QUALITY OF SERVICE OF VOIP OVER DVB-RCS

M. A. Vázquez Castro(*), F. J. González Serrano, A. Martínez Fernández, G. Mohedano Moya

Dept. de Teoría de Señal y Comunicaciones
Universidad Carlos III de Madrid, SPAIN
Avda. Universidad 30, 28911 Leganés, Madrid, Spain
Email: maryan, fran, andresmf@tsc.uc3m.es

**on leave from the University under an ESA fellowship*

***Abstract*- In this work we investigate the use of the DVB-RCS return channel to provide two-way VoIP transmissions over regenerative GEO satellites. Transmission impairments result from congestion and jitter like in terrestrial network but the paramount impairment in a GEO satellite scenario is the large propagation delay. So therefore, we develop a model for the DVB-RCS transmission delays by assuming a MF-TDMA superframe pattern. We also develop a model for the switching matrix on-board that can serve a variable number of users depending on the speech coders under consideration. We finally obtain objective parameters of the quality of voice for different speech coders G.711, G.723 and G.729 and results have shown that some codecs do not achieve sufficient performance in a DVB-RCS scenario.**

I. INTRODUCTION

DVB (Digital Video Broadcasting) Return Channel via Satellite, DVB-RCS, is an ETSI standard that specifies the provision of the interaction channel for satellite interactive networks via Return Channel Satellite Terminals, RCST. It is based on the MPEG transport stream [1][2] standard as transmission platform.

DVB-RCS technology is meant to provide last-mile access to broadband integrated services. Target users are Small & Medium Enterprises (SMEs) in a first deployment and residential users afterwards due to the scalability of the standard and the foreseen transmission frequencies (from Ku to Ka bands).

As an open standard, low-cost terminals may become available soon and compete with terrestrial access technologies such as leased lines, xDSL modems or cable modems. Popular Internet services such as e-mailing or web-browsing can be provided at a very high speed. Another category of applications for DVB-RCS is Voice over IP, VoIP. VoIP can be described as making telephone calls and sending faxes over IP-based data networks with a suitable Quality of Service (QoS). The voice information is not sent via dedicated connections as in the circuit-switched Public Switch Telephone Network (PSTN) but using discrete packets. The broadband links of DVB-RCS make it possible the required constant rate to assure quality of the VoIP connections. A major inconvenience of current satellite DVB-S systems is however that two satellite hops are required for a connection between two access users. Upcoming regenerative satellite systems though will have on-board processing/switching capabilities allowing one satellite hop only.

In this work we investigate the use of the DVB-RCS return channel to provide two-way VoIP transmissions over regenerative GEO satellites. Transmission impairments result from congestion and jitter like in terrestrial network but the paramount impairment in a GEO satellite scenario is the large propagation delay. Therefore, we develop a model for the DVB-RCS transmission delays by assuming a MF-TDMA superframe pattern. We also develop a model for the switching matrix on-board that can serve a variable number of users depending on the speech coders under consideration. We finally obtain objective parameters of the quality of voice for different speech coders G.711, G.723 and G.729 and results have shown that some codecs do not achieve sufficient performance in a DVB-RCS scenario.

II. DELAY MODELLING

The delay experienced in a call takes place on the transmitting side (speech coding and packetization), in the satellite network and on the receiving side.

A. Speech coding delay

The speech encoder converts the digitised (after A/D conversion) speech signal to a bit-stream and it may use a compression algorithm. There are four parameters that thoroughly describe a coder: bit rate, delay, complexity and speech quality. The bit rate can be fixed or variable and the delay is a function of the following factors:

required block of PCM samples (voice frame, T_{fr}) and a "look-ahead" samples/time or future samples needed for prediction tasks, also called algorithmic delay (T_{alg}). Normally, the latency incurred due to processing is typically specified as the frame size in milliseconds and therefore the coding delay (T_{cod},) can be expressed as follows:

$$T_{cod} = T_{fr} + T_{al} + T_{pr} \quad (1)$$

Note that T_{pr} cannot exceed the framing delay T_{fr} otherwise the DSP would not be able to complete processing one frame before the next frame arrived. These delays are represented in Figure 1.

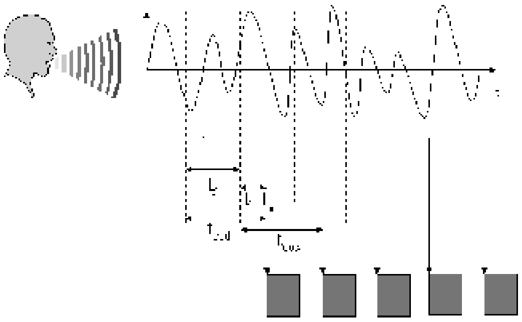

Fig. 1. Speech coding delay contributions.

Decompression time is roughly 10% of the compression time for each block and therefore it will depend on the number of compressed blocks received simultaneously (in one voice packet).

The most commonly used codecs for IP telephony today are G.711, G.729, and G.723.1, since they are all part of the ITU's core standard for multimedia teleconferencing H.323 [6], which utilizes the IETF Real-Time Protocol (RTP/RTCP). The H.323 specification is a comprehensive specification for the implementation of packet-based multimedia over IP networks that cannot guarantee Quality of Service (QoS).

The G.711 speech coder [3] samples the voice signal at 8 Kbit/s and is used in the Public Switched Telephone Network, PSTN. It is commonly called PCM (Pulse Code Modulation). There are two flavours of its algorithm; μ-law used in North America and Japan, and A-law used in the rest of the world. This coder produces a bit rate of 64 kbit/s.

The G.729 [5] produces a bit rate of 8 kbit/s. It is based on the principle of Conjugate Structure-Algebraic Code Excited Linear Prediction (CS-ACELP). The coder works on a frame of 80 PCM speech samples (or 10 msec). For the linear prediction it needs also 40 samples (or 5 msec) of the next voice block of samples, this is called "look-ahead".

The G.723.1 coder [4] generates two bit rates, 5.3 and 6.3 Kbps. Both bit rates share the same short-term analysis techniques for processing the speech. For long-term analysis of speech, the algorithms used are different. The 5.3 Kbps coder uses the Algebraic Code Excited Linear Prediction (ACELP) algorithm, while the 6.3 Kbps coder uses the Multi Pulse-Maximum Likelihood Quantization (MP-MLQ) technique. The coder works on a frame of 240 PCM speech samples (or 30 msec). Besides, there is a look ahead of 60 samples (or 7.5 msec).

Although silence suppression may reduce the bit rate of some of the coders, speech traffic is modelled as constant bit rate traffic.

B. Packetisation delay

Packetisation delay (T_p) is the time taken to fill a packet payload with encoded/compressed speech. This delay is a function of the PCM sample block size required by the vocoder and the number of blocks (*N*) that will be placed in a single packet. The packetised delay is then at least the PCM block size and can be expressed as follows

$$T_{pk} = (N-1)T_{fr} \quad (1)$$

Coder	Algorithm	Bit Rate (kbit/s)	Required PCM block size T_{fr} (msec)	Algorithmic delay T_{alg} (msec)	Packet duration T_{1pk} (msec)	Packetisation delay T_{pk} (msec)	Packet size (bytes)
G.711	PCM	64	0.125	0	20	(N-1)x20	160
G.729	CS-ACELP	8	10	5	10	(N-1)x10	10
G.723.1	MP-MLP MP-ACELP	6.3 5.3	30	7.5	30	(N-1)x30	20 24

Table 1. Summary of VoIP coders and delay contributions

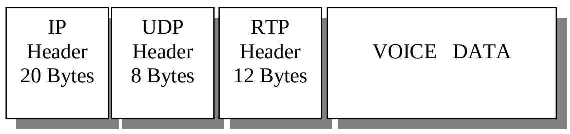

Fig. 2. IP telephone packet.

Table 1 summarizes all the delay contributions described so far and Figure 2 shows an example of packetisation on an IP packet. Note that a trade off arises. For example, if a system is using the G.723.1 coder version producing 20 byte frames every 30 milliseconds, each packet would have 40 bytes of header and 20 bytes of data which means having a 200% of overhead. If in order to reduce such overhead, N voice frames are placed in one single packet, this will add (*N*-1) frames period of latency.

Another option that would not increase the latency is to allow voice frames from different channels to be piggybacked on the same packet. This only can be done when such different voice frames have the same destination. It should be noted that this is not supported by the standard H.323 but it is actually being implemented under proprietary solutions.

In this work, we assume that the packet size is equal to the length of the speech frame, except for G.711 where a 20 msec frame is used.

It should be noted that the packet size shown in Table 1 corresponds only to voice payload. Protocols involved in Voice over IP follow a layered hierarchy, each layer contributing with an additional header preceding the payload.

The Internet protocol, IP, is the lowest layer responsible for the delivery of packets between hosts (unless the IP packet is further encapsulated as it is the case in a MPEG-based system). It is a connectionless protocol and therefore packets may arrive out of sequence. Higher layer protocols should account for this issue for real time applications such as voice. Next layer on top of IP is the transport layer. There are two transport protocols available: Transmission Control Protocol, TCP, and User Datagram Protocol, UDP. TCP is a connection-oriented protocol not appropriate for voice transmission since it relies on packet retransmission to perform its tasks. Retransmission is feasible when handling data but voice should be played back continuously at the receiver side and retransmissions are for that reason not affordable. UDP is more suitable since it is connectionless and just ensures data is routed to the correct port without attempting to achieve the right sequence or data to be integral (no time to ask for retransmission for a missed or corrupted packet).

On top of IP and UDP there is still another layer needed for real applications. It should activate mechanisms to ensure that a stream of data is accurately played back at the receiver. The main task should be detecting delays and jitter so that data can be buffered to be played back at a constant bit rate. The most used protocol to perform this task is Real Time Transport Protocol, RTP. Thus, it can be concluded that the standard method to transport voice payload through an IP-based network requires the addition of three headers as it is shown in Figure 2. An IPv4 header is 20 bytes, a UDP header is 8 bytes and an RTP header is 12 bytes.

When a single VoIP flow is observed, it can be noticed that only a few fields in the RTP/UDP/IP header vary between consecutive packets. Header compression mechanisms reduce header over-head since it is not necessary to send static header fields (e.g., IP addresses and UDP ports) in every packet because these static header fields do not change during the same session. Using RTP header compression, the 40 bytes IP/UDP/RTP header can be compressed to approximately 5 bytes, as shown in Figure 3.

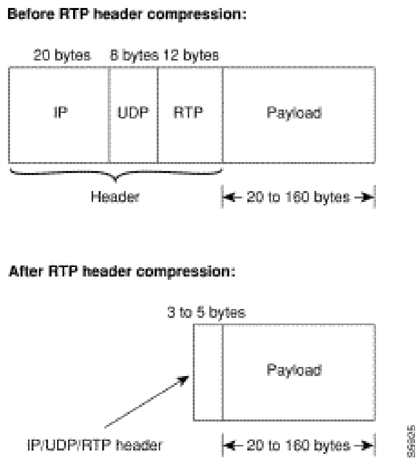

Fig. 3. RTP Header Compression

C. Network delay

In this work, a Multi-Beam MF-TDMA DVB-RCS compliant uplink and Multi-Beam DVB-S Downlink are considered. This way, full Cross-Connectivity between Uplink and Downlink Beams is allowed.

We consider a MF-TDMA consisting of 23 carriers of a minimum of Ri = 518.4 Kbit/s. MPEG-2 Transport Stream is considered with 24 MPEG packets (184 bytes payload and 4 bytes overhead) per frame and a frame length of 69.629 ms.

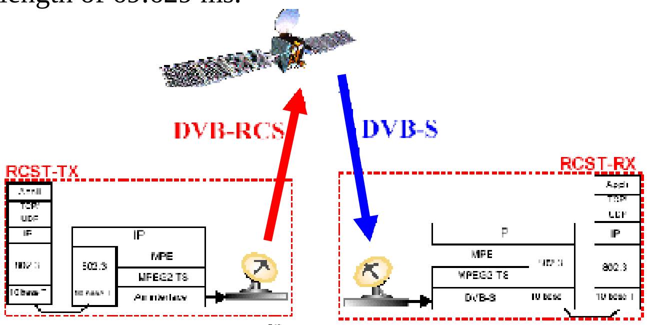

Fig. 4. System overview including Traffic Protocol Stacks.

Several factors contribute to the delay experienced by the voice packets when traversing a communications network. For a GEO network regardless of the architecture, the different contributions can be partitioned as follows:

$$T_{net} = T_{MAC} + T_{trx} + T_{prp} + T_{OBP} \quad (3)$$

where

- T_{MAC}: time the packets must be waiting in the gateways queues before being transmitted on the shared uplink resource
- T_{trx}: transmission time, it depends on the transmission rate of the carrier
- T_{prp}: propagation time which lies between 238 and 278 msec for a GEO system and minimum and maximum elevations respectively.
- T_{OBP}: delay introduced by the on board processor on board.

In this work, we consider that the packetisation delay is included in the MAC delay. For both the packetisation and on board processing delays we first need to set the assumptions for a reasonable implementation of the medium access and the on board processor for a DVB-RCS network with OBP, which are not yet publicly available for the systems currently under development. In addition, a model for the measurement of the quality in terms of delay is also needed. We introduced the model of the packetisation and medium access model, the OBP model and the delay model in the three following sections.

III. PACKETISATION AND MEDIUM ACCESS MODELLING

Voice packets are encapsulated into MPEG-2 TS packets so that the payload includes an integral number of voice packets. Another important issue is the separation in the MF-TDMA frame between two consecutive packets coming from the same user. With R_i=518.4 kbps, up to 7 packets can be allocated within the MF-TDMA frame for other users using a G.711 codec since codec outputs one voice packet each 20 msec. Since it is encapsulated in one MPEG packet, and the duration of one MPEG packet is $\frac{188\times 8}{518400} = 2.901\ \text{msec.}$, the separation between two consecutive packets is $\left\lceil \frac{20}{2.901} \right\rceil = 7\ \text{MPEG packets}$.

In the same way it can be shown that up to 43 packets for G.729 and 73 packets for G.723 can be allocated within the MF-TDMA frame for other users. These figures are summarized in Table 2.

In this work we assume that several packets from different users can be placed together in one single MPEG packet. Recall that this solution is not allowed by the H.323 standard. MPEG packets are not transmitted but wait in the Gateway queues until it is filled with the maximum amount of packets it can carry. We assume fully load system which means that there are always packets waiting to be transmitted.
Destination addresses are randomly generated.

	G.711	***G.729***	***G.723.1***
Voice packets/MPEG	$\left\lfloor \frac{184}{5+160} \right\rfloor = 1$	$\left\lfloor \frac{184}{5+10} \right\rfloor = 12$	$\left\lfloor \frac{184}{5+20} \right\rfloor = 7$
Voice packets/carrier	*1×24=24*	*24×12 = 288*	*24×7 = 168*
Separation (packets)	*7*	*43*	*73*

Table 2. Encapsulation of voice packets into the MPEG-2 TS packets of the uplink's MF-TDMA frame.

IV. ON BOARD PROCESSOR MODELLING

In this work, we assume that the on-board processor (OBP) has the capacity to route data on any of the uplink coverage footprints on to any combination of downlink coverage footprints on a per need basis [8] [10]. The minimum granularity for the switching at the OBP is assumed to be the voice packet. This way, one voice packet may be routed to one destination while the next voice packet may be routed to a different destination or a combination of destinations.
The multiplexing at the OBP operates as follows. A MPEG-2 TS packet is delivered for a given destination whenever its payload is completely filled by voice packets. For this reason, if there are no enough voice packets for a given destination during the processing of an uplink MF-TDMA, the system has to wait for the next MF-TDMA frame before deliver the MPEG packet. It is important to observe that, as it was shown in Table 2, the number of voice packets required to fill up a MPEG packet depends on the used speech coder. For illustration purposes, in Fig. 5 we show how the OBP operates in a system with 4 sources (colors) and 4 destinations. In addition, for the sake of simplicity, we assume that the multiplexing matrix is completely empty and that the downlink frames are ready when they are filled with 5 voice packets. The OBP reads column-by-column the MF-TDMA frame and puts each packet into the buffer corresponding to the downlink destination. The multiplexing matrix outputs a downlink frame whenever it is completely filled (5 packets). Otherwise, the OBP waits for the next MF-TDMA frame.

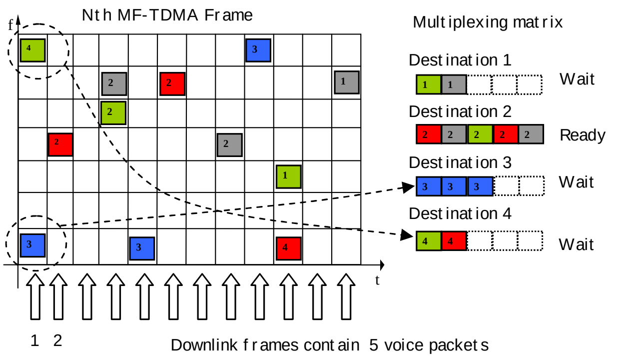

Fig. 5. Example of how the on board switching distributes voice packets onto the downlink frame.

V. VoIP QUALITY OF SERVICE: E-MODEL

In order to measure the voice quality achieved by the different coders when using a VoIP transmission platform as the one described in the precedent sections, we need an objective measure of the subjective voice quality. The so-called E-model provides such a measurement.

This model is given in the ITU-T Rec. G.107 [7]. It predicts the subjective quality of a telephone call based on the characteristic transmission parameters including a number of delay contributions. Namely, it combines such parameters in only one parameter called R factor. It can be used to predict Mean Opinion Score (MOS) from users or the percentage of users rating the voice with a given quality.

The R factor is defined in such a way that the different contributions are additively added as follows

$$R = R_0 - I_s - I_d - I_e + A \qquad (4)$$

where:

- R_o: models noise-related effects such as background noise.
- I_s: models the negative effects inherently present in the voice signal such as the quantification
- I_d: models all the delay-related effects either caused by undesired echoes or by any other source of delay
- I_e: models all the distortion-related effects
- A: models the decrease of quality a user is willing to tolerate by accessing to a system offering instead some other advantage (for example A=20 for satellite connections while A=0 for wired connections).

R takes values between 0 and 100, the higher the better.

VI. SIMULATIONS RESULTS

In order to obtain some results on VoIP quality, a number of preliminary simulations were carried out to investigate a number of related design aspects.
First, we have analysed the efficiency of voice packets encapsulation onto MPEG packets (184+4 bytes) as a function of the codec. Efficiency is given by the ratio between the number of bits in voice packets and the number of bits of a MPEG packet. It is assumed that voice packets are encapsulated into IP/UDP/RTP packets prior to the mapping onto the MPEG frame. Results are shown in Table 3.

	G.711	*G.729*	*G.723.1*
Efficiency (%)	*85*	*65*	*76*

Table 3. Efficiency of the encapsulation of voice packet onto MPEG packets

In general, a voice transmission becomes unacceptable when the total end-to-end delay exceeds the 500 msec bound [7]. This bound limits the VoIP transmission over satellite systems to architectures including OBP since the two-hop propagation time is already more than 500 msec.

In the satellite network, and according to (3), most of the delay stems from the encapsulation on the uplink transport stream (included in T_{MAC}), the on-board processing (switching and multiplexing) T_{OBP}, and the transmission and propagation time. The two first sources of delay in the satellite segment (encapsulation and on-board multiplexing) clearly depend on the number of simultaneous users who share the common resources. In the flowing results, delay refers to total delay, T_{tot} which can be expressed as follows according to (2) and (3):

$$T_{tot} = T_{coder} + T_{net} \qquad (5)$$

We have not included de-jittering delay at the receiver side neither decoding (roughly 10% of T_{cod}) delay.

The effects of the variation in the number of uplink users were simulated by varying the separation between consecutive packets coming from a given user (the more the separation is, the more users are served). In our tests, the separation was increased gradually from a minimum value, imposed by the frame size of the considered codec (see Table 2). These variations in the packet arrival rates can be damped by using a jitter buffer. Results are depicted in Figure. It is important to notice that both G.711 and G.729 codecs provide delays below the 500 msec bound (delays beyond 500 msec are unacceptable for conversational services). However, the G.723.1 codec is not useful for VoIP applications through DVB-RCS systems since its high algorithmic delay (30 msec) raises the total delay higher than the maximum allowable.

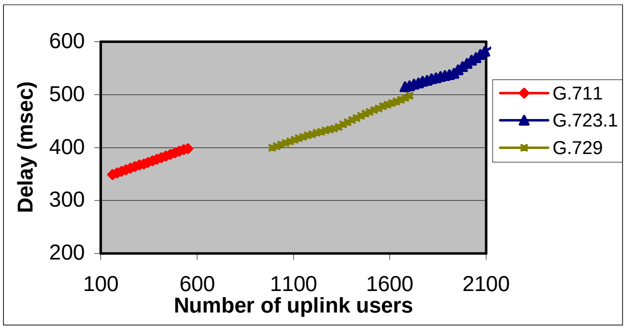

Fig. 6. Delay profile for the G.711, G.723 and G.729 codecs as a function of the number of uplink users.

We have also analysed how the OBP performs when the number of downlink destinations varies. We have assumed that the uplink's frame is full of voice packets. Fig. 7 shows that the total delay increases as the number of destinations does. Fig. 8 shows the VoIP quality factor R obtained for three different codecs as a function of served users.

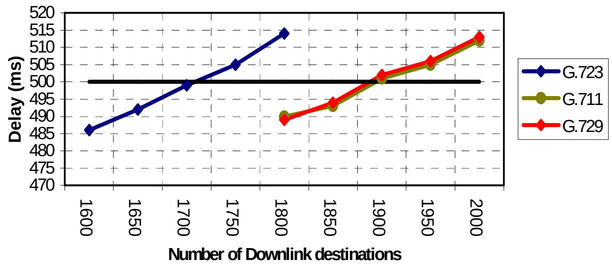

Fig. 7. Total delay as a function of the number of downlink destinations and codec.

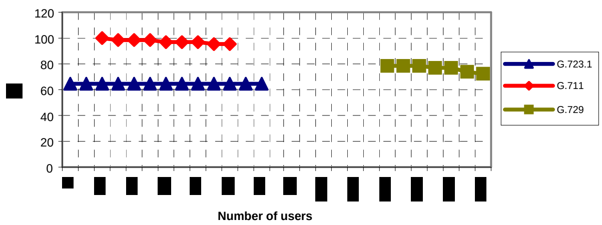

Fig. 8 VoIP quality factor R obtained for three different codecs as a function of served users (rate 100-90 gets best quality and 60-0 poor).

VII. CONCLUSIONS

In this work we have investigated the use of the DVB-RCS return channel to provide two-way VoIP transmissions over regenerative GEO satellites. We have developed a model for the DVB-RCS transmission delays by assuming a MF-TDMA superframe pattern. We have also developed a model for the switching matrix on-board that can serve a variable number of users depending on the speech coders under consideration. Finally, we have used the E-model to determine the quality of voice for different speech coders G.711, G.723 and G.729. Results have shown that the G.723-1 codec does not achieve sufficient performance in a DVB-RCS scenario.

REFERENCES

[1] ETSI EN 301 790, DVB Interaction channel for satellite distribution systems

[2] ETSI EN 300 468, DVB Specification for Service Information (SI) in DVB systems

[3] ITU-T Rec. G.711, "Pulse Code Modulation (PCM) of Voice Frequencies," 1988.

[4] ITU-T Rec. G.723.1, "Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s," 1996.

[5] ITU-T Rec. G.729, "C Source Code and Test Vectors for Implementation Verification of the G.729 8 kbit/s CSA CELP Speech Code," 1996.

[6] ITU-T recommendations H.323: "Visual Telephone Systems for local area networks which provide a non-guaranteed quality of service", Geneve, Switzerland, May 1996

[7] ITU-T Rec. G.107, "The E-Model, a computational model for use in transmission planning," Dec. 1998.

[8] Analysis of IP voice conferencing over EuroSkyWay satellite system Cruickshank, H.; Sun, Z.; Carducci, F.; Sanchez, A.; IEE Proceedings Communications, Volume: 148 Issue: 4 , Aug 2001 Page(s): 202 -206

[9] Y. Le Roy et Al., "The Alcatel 9343 DVB-OBP Product: An On-Board Processor for Digital Television and Internet Data," Seventh International Workshop on Digital Signal Processing Techniques for Space Communications, Portugal 2001.

[10] Chacón S. et Al. "Networking over the IBIS system". IST Mobile and Wireless Communications Summit Greece, 2002