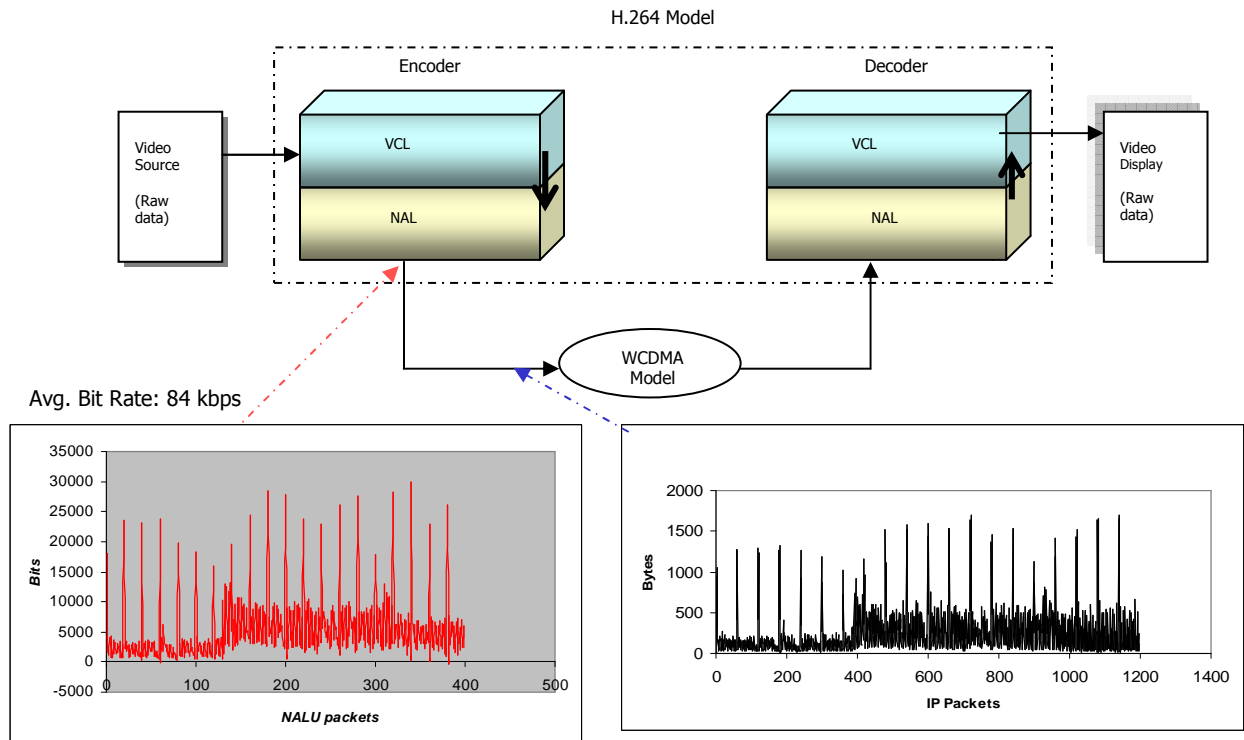

Figure 3: H.264 Models Interconnection

The H.264 source code is written in C and is available in the public domain [6]. The input and output of the source code are in the file format. The uncompressed video is the input to the encoder in a raw file YUV format (4:2:0), Y for luminance and UV for chrominance. The encoder output produces two files, the first one is a compressed file for transmission with extension xx.264 and the second one is for reference purposes, in order to compute the losses statistics and is presented in YUV format. On the other hand, the decoder decodes the received information which is in xx.264 file format and processes the de-packetization and de-codification.

The encoder and decoder have configuration files and need to be set up properly according to the transmission scenario. The majority of this supplied information is constant along the time interval the transmission and reception take place. There is some dynamic setup information that will be adjusted according to the network conditions in an automatic manner, and according to the initial configuration.

4. - WCDMA Link Layer Model

The WCDMA down link reference model includes the transmitter (RNC, Node B), receiver (UE). The corresponding physical layer is characterized with added interferences to the bit stream (Figure 4). As it is known, one of the main functions of the physical layer is to make the signal robust against interferences through the implementation of diverse error correction schemes at transmitter and receiver sides. However, such functions are not implemented here, because the present interest refers only to the number of erroneous bits not recovered at the physical layer. Thus, the main RLC functions considered at the user plane alone are considered. These refer to the segmentation/reassembly, PDU frame length, and diverse ARQ schemes.

In the simulations performed the PDU is typically segmented every 80 octets and transmitted at 64 Kbps within a transmission time interval of 10 ms. The segmentation generates a new PDU frame, which is used to transport the data on the underlying physical layer. Two headers are added in this process (PDCP and RLC) which are used to signal the boundaries of every frame, indicate the frame type, and frame sequence. This information is used to reassembly the frames at the receiver side. The RLC-PDU header is composed of 4 bytes, within the first two bytes is contained the sequence number (12 bits). The next two bytes contain the Length Indicators (of the last octet of each SDU contained in a PDU) plus framing information. The PDCP header maintains the PDCP SDU sequence number for its delivery to lower and upper layers (2 bytes).
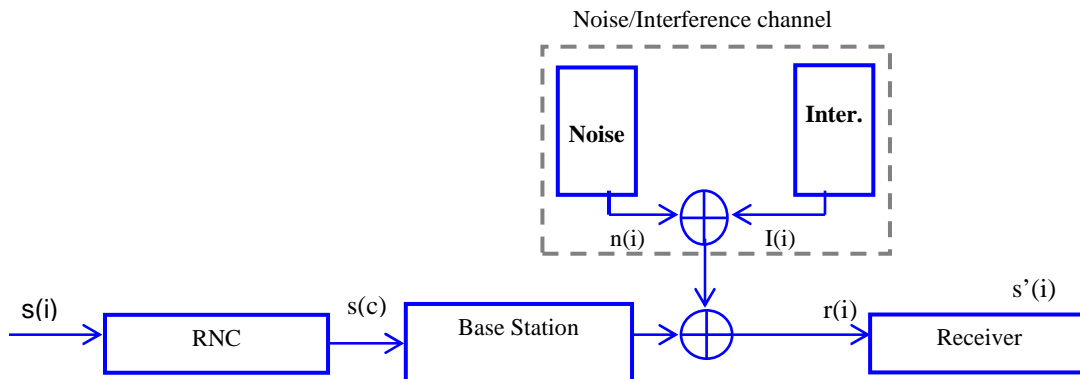

Figure 4: *WCDMA DL Reference Model*

The physical layer is simulated by binary error-patterns encountered at the transport channel in the receiver side. These error-patterns are taken from [13]. These are obtained from measurements at the transport channel under different wireless environments. Following are typical error pattern files characteristics suitable for data streaming transmission and used in this simulation (Table 1). Simulations performed with each BER scenario are repeated 10 times. In each simulation run error patterns of the same BER were added, but with different random starting positions. Then the results are averaged between 10 measurements obtaining more confident results. The diverse bit error patterns basically add errors to the received data at the link layer introducing the respective error rates.

N	File name	Bit rate	Length	BER	RLC PDU size	Mobile Speed
1	18681.3	64 kbps	60 s	9.3_x10^{-3}	640 bits	3 km/h
2	18681.4	64 kbps	60 s	2.9_x10^{-3}	640 bits	3 km/h
3	Wcdma_64kb_3kph	64 kbps	180 s	5.1_x10^{-4}	640 bits	3 km/h

Table 1: Bit Error Patterns (File length: 1480 Kbytes)

Software Model

The link layer model is simulated according to the 3GPP recommendation [10] [11], and it is based on the common test conditions for RTP/IP over 3GPP/3GPP2 reference software model provided in [13]. This software has been adapted to transport the H.264/AVC packets according to diverse transmission characteristics. The link layer model is presented in Figure 5.
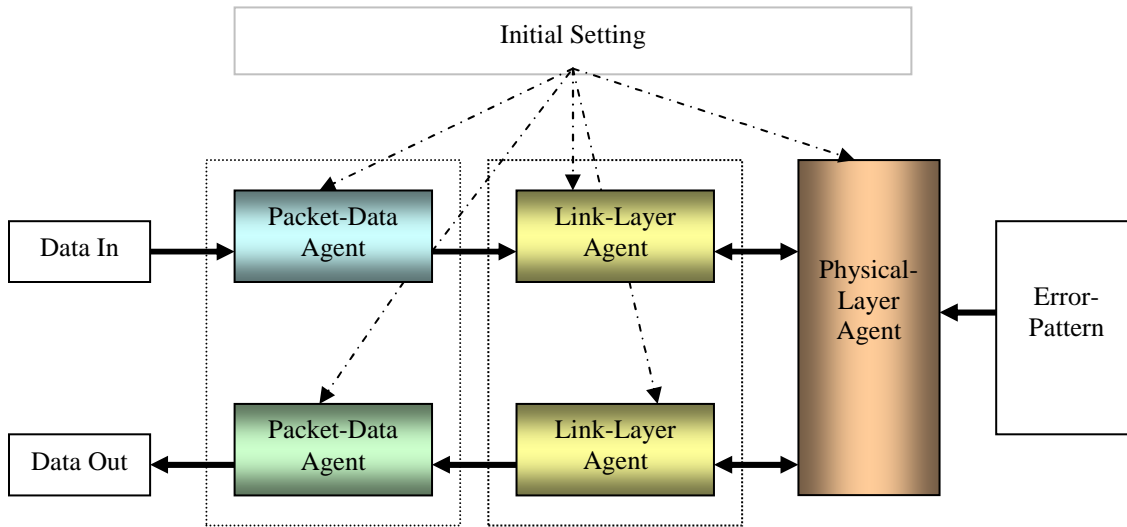

Figure 5: Link-Layer Software Model

This software is composed of three main functions and three external input data. The packet-data agent performs the RTP/UDP routines, specifically the PDCP functions. The link-layer agent performs the main RLC functions, and the physical layer agent adds the error-patterns to the bit stream. The model accepts IP packets and produces at its output video packets as well. The main setting of each function of the model is provided by an external file. The data flowchart of the model at the transmission side is presented in Figure 6 and at the reception side in Figure 7. To account for the transmission delay, the simulator has an incorporated time clock, which is triggered whenever an RLC frame is transmitted and accumulates the transmission and retransmission time of each RLC frame according to the ARQ scheme

and errors reported to the link layer. Figure 6 presents the link layer algorithm flowchart corresponding to the transmitter side up to when the packet is delivered to the physical layer. In case of buffer overflow a signal is generated. The packetization and transmission characteristics are as follows: Overhead compressed RTP/UDP header (3 bytes), PDCP packet header (1 byte), PDCP length (1 byte), and RLC PDU frame header (4 bytes). The RTP/UDP packet's header of each input packet is chopped in the RLC-SDU layer to 3 bytes (no robust header compression is performed).

The link layer function at the receiver side performs frame sequence reordering, unacknowledged (UM), acknowledged (AM), and transparent error correction modes. The error patterns introduced define a specific BER for different transmission scenarios to a mobile user. It is assumed in this model that the link used in the control plane for ACK and NACK signaling is free of errors.
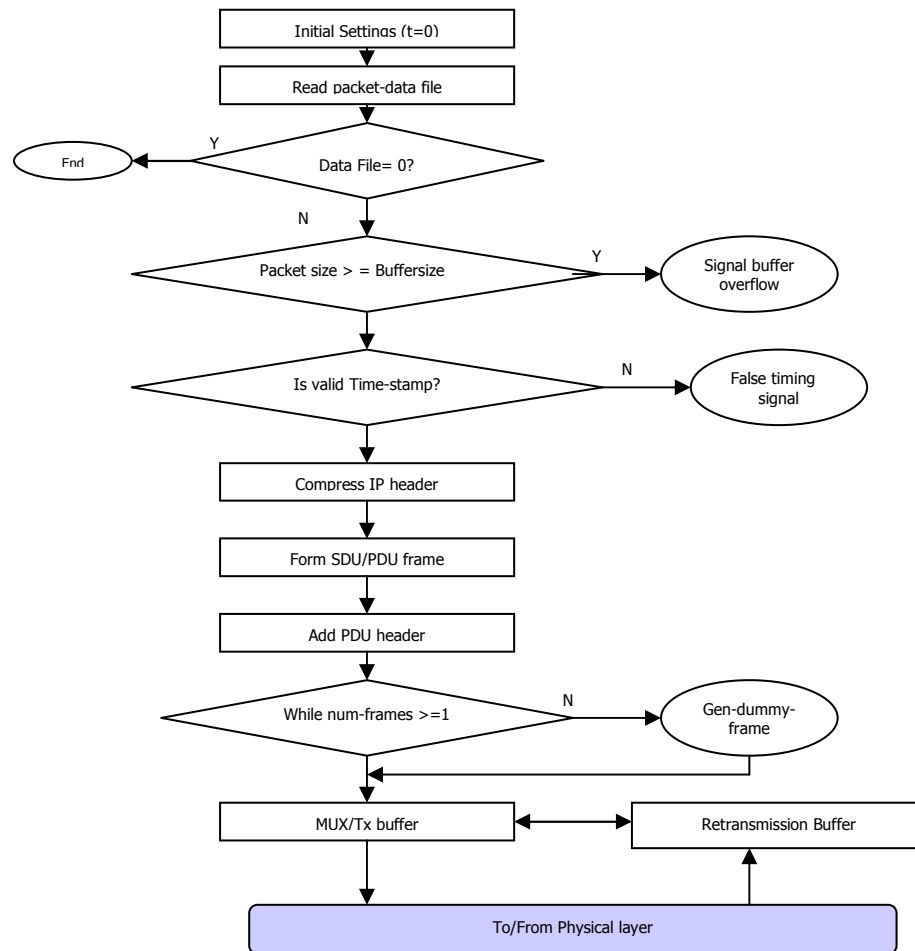

Figure 6: Transmission Flowchart

Hence, the RTP/UDP packets are the input to the WCDMA system. Those packets are encapsulated previously according to the application layer protocol. Then, each RTP/UDP packet is encapsulated into one PDCP packet that becomes a radio link control service data unit (RLC-SDU). The PDCP layer may perform header compression in order to reduce the overhead transmitted on the air interface. An RLC-PDU is of fixed length for each simulation and is determined when the bearer is setup (for example 80 octets). Since the packets arriving into the WCDMA transmitter are of varying length, and generally those packets are greater in length than the required frame size to be transmitted over the physical layer, they should be segmented to the required frame size. The description of this segmentation process is as follows and is presented graphically in Figure 8.

The segmentation is performed if the IP packet size is greater that the RLC-PDU frame size. The first step is to compress the IP header to increase the payload rate, in this simulation the header is set to 3 bytes. To the already chopped header and its payload is added a packet data control protocol (PDCP) header, this new formed frame is called Signal Data Unit (SDU). This SDU packet has the same variability of the IP packet. Thus, according to the PDU size previously set up, the SDU frame is segmented accordingly, forming two or more PDUs. When the segmented SDU frame does not fill an integer number of PDU, padding bits are introduced, increasing the link layer overhead and decreasing the payload rate. Therefore, for a good throughput, the PDU length and the SDU frame might be continuously packetized in order to avoid padding bits.

The RLC-PDU frame is mapped to the physical layer, where it is added a CRC trailer for error detection. The radio link control can perform re-transmission if one PDU frame is corrupted otherwise the frame is discarded along with its corresponding SDU. There are three modes of retransmission: persistent, acknowledged, and un-acknowledged. In the persistent scheme retransmissions take place up until no errors are detected and all erroneous frames recovered. The number of retransmissions is limited according to the transmitted video application, thus in the acknowledged mode the number of retransmissions need to be set up properly. At the receiver side, in Figure 7, the reassembly process is performed, requiring an in-order delivery of the PDU frames. This process introduces delay and jitter proportional to the number of retransmissions used.
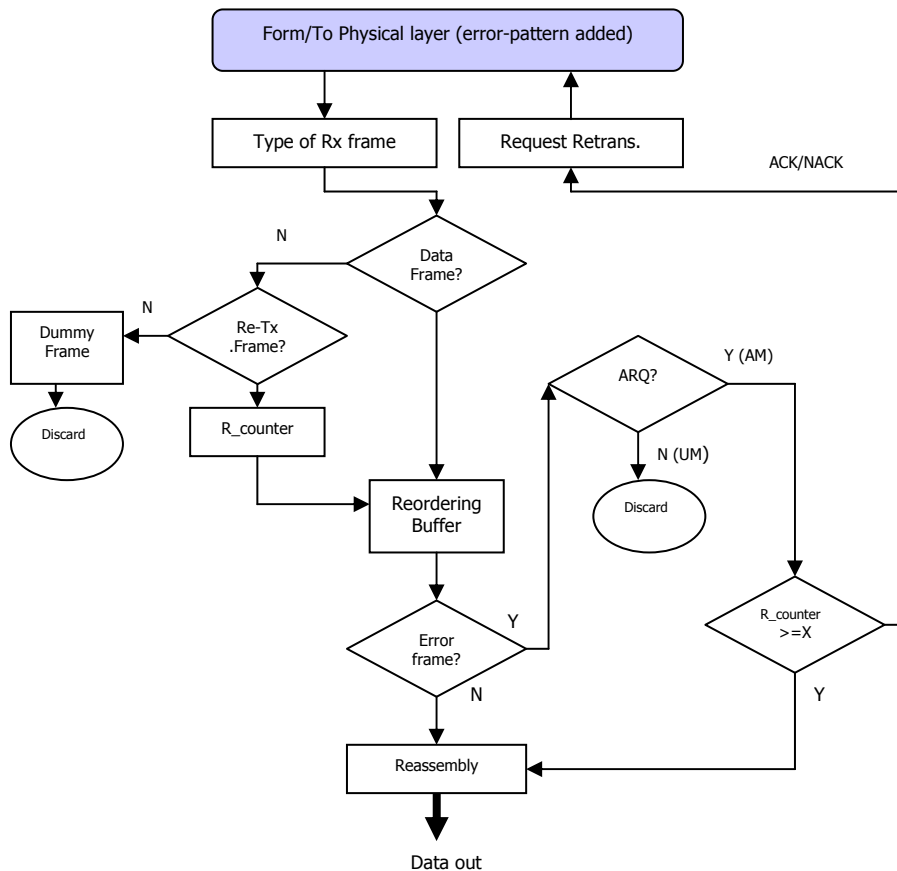

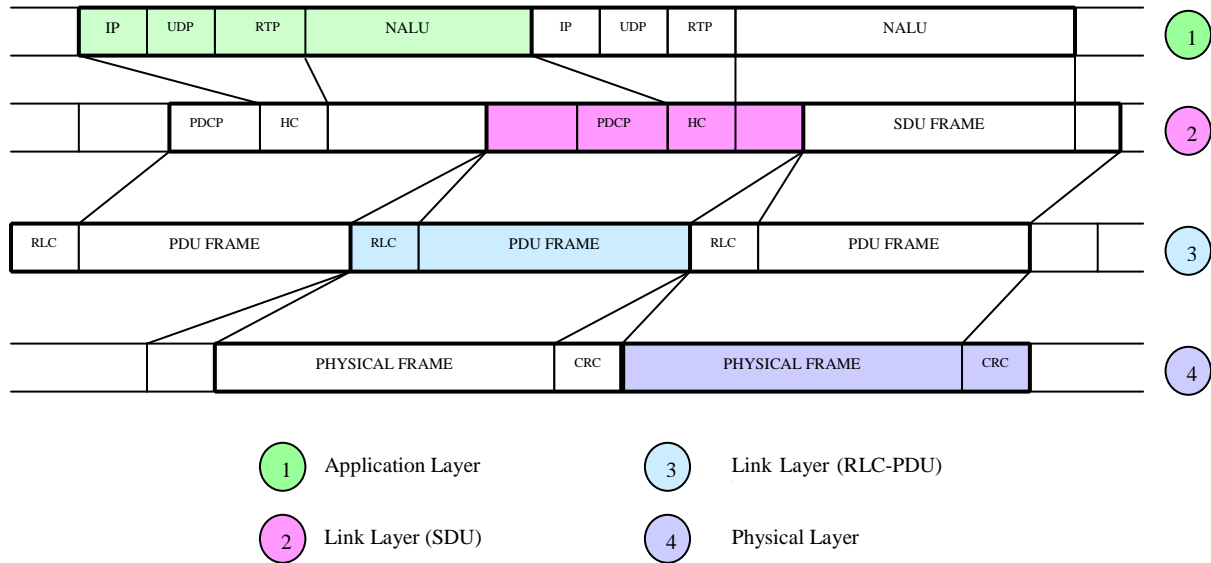
Figure 7: Receiver Flowchart (UM, AM)

IP	UDP	RTP	NALU	IP	UDP	RTP	NALU		1
	PDCP	HC		PDCP	HC		SDU FRAME		2
RLC	PDU FRAME	RLC	PDU FRAME	RLC	PDU FRAME				3
	PHYSICAL FRAME	CRC	PHYSICAL FRAME	CRC					4

1	Application Layer	3	Link Layer (RLC-PDU)
2	Link Layer (SDU)	4	Physical Layer

Figure 8: WCDMA Packetization- User Plane Protocol Stack

5. - Performance Evaluation and Results

The 'foreman' uncompressed video sequence is used as input file to the H.264 encoder, and the number of uncompressed video frames to be encoded and its frame rate are set in a configuration file. The frame width and height of the source are set for mobile devices (176x144 pixels). The extended profile is used and the encoder output is also set for RTP packets delivery. The period of I frames is selected to be 10, it means that the encoder will output one I frame every 10P and 10B frames. In addition, only one I frame has been considered for encoder evaluation purposes. This information is previously set up in the H.264 and WCDMA model configuration files.
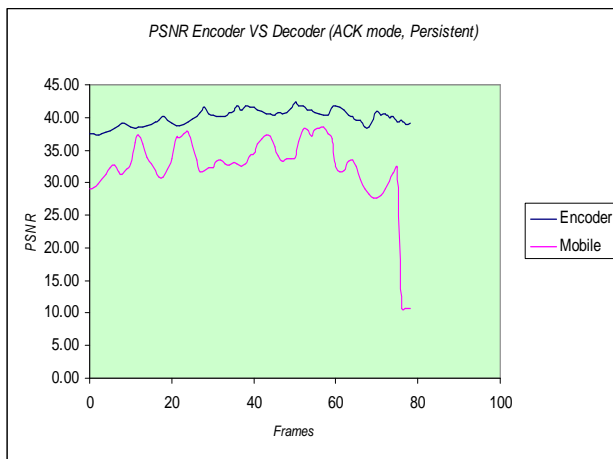

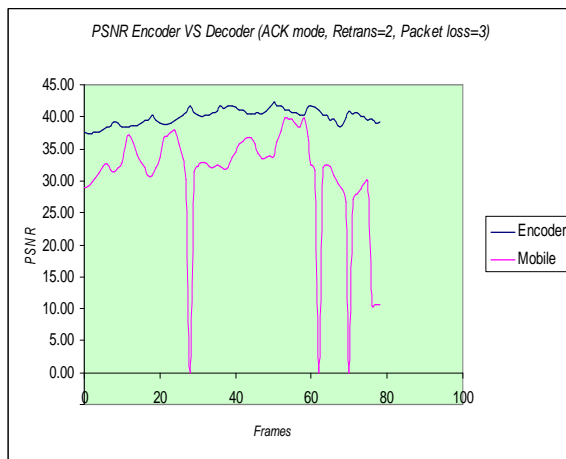*Figure 9: Acknowledgement Mode (persistent)*   *Figure 10: ACK. Mode (two retransmissions)*

As can be seen in figure 9, there is an acceptable degradation of the PSNR at the receiver side, which is originated by the error pattern introduced at the physical layer but without packet losses due to persistent retransmission. The total amount of packets transmitted and received is 81 but at the cost of some delay.

Even though the BER ($3.9x10^{-3}$) introduced is higher, the ARQ capabilities allow to recover the lost packets. Figure 10 presents a packet loss rate of $3.7x10^{-2}$. The number of packets lost in this case is 3, obtained by retransmitting the erroneous frames twice. This scheme presents a high packet loss ratio, and the video presentation is degraded completely from frame 65 to 76. It requires an ARQ=3, 4 implementation to recover the lost frames.
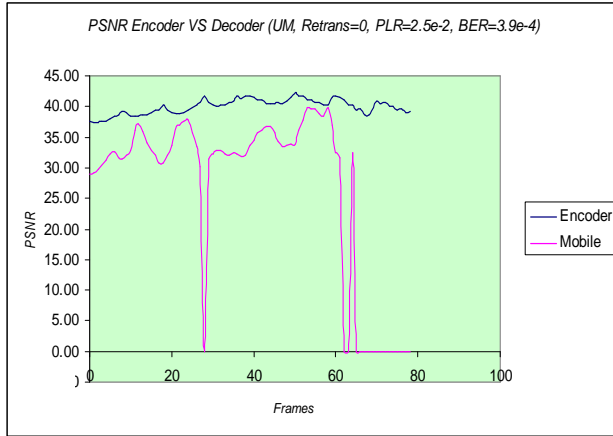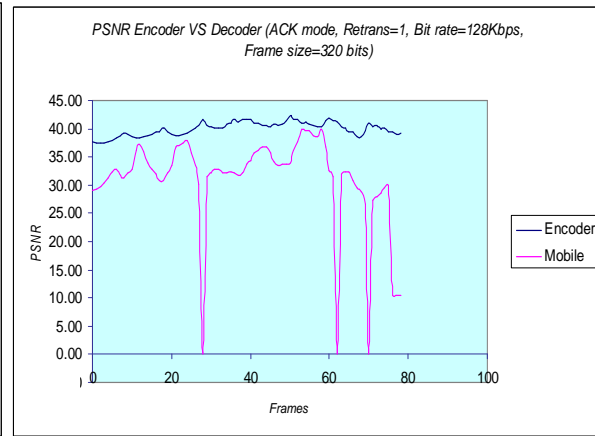

Figure 11: UM vs. PSNR

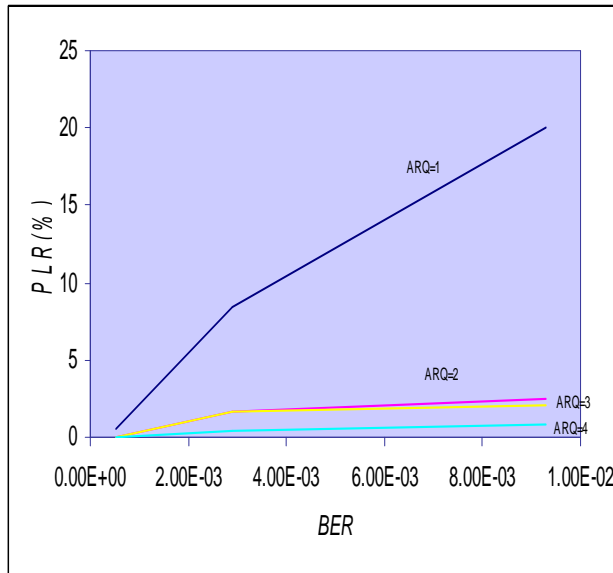*Figure 12: AM (higher bit rate)*

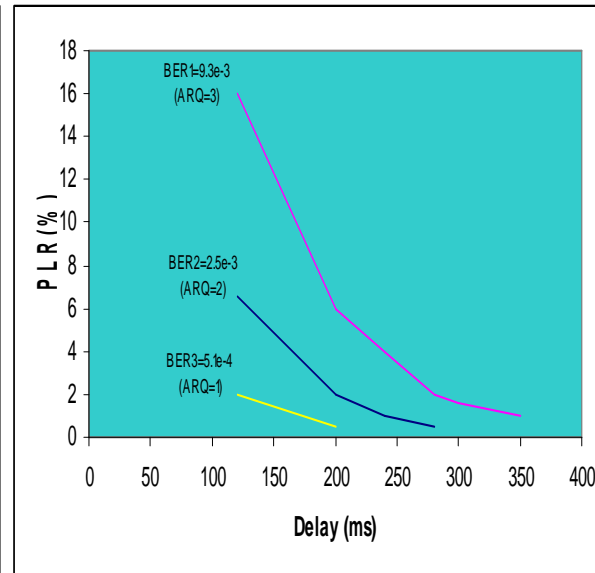*Figure 13: Packet Loss Ratio Vs BER*

Figure 14: End-to-End delay Vs Packet Loss Ratio

Figure 11 shows the quality results for the unacknowledged mode (UM) case. In this case there is no retransmission of corrupted frames. It shows that the number of erroneous frames is higher after packet number 62, thus the video decoder is blocked. Figure 12 was obtained by varying the bit rate and the frame length. The error pattern ($5.0x10^{-4}$) starting position was set to byte 1143200. The result obtained is similar to Figure 10 that was set with worse transmission conditions but with ARQ=2. The reduction of the frame size helps to recover more easily erroneous frames avoiding the loss of packet-data.

Figure 13 presents the variability of the PLR when the BER is increased on the physical layer and when ACK mode is implemented at the link layer. It presents the simulation results for three ARQ

11

configurations. With ARQ=1 the packet loss is higher and it is not suitable for video streaming transmission, but with ARQ=4 an acceptable PLR (1%) is obtained with a BER of $9.3x10^{-3}$. With a BER of 10^{-4} or 10^{-5} and with at least 3 retransmissions most of the information packets are recovered. The effect of implementing ARQ versus delay at the link layer is displayed in Figure 14. It is observed that the end-to-end delay is increased proportionally to the number of retransmissions. When the physical layer introduces high BER, the RLC requires more retransmissions to maintain a desirable PLR. This variation for three cases is also displayed in the previous plot. It can be seen that an application service can specify a BER of $5x10^{-4}$ which may require at most one retransmission.
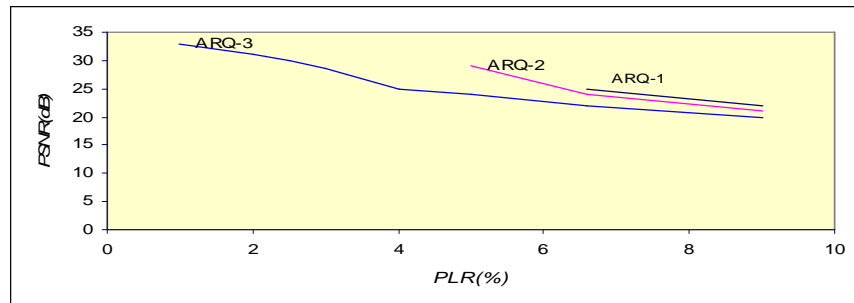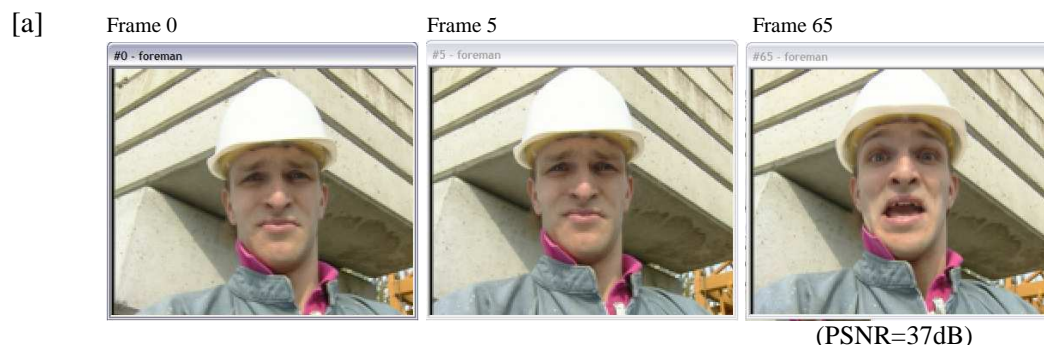

Figure 15*: PSNR Vs PLR*

Figure 15 shows that the received video quality expressed in PSNR (dB) is directly affected by the packets loss ratio. When the number of retransmissions is increased the overall PSNR is consequently increased, reducing the packet loss ratio but at expenses of higher delay. These results were evaluated under a high bit error rate of $2.9x10^{-3}$. The implementation of ARQ=3 allows to obtain an acceptable PSNR at a packet lost ratio of 1%.
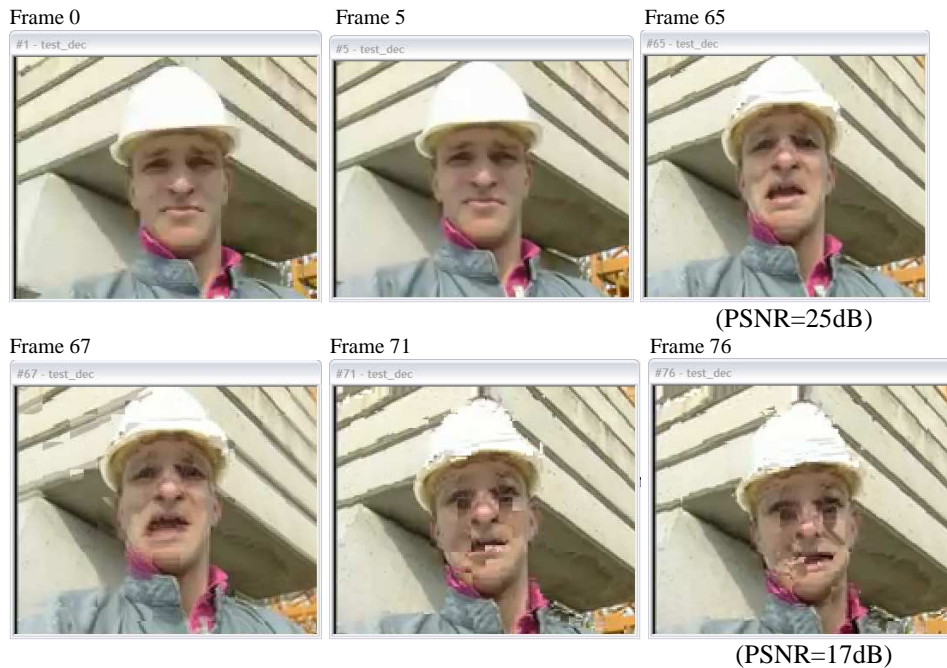
Picture Quality Degradation

Sequence [a] is the picture sequence at the transmitter side for a Foreman sequence - file input, with 15 fps and QCIF of 144X176 resolutions. Sequence [b] is the picture sequence at the mobile receiver (no error concealment is implemented).

The following sequences of pictures are an example of the video quality degradation. The physical layer adds a BER=$3.9x10^{-3}$ with 64 Kbps transmission rate for an ARQ=2. We observe that because frame number 62 is lost (I-frame, Figure 9) the following pictures will be distorted. Since ARQ=2 is not sufficient to recover the erroneous packet necessary for decoding, the decoder is blocked and no output is produced for pictures 76-80. We observe that the PSNR level starting at picture sequence number 67 is of poor quality and considered inadequate for video representation.

[a]   Frame 0   Frame 5   Frame 65

(PSNR=37dB)

[b]   Frame 0   Frame 5   Frame 65

(PSNR=25dB)

Frame 67   Frame 71   Frame 76

(PSNR=17dB)

6. Conclusions

The transmission of packets over the wireless channel needs stringent requirements to comply with minimum QoS specifications. The QoS is in direct relation with the channel bit error rate. Hence, the error-correction schemes are necessary to achieve the minimum required QoS. When the data is compressed video packets, it turns out to be more critical because the loss of one packet may represent the loss of one picture. In such case it is necessary to implement error control techniques at the physical, link, and application layers.

The specifications of the WCDMA system make suitable this system for packet video transmission. It provides by its wide-band transmission and CDMA modulation at the physical layer robustness against interferences and even security. At the link layer, the implementation of ARQ (AM) is another technique of error control, and can be set-up according to the transmission scenario. Implementing the RLC-ACK scheme at the link layer makes possible the transmission of packet-switched video, inclusive at corresponding BERs of 10^{-3}, 10^{-4} with an acceptable video quality presentation at the end-user. This is possible because streaming applications allow end-to-end delay (< 2 seconds) for retransmissions, buffering, and processing. Therefore, simulation results show that the RLC-ACK mode can be flexibly configured to trade off the required reliability and maximum delay allowed in the RLC layer.

References

[1] ITU-T Recommendation H.264, "Audiovisual and multimedia system", May 2003.
[2] T. Rappaport, *Wireless Communication Principles and Practice,* 2nd Edition, Prentice Hall, 2002.
[3] T. Ojampera and R. Prasad, "An overview of air interfaces multiple accesses for IMT- 2000/UMTS.", *IEEE Communication Magazine*, Vol. 39, pp. 82-95, Sept. 1998.
[4] 3GPP TR25.885, "UMTS Technical report rel.5, Specification", Sept. 2001.
[5] S. Wenger and Hannuksela, "RTP payload format for H.264 payload", *RFC-3984*, Feb. 2005.
[6] JVT Reference software: *http://bs.hhi.de/~suehring/tml/download/*
[7] 3GPP TR26.937, Transparent end-to-end PS streaming service, RTP usage model, 3/2004.
[8] 3GPP TS 25.301 V6.3, "Radio interface protocol architecture", 2005.

[9] 3GPP TS 25.321 V6.5, "MAC protocol specification", 2005.
[10] 3GPP TS 25.322 V6.2, "RLC protocol specification", 2005.
[11] 3GPP TS 26.234 V6.0, "Trans. end-to-end PS streaming service (PSS): Protocol and codecs", 2005
[12] ITU-T H.323, "Packet based multimedia communication systems", 2003.
[13] V. Varsa, M. Karczewicz, G. Roth, R. Sjöberg, T. Stockhammer, G. Liebl, "Common test conditions for RTP/IP over 3GPP/3GPP2", ITU-T SG16, VCEG-N80, Santa Barabara, CA, USA, Sept. 2001.

Authorization and Disclaimer