Performance of VoIP Call Set-up Over Satellite-UMTS Using Session Initiation Protocol

V. Y. H. Kueh, R. Tafazolli, B. G. Evans

Centre for Communication Systems Research (CCSR), University of Surrey, Guildford, Surrey, United Kingdom E-mail: V.Kueh@eim.surrey.ac.uk

Abstract: Session Initiation Protocol (SIP) is an application layer signalling protocol used in the IPbased Universal Mobile Telecommunication Systems (UMTS) network for establishing multimedia sessions. With a satellite component identified to play an integral role in UMTS, there is also to support SIP-based session need a establishment over Satellite-UMTS (S-UMTS) to achieve end-to-end seamless **IP-based** terrestrial/satellite network integration. Due to the inherent characteristics of SIP, the performance of SIP-based session set-up is largely compromised when transported over an unreliable wireless link with a large propagation delay. This paper presents the work done in incorporating a link layer retransmission based on the UMTS Radio Link Control acknowledgement mode (RLC-AM) mechanisms to improve the call set-up performance. Specifically, our investigation focuses on the impact of the poll prohibit timer on the system performance and via simulations in OPNET, some insight into the configuration of this timer is identified.

1. Introduction

Third generation mobile communication systems, such as Universal Mobile Telecommunication Systems (UMTS), are paving the way towards the much anticipated integration of the most successful technologies of the last decade - the Internet and cellular mobile telephony. As we enter the 21st century, we will see wide-area wireless Internet access on a global scale to a large variety of services such as web browsing, multimedia messaging, e-commerce, video telephony and streaming multimedia. The provision of IP-based multimedia services in UMTS is made possible through the introduction of the IP Multimedia Subsystem (IMS) as part of the 3GPP Release 5 set of standards. IMS is an overlay control network, which makes use of the underlying packet-switched domain for the transport of signalling and user data. Session Initiation Protocol (SIP) [1], a protocol developed within the Internet Engineering Task Force (IETF), has been selected by the Third Generation Partnership Project (3GPP) as the end-to-end signalling protocol for establishing IP-based multimedia sessions such as voice over IP (VoIP) between the user equipment (UE) and the IMS, as well as between the components within the IMS and with other end users over the Internet [2].

At the same time, a satellite component has been identified in UMTS in order to provide a true global seamless - anytime, anywhere - mobile multimedia communication. Rather than as a stand-alone system as in the 2nd generation mobile global satellite systems (Iridium, Globalstar, ICO), whereby the terrestrial and satellite mobile systems were developed independently, satellite-UMTS (S-UMTS) is expected to play a complementary role to the terrestrial-UMTS (T-UMTS). In addition to its fast service deployment and coverage extension capability, S-UMTS, as a direct consequence of its broadcast nature and ubiquitous coverage, offers a natural way to provide multicast and broadcast services in the most cost-efficient manner. To achieve a high degree of commonality with the IPbased T-UMTS network, there is a need to support SIP-based session establishment over the satellite component as well.

Nevertheless the transport of SIP over the radio interface is not efficient due to the inherent characteristics of SIP being transactional-based and generous in size. This is made worse when transversing an error-prone wireless link with a larger satellite propagation delay. One of the techniques to reduce the call set-up delay is by compression of signalling messages, such as the text-based message compression technique proposed by [3]. In this paper, we investigate the implications of incorporating a link layer retransmission based on the acknowledgement mode (AM) mechanisms defined in the UMTS Radio Link Control (RLC) specification [4] as a method to conceal the link-related losses from the upper layers so as to improve the call establishment performance. RLC is a very versatile protocol with a range of parameters and timers, which can be reconfigured. More specifically, the impact of the setting of the poll prohibit timer on the system performance under different channel conditions is studied.

The paper is organised as follows. In Section 2, the end-to-end session establishment procedure for an IMS voice call over S-UMTS is explained. This is followed Section 3 by a brief description of the in retransmission schemes as defined for RLC-AM in 4 describes the simulation UMTS. Section environment developed using OPNET¹ and the assumptions used. Section 5 presents the results as well as discussions, and final remarks are addressed in Section 6.

¹ OPNET is a trademark of Opnet Technologies Inc.



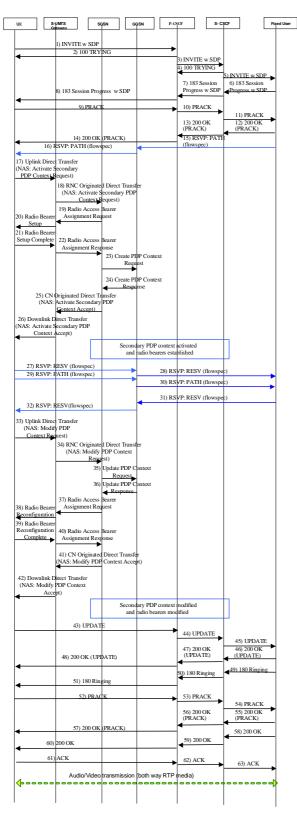


Figure 1: SIP call set-up message sequences for mobile originated call to fixed user

There are several key procedures involved before an IMS session can be established. Upon switch on, a UE performs an RRC Connection, an elementary procedure to establish a radio control connection between the UE and the radio access network [5].

After having transited from RRC Idle mode to RRC Connected mode, the UE performs an Iu signalling connection set-up sequence for sending Non-Access Stratum (NAS) messages (these include Service Request/Response via the Initial Direct Transfer procedure, Authentication and Ciphering Request/Response via the Direct Transfer procedure; and Security Mode procedure) between the UE and the GPRS network. The UE then needs to perform registration, which is mandatory in UMTS before the UE can initiate or terminate a session. There are 2 levels of registration, namely registration at the bearer level, i.e. with the GPRS network, whereby the UE obtains its IP address, and registration at the application layer, i.e. with the IMS. Registration with the IMS is done so that not only can the UE be reached for terminating sessions and services (user availability), but also the UE can be pre-authenticated early and assigned to a particular serving proxy, whereby the user service profile is downloaded for that user to trigger. The current working assumption in 3GPP is that authentication is done during registration (i.e. prior to session establishment) so that the authentication procedure does not contribute to the overall session set-up time.

Figure 1 depicts an example of an end-to-end session establishment signalling flow for a mobile originated VoIP session, whereby the UE is accessing another fixed terminal (a SIP user agent) reachable in the Internet through the UMTS Satellite Radio Access Network (USRAN), the UMTS IP Core Network and the IMS. The reference architecture employed herein is consistent with the UMTS Release 5 IP-based network architecture [2]. The USRAN consists of the satellite and the S-UMTS gateway, whereby in the selected scenario, the satellite is considered to be a transparent GEO satellite while the Node B and Radio Network Controller (RNC) are collocated in the gateway. It is assumed that the UE is located within the service area of the network operator to whom the UE subscribes and that the UE has already set-up an RRC connection and performed registration with both the GPRS network as well as with the IMS network. Note that a primary PDP (Packet Data Protocol) context is assumed to be activated and is used to carry the IMS related signalling.

As can be seen, the call establishment can be complicated as it involves a great deal of signalling exchange, which involves not only SIP related messages (for e.g. INVITE, PRACK, 180 ringing), but also RSVP (for e.g. PATH and RESV) and UMTSspecific procedures (for e.g. Radio Bearer Set-up and Modify PDP Context Request). SIP session set-up basically comprises of four distinct phases - the session invitation, resource reservation, session offering and session connection. The session invitation phase (steps 1-5) starts with the calling party sending a SIP INVITE to the called party, and the calling party receiving a 100 Trying response from the SIP servers as an indication that the network is in the process of routing the invitation to the destination. This is followed by the reservation phase (steps 6-48), of which the necessary resources are reserved to ascertain that the transport bearer for the media stream is available when the called party answers, so as to avoid the annoying case of a user answering a ringing phone only to find that there is no speech path available. Furthermore it enables early tones and announcements (for e.g. ring tone or busy tone) to be played back to the calling party (using the bearer set-up for the media stream) prior to the call being answered. This resource reservation phase, deemed to be the most complex part of the session establishment, is necessary in order to achieve the quality of service (QoS) needed in UMTS for conversational calls. Once the appropriate resources for the network and radio access bearers are available, the session offering phase begins (steps 49-57) with the called user alerted to the incoming call, and the calling party being informed by the 180 Ringing provisional response. Finally, the session is connected (steps 58-63) when the called party answers the call, and a 200 OK final response is sent and the calling party acknowledges it by sending an ACK message.

3. UMTS Radio Link Control Protocol Acknowledgement Mode Overview

The UMTS RLC-AM is based on a hybrid sliding window ARQ protocol with selective acknowledgments (SACK) and negative acknowledgments (NAK). It provides segmentation and retransmission and is capable of in-sequence delivery, duplication detection and piggyback control information.

Segmentation is performed if the received RLC service data unit (SDU) is larger than the length of available space in the AM mode data (AMD) protocol data unit (PDU). The AMD PDU size is a semi-static value that is configured by upper layers and can only be changed through re-establishment of the RLC-AM entity by upper layers.

Retransmission of PDUs, which have not been successfully received at the receiver, is performed when the sender receives a feedback from the receiver in the form of a STATUS report. Each STATUS report consists of one or more STATUS PDUs. Acknowledgment confirmation of received PDUs as well as those that are not successfully received is indicated in this report. Status report is sent as a control PDU, which has a higher priority for transmission than the AMD PDUs; STATUS PDUs can also be piggybacked onto an AMD PDU if space permits. A STATUS report is triggered when either a polling request, made by marking the polling bit of outgoing AMD PDUs, is received or a missing AMD PDU is detected when a 'missing PDU indicator' option is configured. The former method is known as solicited STATUS report, while the latter is known as unsolicited STATUS report. There is a third trigger based on a timer, known as the timer based STATUS transfer, which is not considered in our study. Note that an AMD PDU can only be transmitted up to a maximum number of times equal to (MaxDAT - 1), after which an SDU discard procedure is initiated. Note that if the 'No_discard after MaxDAT number of transmissions' option is configured, then the RLC reset procedure is initiated instead.

There are various polling triggers, of which we only consider the following in our simulations: last PDU in transmission buffer, last PDU in retransmission buffer and timer based (Timer_Poll). The triggering of these different polling mechanisms should be configured properly to avoid deadlock situations. Also there is a trade-off in the frequent sending of these polling requests. On the one hand, a fast polling request can improve the delay performance, as the acknowledgement feedback delay is reduced. However, on the other hand, extra bandwidth is consumed since STATUS report will be generated more often and these occupy the link bandwidth on the reverse link. Also this overhead can degrade the throughput as well as the delay of the AMD PDU sent on the reverse link since control PDU has precedence over them.

Since there is a potential risk of the network being overwhelmed by excessive polling requests (as the different polling options can be present simultaneously), a poll prohibit timer can be configured to prohibit too frequent polling. The poll prohibit timer is started along with the poll timer when an AMD PDU with the poll bit is sent. No polling is allowed until the poll prohibit timer expires; only one poll is transmitted when it expires, even if several polls were triggered during the time it was active.

To avoid buffer overflow and to reduce the maximum acceptable delay, an SDU discard function is performed. There are several alternative operational modes detailed in the specifications for RLC-AM, of which the 'SDU discard after MaxDAT number of transmissions' is chosen in our study. This option keeps the SDU loss rate constant at the cost of a more variable delay compared to other options.

As can be seen, RLC is extremely flexible and can be configured in several ways. Nevertheless, with a multitude of different options, parameters and timers, it can be a formidable task in fine-tuning them given their close interactions.

4. Simulation Model and Assumptions

With the multitude of protocols involved in the session set-up procedure, and also with various timers and mechanisms interacting with each other in a single protocol, e.g. for the RLC, an analysis of the combination of these issues in a variable channel condition is difficult to be quantified analytically, if at all possible. Therefore we have resorted to simulation, which is also capable of exploring the system behaviour in the presence of very complex parameter settings, instead of having the need to make simplified assumptions as is usually done in an analytical approach. So to assess the performance of SIP-based call establishment over S-UMTS incorporating RLC-AM, a system level simulator was developed in OPNET.

The signalling sequences, as depicted by Figure 1, were modelled according to their protocol behaviour as detailed in their respective specifications with the message sizes following the ones in [6] and are listed in Table 1. Since IPv6 is adopted in IMS, all the messages sent on the Radio Bearer (RB) have an UDP/IPv6 header of 48 bytes and a Packet Data Convergence Protocol (PDCP) header of 3 bytes, while messages sent over the Signalling Radio Bearer (SRB) do not incur these header overheads. These higher layer messages will be passed to the RLC, where the AMD PDU size is set to 320 bits. RB and SRB messages are sent to the MAC layer via the Dedicated Traffic Channel (DTCH) and Dedicated Control Channel (DCCH) logical channels, respectively. These dedicated logical channels are later mapped to the Dedicated Channel (DCH) transport channel, before being finally sent over the air interface via the Dedicated Physical Data Channel (DPDCH). Note that with our choice of configurations, there is always a one-to-one or one-to-many SDU to PDU(s) mapping, and the RLC PDUs are mapped one-to-one onto the transport blocks (only one RLC PDU is assumed to be transmitted per TTI).

The round-trip delay over the UMTS IP Core Network was assumed to be 150 ms [7], while the mean one-way Internet delay and its standard deviation are 50 ms and 7 ms, respectively [8]. The processing time per SIP message in each server is 25 ms [3] while the RLC processing time was assumed to 15 ms for both uplink and downlink be communications [9]. The mobile satellite channel model used was a simplification of the Lutz statistical model [10], where the fading process is switched between a good state (unshadowed areas) and a bad state (shadowing areas). In the bad state, everything that is sent is assumed to be corrupted, while in the good state, the successful reception of the packet depends on the block error rate (BLER), which is uniformly distributed. Thus the channel model used was essentially an ON-OFF model with a certain BLER characterising the ON state, and the transition between the good and bad states is characterized by a two-state Markov chain. It was assumed that there is no loss in the fixed network. The rest of the parameters used in the simulations are summarized in Table 2.

The performance metrics measured in this study are based on the call set-up quality, which relates to what the user experiences when a call is made [11]; they are the call set-up delay and call blocking probability. Also parameters measured at the link layer, i.e. the RLC SDU and PDU delays, are also used as performance indicators. The call set-up delay is defined here to be the interval between entering the last dialled digit and receiving a ringback [12], i.e. from the time the INVITE is sent until a 180 Ringing message is received. As defined in [11], the call blocking event² occurs when the user decides to abort the call after the set-up delay becomes excessively long; here it is governed by both the protocol timers and the maximum attempt of the transmission of each higher layer messages. The RLC SDU delay³ is measured from the time at which the messages from higher layers are delivered to the RLC at the transmitter until the time the messages are correctly reassembled at the receiver. The RLC PDU delay³ is defined from the time an RLC AMD PDU is first transmitted to the time it is correctly received at the receiver or aborted after the maximum number of allowed retransmissions.

Note that in 95% of all cases, confidence measurements confirmed that the confidence interval of the simulation results is within 3% of the expected value.

Message	Radio Bearer	Length
0	Туре	(bytes)
1. INVITE	RB	620
8. 183 Session Progress	RB	500
9. PRACK	RB	250
14. 200 OK (PRACK)	RB	300
16. RSVP: PATH (flowspec)	RB	128
17. Uplink Direct Transfer	SRB	275
(NAS: Activate Secondary PDP		
Context Request)		
20. Radio Bearer Set-up	SRB	10
21. Radio Bearer Set-up Complete	SRB	5
26. Downlink Direct Transfer	SRB	20
(NAS: Activate PDP Context		
Accept)		
27. RSVP: RESV (flowspec)	RB	148
29. RSVP: PATH (flowspec)	RB	128
32. RSVP: RESV (flowspec)	RB	148
33. Uplink Direct Transfer	SRB	275
(NAS: Modify PDP		
Context Request)		
38. Radio Bearer Reconfiguration	SRB	10
39. Radio Bearer Reconfiguration	SRB	5
Complete		
42. Downlink Direct Transfer	SRB	20
(NAS: Modify PDP		
Context Accept)		
43. UPDATE	RB	620
48. 200 OK (UPDATE)	RB	450
51.180 Ringing	RB	230
52. PRACK	RB	250
57. 200 OK (PRACK)	RB	300
60. 200 OK	RB	450
61. ACK	RB	230

Table 1: Typical message size for session establishment

² Call blocking event can alternatively occur when the call attempt is denied by the network admission controller due to insufficient resources.

³ Note that RLC SDU delay only recorded results for successful RLC SDUs, while RLC PDU delay also included the delay for unsuccessful PDUs, i.e. those that have reached their maximum attempt of transmissions.

SIP Timer, T1	2 s	Timer T3380	30 s
SIP Timer, T2	16 s	Timer T3381	8 s
SIP Timer, T4	17 s	RLC Window Size	1024
RSVP R Value	30 s	RLC Poll Timer	0.8 s
RSVP K Value	3	RLC Poll Prohibit	1.6 s
		Timer	
Max. Tx. of		RLC MaxDAT	4
Activate Sec. PDP	5	TTI	10 ms
Context Request &		Data Rate	32 kbps
Modify PDP			*
Context Request			

Table 2: Simulation parameters settings

5. Simulation Results and Discussion

Figure 2 to Figure 9 compare the performance when the poll prohibit timer is set with when it is disabled for T_{good} (mean sojourn time in good state) ranging between 0.5 and 10s, and T_{bad} (mean sojourn time in bad state) equals to 0.5, 2 and 4s for different BLER encountered in the good channel state. The results presented here are with the solicited STATUS report feedback.

The results illustrate that better performance is achieved when the poll prohibit timer is not configured

Figure 2: Session set-up delay comparison with and without poll prohibit timer for BLER = 1% in good state

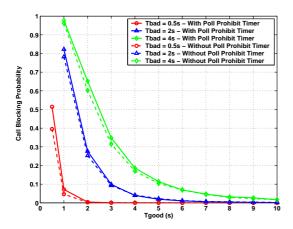


Figure 4: Call blocking probability comparison with and without poll prohibit timer for BLER = 1% in good state

(similar to the poll prohibit timer set to a minute value). This is because with the poll prohibit timer configured, the recovery of missing PDUs is delayed since a STATUS report, which indicate the missing PDUs to be retransmitted, cannot be sent until a poll is received, and that request cannot be made until the poll prohibit timer expires. This resulted in a higher PDU delay, which essentially leads to a higher SDU delay and the overall session set-up delay. In addition when retransmission at the link layer is deferred for too long, higher layer protocols timers may timeout and if the maximum transmission attempt of these higher protocol is reached, the call set-up cannot be completed successfully and this leads to a higher blocking probability (call set-up failure). Also it can be seen that the performance improvement for the case without the poll prohibit timer set over the one with the poll prohibit timer configured increases as the channel condition gets worse due to the more rapid recovery process of the missing PDUs.

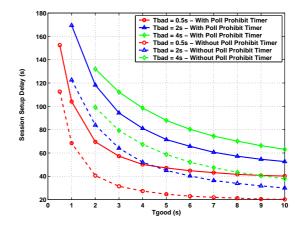


Figure 3: Session set-up delay comparison with and without poll prohibit timer for BLER = 10% in good state

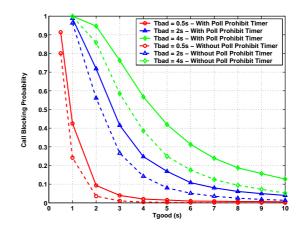


Figure 5: Call blocking probability comparison with and without poll prohibit timer for BLER = 10% in good state

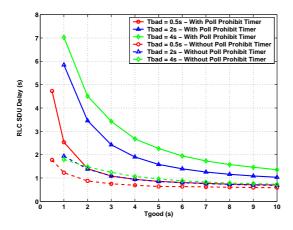


Figure 6: RLC SDU delay comparison with and without poll prohibit timer for BLER = 1% in good state

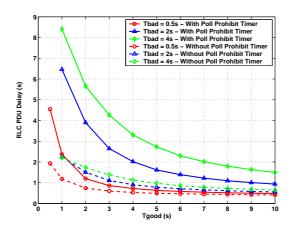


Figure 8: RLC PDU delay comparison with and without poll prohibit timer for BLER = 1% in good state

Having compared the effect of having the poll prohibit timer disabled and enabled, the impact of setting different values for this timer is now investigated. Figure 10 to Figure 13 show these results for different values of poll prohibit timer (expressed in terms of poll timer) for different BLER experienced in the good state when T_{good} is 6s and T_{bad} equals to 2s.

It can be seen that poll prohibit timer with lower values gives a better performance in terms of delay since the error recovery can be performed faster as it is not impeded by the prohibit timer as much as the ones with higher values. Note that, essentially the recovery speed of the missing PDUs is a function of the poll prohibit timer and the propagation delay; with the satellite propagation delay already fixed, a low poll prohibit timer can cause a considerable decrease in the total session set-up delay. For example when the BLER is 10%, the session set-up delay is reduced by more than 27s when a poll prohibit timer of 3 times the poll timer value is used, as compared to a poll prohibit timer of 5 times the poll timer value. Also, the call blocking probability is higher with a higher value of poll prohibit timer; this can be explained with the same reasoning as before, i.e. the higher layer

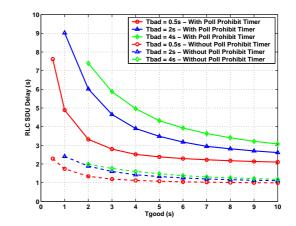


Figure 7: RLC SDU delay comparison with and without poll prohibit timer for BLER = 10% in good state

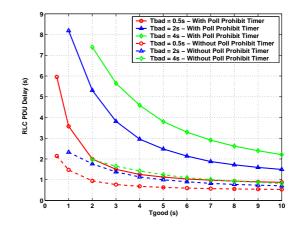


Figure 9: RLC PDU delay comparison with and without poll prohibit timer for BLER = 10% in good state

protocols may time out before the missing PDUs can be fully recovered by the link layer, and this could lead to call set-up failure when the higher layer protocols reach their maximum transmission attempt. It is further observed that the performance degradation is more severe when the channel condition worsens; in those conditions when the transmissions of the AMD PDUs, which may contain the polling requests, and the STATUS reports are more susceptible to wireless errors, more retransmissions are needed before the SDUs are successfully reassembled, and therefore make the effect of a large poll prohibit timer more pronounced. Note that the observations made here that lower values of prohibit timer give better performance are different from the ones made in [13]; the reason being that in [13], the source activity is high and hence frequent polling causes more STATUS reports to be generated and these feedback, which have higher priority than new PDUs, causes the TCP delays to increase. This adverse effect of higher frequency polling is not observed here since, unlike TCP traffic, the source activity for session set-up signalling sequences is low as they are transactional based.

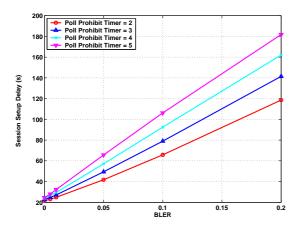


Figure 10: Session set-up delay for different prohibit timer values (prohibit timer values are given in terms of poll timer value)

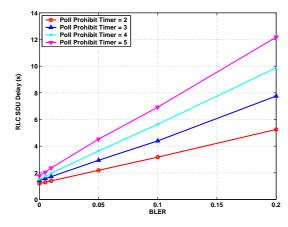


Figure 12: RLC SDU delay for different prohibit timer values (prohibit timer values are given in terms of poll timer value)

6. CONCLUSIONS

With SIP being the official end-to-end IP signalling protocol for establishing multimedia sessions in IPbased terrestrial UMTS networks, it is therefore essential to support SIP over S-UMTS as well so as to achieve maximum commonality with the terrestrial systems. Due to the inherent characteristics of SIP signalling being transactional-based and generous in size, the transport of SIP over the radio interface is not efficient and when transversing the error-prone wireless link with a larger satellite propagation delay, the call set-up performance is severely compromised. In this paper, we have addressed the issue of incorporating a link layer retransmission based on the RLC-AM mechanisms so as to improve the SIP-based call set-up performance.

Towards that end, the effect of the poll prohibit timer configuration has been studied. It was shown that the error recovery could be speeded up by having smaller values for the poll prohibit timer (or total disabling the poll prohibit timer in the extreme case). This configuration brings about a more frequent polling request, which is crucial in a bad channel

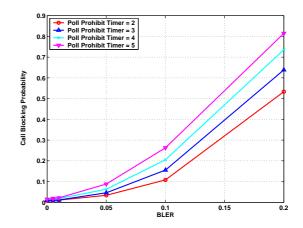


Figure 11: Call blocking probability for different prohibit timer values (prohibit timer values are given in terms of poll timer value)

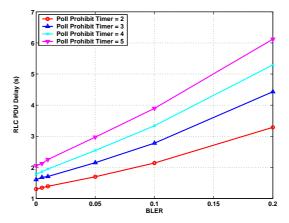


Figure 13: RLC PDU delay for different prohibit timer values (prohibit timer values are given in terms of poll timer value)

condition as it enables the sender to know the status of its PDU transmission more promptly and consequently take the appropriate action when missing PDUs are reported. Higher frequency of polling comes at a price though, in that with the STATUS reports being generated more often, the throughput and the delay of the AMD PDUs sent on the reverse link can be degraded (since control PDU has higher priority). However this adverse effect is not observed here since the call set-up signalling flow has a low source activity due to its request-response nature.

REFERENCES

- J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley and E. Schooler, "*SIP: Session Initiation Protocol*," Request for Comments 3261, IETF, June 2002.
- [2] 3GPP TR 23.821 V.1.0.1: "Architecture Principles for Release 2000, (Release 2000)," July 2000.
- [3] G. Foster, M. Pous, A. Sesmun, V. Kenneally and D. Pesch, "Performance Evaluation of UMTS Packet Voice Call Control," 5th European Personal Mobile Communications Conference

(EPMCC 2003), Glasgow, Scotland, April 2003, pp. 136-140.

- [4] 3GPP TS 25.322 V5.3.0: "Radio Access Networks; RLC Protocol Specification; protocol specification (Release 5)," December 2002.
- [5] 3GPP TS 25.331 V5.3.0: "Radio Access Networks; Radio Resource Control (RRC); protocol specification (Release 5)," December 2002.
- [6] 3GPP TSG GERAN, "A Comparison Between Packet-Switched Call Setup Using SIP and GSM Circuit-Switched Call Setup Using RIL3-CC, RIL3-MM, RIL3-RR, and DTAP," GP-000508, Nortel Networks, November 2000.
- [7] H. Holma and A. Toskala, "*WCDMA for UMTS*," John Willey & Sons, 2002.
- [8] T. J. Kostas, M. S. Borella, I. Sidhu, G. M. Schuster, J. Grabiec, and J. Mahler, "*Real-time voice over packet-switched networks*," IEEE Network, no. 1, January/February 1998, pp. 18-27.
- [9] 3GPP TR 25.853 V.4.0.0: "Delay Budget within the Access Stratum (Release 4)," March 2001.
- [10] E. Lutz, D. Cygan, M. Dippold, E. Dolainsky and W. Papke, "The land mobile satellite channel recording, statistics and channel model," IEEE Transactions on Vehicular Technology, May 1991, pp 375-386.
- [11] J. Laiho, A. Wacker and T. Novosad, "Radio Network Planning and Optimization for UMTS," John Willey & Sons, 2002.
- [12] T. Eyers and H. Schulzrinne, "Predicting Internet Telephony Call Setup Delay," IPTel 2000 (First IP Telephony Workshop), Berlin, Germany, April 2000.
- [13] Atul S Joshi, R. Sujatha, Anoop Kulkarni, "Performance Evaluation of TCP over Radio Link Control (RLC) Protocol for WCDMA Physical Layer Based Third Generation Wireless Networks," in Proc. IEEE's 3GWireless'00 Conference, Silicon Valley, USA, June 2000.