Performance of Video Telephony Services in UMTS using Live Measurements and Network Emulation

Matthias Malkowski
Communication Networks, RWTH Aachen University, Faculty 6
Kopernikusstr. 16, 52074 Aachen, Germany
Email: mal@comnets.rwth-aachen.de

Daniel Claßen
P3-Solutions GmbH
Dennewartstr. 25-27, 52068 Aachen, Germany
Email: dcn@p3-solutions.de

***Abstract*— While speech services in mobile communication systems are investigated quite well, video telephony services are relatively novel in this sector. Because of the market penetration of camera equipped mobile phones the video telephony service is expected to become a widely used service. In this paper an introduction of the protocols used for video telephony in UMTS is given. Concepts for the performance evaluation in terms of video (PEVQ) and audio quality (PESQ) are presented and utilized. Evaluations are performed by both live measurements and network emulation. The results show that there is quite potential for improvements related to video telephony in UMTS in terms of video quality and channel setup time. Finally, an improved radio bearer configuration is provided which aims at a better integration of video telephony services into the UMTS architecture.**

I. Introduction

New services are the major key to the success of upcoming and current communication systems. Video telephony is one of these services which is recently introduced to the UMTS User Equipment (UE) and network. For new (mobile) networks the support for such services with an adequate Quality of Service (QoS) is of primal interest for both customer and operator. Because recent communication systems have a large set of parameters and a rather flexible resource management the network configuration for operators is quite sophisticated and not straight forward. For equipment and service developers the new networks offer many implementation dependent aspects, e.g. algorithms and protocol parameters, which might influence performance dramatically. In order to make performance evaluations of deployed and upcoming communication systems and services several tools have been developed. In this paper a framework for performance evaluation of video telephony services in UMTS is presented and utilized.

Section II gives a short introduction of the protocols that are specified for video telephony services in UMTS. Furthermore, the developed tools that implement those protocols are presented. Section III presents the performance evaluation concepts which include live measurements and network emulation. The performance measures used for the evaluation of video telephony are presented, too. Results using both concepts are presented in section IV. Furthermore, a proposal for a tighter coupling of the video telephony stack with ths UMTS stack which improves the Quality of the video telephony service is given.

II. Video Telephony Principles

Video telephony in UMTS is implemented using the protocols of the ITU-T Recommendation H.324 on top of the protocols of the data link layer specified by the 3GPP.

A. 3G-324M Standard

This section provides a brief overview of the 3G-324M standard [1] to which video telephony in UMTS conforms [2]. Figure 1 shows the architecture of the protocol stack. The control system, also called H324M System Control, is the main entity which handles the whole call setup in the application layer. If the physical connection is established by the lower layer, the System Control initiates sending first control messages. These commands conform to the H.245 standard [3] and they are responsible for establishing logical channels for audio and video data transfer. The control channel segmentation and reassembly layer (CCSRL) segments these massages, while the numbered simple response protocol (NSRP) handles retransmissions, error protection and the responses of incoming control commands. It is a kind of data link layer for the control plane in this context.

In the H.223 entity [4] the outgoing data from the control channel as well as the audio and video data from already opened media channels is multiplexed into common packets called mux PDUs (Protocol Data Units). In the PDU header whose error protection is dependent on the chosen mux level (0,1,2,3), it is defined how many bytes of each channel the PDU consists of.

After PDUs are formed they are sent to the lower layers and subsequently to the remote party. Incoming packets are also demultiplexed in the H.223 entity and then distributed to the according decoder after the adaptation layer (AL) handled the error control. The AL provides three different error protection methods for the data of the regarded logical channels. AL1 contains no error protection, AL2 provides only a CRC error detection and AL3 provides the opportunity of retransmitting erroneous packets.

For coding and decoding audio and video data mobiles currently use AMR speech codec [5] and MPEG4 [6] or H.263 video codec [7]. The video codec to use is determined by a capability exchange procedure handled by the control plane.

B. UMTS Specifications

The protocol stack configuration which is used for video telephony services in UMTS is rather simple. The PDUs of the H.223 entity have a fixed size and are transmitted using the transparent mode of the RLC layer [8]. In the MAC layer [9] the RLC PDUs are mapped solely on one transport channel. Hence, no multiplexing in MAC layer is required. The Signalling Radio Bearers (SRBs) of the control plane are multiplexed with the video telephony Radio Bearer (RB) in the physical layer and do not limit the theoretical data rate available for the video telephony Radio Bearer (RB). The detailed configuration of the circuit switched conversational radio bearer which is used for video telephony in UMTS ([10], [11]) is illustrated in figure 2.

C. Emulation of 3G-324M Terminals

The company P3 Solutions GmbH developed a video telephony emulator. With this Video Telephony Stack (VT Stack)

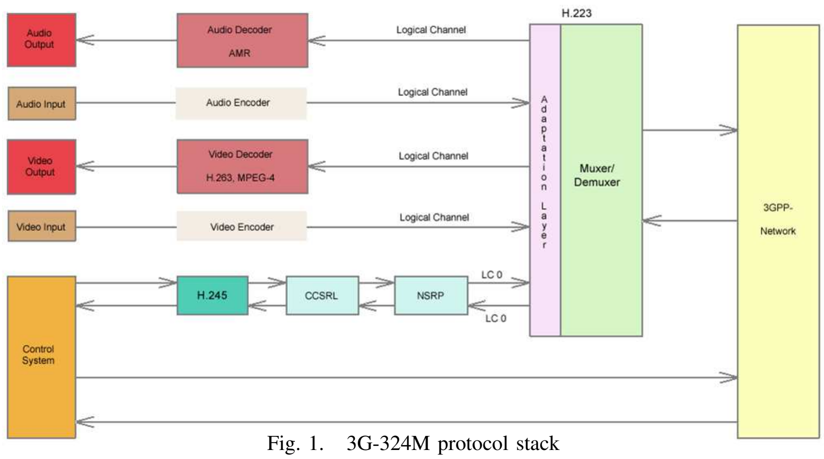

Fig. 1. 3G-324M protocol stack

it is possible to make a video call over the UMTS network against a mobile or another instance of the Stack. Furthermore, a video call between two VT Stacks connected over the internet protocol using TCP or UDP is possible.

The implementation of the stack corresponds to the architecture illustrated in figure 1. The only difference is that no encoders are used. The transmitted media data is read from precoded audio and video files. Reference streams exist to measure audio and video quality in a post processing by comparing them to the received and recorded streams.

D. Implementation of the UMTS Protocol Stack

A detailed implementation of the UMTS protocol stack of the Radio Access Network (RAN) is provided by the Chair of Communication Networks of RWTH Aachen University. The bit accurate protocol implementation is part of the UMTS Radio Interface Simulator (URIS).

III. Concepts

This paper focuses on two general principles in which the evaluations are performed. The first one uses live measurements in 3 configuration setups. The second principle is based on a network emulation.

The benefits taken from the complementary advantages and disadvantages of live measurements and network emulation are used to validate and calibrate the models which are included in the simulator by the live measurements and to improve the accuracy and reduce the variances which are common to live measurements.

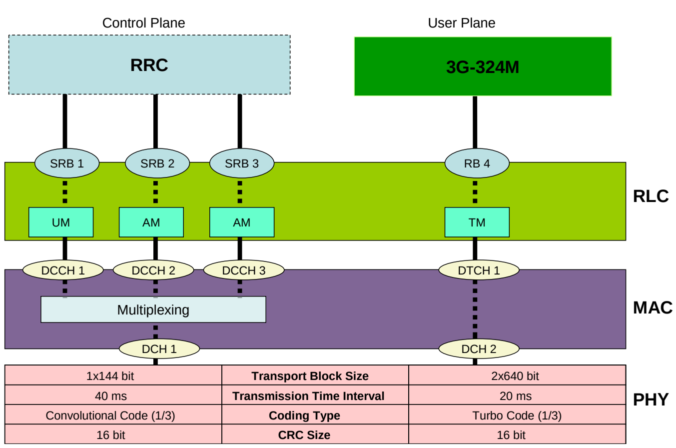

Fig. 2. Integration of video telephony in UMTS

A. Live measurements

Figure 3 demonstrates three different possibilities of live measurements. In a) two VT Stacks running on a PC transmit the video telephony data over a UMTS data card. Doing so a video call between two mobiles is emulated. In this case both directions uplink and downlink are under test. Figure 3b) shows another method to measure only one transport way. Here the remote party is connected to ISDN. In c) the VT Stack is running against a real UMTS mobile.

The main advantage of live measuring in general is that tests are always made under network conditions as users also find in real life.

B. Network Emulation

In order to eliminate certain factors found in real networks a network emulation has been chosen as a second approach for performance evaluation. Those factors primarily include changing channel conditions caused by varying traffic load/interference. For comparisons using different parameter sets fixed reference scenarios are very useful. Furthermore, the network emulation allows full control over parameters, where only operators or even vendors have access to. The analysis of protocol enhancements, which are not yet specified, are also possible using such a network emulation.

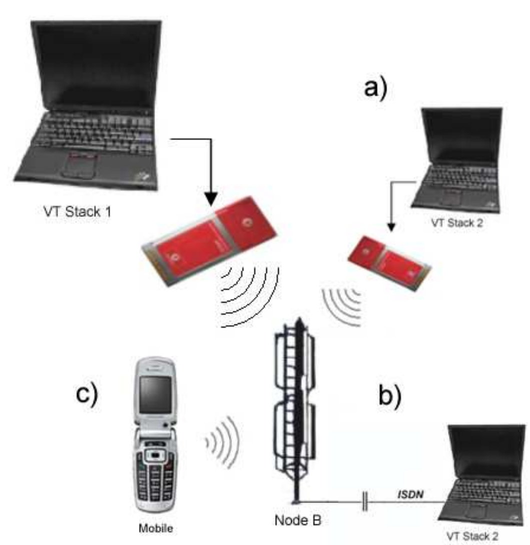

Fig. 3. Live measurement scenarios

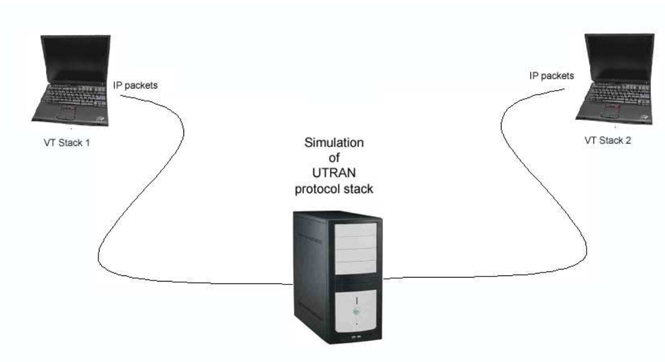

Fig. 4. Scenario with network emulator

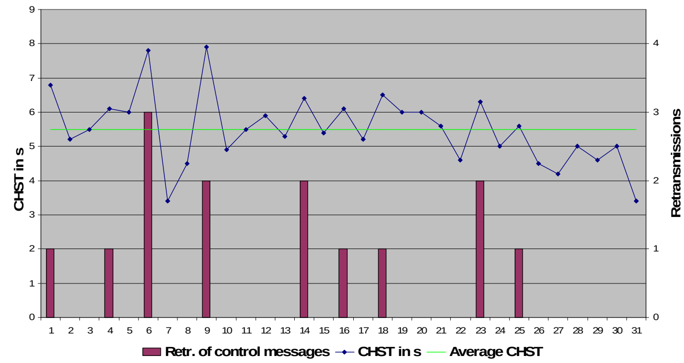

Fig. 5. Channel setup time

The traversing packets are realistically influenced by the following aspects. Traffic is delayed according to the lower layer protocols of UMTS (e.g. queueing, processing, retransmissions). Packets can be dropped in case a maximum number of retransmissions have been performed or no retransmissions take place (unacknowledged or transparent mode). Furthermore, packets can be modified to contain erroneous bits in case they are transmitted using the transparent mode and the decoding in the physical layer failed.

C. Measuring audio and video quality

The company OPTICOM GmbH provides a tool for measuring audio quality called PESQ [12]. PESQ stands for "Perceptual Evaluation of Speech Quality". The algorithm conforms to the ITU-T standard P.862 [13]. The received test signal is compared to the undistorted reference signal by modeling the human acoustic perception. As result a value similar to a Mean Opinion Score (MOS) is calculated. It ranges from -0.5 (worst) to 4.5 (best). The algorithm also provides a voice activity detection (VAD) and a silence detection.

OPTICOM also developed an algorithm for video quality evaluation called PEVQ ("Perceptual Evaluation of Video Quality") [14]. In contrast to PESQ it is not standardized, yet. Nevertheless, PEVQ standardization is in progress.

The functionality of PEVQ is similar to the PESQ algorithm. Here the human visual system (HVS) is modeled. The result is also a MOS like score with the difference that the maximum value is 4.759.

In our experiments we recorded an own reference video. It takes nearly 30 seconds and consists of three different parts which are typical for video telephony. The first ten seconds only show somebody's mimic and movements of his head, while in the second part the camera is swayed around. In the last ten seconds one stationary object is filmed with some hand shiver.

IV. Results

In this chapter we will present exemplary results of the different measurement methods introduced before. First, we will present channel setup time measured in live networks. The next results depict the effects of discarding erroneous transport blocks. Furthermore, the applicability of the acknowledged mode in the RLC layer is studied. Finally, we propose a new configuration which improves certain aspects of the UMTS video telephony standard.

A. Channel setup time

The channel setup time (CHST) is one essential quality aspect that is very important to minimize. The channel setup time is the time interval between accepting an incoming call and succeeding the establishment of the logical media channels which not includes the radio bearer establishment procedure of UMTS. This corresponds to the time that goes by after a user presses the button to answer an incoming call and he then could see the first video frame. During this period of time all the control messages are exchanged. The reduction of this time would be interesting inasmuch as the user even has to pay for this time.

Figure 5 displays the measured CHST of each connection establishment during a drive tests. The average channel setup lasts about 5.5 seconds. The resulting time intervals range from a minimum of 3.4 seconds up to the maximum of 7.9 seconds. To point out the influence of the retransmissions the number of occurred control messages that are transmitted more than once is also inserted into the figure. It is obvious that the peak values are resulting from those retransmissions. Each session that has a CHST higher than 6 seconds contains retransmissions of control frames. Without these peaks the resulting average time would be nearly five seconds.

Generally, the CHST is very high in all measurements, even though the highest possible segment size of the NSRP protocol was used. The primary reason for the high values is the stop-and-wait nature of the NSRP protocol. Especially if retransmissions are needed the complete end-to-end delay affects the setup time.

B. Discard of erroneous transport blocks

In this section we analyze the influence of the UMTS feature to send transport blocks which still contain bit errors to the higher layers. In contrast to audio the video quality seems to react very sensitive to absent data. With high BLERs the quality is really bad and the decrease of quality occurs very fast. In order to abide video quality, packets should not get lost in lower layer. Even corrupted data packets should arrive in application layer. Hence, it appears like the behavior of the data card and of most mobiles is not optimal where disturbed packets are dropped although the transparent mode is chosen for video telephony. The two demonstrated PEVQ curves in figure 6 show that in this case it is better not to drop packets where the checksum failed. A maxmimum improvement of 1 quality step could be reached here.

C. Acknowledged mode for video transmission

In the H.324M standard the AL3 is using an automatic repeat request (ARQ) for video transmission. Results showed that such an end-to-end ARQ suffers from the large round-trip-delay (RTD) of the complete network and is not useable for video telephony services.

In this section we analyzed the video quality if the video information is transmitted on a logical channel which uses

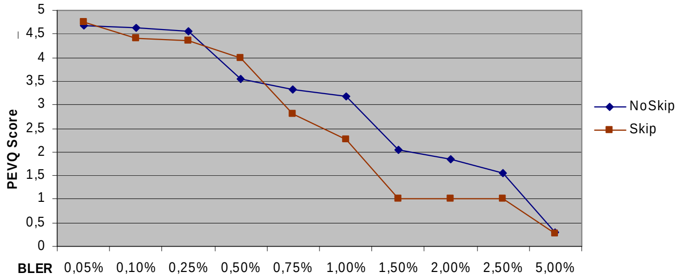

Fig. 6. PEVQ scores for simulator measurements

the acknowledged mode in the RLC layer. It can be expected that the delays become essentially smaller because the ARQ mechanism is directly located at the Radio interface.

The discard function of the AM is enabled and its timer is set to 200ms. Hence, a maximum delay is limited to nearly one frame for 4 to 5 frames per second. Of course, frames may get lost if packets that include the VOP startcode are discarded because of being buffered longer than 200ms. The maximum number of retransmissions is set to 4 in these example measurements.

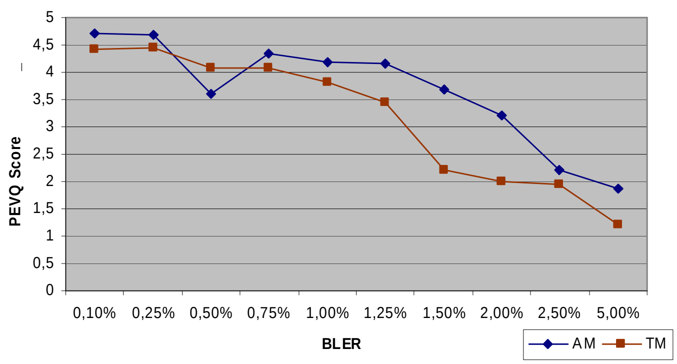

Fig. 7. PEVQ using acknowledged mode

Figure 7 shows the resulting PEVQ scores for each measurement. In general, the video quality is better with the configured AM. To explain the exception of the third result and to analyze the interesting delays caused by the retransmissions, the following figure 8 has to be considered.

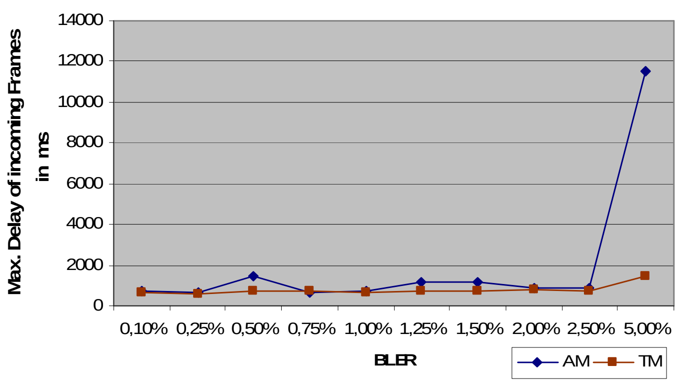

Fig. 8. Maximum delay using ackowledged mode

The ARQ method with the configured discard timer tends to result in loosing more frames, but the incoming frames then contain much less artifacts. This explains the better video quality.

The more interesting values are the maximum delays of single frames that are caused by retransmissions of single packets. Here the values of the AM-curve are always equal or greater than those of the TM-measurements.

The discarded frames are responsible for the worse quality result of the third video call. A high number of continuous lost frames leads to the maximum delay and thus to a bad PEVQ score.

The measurement results show that the video quality can be improved by changing the transport mode in the RLC layer. The disadvantage of the AM is obviously the frame delay caused by retransmissions. These delays are not very significant for low BERs. But for higher BERs the duration of freezed images are very high. It has to be deliberated if such delays are more annoying than totally distorted video frames that appear in transmissions in the TM. Of course more measurements have to be done in the future with different configurations to the discard functions to confirm if AM is the best solution. Unfortunately, a selective-repeat ARQ together with passing corrupted packets after a specified number of retransmissions is not available in UMTS. The AM is not allowed to pass corrupt data. Otherwise, a mixture of the AM and the TM functionality would be the best solution.

D. Proposal of a new RB configuration

At the end of this work an idea for an improved RB configuration is proposed in figure 9. The first visible difference to the existing architecture is a higher number of radio bearers. Each RB is configured to address the specific demands of the protocols and codecs which are directly connected to them. Using such a configuration the multiplexing functionality of H.223 would not be needed anymore.

In the proposed architecture the AMR codec is configured similar to the speech only service. Here, the AMR encoder separates its data in three different classes according to their importance. The data of each class is transmitted on a different transport channel which differ in the channel coding techniques. While the first channel with the highest importance and the second channel are encoded with a convolutional code with a rate of $\frac{1}{3}$, the last channel is encoded with a half-rate convolutional code. Additionally, a 12 bit CRC checksum is attached to the first channel. Furthermore, the rate matching attributes allow to control the puncturing of the coded information regarding their importance [15].

The described speech radio bearer is very flexible and shows the opportunities of UMTS. Data of different importance can be transmitted with different coding mechanisms and with different protection methods. Also adaptations to varying bandwidth conditions are possible.

In [16] tests with a H.263 video codec that provides data on two different prioritized layers have been introduced. The effects of an unequal error protection (UEP) were analyzed within this article.

The encoder operates with an average rate of 51.2 kbit/s for coding. The encoded data of one picture is divided into two layers. The high-priority layer, for example, contains the coding information, motion vectors and the lower frequency DCT coefficients. This layer produces data with an average rate of 33 kbit/s. The low-priority layer contains the high frequency DCT coefficients and operates with an average rate of 18.2 kb/s. The lower-priority data is less error sensitive. If errors occur in this layer, the data can simply be dropped and the image is reconstructed with a smaller resolution. The high priority layer contains the most important data that is also very error sensitive. In this article two different ways of channel coding were tested for both layers. The conclusion including the consideration of the previous tests of an enabled AM leads

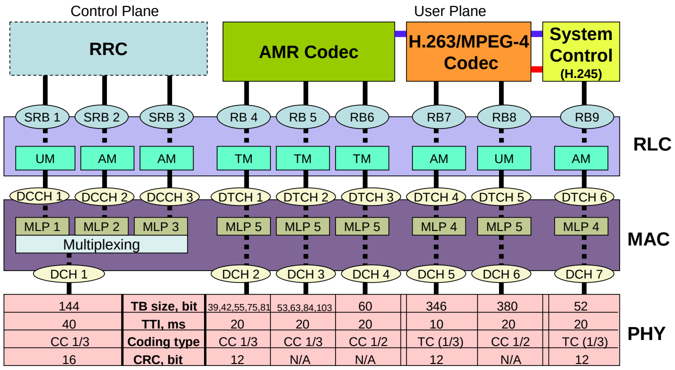

Fig. 9. Proposed video telephony RB architecture

to a RB configuration illustrated in figure 9. Beyond the three audio transport channels, two video transport channels are added as well as one control channel that is necessary for the configuration of the codecs. This is demonstrated by the channels that connect the system control to the audio and the video codec. The video transport channels are adapted to the measurements of [16] and to the previous results. The high priority data of the video codec is transmitted in the AM and is protected by a one third turbo code and an attached checksum. The TB size is marked with * because these values only correspond to the data rate of the respective H.263 codec. If an MPEG-4 codec is used that is also able to divide its data on several object planes, other TB sizes have to be configured. The TTI of 10ms is selected to be able to react faster to required retransmissions. For the second video channel the UM is chosen. Transmission errors could be detected and the data would be discarded. This may not have so dramatic consequences as explained above. The control channel is also protected very well. The selected transport block size refers to currently used control messages. Most commands and responses could then be transmitted within a few TTIs. With such a configuration the original NSRP can be replaced. The channel setup time would be also reduced with this architecture. The capacity of the control channel is chosen a little bit higher so that the summation of all application channels capacities would exceed the actually used rate of 64 kb/s. Here, the idea is that most of the control messages are only transmitted at the beginning of each session. Later on the control channel does only need a small bandwidth. This is defined in detail in the transport format combination set (TFCS).

For all channels the MAC logical channel priorities (MLPs) are proposed. The data of the control channel and the high-priority video channel should be preferred when the respective RLC buffers are checked for data available for transmission. Both channels have the same priority, but do not influence each other because the control channel mainly sends commands at the beginning during the channel setup.

The most important question that has to be answered before introducing such a new architecture is the compatibility to the existing systems. One possible solution adapted to the proposed RB should be described briefly. Therefore, the new proposed system configuration has to be considered as an extension to the actual 3G-324M standard.

At first, when establishing a new connection the old configuration should be used. That means that one conversational 64 kbit/s transport channel in the TM has to be established. With a new H.245 control message for example it could be requested if the terminal supports the new mode. If this request is acknowledged, the application has to request RRC messages to reconfigure the VT radio bearers. Then the respective dedicated channels could be opened. Of course for such a solution new H.245 messages would be required. At the end the vendors have to discuss if such a new system would be economic and if it is realizable to introduce to their equipment.

V. Conclusion

In this paper several standards for UMTS and video telephony services in UMTS as well as measurement tools which are conform to these standards were presented. With the VT Stack the emulation of mobile terminals for video telephony is possible. With a connection to the UMTS network emulator the whole video call, including different network conditions, can be emulated. With the introduction of PESQ and PEVQ exists the opportunity to measure audio and video quality.

By way of examples the results of analyzing several aspects of improvement have been shown. Here the handling of erroneous received packets in the VT Stack as well as the effect of discarding packets in the network emulator are presented. As a result it seems to be better to forward every received video packet to the video decoder to abide video quality and not to discard them in lower layers for any reason. For disturbed packets this even happens in real mobiles and the 3G data card. Furthermore, the applicability of the AM for video telephony has been shown. Finally, a proposal for a video telephony architecture which elliminates several disadvantages of the currently implemented video telephony system in UMTS has been presented.

References

[1] ITU-T Recommendation H.324, “Terminal for low bit rate multimedia communication,” 2005.
[2] 3GPP TS 26.911, “Modifications to H.324,” 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.1.0, 2004.
[3] ITU-T Recommendation H.245, “Control protocol for multimedia communication,” 2005.
[4] ITU-T Recommendation H.223, “Multiplexing protocol for low bit rate multimedia communication,” 2001.
[5] 3GPP TS 26.073, “Adaptive Multi-Rate (AMR) speech codec; C-source code,” 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.0.0, 2004.
[6] ISO/IEC 14496-2, “Information technology - coding of audio-visual objects - part 2: Visual,” 2005.
[7] ITU-T Recommendation H.263, “Video coding for low bit rate communication,” 2005.
[8] 3GPP TS 25.322, “Radio Link Control (RLC) protocol specification,” 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.5.0, 2005.
[9] 3GPP TS 25.321, “Medium Access Control (MAC) protocol specification,” 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.6.0, 2005.
[10] 3GPP TS 25.331, “Radio Resource Control (RRC) protocol specification,” 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.7.0, 2005.
[11] 3GPP TS 34.108, “Common test environments for User Equipment (UE) conformance testing,” 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.0.0, 2005.
[12] OPTICOM GmbH, “User Manual PESQ OEM,” Manual Version 1.6, 2003.
[13] ITU-T Recommendation P.862, “Perceptual evaluation of speech quality (pesq), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs,” 2001.
[14] OPTICOM GmbH, “PRELIMINARY User Manual PEVQ OEM,” Manual Version 0.9.2, 2005.
[15] B. Sebire, K. Pedersen, T. Bysted, “Flexible Layer One for the GSM/EDGE Radio Access Network (GERAN),” *International Conference on Telecommunications*, vol. 1, pp. 310–316, December 2003.
[16] C. Dubuc, D. Boudreau and F. Patenaude, “The Design and Simulated Performance of a Mobile Video Telephony Application for Satellite Third-Generation Wireless Systems,” *IEEE Transactions on multimedia*, vol. 3, no. 4, pp. 424–431, December 2001.