On the Performance of Push-to-Talk over EDGE Technology

Luis Claudio dos Santos e J. Roberto Boisson de Marca [(1)]

Resumo— **O serviço de *Push-to-Talk* sobre tecnologia celular (PoC) é um serviço de voz, half-duplex, empregando protocolo IP, que tem recebido muita atenção devido a sua eficiência em termos de custo e de utilização de recursos. Nesse artigo é feita uma avaliação da capacidade deste serviço quando a tecnologia de rádio adotada é a GSM/EDGE. Os resultados são obtidos através de simulação orientada a objeto adotando uma política de adaptação do enlace que limite a taxa de erro de blocos a um valor de 3%. O parâmetro de qualidade empregado para estabelecer a capacidade foi um retardo total para transporte da informação não superior a 1,6 s.**

Palavras-Chave— **Serviço Push-to-Talk (PoC); Tecnologia EDGE; SIP - Session Initiation Protocol; Serviço Celular 3G; Protocolo RTCP**

Abstract— **Push-to-Talk over cellular (PoC) is a two-way on-demand voice and data communications service which gained a lot of attention due to its resource and cost efficiency. In this paper we evaluate the capacity of the GSM/EDGE technology to offer PoC services. The evaluation is done through an object oriented simulation and for a link adaptation policy that keeps the block error rate limited to 3%. The capacity is computed under the constraint that the total transport delay is equal or less the recommended figure of 1.6s.**

Index Terms— **Push-to-Talk (PoC); EDGE; Session Initiation Protocol; 3G Cellular Service; Real Time Control Protocol**

I. INTRODUÇÃO

Push to talk over cellular (PoC) is an extension over the cellular communications infrastructure of the traditional walkie-talkie service [1]-[3]. Push-to-Talk emergence happened in the mid 90´s over proprietary mobile communications networks and has been quite popular since then. However, until recently the lack of standardization did not allow widespread offerings of PoC services. PoC can be seen as a half-duplex near real time voice over IP (VoIP) service. It is activated by a user pushing a specific button (TALK button) in his mobile and waiting until he/she receives an indication to talk. From that moment on the user has exclusive right over the channel until the button is released. The conversation floor (traffic channel) can then be acquired by someone else.

Network elements control the allocation of right-to-talk to a particular user during a conversation. Within the OMA (*Open Mobile Alliance*) architecture [6] this task is assigned to the PoC server who is then responsible for user management, floor control activities and media forwarding in case of one to one communication and media replication for group calls. The user equipment (UE) carries PoC client software and communicates with the PoC server through the Core Network (CN) and Radio Access Network (RAN). The use of IP technology guarantees an efficient use of RAN and CN resources which are reserved only for the duration of the talk spurts instead of for the duration of the call. The *Session Initiation Protocol* (SIP) [4] is recommended for call set up and teardown. The Real Time Transmission (RTP) Protocol [5] is employed for the transmission of voice packets. The RTP Control (RTPC) protocol [5] is adopted for determining which call member is allowed to talk at any given time during a call.

In this work an object oriented simulation tool is employed to evaluate the user capacity of the PoC technology when the underlying radio interface technology is the GSM/EDGE.. The PoC specification recommends a particular value for the maximum transport delay. This value is established as a constraint in computing the user capacity.

Contributions in the open literature that report on the performance of the Push-to-Talk technology are limited in number. There are however at least two recent publications on this topic. In reference [14], a study was presented addressing the performance of the push-to-talk service over the Bluetooth air interface. The authors in [15] develop an analytical PoC delay analysis for the GPRS air interface. They also consider sharing of traffic channels between the PoC service and the traditional telephony service and the assignment of priorities. However mode switching was not considered and no specific assumption was made regarding propagation characteristics

(1) Luis Claudio dos Santos, YIP do Brasil, Sorocaba, SP, Brasil e J. Roberto Boisson de Marca, Centro de Estudos em Telecomunicações, PUC-Rio, Rio de Janeiro, Brasil, E-mails: lcs@cetuc.puc-rio.br; jrbm@ cetuc.puc-rio.br .

The GSM/EDGE technology, that is briefly described in the following Section, has as one of its main features the ability to adapt the transmission rate to the instantaneous channel condition. The results presented in Section IV focus on a modulation and coding mode adaptation that aims at maintaining the residual block error rate (BLER) below 3%. The capacity of the system is determined under the constraint that the total transport delay cannot surpass the recommended value of 1.6s The numerical results were obtained with a simulation model that is described in Section III. Concluding remarks are offered in Section V.

II. GSM/EDGE FEATURES

EDGE (*Enhanced Data for Global Evolution*) is a 3G technique which evolved from GSM, the immensely successful second generation technology [7]-[8]. It uses the same frame structure, eight time slots per carrier, and also the same 200 KHZ carrier spacing as specified for GSM. EDGE is designed to offer packet mode high bit rate transport of information and can be considered an enhancement of the initial evolution of GSM towards packet data service, the GPRS (*General Packet Radio Sevice*). Both GPRS and EDGE are packet overlays on top of the standard GSM circuit switch (CS) architecture. GPRS/EDGE service is implemented through the addition of two packet switching nodes (Serving and Gateway GPRS Support nodes). However, GPRS and EGPRS have different protocols and different behavior on the base station subsystem side. EDGE incorporates a new modulation technique (PSK-8) and error-tolerant transmission methods, combined with improved link adaptation mechanisms that make possible to offer data transmission rates at least three times larger than GPRS. GPRS employs four coding schemes, designated CS1 through CS4. All four schemes employ MSK modulation but different amounts of error-correcting coding that are optimized for diverse radio environments. On the other hand EDGE has available nine modulation and channel coding schemes, MCS1 to MCS9. The lower four EDGE modulation and coding schemes (MSC1 to MSC4) adopt GMSK, whereas the upper five (MSC5 to MSC9) use 8PSK modulation. EDGE slot bit rates vary then from 8.4 (MCS1) to 59.2 Kb/s (MCS9). The higher the bit rate the weaker is the protection afforded by the coding scheme. Therefore it is expected that the use of the higher MCSs will only be possible in very favorable propagation and interference conditions.

The actual average bit rate achieved in a particular transmission will depend on several factors such as the frequency planning, the morphological scenario and the system loading. Fig. 1 [9] illustrates the cumulative distribution function of the carrier-to-interference ratio (C/I) for an EDGE system operating in a (4,12) arrangement (frequency reuse 4 and three sectors per cell) and ITU-R propagation model TU-3 (Urban environment, pedestrian speed of 3km/h). As can be seen the median of the distribution is between 17 and 18 dB. The data provided by this curve is

used later on as propagation characteristic for the simulation experiments.

As mentioned previously, each EDGE MCS scheme contains a different level of error protection. Therefore the net bit rate (taking into account the observed radio block erasure rate - BLER) will vary with the value of C/I at that particular instant. Fig. 2 [10] shows the net transmission bit rate per slot as a function of C/I when each modulation and coding scheme is used. The outer dark curve corresponds to the optimal link adaptation policy. This ideal policy would achieve the best performance by switching between the appropriate MCS modes at the optimal C/I level. The gain achieved by the ideal adaptation with respect to a scheme that is more suitable for implementation can be substantial, as demonstrated in Fig.3 [11]. In this Figure it is shown that the throughput per slot advantage of the ideal switching strategy can be larger than 20 kb/s with respect with a strategy that changes between MCS modes so that the BLER value never exceeds 3%. The capacity results reported in the sequel were obtained implementing the less efficient but implementable approach as opposed to the ideal one.

III. SIMULATION DESCRIPTION

A simulation model was developed to evaluate PoC over EDGE capacity for a given transport delay constraint. The simulation results presented in Section IV were obtained using the object oriented simulation tool Tangram-II described in [12]. Our PoC_Max model is composed of four objects. The first object (CORE) represents the behavior of the fixed network. The second object (BSS) represents the functionalities of the Base Station Subsystem whereas the MS object models the behavior of the mobile stations. Finally, the characteristics of the air interface are included in the object EDGE_Channel. Two base stations were simulated as shown in Fig. 4.
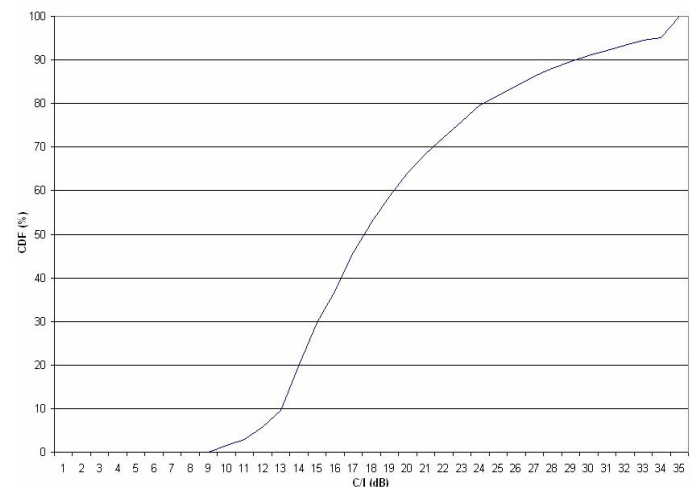

Figure 1 : Cumulative Distribution Function of C/I for EDGE with 4/12 frequency arrangement and TU-3 propagation model [9].

The radio channel behavior is represented by a two state Markov chain. The first state (G) models good propagation and interference conditions which are here represented by a value of C/I ≥ 17db. On the other hand, values of C/I in the bad (B) state are never higher than 17dB. The two states are made equally likely and the decision whether a mobile will change state or not is made at every 1 second of simulation time. The probability of moving to a different state at each decision instant is made equal to 0.1, i.e., once a user starts in a given channel condition there is a high probability of continuing in that condition for a period of time. This choice reflects the expectation that most users of this service will be pedestrians. Once the channel state is determined the shape of
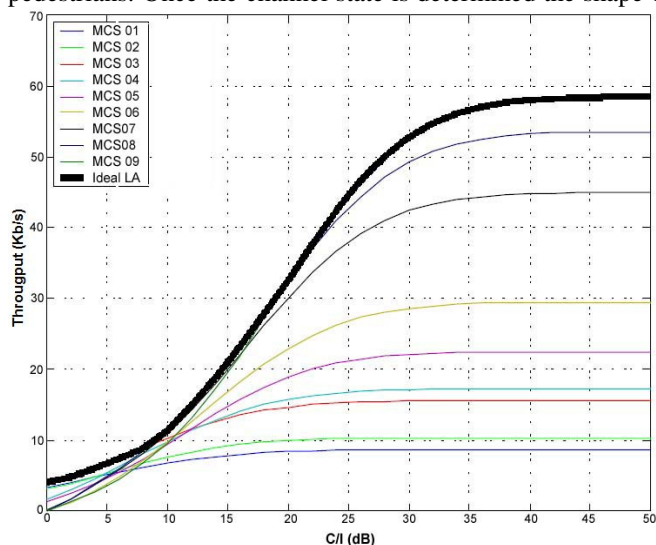

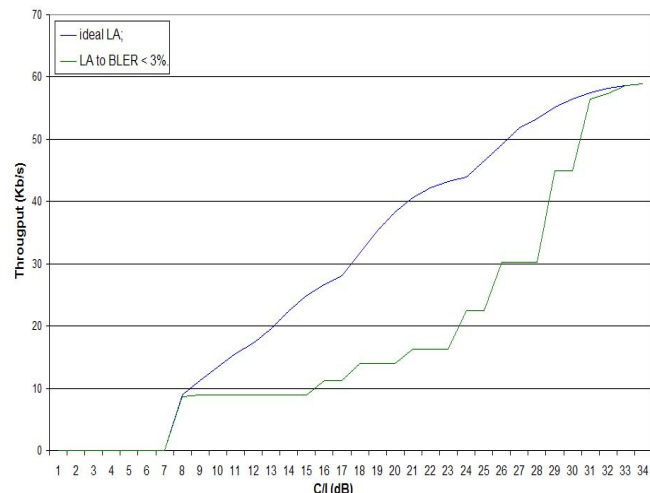**Figura 2 : Throughput vs. C/I for each EDGE MCS and Ideal Link adaptation [10].**

Figura 3 : Throughput vs C/I for Ideal Link Adaptation (iLA) and Adptation based on BLER < 3% rule [11].

the cumulative distribution function follows that in Fig.1 in the respective range.

The delay results discussed in the sequel assume that the PoC session between the two users has already been set up. Therefore the delay incurred in exchanging SIP messages, as

described in [2] was not included. The state diagram for a PoC station is shown in Fig. 5 and follows the PoC 2004 recommendation [6]. Once a user decides to establish communication with another person or group of persons, and pushes the TALK button, he/she has to wait the completion of two steps. The first one is the activation of a temporary block flow (TBF) and the second is the communication by the BSS that the floor has been granted to that request and therefore every other user involved in that communication will move to the LISTENER state. The protocol used for floor management signaling is the RTCP [5]. The total transport delay in this system is then an aggregate of six contributions:

$$T_D = T_E + T_F + T_L + T_t + T_Q + T_P \qquad (1)$$

T_F is the delay contribution due to the floor management operations. In this work the time it takes to acquire the floor is considered to be an exponential random value with average 700 ms [6]. Similarly, the time to release the floor is also exponentially distributed but with a mean equal to 600ms. The time (T_E) to activate a TBF is also modeled as an exponential random variable, both in the uplink and in the downlink, with a mean value of 360ms in the uplink and a somewhat smaller value of 240ms for the downlink [2].

The total latency T_L models the delay due to the processing in the terminal as well as the delay encountered in going through all the network entities. Here the terminal processing delay was considered to be fixed and equal to 180 ms whereas the delay attributed to the core network was assumed exponentially distributed with a mean of 280ms.

The transmission delay (T_t) depends on the particular MCS mode being employed in the packet transmission and as stated will depend on the instantaneous channel condition. The queueing time (T_Q) is the time the packet is kept in the Packet Control Unit (PCU) queue waiting for transmission.
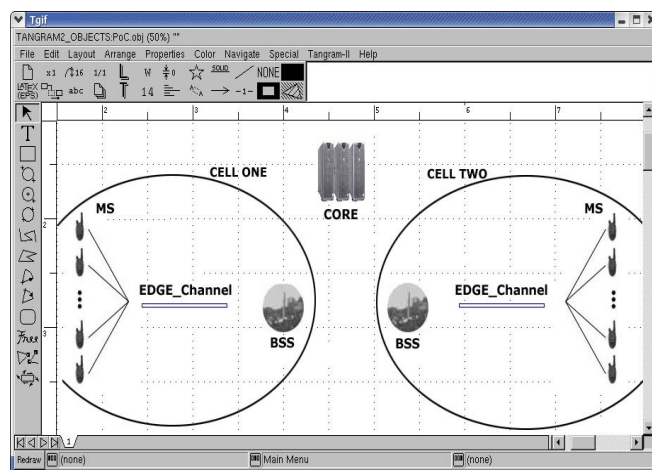

Figura 4 : PoC simulation model in Tangram-II environment.

Following the PoC specification the speech coder adopted in this work was the AMR (*Adaptive Multi-Rate*) codec. The AMR [13] is actually a family of encoders employing the

ACELP (*Algebraic Code Excited Linear Prediction*) algorithm and allowing the operation in eight different bit rates, varying from 4.75 to 12.2 Kb/s in addition to a comfort noise configuration. The AMR solution also allows for link adaptation as a function the channel quality. Nevertheless it seems to be typical that PoC implementations make use of just three of those modes. Hence, in this article we chose to work with the 4.75, 7.4 and 12.2 speech coding rates. The block size of the speech codecs is 20ms which results in 95, 148 and 244 bits being generated per speech block, respectively, for the chosen speech codec modes. In addition to these information bits, protocol header information must also be included in each transmitted packet. In the results described in the following section a total of 40 bytes is added per packet, corresponding to the headers inserted by UDP, IP and RTP protocols. Depending how the speech bits are placed into a packet, stuffing bits may also be required to guarantee an integer number of bytes in an IP packet. Note that to increase efficiency it is possible to place several speech blocks in one transmission packet, reducing the effect of protocol header overhead. On the other hand, there is a delay introduced due to the wait for the generation of several speech blocks, represented by the term T_P in Eq.(1), to fill a packet. The effect of this trade off is analyzed in the sequel by varying the number of speech blocks per packet (this number is sometimes referred in the literature as PTIME).
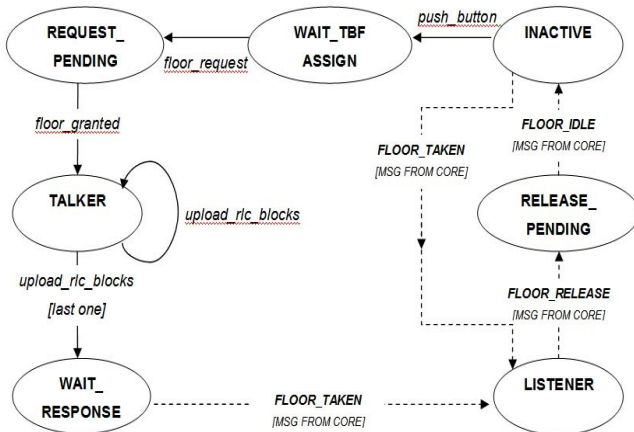

Figura 5 : State diagram for user sation during a PoC session

The time interval between the floor being released and the user pressing the TALK button, referred here as response time - t_R, was considered to be a random variable obtained from a log-normal distribution with variance 0.5 and with its average value varying between 4 and 10s. The lower bound corresponds to a conversation were the exchange between the speakers is happening at a very fast pace and therefore stressing the system. The speech burst duration was modeled according to a Weibull distribution with scale equal 8 and shape 6, translating into an average of 7.4s. The choice of this distribution is justified by the expectation that in a half duplex communication the duration of speech utterances will be shorter and with smaller variations [1]. Hence it was judged inadequate to use distributions exhibiting significant mass

around the origin and somewhat longer tails such as the exponential distribution.

In the following Section the results obtained with the simulation tool and the model just described are discussed.

IV. PERFORMANCE RESULTS

The numerical results here presented were obtained assuming that there are two slots in a single GSM carrier allocated to the PoC over EDGE service. No dynamic allocation of the remaining traffic channels was considered. The GPRS/EDGE specification allows for sharing of a single traffic channel by multiple data connections. In this work the number of sessions sharing one full rate traffic channel was limited to 6. However as will be seen by the results obtained this limit was not often reached. The total simulation time used to derive each performance point was 450 seconds. Capacity evaluations were computed by assuming that the quality of experience requirement on the end-to-end channel delay (QoE3) [14] must be met, i.e., $T_D \leq 1.6s$.

Initially we consider the use of the 4,75 Kb/s speech coder and investigate the difference in performance due to the EDGE transmission mode switching strategy. As explained in Sec. II, it is unquestionable that if the change between MCS schemes could be made at the ideal channel quality (measured by the instantaneous value of C/I) points the performance would be considerably better, as emphasized by Fig.6. When the mode adaptation is performed according to the criterion of limiting the radio block erasure rate (BLER) to 3% the number of users per slot is substantially diminished. Considering that the QoE3 requirement must also be met the number of users per slot is limited to three. If the mode switching could be effected at the ideal points six users could easily share a traffic channel. In the remainder of this article the ideal strategy will no longer be considered since it is not suitable for a practical implementation. Results will be given only for the strategy that attempts to keep the BLER below 3%.

Next the effect of the variation of the parameter PTIME in the PoC service capacity is addressed in Figs. 7-9. In every curve it can be seen that PTIME =20 allows the system to serve a higher number of users per slot than when lower values for PTIME are used. However the difference in performance varies with the bit rate of the speech coder being used. The curves in Figure 7 were obtained using the 4,75 Kb/s coder. It can be seen that for light loads and up to 2 users per slot, the difference in performance verified when the value of PTIME varies is negligible, except for PTIME=20. Encapsulating 20 speech blocks in each packet yields a system performance of three users per slot (or equivalently 6 users per sector since it is being assumed that there are 2 traffic channels devoted to PoC) while still meeting the constraint on T_D. Lower values of PTIME allow at most two users per slot before the delay
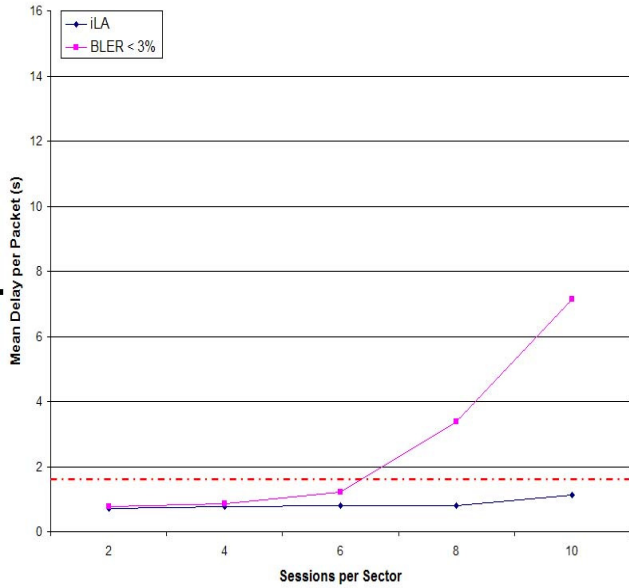
Figura 6 : Average delay per packet vs. number of conversations per sector for both ILA and BLER < 3% scheme. AMR 4.75 Kb/s codec.
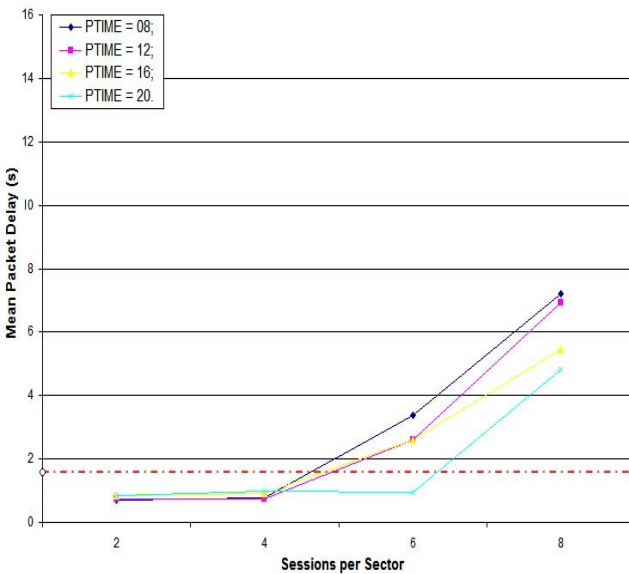

Figura 7 : Average delay per packet vs. number of conversations per sector for different values of PTIME. AMR 4.75 Kb/s codec.

constraint is violated. When the bit rate of the speech coder is increased to 7,4Kb/s (Fig.8) the congestion also increases and the advantages of a higher value of PTIME becomes more apparent. However, the relative advantage of PTIME=20 becomes smaller. The same observations are valid when the speech coding mode is further increased to 12.2Kb/s, as shown in Fig.9. There are clearly noticeable benefits of working with a higher value of PTIME to reduce overhead inefficiencies. However this approach inherently adds a delay component and just by adjusting PTIME is not possible to produce a good performance. When the 12.2 Kb/s codec is employed even a moderate average capacity of 3 users per sector (or 1.5 user/slot) is not attainable. As a matter of fact, considering the QoE3 requirement and for this high rate

speech mode, it is only possible to achieve an average capacity between 1 and 1.5 user/slot, regardless of the value of PTIME.

Fig. 10 illustrates the performance dependency on the average value of the log-normal distribution that regulates the response time t_R. When the mean of this distribution is increased from 4 to 10 seconds the traffic demand in the system decreases. The corresponding increase in capacity is substantial, actually it more than doubles for this particular implementation of the AMR coder, which allowed for dynamic adaptation of the speech coding rate.
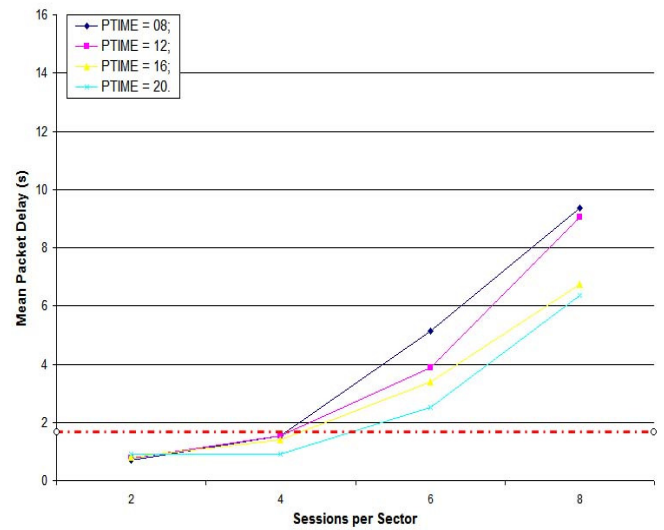

Figura8 : Average delay per packet vs. number of conversations per sector for different values of PTIME. AMR 7.4 Kb/s codec.
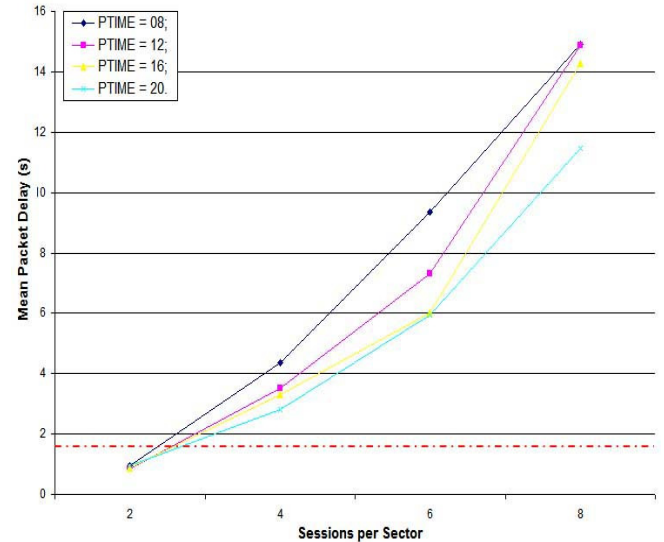

Figura 9 : Average delay per packet vs. number of conversations per sector for different values of PTIME. AMR 12.2 Kb/s codec.

V. CONCLUDING REMARKS

In this work we initially reviewed the main features of the GSM/EDGE technology and of the Push-to-Talk over cellular (PoC) service. Then the capacity of the PoC service when the underlying radio interface technology is the EDGE was estimated. This performance computation was evaluated through computer simulation and it used existing results for the propagation and interference characteristics of EDGE. The results are given for three different bit rates of the AMR speech codec. It was shown that for a fast response time by the communicating users, the capacity achieved by the 12.2 Kb/s codec is quite low. However if lower rate codecs are employed it is possible to operate with at least 2 conversations per traffic slot. Improvements are also likely to be obtained if enhancements of the GERAN specifications, such as the Extended Uplink TBF feature, reported in [16], are adopted.

VII.   REFERENCES

[1] S. K. Raktale, "3PoC: An Architecture for Enabling Push-to-Talk Services in 3GPP Networks," 2005

[2] P. Kim, A. Balazs, E. van de Brock, G. Kieselman and W. Böhm, "IMS-based Push-to-Talk over GPRS/UMTS", Records, *IEEE Wireless Communications and Networking Conference – WCNC*, pp. 2472 – 2477, New Orleans, March 2005.

[3] E. O´Reagan and D. Pesch, "Performance Estimation of a SIP based Push-to-Talk Service for 3G Networks", Records, *Fifth European Wireless Conference*, Barcelona, February 2004.

[4] Internet Engineering Task Force – IETF, "SIP: Session Initiation Protocol,", RFC 3261.

[5] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: A Transport Protocol for Real Time Applications," IETF RFC 3550, July 2000.

[6] OMA Push to Talk over Cellular (PoC) - Architecture, Candidate Version 1.0 – 28 April 2005, OMA_AD _ PoC – v 1.0 – 20050428-C, 156p.

[7] K. Zangi, A. Furuskar and M. Hook, "EDGE: Enhanced Data Rate for Global Evolution of GSM and IS-136," Proc. MDMC´98.

[8] T. Halonen, J. Romero and J. Melero, *GSM, GPRS and EDGE Performance (Evolution Towards 3G/UMTS)*, New York: Wiley, 2003.

[9] P. Schefczik, "RLC/MAC Simulation of GPRS/EDGE," Lucent Technologies, http://www.winlab.rutgers.edu/pub/symposiums/GPRS/Agenda.html

[10] H.M. Queiroz and L.G.R. Guedes, " Simulation Environment for EDGE System (E-GPRS) with Decision Mechanism based on Error Threshold Coding," (in Portuguese), I Encontro Regional de Modelagem e Análise de Sistemas, UCG, October 2004.

[11] H. Olofsson, "Requirements for GPRS Evolution Towards Providing Third Generation Services," Ericsson Radio Systems, http://www.winlab.rutgers.edu/pub/symposiums/GPRS/Agenda.html

[12] E. de Souza e Silva and R. M. Leão, "The TANGRAM-II Environment,", Proc. 11th Intl. Conf. on Computer Performance Evaluation, Modelling Techniques and Tools, vol. LNCS 1786, pp. 366 – 369, March 2000.

[13] K. Järvinen, "Standardization of the Adaptive Multi-Rate Codec," *Proc. European Conf. on Signal Processing*, Tampere, 2000.

[14] V.Rönnholm, "Push-to-Talk over Bluetooth", Proc.,39th IEEE Hawaii Intl. Conf. on Systems Science, 2006.

[15] Y.Wang, M. Zuckerman and R.J. Harris, "PTT Packet Delay Analysis for GPRS/GSM Links", IEEE Communication Letters, vol.10 (6), pp. 456 – 458, 2006.

[16] C.F. Ball, C.Masseroni and R. Trivisonno, "Introducing 3G like Conversational Services in GERAN Packet Data Networks", Records, IEEE Vehicular Technology Conf. Spring, 2005.
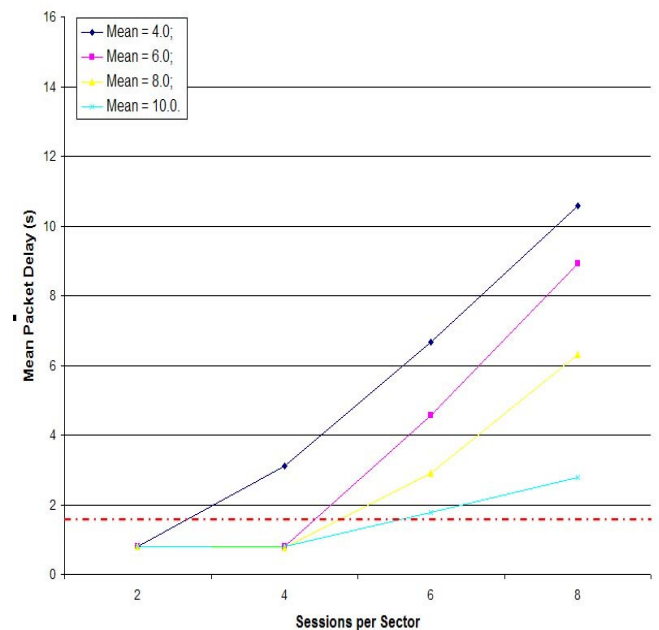
Figura 10 : Average delay per packet vs. number of conversations per sector for different mean values the log-normal distribution. AMR with coding rates adapted acording with C/I.