SIP-based Applications in UMTS: A Performance Analysis

Maria Isabel Pous[1], Dirk Pesch[1], and Gerry Foster[2]

[1]Adaptive Wireless Systems Group
Cork Institute of Technology, Ireland
e-mail: dpesch@cit.ie

[2]Motorola GSM Networks
Swindon SN5 8YQ,
United Kingdom

Abstract: **With the ever increasing penetration of IP technologies and the tremendous growth in wireless data traffic, the wireless industry is evolving the mobile core networks towards all IP technology. The 3rd Generation Partnership Project (3GPP) is specifying an IP Multimedia Sub-system (IMS) in UMTS Release 5/6, which is adjunct to the UMTS Packet Switched (PS) GPRS CN. This IP-based network will allow mobile operators to provide commonly used Internet applications to wireless user. The UMTS IMS uses the Internet Engineering Task Force (IETF) defined text-based Session Initiation Protocol (SIP), to control a wide range of anticipated IP-based services offering new services such as multimedia calls, chat, presence services. Initial indications as to the signalling delay associated with SIP messages have concerned operators about the viability of such services over the UMTS air interface. This paper provides an insight into the UMTS system performance, focusing on the UMTS SIP-based service where typical delay-sensitive and non-sensitive applications, such as chat and messaging services are studied. Furthermore we discuss and analyse the requirements and possible solutions for the efficient use of SIP in a wireless environment, such as protocol compression.**

1. Introduction

Second-generation wireless systems, such as GSM, were primary designed to provide voice services to the end user with an acceptable quality. This has been achieved with remarkable success. Moreover, short messages and low-rate (9.6kbps) data services were added to speech services. Lead by the demand for mobile data access and the explosive growth of Internet data services over the past 10 years, wireless data applications are seen as the major new revenue stream for next generation mobile networks, i.e. 3G mobile networks.

Presence and Instant Messaging services have a strong following on the Internet, with services such as AOL IM, Windows Messenger, Yahoo Messenger, Jabber, and ICQ. A similar service does not exist in the mobile domain yet, but efforts by 3GPP are underway to define such a service, which will utilise the IETF SIP protocol and its SIMPLE extensions for Presence and Instant Messaging [3, 4, 5]. This messaging service combined with the presence awareness (always on-line paradigm) will compliment and may even replace the present day SMS.

In order to provide insight into the performance that can be expected from such as service, a system model has been implemented in a computer simulation environment. Initial results indicate that this service will put a significant burden on the UMTS Radio Access Network (RAN) as well as the Core Network (CN) to the large message sizes of the text based signalling protocols used.

2. UMTS Network Architecture Rel. 5/6

Third-generation mobile systems evolve the mobile core network towards an all IP technology with a new radio network that provides higher capacity and data rates required for the support of advanced multimedia services. The 3G evolution is taking place in different phases in which both radio and core networks are upgraded from those in GSM. While the first phase of UMTS based on Release 99 still includes tow distinct core networks, one for circuit-switched (CS) and one for packet-switched (GPRS) support, UMTS Release 5/6 moves towards an all IP Multimedia Core Network Subsystem (IMS), with full IP packet support. Figure 1 depicts the main components of the UMTS Rel6 architecture, including the UTRAN, the PS CN elements and the elements of the IMS.

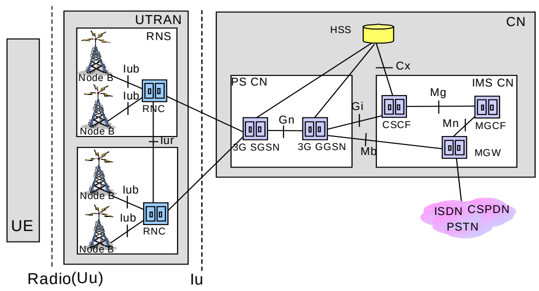

Figure 1 - UMTS functional architecture

The IMS introduces three main logical network elements to the existing infrastructure: the Call Session Control Function (CSCF), the Media Gateway Control Function (MGCF), and the Media Gateway (MGW). Basically the MGCF controls the MGW used in connections to external networks. The CSCF provide the logic of how transactions using IMS are treated. The CSCF may assume several functions, depending on whether it is operating as a Proxy, Interrogating or Serving CSCF [1, 2]. The Home Subscriber Server (HSS) is also introduced providing user profile information similar to that of today's HLR.

3. SIP Presence and Instant Messaging

3.1 SIP based Call Control

The primary function of SIP [3], as its name implies, is the establishment of a session. The session set-up starts with the *'SIP INVITE'* message and finish with the *'SIP ACK'*. The two end parties negotiate the media characteristics for the session and make a decision on the media streams they will support during the session using the Session Description Protocol (SDP). After the media characteristics have been determined, the network reserves the necessary resources for supporting this session. The resource reservation phase entails creating a secondary PDP context for transport of the required media, and setting up the corresponding radio access bearers and radio bearers. In the case presented here, different sessions are to be preformed, voice calls and instant messaging sessions therefore the secondary PDP context is activated depending on the service traffic characteristics. Once the resource reservation is completed successfully, the terminating point sends a *'SIP 200 OK'* final response and the originating mobile replies with a *'SIP ACK'* message to confirm the session set-up.

In this paper, we have modelled and simulated different SIP call/session set-up scenarios, according to the 3GPP specifications [1].

- **Mobile Originated (MO):** the mobile (caller), assumed located in the home network, initiates a session destined to a fixed phone (callee). A single operator performs both parts origination and termination.
- **Mobile Terminating (MT):** the fixed phone (caller) starts the call to the mobile part, which is considered attached to the network.
- **Mobile-to-Mobile (M2M):** the session involves two mobiles, located in the same network, being the home network for both entities.

3.2 Presence & Instant Messaging services

SIP capabilities have been extended to handle other services already in use in the Internet domain, Presence and Instant Messaging. The presence service defined in RFC 2778 [6] by IETF is being standardised in 3GPP Release 6 [4, 5] for its support in UMTS. Presence service allows users to subscribe to each other and be notified of changes in their state (e.g. going off-line, changing contact details, etc.). Its combination with a messaging service will provide a simple and fast way of real-time communication between online users.

3.2.1 Presence Service Overview

Presence conveys the ability and willingness of a user to communicate across a set of devices (presentity). Figure 2 shows the presence model architecture as defined in 3GPP TS 23.141 [4] and IETF RFC 2778 [6]. The Presence Server, which resides in the presence entity's home network, manages and distributes the information to interested parties, called watchers. The two sets of entities involved, *presentity* and *watchers*, are either internal or external to the home network and access network. Watchers access the server through presentity *proxies*

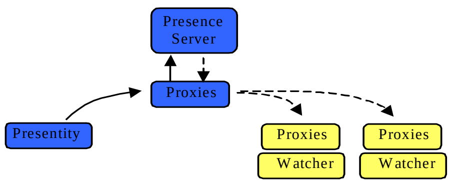

Figure 2 - Presence service

3.2.2 Instant Messaging Service Overview

The exchange of content between the participants in near real time is realised with instant messaging. Generally, the content is short text messages and its transfer fast enough in order to maintain an interactive conversation. Each message can be sent independently using the SIP MESSAGE method, or messages can be associated into sessions that are initiated using SIP INVITE. The first approach is often referred as *pager-mode*, due to its similarity to the behaviour of two-way pager devices, and is used when small short IMs are sent to a single or reduced number of recipients. On contrast the second approach, called *session-mode messaging*, is required for extended conversations, joining chat groups, etc. Both approaches, defined by SIMPLE, are considered in our model.

- **Message Session Model**: In this model the IM traffic is viewed as a media stream, which is part of a session established with a SIP INVITE method. Before user communication can start, a SIP INVITE is used to set-up the session, describing the IM stream in the SDP part of the message. As the data is always sent over a reliable link, the message size is not restricted. This model offers advantages when the number of messages processed increases. Once the initial INVITE request is processed, the subsequent SIP messages sent within the established session, bypass any intervening SIP proxy. Therefore the message load decrease on those network elements. The model is used in text conferencing and chat applications where it is useful and more efficient to have messages associated.
- **Paging Model**: Here no set-up or session establishment is required before sending a message. Therefore each message is sent independently using the MESSAGE method. It mimics the operation of SMS in today's wireless network. This method has limitations on the message size (<1300 bytes) due to network congestion concerns.

4. UMTS Signalling Procedures

At the start of a packet-switched user application, a Bearer Service connection (PDP context with specific Radio Access Bearer and Radio Bearer) needs to be established to enable transfer of data. However, before a RNC can control any requested bearer, it needs to create a signalling connection between the UE and the CN. This connection transfers the higher layer information between entities in the Non Access Stratum. Between the UE and the UTRAN, RNC uses the Radio Resource Control (RRC) connection services in creation the Signalling Radio Bearer (SBR), and through the Iu interface a signalling bearer is then created.

4.1 Signalling Connection

Upon power on, the UE establishes at most one radio control connection in order to access the UTRAN. The set-up procedure, as shown in Figure 3, is always initiated by the UE with the '*RRC Connection Request*' message. Upon receiving this message, the RNC transmits a '*RRC Connection Set-up*' message to the UE and then the UE changes its RRC state from IDLE to CONNECTED. Finally the UE confirms the RRC connection establishment by sending the '*RRC*

Connection Set-up complete' message indicating its capabilities. With the Radio Resource Control (RRC) connection one or more **Signalling Radio Bearers** (SRBs) are created to transmit RRC signalling.

Once the RRC connection has been established, the UE sends the message '*RRC Initial direct transfer*' to RNC which in turn maps it in the SGSN into a RANAP message (RAN application part). After that, Authentication is performed and the Bearer Service set-up is triggered.

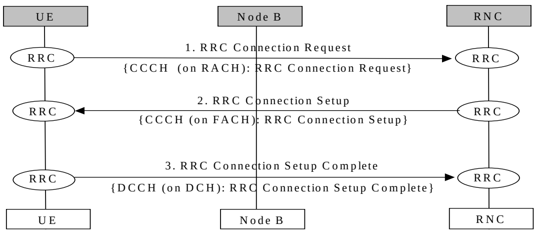

Figure 3 - RRC connection establishment

4.1.1 UMTS Bearer Service: PDP Context Activation

In UMTS, in order to enable any transfer of data in the PS domain, a PDP context must be established between the UE and the GGSN using the PDP Context Activation procedure. This procedure may be initiated by the UE or by the network depending on the direction of the session. A PDP context establishes an association between the UE and the CN for a given QoS on a specific NSAPI, UMTS Bearer Service. It contains routing information that is used to transfer the PDP PDUs between the UE and the GGSN. Activation of PDP context entails checking of the UE's subscription selection of the APN and the host configuration. Once a primary PDP context has been established for a given PDP address, a secondary PDP context can be activated re-using the PDP address and other information associated with the already active PDP context, but with a different QoS profile. Figure 4 shows the signalling message exchange for PDP Bearer Activation.

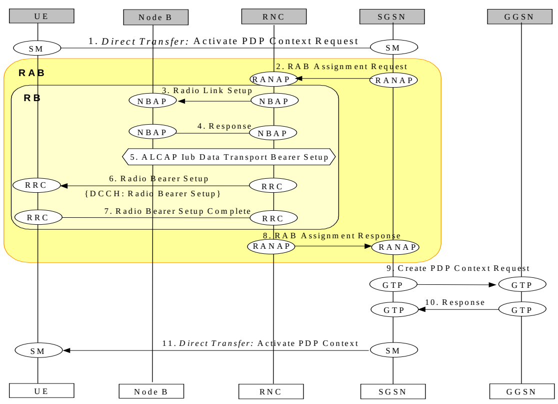

Figure 4 - UMTS Bearer Service: PDP Activation

4.2 IMS Signalling Procedures

Once the connection is established, the UE needs to access the IM sub-system. IMS makes use of SIP signalling flows and procedures required for the provision of presence and IM service detailed below.

4.2.1 Proxy CSCF Discovery

In the PDP context activation procedure, besides acquiring a PDP context within the PS CN, the UE also identifies a Proxy CSCF. This is a SIP proxy, as defined before, and is the contact point of the UE and is located in the same network as the GGSN, i.e. in the home or visited network, depending on whether the mobile is or is not roaming.

4.2.2 Application Level SIP Registration

In order to request the services provided by the IM domain, the user must perform an application level registration. This can only be done after registration with the access network is complete and after a signalling connection has been established for transfer of IP signalling. In other words, the user needs to activate a PDP context to transfer of IM related SIP signalling. The QoS parameters specified in activation of the context are appropriate for IM subsystem related signalling.

Figure 5 shows the flow of messages for registration of the UE with its Serving CSCF, assuming the UE was not previously registered. As shown, the S-CSCF authenticates the mobile before registration is successful.

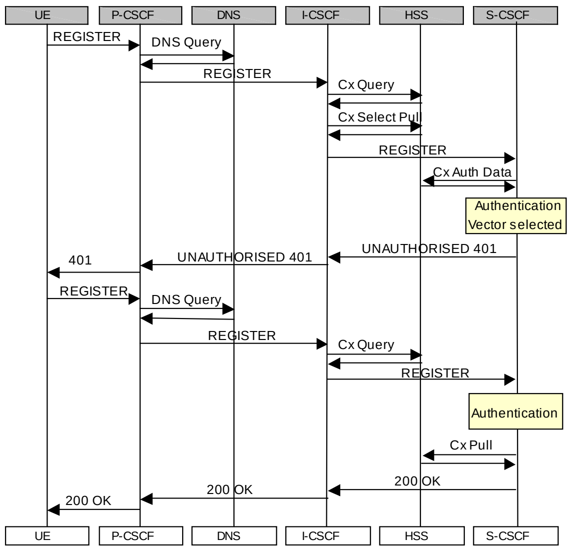

Figure 5 –SIP Register method

4.2.3 Subscription

Once the mobile is connected to IMS, subscription to the presence provider servers is required for those users using its capabilities. First the watcher entity, subscribe to his presence list server (PLS). The PLS will then forward the subscribe request to the desired presentity server (PS) if available. As soon as the message arrives to the PS, a notify request is purchase for the watcher's PLS with the required presentity's detailed information. Finally, the PLS will notify the watcher entity with the latest information. The subscription message flow is shown in Figure 6.

4.2.4 Session initiated

When the IM session follows the session model, the mobile initiates a session with an '*INVITE*' transaction in order to create the required association between the sequence of messages. The signalling required for the establishment of a session is analysed in two individual procedures, the Mobile Origination (MO) and the Mobile Termination (MT).

The session establishment starts with the *'INVITE'* message being sent for the caller to the callee. The two end parties negotiate the media characteristics that will be supported for the session. After these have been determined, resource reservation is required, which entails creating a secondary PDP context for transport of the required media, and setting up the corresponding radio access bearers. If resource reservation is successful, the terminating point sends a SIP *'200 OK'* final response and the originating mobile replies with a 'SIP ACK' message to confirm the session set-up. The session initiation message flow is shown in Figure 7.

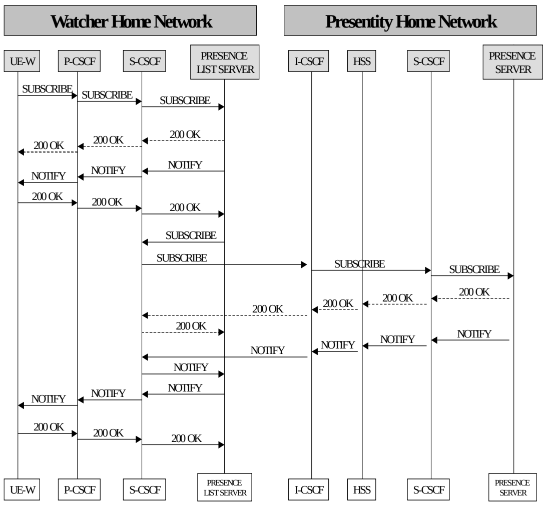

Figure 6 - Presence Subscription method

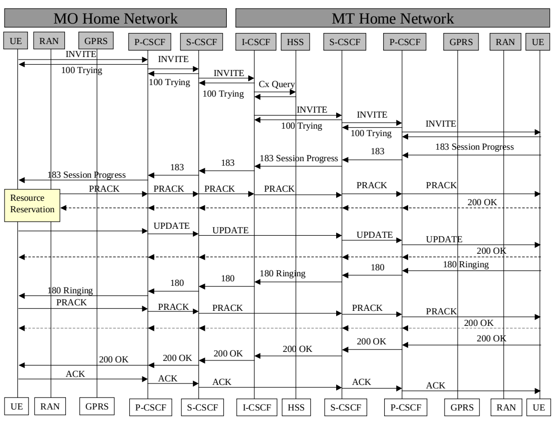

Figure 7 - Session set-up Mobile to Mobile

5. Simulation Model and Environment

The Dynamic Signalling Simulation Environment, 'SigSim', shown in Figure 8, is designed to estimate the end-to-end signalling load in terms of number of messages handled per network element and procedural delays. The dynamic nature of 'SigSim' derives from the stochastic modelling of users mobility within a particular environment as well as user behaviour in terms of accessing different services. The simulator implements a model of cell layout and UMTS network. Even though only signalling traffic is simulated, traffic models are implemented and accounted for the period of time a user is using a particular service.

5.1 Network Model

Figure 1 presents the functional architecture of the IMS as defined in the UMTS Release 5/6. However, this model does not provide the representative physical model that is required to represent delays realistically, therefore a more practical realisation is proposed here. First the model reference for the basic signalling services, such as packet-switched call control, according to Release 5 specifications is presented. This model is then adapted for the enhanced services and application capabilities introduced in Release 6 with special attention on the Presence and Instant Messaging services.

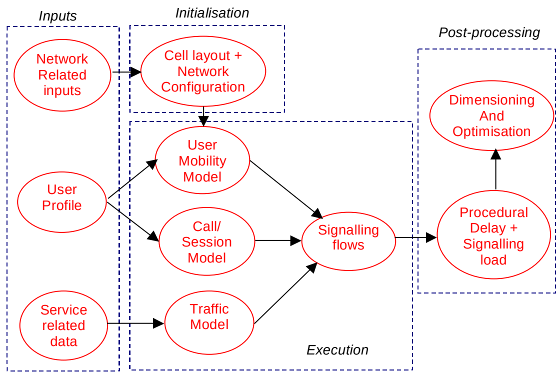

Figure 8 – SigSim Simulation Environment

5.1.1 Basic Model Reference Approach

According to 3GPP specifications basic sessions between mobile users always involve two S-CSCFs (one for each user) and an I-CSCF to select them. On the other hand, a session between a user and a PSTN endpoint involves an S-CSCF for the UE, a BGCF to select the PSTN gateway and a MGCF for the PSTN. Therefore, SIP messages are routed through four SIP proxies in the mobile to fixed scenario, i.e. P-CSCF, S-CSCF, BGCF and MGCF. This is worse for the mobile-to-mobile case where a P-CSCF and a S-CSCF are required for both entities in addition to the I-CSCF, adding up to a total of five SIP proxies or servers.

The more SIP proxies the message has to traverse the greater the transmission delay. Consequently, in order to reduce the transmission delay, we propose to collocate IMS logical network elements with similar functionality into three physical nodes as illustrated in Figure 9.

- **SIP Server Node:** integrates the P-CSCF and the S-CSCF in a common node within a particular operator's network. Every mobile contacts the IMS through a Proxy-CSCF. After registration the P-CSCF routes the SIP messages to the Serving-CSCF SIP control element. The P-CSCF resides in the network where the mobile resides, visited or home network, whereas the S-CSCF always resides in the home network. The scenarios considered here assumed that all mobiles are in their home network, therefore by collocating those two entities the number of messages transmitted through the network is considerably decreased by 34%, thus also reducing transmission delay.

- **IMS Gateway Node:** when the session is established between a mobile user and a PSTN endpoint such as a fixed telephone user, the BGCF and MGCF handle the SIP signalling for the PSTN endpoint. The BGCF, at the start of the session set-up, selects the PSTN network with which the inter-working is to occur and forwards the message to the corresponding MGCF. Although the BGCF has not considerable impact on the session set-up, as is not included in the SIP message path, the collocation with the MGCF contributes minimising message transaction time.
- **Database Node:** host the I-CSCF and the HSS, which is a large database with extended HLR capabilities. The I-CSCF functionality for a non-roaming user is reduced to contacting the HSS for information. It queries the HSS to assign the Serving-CSCF at the registration point and also obtains the S-CSCF address of the terminating counterpart during session set-up. Therefore it seems reasonable to collocate both.

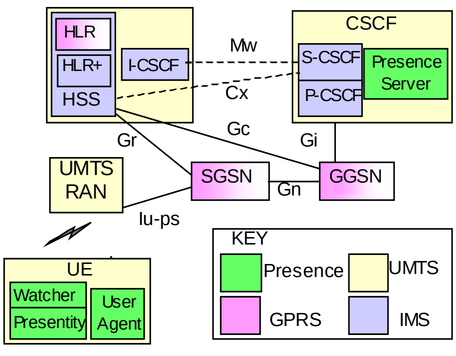

Figure 9 - Model Reference

We assume that the SIP Server platform provides a 25ms SIP to SIP message turnaround duration for up to 800k subscribers per network element. The interrogation and Cx data retrievals to/from the Database Node platform can take 55ms per transaction (read, read/write and forward average).

5.1.2 Enhanced Model Reference Approach

The Presence service, which resides in the IP Multimedia Sub-system, is being standardised in Release 6 3GPP standards. The presence server manages the presence information of a user (presentity) that is uploaded by different agents (network elements, terminals or external elements) and combines it into a single presence document in a standardised format. Furthermore, the server allows other users (watchers) to subscribe to it for receiving presence information. For simplicity, we consider that both watcher and presentity entities reside in the same network, the home network. As such, they communicate through the home network's SIP CSCFs proxies and no external agents are involved. Based on this simplified architecture, a practical realisation of the UMTS presence service model is proposed here, where different elements are collocated in order to reduce the message transmission delay.

The presence server is collocated with the register server, i.e. the S-CSCF. Furthermore, the watcher and the presentity entities reside on the User Equipment and communicate with the server across the SIP proxies, P-CSCF, S-CSCF and I-CSCF. Figure 9 shows the considered reference UMTS architecture including the two introduced approaches.

5.2 Session Model

Any user requesting packet services needs firstly to activate a PDP context in order to establish a link (*session*) between the UE and the core network, in order to transmit non-UMTS signalling (SIP signalling) and bearer data. Each session may hold one or more services (*user sessions*) and if their QoS differ a modification in the session is undertaken.

5.3 Traffic Model

Traffic models govern the generation of bearer traffic within a user session. Each service is characterised by a traffic model. A service session may last for the duration of the PDP session or several service sessions may be initiated within a PDP session.

In order to characterise the complex nature of the packet data traffic services a *structural* or *hierarchical* model is considered. The *hierarchical* model presents multilayer or multilevel processes that characterise the different levels existing behind the packet service. Figure 10 shows those different levels of granularity (session, packet connection and packet) as described in the ETSI packet data model and generally adopted for data services modelling such as the world wide web.

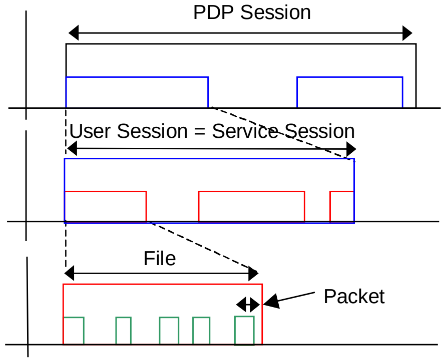

Figure 10 - Packet session traffic model

An IM session consists of several files down/upload. Each file is further made up of packets. The traffic model characterising SIP IM defines the average number of files within an IM session, the mean time between file down/uploads, the average size of a file, the average packet size and the average time between packets.

Considering the similarity of "***Message IM***" service and Internet chat, the traffic characteristics have been obtained from the UMTS Forum *'Telecompetion, Inc. report'* for mobile Intranet chat. The ***"Paging IM"*** is a person-to-person service that mimics SMS messaging. Therefore, at the session level we approximate the Paging IM behaviour using SMS. As part of this work a survey among Cork Institute of Technology students was carried out in order to identify the frequency of SMS usage. SMS usage among young people in the Republic of Ireland is the highest in Europe and the messaging rates obtained through the survey represent a somewhat upper bound of what can be expected for initial IM service usage once introduced in UMTS. At the packet and file levels the considered service differs from the SMS as the size of the messages is not limited to 160 characters. The packet level is therefore based on an e-mail service.

6. Performance Characteristics

Results derived form the Signalling Simulator, '*SigSim*', are presented for two different application services, "***Paging IM"*** (interactive) and "***Message IM"*** (streaming). A typical UMTS network topology is analysed for a dense urban environment, where the UTRAN consist of 784 NodesBs and 4 RNCs. The PS domain consists of 2 SGSNs and 1 GGSN. With this configuration, all core-network related and mobility signalling are accounted for. In the IMS, it is assumed that there are 1 P-CSCF, 1 S-CSCF, 1 I-CSCF and 1 MGCF, which are located as previously defined. The analysis is limited to the SIP signalling message load and time delay for different services and scenarios considered. The model keeps track of the Radio Access Network delay (RAN delay) and the Core Network delay (CN delay). In this study we assume that a single operator performs both originating and termination part, therefore the mobile users are always located in their home network and roaming is not considering. The two counterparts, the mobile initiating the session and the mobile receiving it, are modelled separately and their combination provides the end-to-end analysis.

We considered a total simulation duration time of 5 hours, where the usage of the service is considered as 100% for both IM services modelled in each case. The percentage uplink traffic is set to 0.5, i.e. one-to –one symmetric conversation. The results provided are statistics in terms of message loads and end-to-end time delays introduced by the SIP signalling.

6.1 SIP Message Sizes

The IM service with presence capability requires several SIP procedures for the establishment of a session. We analyse in this section the considered SIP messages sizes for each procedure and also provide message sizes for compressed SIP messages using TCCB based compression [8]. Four main procedures are analysed, 'SIP Register' 'SIP Subscribe' to the List Server and to the presentity and finally the 'SIP Invite'. The messages considered are based on [2, 5], however the following assumptions were considered:

- The SIP Invite procedure examined is assumed to contain only one media type, text. This decreases the message size due to a shorter SDP part.
- The TCCB compressor does not compress the SIP message body.
- The *'SIP NOTIFY'* message body size contains several attributes. The size of which is determined by a linear function, which depends on the number of tuples and attributes present in the presence information message body.

The following tables show the size per message for each SIP procedure we considered. Table 1 illustrates the SIP messages exchanged in an application registration. Table 2 provides the SIP subscription messages and finally Table 3 shows the size of the SIP messages exchanged during the set-up of a session. An averaged compression rate of 40% can be achieved, when using the presence capability.

UPLINK	***UE*** Uncompressed	Compressed	*% **Compressed***
REGISTER	534	429	20
REGISTER Auth.	639	405	37
DOWNLINK	*P-CSCF* Uncompressed	Compressed	*% **Compressed***
401 Unauthorised	363	203	44
200 OK	426	218	49

Table 1 - SIP Register method messages

UPLINK	***UE*** Uncompressed	Compressed	*% **Compressed***
SUBSCRIBE	472	296	37
200 OK	215	112	18
DOWNLINK	*P-CSCF* Uncompressed	Compressed	*% **Compressed***
200 OK	302	102	66
NOTIFY	4220	276	34
NOTIFY (state)	458 + f(x)	322 + f(x)	30 + f(x)

Table 2 - SIP subscribe method messages

UPLINK	***UE*** Uncompressed	Compressed	*% **Compressed***
INVITE	608	205	52
UPDATE	484	326	33
PRACK (180)	288	147	49
ACK	263	137	48
DOWNLINK	*P-CSCF* Uncompressed	Compressed	*% **Compressed***
183-Session Progress	684	486	29
180-Ringing	293	103	65
200-OK	263	102	61

Table 3 - MMO SIP Invite method message

6.2 Delay Analysis

Finally in this section we present the end-to-end procedural delay for several UMTS-specific and SIP signalling flows. Procedural delays consist of the transmission delays across interfaces and processing and queuing delays at network elements. The mean and 95th percentile delays are provided. The simulation results are obtained assuming a subscriber population of 10,000 users, accessing the service. However for the ***"Message IM"*** session only 30% of the mobiles establish a session, due to the large session inter-arrival time. For the ***"Paging IM***" service nearly all mobiles, 93%, established a successful session.

6.2.1 Message Instant Messaging Session

We considered that all UMTS (PMM) signalling uses a 3.4 kbit/s Radio Bearer. However both IMS (SIP) and data bearer (i.e. IM exchange) use a DCH at 64kbit/s. Table 4 illustrates the UMTS signalling and SIP signalling delays, which are required for the establishment of the first session.

Flow ID	TCCB	***Mean delay*** ***RAN***	***Core***	***Total***	*95th Percentile*
PS Session Set-up	N/A	1.46	0.77	2.23	3.36
SIP Registration	Off	0.41	0.88	1.28	1.48

SIP Subscription LS	Off	0.24	0.53	0.76	0.89
SIP Subscription Pres.	Off	0.16	0.35	0.51	0.56
MMO SIP Invite	Off	0.82	1.25	2.07	2.39
Secondary PDP activ.	N/A	1.18	0.76	1.94	-

Table 3 - MMO *"Message IM"*

For the Mobile Terminated case (MMT), the mobile is considered registered and subscribed to the network and only the 'SIP Subscribe' to the presentity and 'SIP Invite' coming from the originating part are modelled as shown in Table 4. The presentity SIP presence server controls the admission for the subscription, and therefore there is no RAN contribution for the SIP Subscription.

Flow ID	TCCB	***Mean delay***			*95th Percentile*
		RAN	***Core***	***Total***	
NI PDP activation	N/A	2.74	4.54	4.28	5.65
SIP Subscription Pres.	Off	0	0.05	0.0.5	0.0.5
MMT SIP Invite	Off	0.94	1.03	1.97	197

Table 4 - MMT *"Message IM"*

As expected the delays introduced by the SIP signalling flows are large. An end-to-end ***"Message IM"*** session between two mobiles takes a total of 12.86 seconds (6.49s on the Ran side and 6.39 on the CN one) increased by 15% for the 95% quantile delay value. However, this result refers to the first session established on the PDP context. Subsequent sessions are set-up within 10.79 seconds.

If TCCB compression is applied, the RAN delay for the first session is reduced by 12% with a reduction of only 6% on the total delay, as the core network delay is the main contribution of the overall delay. On subsequent sessions the RAN and total delay reduction is similar.

Flow ID	TCCB	***Mean delay***			*95th Percentile*
		RAN	***Core***	***Total***	
SIP Registration	On	0.31	0.88	1.19	1.33
SIP Subscription LS	On	0.17	0.53	0.70	0.78
Subscription Pres. MO	On	0.14	0.35	0.49	0.51
MMO SIP Invite	On	0.58	1.25	1.83	2.01
MMT SIP Invite	On	0.71	1.03	1.74	1.74

Table 5 - MM *"Message IM"* TCCB compressed

6.2.2 Paging Instant Messaging Session

We obtained the same set of results for the "Paging IM" service except it does not require a 'SIP Invite'. The number of flows activated, for release and establishment of the connection, are higher in the ***"Paging IM"*** case, however. Within a ***"Paging IM"*** session every mobile triggers 2.7 times the release procedures, whereas in the ***"Message IM"*** case only once per session release procedures were activated. The increase in those signalling procedures indicates the interactive character of the ***"Paging IM"*** service thus the bearers are released more frequently.

7. Conclusions

In this paper we presented a performance analysis of IP based packet-switched UMTS services. As specified by 3GPP standards, the considered services use the SIP protocol as their main session control protocol. We focused on the effect that such text-based protocol has on the service performance in a UMTS network. The end-to-end delay simulation results show that instant messages are not necessarily transmitted in near instant fashion but that substantial delays, with an averaged of about 12.86 sec are encountered for the first ***"Message IM"*** session establishment. The results improve however for subsequent sessions as they do not require transmitting all SIP signalling again. Consequently further reduction in the transmission delay is obtained (10.79 sec). The results presented show that SIP signalling introduces a large transmission delay in the network. The TCCB SIP message compression method and the use of higher data rates decrease the transmission delay on the radio access side. However the time delays on the core network are high which is due to the high number of messages that are sent through the network in each SIP flow and the number of network elements the messages traverse through. We see two main approaches as possible solutions to decrease the core delay, decrease the number of messages exchanged during the SIP procedures and reduction in the number of network elements by co-locating them. However those solutions imply the amendment of the UMTS specification and modification of some of the present assumptions.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Irish Department of Education and Science Technological Sector Research Programme Strand 3 in funding parts of the work reported in this paper under grant CRS/00/CR02.

REFERENCES

[1] 3GPP TS 23.228, "IP Multimedia System, Stage 2"

[2] 3GPP TS 24.228, "Signalling flows for the IP multimedia call control based on SIP and SDP, Stage3"

[3] IETF RFC 3261, "SIP: Session Initiation Protocol"

[4] 3GPP TS 23.141, "Presence Service; Architecture and Functional Description"(Release 6)

[5] 3GPP TS 24.841, "Presence service based on Session Initiation Protocol (SIP); Functional models, information flows and protocol details"(Release 6)

[6] IETF RFC 2778: "A Model for Presence and Instant Messaging"

[7] IETF RFC 2779: "Instant Messaging / Presence Protocol Requirements"

[8] V. Kenneally, D. Pesch, I. Majumdar, "Evaluation of SIP Compression for IP based Wireless Multimedia Communication", Proc. IT&T Conference, Waterford, Ireland, Oct. 2002