

# Voice and Video Telephony Services in Smartphone

**Valeria Loscri', Mauro Tropea, and Salvatore Marano**

*Department of Electronics, Informatics and Systems Department (DEIS), University of Calabria via P. Bucci, 42C, 87036 Arcavacata, Rende (CS), Italy*

Received 3 August 2005; Revised 13 February 2006; Accepted 2 March 2006

Recommended for Publication by Fary Z. Ghassemlooy

Multimedia telephony is a delay-sensitive application. Packet losses, relatively less critical than delay, are allowed up to a certain threshold. They represent the QoS constraints that have to be respected to guarantee the operation of the telephony service and user satisfaction. In this work we introduce a new smartphone architecture characterized by two process levels called application processor (AP) and mobile termination (MT), respectively. Here, they communicate through a serial channel. Moreover, we focus our attention on two very important UMTS services: voice and video telephony. Through a simulation study the impact of voice and video telephony is evaluated on the structure considered using the protocols known at this moment to realize voice and video telephony.

Copyright © 2006 Valeria Loscri' et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

In the last few years, with the downturn of the new economy and particularly the telecommunications sector, mobile operators have been rethinking ways to deliver innovative and cost-effective services by providing IP connectivity to every mobile device. Moreover, new components and new modules have been created as smartphones. The latter are characterized by an opportunistic choice of operating systems. The more common operating systems developed for these components are *Windows Mobile*, *Linux Mobile*, and above all *Symbian*. Moreover, other technologies have been studied to enable similar terminals to add the computer functions to telephony functions [1, 2]. In this work we consider an architecture implying the subdivision of the modules around the applications: the interface in a unique component and telephony functions in another module. The two modules (each equipped with its own processor) communicate through a serial channel. The standard considered here to have the serial communication is the RS-232 [3]. Specifically, an extension to this last standard is considered, European telecommunication standards institute (ETSI) 07.10. The standards to support the multimedia telephony services on a smartphone: H.300s, G.700s, H.260, T.120s [4] were analyzed. Then in order to validate our approach, accurate traffic models are described for simulations before presenting and analyzing simulation results.

## 2. MULTIMEDIA TELEPHONY ISSUES: THE STATE OF THE ART

Multimedia telephony is a delay-sensitive application: an upper limit of 150 ms of end-to-end delay with low variation must be ensured to guarantee the operation of the telephony service and user satisfaction [5]. Packet losses, relatively less critical than delay, are allowed up to a certain threshold since they can be compensated by loss recovery mechanisms at the codec level. For example, the G.729 codec with good voice quality requires packet loss of less than 1 percent to avoid audible errors [5]. The standards used to ensure the voice and video telephony services are *H.300s*, *G.700s*, *H.260s*, *T.120s*. These standards cover all the categories of coding-decoding. The main video standards are

- (i) H.261: video codec for audio-visual services (64 Kbps);
- (ii) H.263: video codec for telecommunications less than 64 Kbps.

Audio standards:

- (i) G.711: pulse code modulation (PCM) for audio frequencies, use a B channel 64 Kbps;
- (ii) G.722: 7 KHz for audio codified to 64 Kbps;
- (iii) G.723.1: dual rate speech codec for telecommunications to 6.4 Kbps and 5.3 Kbps;
- (iv) G.728: codec a 16 Kbps with small delay using linear prediction.

Data standards:

- (i) T.120: protocols and services for multipoint Data conferencing.

Video window sizes:

- (i) NTSC—National Television Standards Committee, used in USA, Canada, and Japan.  $640 \times 480$  pixels;
- (ii) PAL—phase alternation by Line, used in Europe, Africa, and the Middle East.  $768 \times 576$  pixels;
- (iii) SECAM—sequentielle couleur avec memoire, used in France and Russia;
- (iv) CIF—common interchange format; optional with H.261,  $352 \times 288$  pixels;
- (v) QCIF—quarter common interchange format; used with the standard H.261,  $176 \times 144$  pixels.

The H.323 contains the audio, video, and data specifications applied to every telephony system. Examples are explained as follows:

- (i) integrated digital service network (ISDN): H.320 including: Video: H.261; Audio: G.711, G.722, G.728; Data: T.120;
- (ii) local area network (LAN): H.323 including: Video: H.261; Audio: G.711, G.722, G.723, G.728; Data: T.120;
- (iii) public switched telephone network (PSTN): H.324 including: Video: H.263; Audio: G.723.1; Data: T.120;
- (iv) internet: H.323 including: Video: H.263; Audio: G.723; Data: T.120.

Specifically, attention is focused on the ITU standard H.324 [6] describing multimedia terminals operating over PSTN (*public switched telephone network*). This is the point for the evolution of the new standard supporting video telephony services over mobile terminals [7].

### 3. PROPOSED ARCHITECTURE OF SMARTPHONE

In this work, a new type of smartphone characterized by a network interface GSM/GPRS/EDGE/UMTS, based on an architecture with two processors is considered. A discrete event-driven simulator was realized to evaluate the performances of the multimedia services on our smartphone. A simplified version of our terminal is shown in Figure 1. Two processors can be distinguished called, respectively, AP (*application processor*) and MT (*mobile termination*). A platform based on the GNU/Linux system was considered. In Figure 1 note how the AP and the MT are linked.

Two different operating systems are considered for the *mobile equipment* (ME) and the *terminal equipment* (TE). They are equipped with two different processors and the physical link is a serial channel. It is clear that ME and AP represent the same module, the *application processor* and TE or MT identify the *mobile termination*. The WTM (*wireless telephony manager*) is the software module that permits communication between the application layer and the mobile termination (the serial channel mentioned above). Through the WTM module it is possible that applications

running on a mobile terminal communicate with other remote applications through an available network communication (GSM/GPRS/EDGE/UMTS/Bluetooth...). The WTM has to manage the traffic and the resources available in the MT module. This is because the applications considered above can be concurrent. The WTM was designed as a versatile module permitting different types of communications to be managed through standard protocols (TCP/IP) or other types of protocols (files, data-streams, etc.).

Between the two modules AP and MT there are different types of interfaces:

- (i) AT commands (standard and nonstandard)
- (ii) *interprocessors communication* (IPC) through a multiplex serial protocol based on 3GPP TS 07.10 standard.

The latter solution to design our terminal was chosen. In this way, through the standard 3GPP TS 07.10, communication between AP and MT modules can be obtained. This standard permits some number of sessions over an asynchronous serial channel to be established. Each session can be used to transfer data, voice, fax, SMS, GPRS, and so forth. In this way it is possible to execute different applications simultaneously. Naturally the multiplexer protocol is not dependent on the specific AP and MT modules and it is designed for mobile terminals with a battery and for this reason it is equipped with *power saving* functions. The multiplexer protocol is characterized by different types of functionalities:

- (i) base;
- (ii) advance without error recovery;
- (iii) advance with error recovery.

The first option was chosen for consideration. This is due to the specific characteristics of the serial link considered in this work. In effect it is a simple physical serial channel and for this reason it is not necessary to consider error recovery. Through the use of the standard 3GPP TS 07.10 it is possible to have a virtualization of the serial channel. In fact, it is based on some number of virtual channels called *data link connection* (DLC). The standard does not specify the number of channels that has to be opened. In fact, in the standard it is only specified that the total number of channels must be greater than 63 and the 0 channel is a specific channel defined as control channel. Channels 1–7 have the same priority. An application was associated with each channel. Examples of applications are

- (i) SMS,
- (ii) voice call,
- (iii) GPRS data connections,
- (iv) UMTS data connections,
- (v) video Call (UMTS).

Based on these considerations we chose to consider 5 channels and the control channel. Hence, 6 channels were considered in all. In this work attention is focused on the specific traffic generated by a video calling and the performances of our smartphone considering this specific service will be evaluated.

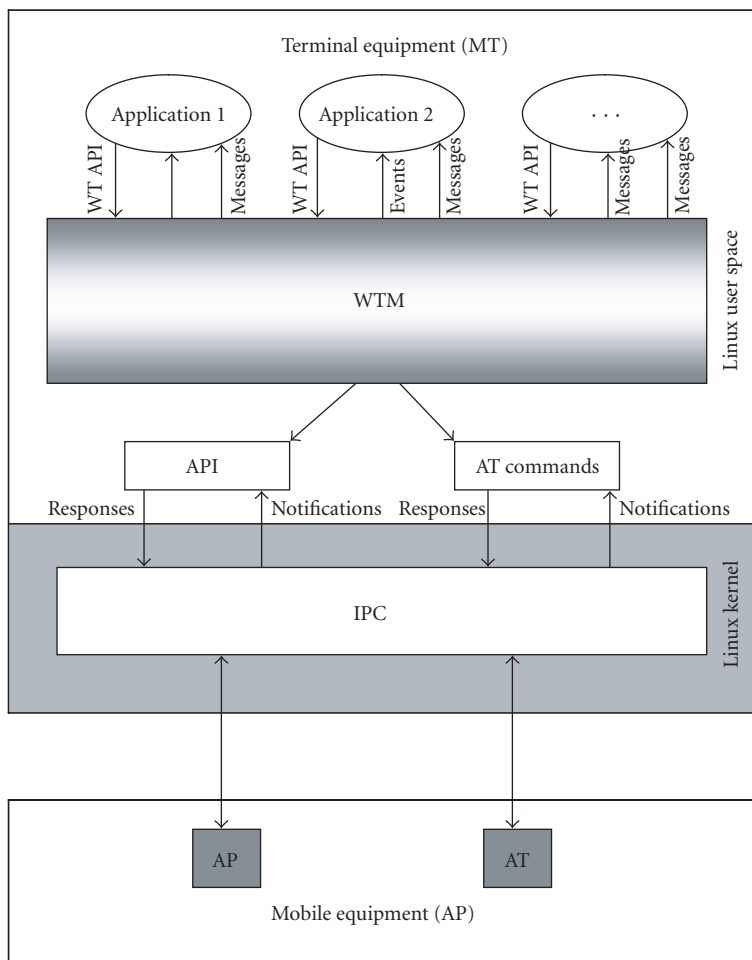


FIGURE 1: AP and MT.

## 4. VIDEO CALL ON SMARTPHONE

### 4.1. Overview

The video telephony service is characterized by delay requirements similar to those of voice services; due to the nature of the video compression BER requirements are more constraining than that of the voice. Specific UMTS have provided for video telephony services on a circuit switched connection where they have to use the ITU-T recommendations. *H.324M* or, as called by the 3GPP, *3G-324M* [4]. It is a specific case of the H.324 version that follows the annexed C. This recommendation covers the technical requirements for very low bit-rate multimedia telephone terminals operating over the general switched telephone network (GSTN).

H.324 terminals provide real-time video, audio, or data. The multimedia telephone terminals defined in this recommendation can be integrated into PCs or workstations, or be stand-alone units.

The terminal user specified from this standard has the structure as shown in Figure 2.

The general structure of the system is very similar to that of the original standard (H.324) [6]. Substantial differences

are present in the specific component use in order to fulfill every base function.

The annex C of the H.324 standard describes specific issues to allow the use of H.324 terminals in error-prone transmission environments. These issues include specific options for H.324 terminals:

- (i) the mandatory use of NSRP (numbered simple retransmission protocol);
- (ii) the use of robust versions of the terminal multiplexer;
- (iii) procedure for level set-up;
- (iv) procedure for dynamic change between levels during a session.

The 3GPP standard defines the UMTS/WCDMA requirements and also the structure and implementation of the 3G-324M standard as defined in TS 26.111 [8].

The network 3G-324M components include end-point, cellular terminals or PDA wireless terminal, base stations that support the circuit switched services and gateway that permit the interaction with the Internet network and a server that permits the supply to multimedia services on demand. The 3G-324M requirements that use the circuit switched

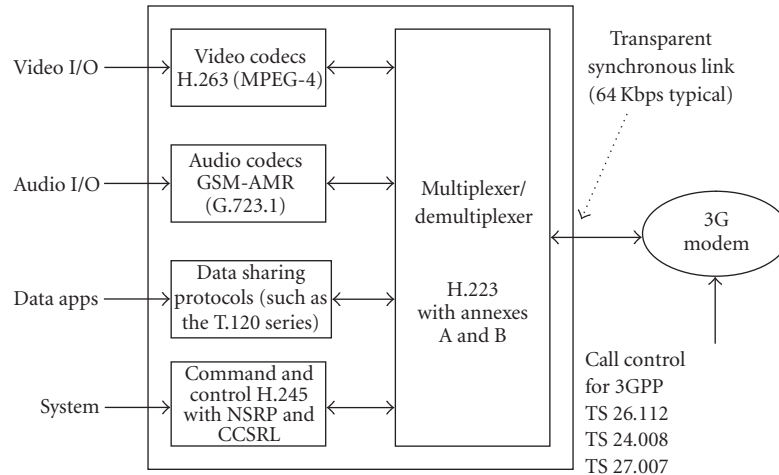


FIGURE 2: 3G-324M system diagram.

network allow multimedia conversational services with delay sensibility to be obtained, such as “video conferencing for personal and business use,” “multimedia entertainment services,” telemedical services, Surveillance, live video broadcasting and Video-on-demand (movies, news clips), besides the normal video call.

An appropriate interface has to be implemented so that a terminal can be interfaced with the external network. The UMTS/WCDMA network provides for the use of a specific UMTS modem that works with specific commands allowing multimedia applications to be set up and used. 3GPP defines a set of AT commands [9] that are used to set and manage the modem over 3G-324M terminals.

After a connection is successfully established then a communication channel for data which will travel to 64 kbps will be used. The call set-up is a time that the user must wait to have an audio-video connection.

Fundamental operations that a video call have to follow are as follows.

Audio-video transmission:

- (i) acquisition from video camera,
- (ii) codify video with H.263 encoder,
- (iii) codify audio with G.723.1 or ARM encoder,
- (iv) multiplexing audio/video H.223,
- (v) H.245 (to do controls),
- (vi) adaptation to UMTS network target for transmission to the outside,
- (vii) framing 07.10 for sending on serial channel.

Audio video reception:

- (i) 07.10 “unpacking,”
- (ii) GSM/GPRS/EDGE “unpacking,”
- (iii) Audio/video demultiplexing,
- (iv) H.263 decoding,
- (v) audio decoding,
- (vi) dimension display scaling, color conversion e RGB16 packeting,
- (vii) display on LCD.

#### 4.2. Video codec

QCIF is an image format adapted to videoconference which has an acquisition bit rate that can vary from 10 to 30 frames per second (fps). The dimension in pixel is  $144 \times 176$ . These requirements, in agreement with ITU H.263 standard, are used commonly for the video codec that have to be transmitted on a channel with a bandwidth inferior to 64 Kbps. The QCIF is correlated to CIF (common intermediate format) because it represents a quarter of it, in fact, the CIF pixels are  $288 \times 352$ . The encoding is operated on an intraframe and interframe. The interframes are those images that are correlated to the previous ones, in particular they contain information only on the image differences with precedent images and then it is impossible to decode these images without, beforehand, having decoded the previous images. The intraframe, instead, can be decoded without the need for outside information.

CIF and QCIF images are subdivided into blocks, macroblocks, group of blocks, and complete images. Every block is formed from a square of  $8 \times 8$  pixel and every macroblock (MB) is formed from four blocks, therefore it is a square with a side of 16 pixels. Generally, these are luminance pixels. For every four luminance pixels there are a CB pixel and a CR pixel, then a macroblock is formed from four luminance blocks and one of chrominance CB and one of chrominance CR. Every group of blocks (GOB) is formed from  $3 \times 11$  macroblocks, for which reason, it can finally be asserted that an image in CIF ( $352 \times 288$ ) format is composed of 12 GOB and one QCIF ( $176 \times 144$ ) from 3 GOB.

#### 4.3. Audio codec

The audio codec represents an irreplaceable manner for the transmission and the computing of any audio signal, from a simple vocal signal to a complex musical one.

The H.324 standard previews the support to the audio codec G.723.1 [10] which permits the encoder and decoder audio to 5.3 Kbps and 6.3 Kbps.

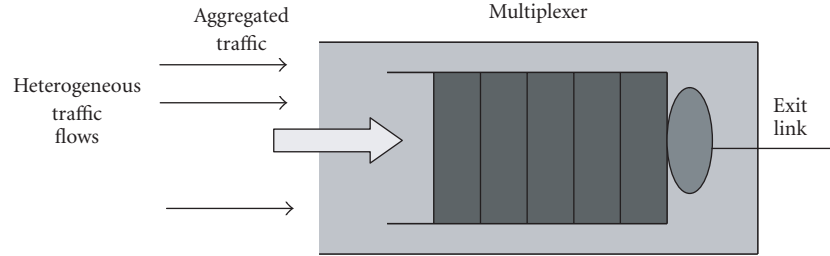


FIGURE 3: Markovian overlapping.

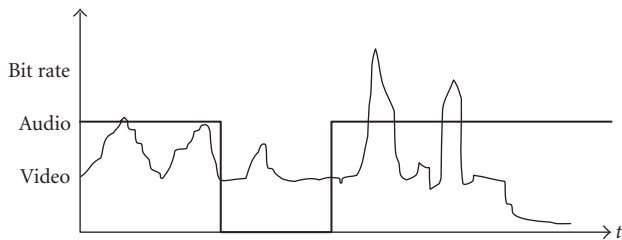


FIGURE 4: Audio and video traffic data overlapping.

The UMTS codec adopts a technique called AMR (adaptive multirate) [11]. The vocal encoder is a single vocal encoder integrated with eight possible speed of source: 12.2, 10.2, 7.95, 7.40, 5.90, 5.15, and 4.75 Kbps. The AMR bit rate is controlled from the radio access network and does not depend on the source activity. In order to make interoperability easy with the existing cellular networks, some speeds are equal to those already present in the networks before UMTS. The vocal encoder AMR switches its own bit rate every 20 ms of vocal frame.

The connection vocal AMR bit rate can be controlled from the network access as a function of the radio interference and vocal connection quality. For example, it is possible to use an inferior bit rate, during traffic peaks, like high traffic hours, so as to offer greater contemporary connection capability despite slightly inferior vocal quality.

## 5. SIMULATION RESULTS

The video call on the serial channel was simulated through the construction of an ad hoc simulator. The tests were conducted considering a number of applications running simultaneously on the smartphone for estimating the worst case of the serial channel. The multitasking on the serial channel is possible using the ETSI GSM 07.10 protocol. Video call simulation is difficult for the heterogeneity of the traffic; in fact, in our experiments, two different types of traffic are presented: audio and video traffic.

The first one is modeled with an ON/OFF model; instead, the second one is characterized by a great variability of bit rate without silence moment like in the audio traffic. This variability is, in effect, considered constant because the video

call reserves a fixed bandwidth of 64 kbps for the entire call duration.

The maximum dimension of the package that comes outside from multiplexer H.223 [12] is 254 with maximum of 4 bytes of the header.

In our simulations two different scenarios were considered:

- (i) fixed dimension of package. The simulation of this traffic is performed in order to consider the worst case for the serial channel performance;
- (ii) variable dimension of package. An algorithm is implemented that creates a package with a dimension that goes from 100 bytes to 254 bytes.

The parameters considered for the performance evaluations are as follows.

- (i) Serial channel bandwidth occupation: it is calculated as the number of bits inside the channel buffer.
- (ii) Number of packet loss: it is calculated as the number of packets that is not possible to put inside the buffer.
- (iii) Transmission overhead: it is calculated as the overhead introduced by the 07.10 protocol for transporting the information.

### 5.1. System model

It is very hard to generate traffic that well simulates a video calling, because the data represented are very heterogeneous. This heterogeneity is well represented by Figure 3.

In our simulation we considered Markovian overlapping in which two different kinds of traffic, video, and audio, with different bit rates, are overlapped. In Figure 4 the overlapping is shown.

The audio was modeled as an ON-OFF [13] source traffic, vice versa the video cannot be modeled in this way because there is a continuously bit rate variation and the transmission does not have silent moments as in the audio. Also, if there is a variability of the bit rate, since the video calling sets a bandwidth to 64 kbps in circuit switching, the total traffic will be a constant bit rate. The maximum dimension of packet is constant and it is fixed through the multiplexer H.223 and it is 254 bytes with a maximum of 4 bytes of header; the traffic can be simulated in two different ways.

TABLE 1: Simulation parameters.

Simulation parameters	
First campaign	
Channel velocity (kbps)	57.6, 115.2, 230.4
Video call duration (s)	120, 240, 360, 720
Packet dimension (bytes)	258
Second campaign	
Channel velocity (kbps)	57.6, 115.2, 230.4
Video call duration (s)	120 (da webcam)
Packet dimension (bytes)	Variable (100–258)

### (i) Fixed dimension packet

The dimension of the packet is maintained fixed based on the dimension of the multiplexer H.223. This dimension has been fixed as the maximum dimension of the packet. This assumption is not realistic because this signifies that there are continuous and rapid scene changes and consequently continuously coding within the audio codifier that overcharges the packet. In terms of simulation it is interesting to evaluate it because we consider the worst case channel condition. In practice, this kind of traffic generation is implemented allocating the necessary bandwidth permitting 258 bytes of traffic to pass on the serial channel with a bit rate of 64 kbps.

### (ii) Variable dimension packet

In order to generate this traffic a 3GPP file was generated. Audio and video dimension frames were extracted from this file and they were used as input of the serial channel in our simulator.

Audio and video data traffic were evaluated together in terms of bandwidth occupation, overhead, and so forth, because the main objective in this work is to establish the correct dimensioning of the buffer of the serial channel to permit the video calling to work well in a similar structure as considered above. The correct dimensioning of the channel and the simulation of the video calling data traffic permit a data transfer to be realized with a delay that is represented only from the propagating delay on the data channel. In fact, it cannot introduce another kind of delay for this kind of traffic, otherwise the same structure cannot be considered to realize a similar device.

## 5.2. First simulation modality

In the first simulation type a set of video calls are simulated that are generated with a fixed packet dimension. The simulation parameters are shown in Table 1.

It is interesting to study the video call channel bandwidth occupation (Figure 5). It is possible to observe an increase of occupied bandwidth with the increase of channel speed. This is observed for all the durations considered. Here only the case of a duration of 120 seconds is shown, because the

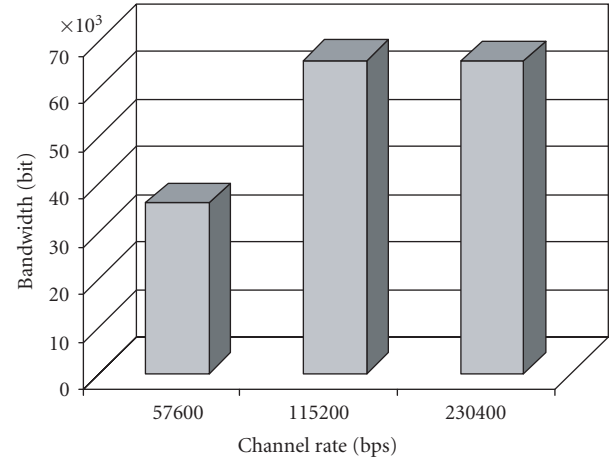


FIGURE 5: Bandwidth variation TX (120 sec.).

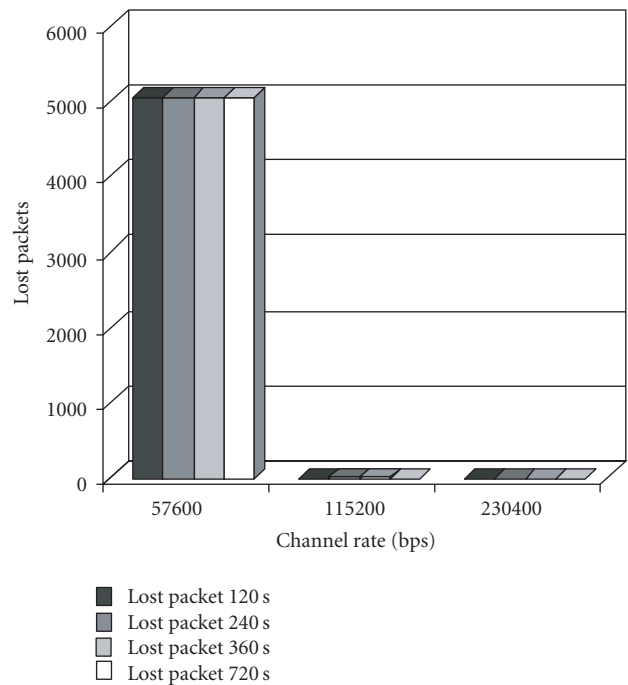


FIGURE 6: Packet Loss for different rate values.

graphic slope is equal also for the duration of 240, 360, and 720 s.

Figure 6 shows the packet loss for different types of channel speed and different video call durations. In this case it is possible to observe that for the velocity of 57.6 kbps the system presents a packet loss that has been calculated at about 2% with respect to the overall packet sent in the channel. This is due to an inferior channel speed versus the standard speed for this type of application in an UMTS networks of 64 kbps. Then, for the channel rate of 57.6 kbps it is normal to observe a little loss. Instead, in the other cases, as it is possible to observe in the graphic, the packet loss is zero, because the channel speed is greater than UMTS speed channel.

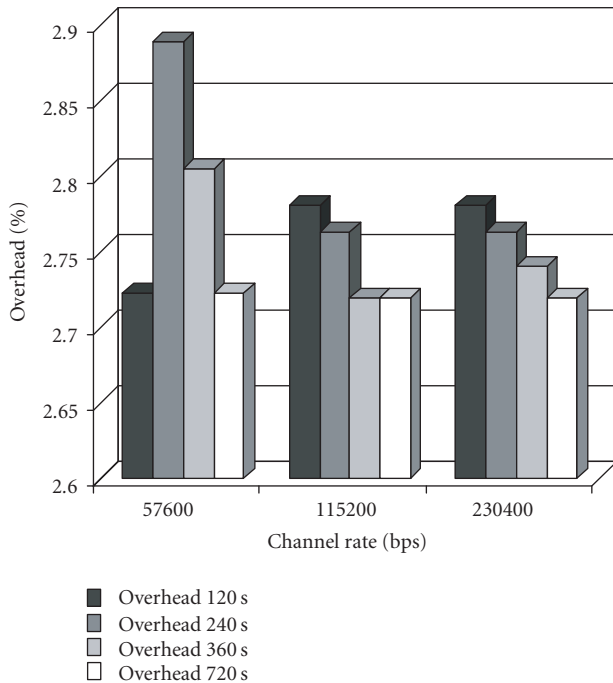


FIGURE 7: Overhead (%).

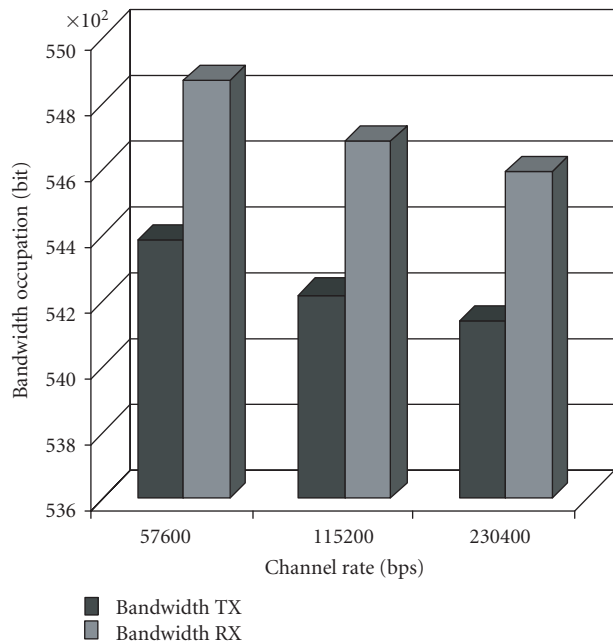


FIGURE 8: Bandwidth variation for different rate values.

In Figure 7 the overhead introduced owing to the effect of the packetization can be observed. In practice the protocol 07.10 introduces some overheads into the serial channel to transfer data from AP to MT and vice versa. This overhead has to be taken into account and it is  $\approx 3\%$  of overhead for each packet. In this way, we have 7 control bytes for 258 data bytes; we can conclude that it is an acceptable overhead.

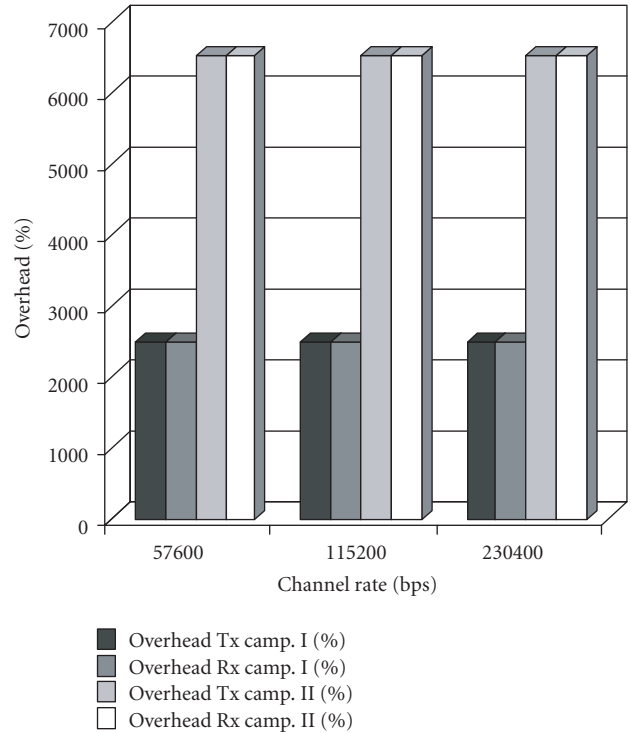


FIGURE 9: Campaign I versus campaign II overhead comparison.

### 5.3. Second simulation modality

In this second campaign traffic generation with variable packets was used. The scope of this simulation campaign is to show the same simulation parameters, like that in the first campaign, randomly varying the packets dimension. This type of scenario is more realistic than the first one considered above. It shows the behavior of a terminal that performs a video call, as known through a 3GPP software tool. This tool showed that for a video call a bandwidth of 49.5 kbps is sufficient, which is a smaller velocity than that of the UMTS standard.

It is interesting first of all to see the slope of the total bandwidth on the channel.

The bandwidth occupation is constant for different velocity values, but it is saturated for rates of 57.6 kbps. This means that it is a limit velocity and that only thanks to the buffer dimensioning there is no packet loss (Figure 8). In this case the channel is strongly stressed. It can be seen that, in this second campaign, there is an increase of the overhead in respect to the first campaign. As can be seen from the graphic it is approximately doubled (Figure 9). This shows a considerable increase of the resource waste. This is due to the packet variable dimension.

Then, it can be concluded that to have a packet with a constant dimension it is useful in terms of waste, but unfortunately is not very realistic.

A true video call generates a set of variable packets, then it is unforeseeable to know how much bandwidth waste there will be on the channel. This leads to the decision of giving a

sufficient bandwidth to the application. From the study carried out it seems that a bandwidth of 115.2 kbps is the bandwidth deputed for performing video calls on a smartphone terminal.

## 6. CONCLUSIONS

The study performed in this paper points out what are the specific features of a video call, generating a traffic that can simulate the real behavior of this type of application over smartphone terminals. It is useful to emphasize how *the video call* is not still an optimized service. In fact, it travels on a circuit switched connection and this leads to some difficulties like a fixed bandwidth allocation, with the problem of waste and a slowness in video audio synchronization. The characteristic parameters of the video call have been taken into consideration in traffic generation. The main information that characterizes this service is a fixed bit rate of 64 kbps, but the typical video traffic, as we have seen, is highly variable, since a great part of the weight of the packets is given from the data video codified with H.263.

## ACKNOWLEDGMENT

This work was supported in part by the Enteos Mobile Business Society of Trieste (Italy).

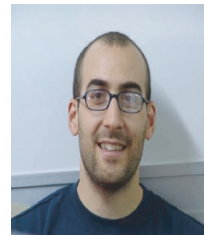
## REFERENCES

- [1] R. B. Ali, S. Pierre, and Y. Lemieux, "UMTS-to-IP QoS mapping for voice and video telephony services," *IEEE Network*, vol. 19, no. 2, pp. 26–32, 2005.
- [2] I. Maniatis, G. Nikolouzou, and S. Venieris, "QoS issues in the converged 3G wireless and wired networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 44–53, 2002.
- [3] [http://www.camiresearch.com/Data\\_Com\\_Basics/RS232\\_standard.html](http://www.camiresearch.com/Data_Com_Basics/RS232_standard.html).
- [4] 3GPP TS 26.110, "Codec for circuit switched multimedia telephony service; General description".
- [5] Cisco Systems, "Quality of Service for VoIP" 2001, <http://www.cisco.com/universcd/cc/td/doc/cisintwk/qosol/qosvoip.pdf>.
- [6] ITU-T H.324 Series H, "Audio Visual and Multimedia System Infrastructure of audiovisual service—System and terminal equipment for audiovisual service. Terminal for low bit rate multimedia communication".
- [7] UMTS 22.72 v 0.0.0 (1999-01), "Universal Mobile Telecommunications System (UMTS); Real-time Multimedia in UMTS".
- [8] 3rd Generation Partnership Project, "Technical Specification Group Service and System Aspect; code for circuit switched multimedia telephony service; modifications to H.324".
- [9] ETSI TS 127007 v 3.12.0 (2002-12), "Digital Cellular Communications System (Phase2+); Universal Mobile Telecommunication System (UMTS)".
- [10] <http://www.ncoretch.com/speech/pdf/g7231c54x.pdf>.
- [11] 3G TS 26.071 v 3.0.1 (1999-08), "3<sup>rd</sup> Generation Partnership Project; Technical Specification Group Service and System Aspects; mandatory speech codec speech processing functions AMR speech codec; General Description (3G TS 26.071 v 3.0.1)".
- [12] ITU-T Recommendation H.223, "Multiplexing Protocol for low bit rate multimedia communication".
- [13] P. T. Brady, "A model for generating on/off speech patterns in two way conversations," *Bell System Technical Journal*, vol. 48, 1969.

**Valeria Loscri** received the degree in computer engineering from University of Calabria, Italy, in 2003, she has been with the telecommunications research group of the University of Calabria where she is fully involved in a number of projects concerning the multimedia wireless communications. She is currently working toward the Ph.D. degree in the Telecommunication Laboratory, Department of Electronics, Informatics and Systems (DEIS). Her research interests include quality of service, and medium-access control, performance analysis, ad hoc networks, and wireless sensors networks.



**Mauro Tropea** graduated in computer engineering at the University of Calabria, Italy, in 2003. Since 2003 he has been with the Telecommunications Research Group of DEIS in the University of Calabria. In 2004, he won a regional scholarship on satellite and terrestrial broadband digital telecommunication systems. Since November 2005 he has been a Ph.D. student in electronics and communications engineering at the University of Calabria. His research interests include satellite communication networks, QoS architectures and interworking wireless and wired networks, mobility model.



**Salvatore Marano** graduated in electronics engineering at University of Rome in 1973. In 1974 he joined the Fondazione Ugo Bordoni. Between 1976 and 1977, he worked at ITT Laboratory in Leeds, United Kingdom. Between 1977 and 1979 he was with Face Standard Central Laboratory in Pomezia (Rome). Since 1979 He has been an Associate Professor at the University of Calabria, Italy. His research interests include performance evaluation in mobile communication systems, enhanced wireless and satellite systems, personal communications systems.

