SIP as a Unified Signalling Solution in a Beyond 3G System

Georgios V. Lioudakis, Vaggelis Nikas, Christoph Pollak, Richard Wisenöcker, Nick Dellas

Abstract—The adoption of packet-switched technologies in mobile communication systems has enabled the provision of IP-based services to mobile users. Yet, these systems, such as the UMTS network, are unable to meet the stringent delay requirements imposed by the enhanced multimedia services due to the anchor point that introduce into the user's data path. Moreover, such applications will rely on IP-based control protocols for session and mobility management, resulting in duplication of functionality in the network at both the bearer and the application level. For the above reasons, the SAILOR network architecture is presented in this paper, proposing an evolution to the standard UMTS network, where GSNs are integrated into a single entity. Furthermore, the SIP protocol is adopted, for performing mobility and session management procedures currently undertaken by the UMTS Non-Access Stratum protocols. Both modifications to the UMTS network and protocol architecture, result in the smooth evolution of the UMTS core network towards IP as well as accomplish a significant performance gain.

Index Terms—RASN, UMTS, SIP, All–IP architecture

I. INTRODUCTION

THE penetration of packet-switched technologies in mobile communication systems has paved the way for the integration of the telecom with the IP world, in order to ease the provision of IP-based services to users on the move. The introduction of a Packet Switched (PS) Core Network into legacy mobile networks has been the first step towards the provision of IP access to mobile users.

The strict hierarchy adopted by such systems as well as the integrated protocol stacks used for the implementation of mobility/session management, Quality of Service and security support, does not leave enough space for IP-based protocols to dominate. For example, in the UMTS network, the GGSN serves as a fixed IP gateway for mobile hosts, introducing significant anchoring when hosts move far away (as is the case with the Home Agent in Mobile IP). However, the application set supported by such systems will encompass a variety of real-time multimedia applications with stringent delay requirements, obviating the need for removing anchor points from the user's data path. Moreover, current 3GPP standardization provides two instances for the control functionality of PS services; at bearer level for signalling between TE, GSNs and HSS/HLR, and at application level for interaction with the IMS by means of the SIP protocol [1], leading to the repetitive transmission of information serving the same purpose, such as User Identity, QoS parameters etc. As a result, the duplication of functionality in several cases leads to a considerable consumption of resources.

In the context of the IST project SAILOR [2], a smooth evolution to the standard UMTS network architecture is proposed, identifying the need to loosen the fixed IP gateway point in the UMTS network and to integrate network functionalities in a greater extent. This new architecture aims at eliminating the current network sufferings, without compromising on the capabilities and the overall performance of the network. A fundamental innovation in the proposed architecture is the integration of the core network GSNs into a single one, denoted as Radio Access Supporting Node (RASN). The RASN is placed at the border between the UMTS Access Networks (AN) and the UMTS Core Network (CN) and will enable the UMTS CN to become fully IP compliant.

The rest of this paper is organized as follows: Section II provides a general description of the RASN, as well as of the overall SAILOR architecture. Sections III and IV provide the functional and design description of the RASN, respectively, while section V concludes this work.

II. GENERAL DESCRIPTION OF RASN

From the ANs point of view, the RASN plays the role of a 3GPP CN, while, from the CN point of view, the RASN completely hides the specific aspects of UMTS radio access technology. In other words, the RASN provides similar functionality as a standard UMTS CN to the T-S-UMTS ANs and looks like an (enhanced) IP router to the other IP routers of the fully IP fashioned UMTS CN.

With the adoption of the RASN, the GTP protocol [3], performing SM/MM functionalities within the UMTS CN, is eliminated. Instead, IP-based protocols are used for this

Georgios V. Lioudakis, Vaggelis Nikas and Nikos Dellas are with Ericsson Hellas (40.2 Km Attica road, 19002 Peania Municipality, Athens, Greece; tel: +302107722424; email: gelioud, enikas, ndellas@telecom.ntua.gr).

Christoph Pollak and Richard Wisenöcker are with Siemens AG Österreich (Autokaderstraße 29, A-1210, Vienna, Austria; email: christoph.pollak, richard.wisenoecker@siemens.com).

purpose, while the traditional PS domain of the UMTS CN is replaced by a fully IP-compatible backbone. In this converged UMTS-IP environment, the RASN comprises the first-hop IP router for mobile hosts located within its serving area. Fig. 1 presents the proposed architecture.
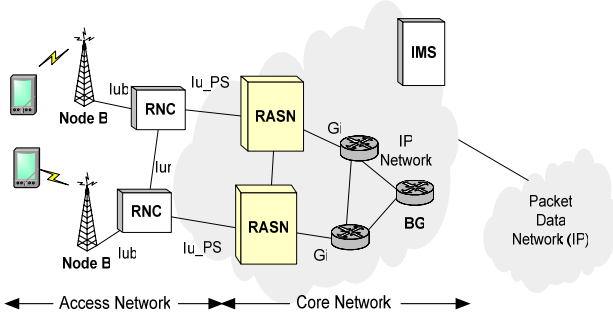

Fig. 1. Proposed SAILOR Architecture.

Additionally, the Non-Access Stratum (NAS) procedures carried out by legacy SM/MM protocols [4] are replaced by the IP-oriented SIP protocol. SIP is proposed since, in addition to session control and terminal mobility support [5], it inherently covers the other mobility types, i.e. personal and service mobility.

In the proposed architecture, the RASN assumes the role of providing radio access unawareness to the IP-based CN by hiding the specific aspects of the radio access technology. For that reason, it terminates the access stratum UMTS protocols as well as the overlaid so-called SIP_{RAN} layer. SIP_{RAN} can be viewed as conventional SIP signalling augmented with the appropriate NAS content [6], when needed, while it caters for SM/MM functionality at the bearer level, making the respective UMTS procedures redundant. Moreover, the RASN is equipped with standard SIP proxy functionality for forwarding SIP requests, addressed to SIP entities within the CN and beyond (external PDNs or other PLMNs). SIP_{RAN} to SIP relaying functionality is also required for providing transparency to entities running the standardized SIP. It is therefore ensured that the SIP messages transmitted by the RASN towards the IP network are fully compliant to the current SIP standardization.

In the introduced protocol architecture, the extended SIP messages are transferred over the UMTS control plane
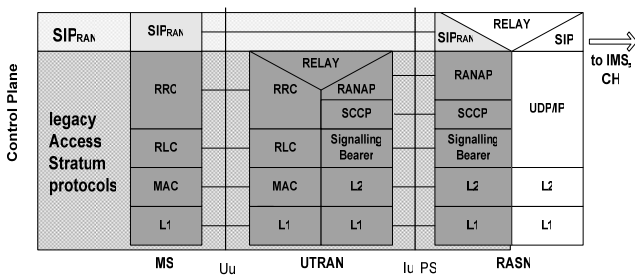

Fig. 3. Control Plane Protocol Stack.

protocol stack (Fig. 2), in contrast to the current UMTS

specifications [4], [7], [8], where SIP messages are conveyed over the user plane protocol stack (Fig. 2), presupposing that the data bearers have been established prior to the service provision.
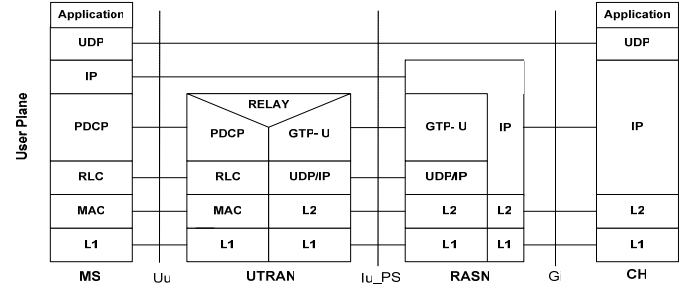

Fig. 2. User Plane Protocol Stack.

III. FUNCTIONAL DESCRIPTION OF RASN

As described above, the main goal of the SAILOR architecture is to substitute the legacy 3G signalling processes (MM/SM) with IP based ones, by means of the SIP_{RAN} signalling transmitted over the UMTS control plane. To that respect, the most important operations provided over the NAS protocols are mapped to corresponding SIP procedures. This mapping is roughly overviewed in Table I. A more detailed description is provided to the following subsections. The complete analysis is described in [9].

TABLE I
MAPPING OF NAS MESSAGES TO SIP MESSAGES

3G NAS procedure	SIP procedure
GPRS Attach	SIP Registration
GPRS Detach	SIP de-Registration
Authentication	SIP Registration
	SUBSCRIBE/NOTIFY mechanism
Routing Area Update	SIP re-Registration
Serving-RNC Relocation	SIP re-Registration
PDP Context Activation	SIP Session Initiation
PDP Context Deactivation	SIP Session Termination

A. GPRS Attach/Detach

The GPRS Attach is performed by means of a SIP REGISTER request addressed to the home domain's registrar (i.e. the IMS node of the Mobile Service Provider, MSP). A fundamental differentiation of the SAILOR approach to the legacy UMTS operation is the IP address allocation process. In particular, a successful GPRS attach procedure in NAS yields a Mobile Station (MS) being reachable for incoming traffic addressed to its MS-ISDN, but no IP address has been assigned to it yet[*]. Instead, the GGSN dynamically allocates an IP address for the MS prior to the data transfer, in the framework of a PDP context activation procedure. A successful SIP REGISTER, on the other hand, yields a client being reachable by its SIP URI and its current valid IP

---

[*] In the typical case this is the standard UMTS network operation. However, UMTS systems can also support static-defined IP addresses, so that a subscriber is reachable without active PDP context.

address, allocated and assigned to the MS by the RASN in the context of the registration procedure.

The procedure (Fig. 4) is briefly executed as follows: the MS transmits a REGISTER message towards the IMS. The RAN specific information (RA Identifier etc) is included in the so-called NAS vector, contained in the body of the SIP message. No IP address has been assigned to the MS yet. Upon receipt of a REGISTER, the RASN reserves an IP address belonging to its address space for the MS and proxies the message to the SIP registrar, after including the MS's new IP address. The 401 Unauthorized response is demanded as part of the authentication procedure. In particular, the first REGISTER request causes the failure of the authentication algorithm in the registrar, since the credentials containing authentication information for the MS have not yet been included. However, the 401 Unauthorized response challenges the MS for providing such credentials, which are subsequently included in the second registration request. As illustrated, the MS becomes aware of its new IP address by means of the SIP 200 OK message, indicating the success of the registration. Consequently, it configures its IP stack for being able to receive incoming traffic addressed to this IP address.

The SUBSCRIBE message after the registration serves for authentication purposes. In particular, with regards to security aspects, the network in NAS operation may at any time require the authentication of the subscriber, based on the chosen security policy. In order to be in line with this functionality, the SIP approach introduces a SUBSCRIBE/NOTIFY mechanism. After successful registration, an MS communicates a SUBSCRIBE request to the attached domain's registrar (i.e. IMS), in order to be subscribed for network-requested authentication. The IMS may at any time request the authentication of a subscribed MS by sending a NOTIFY message that contains the necessary parameters the latter has to use for its authentication. The MS should, subsequently, authenticate itself by sending a REGISTER message within a predefined period. In order to save resources at the air interface, the RASN is responsible for sending the SUBSCRIBE messages on behalf of the MS.
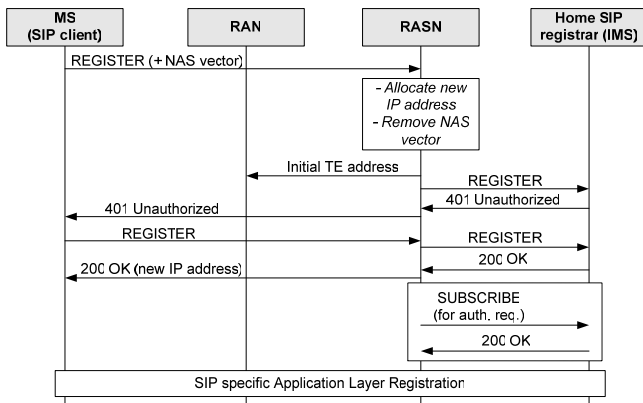

Fig. 4. GPRS attach.

Incoming packets addressed to the MS (either to the MS's IP address or the user's SIP URI) finally reach the serving

RASN by following the established downstream IP path. Since the registrar is aware of the MS's current IP address, the standard SIP session initiation procedure guarantees that INVITE requests will reach the MS's current IP subnet. The serving RASN, acting as an edge router from the CN's point of view, will subsequently deliver the packets after paging the MS through the standard UMTS paging procedure.
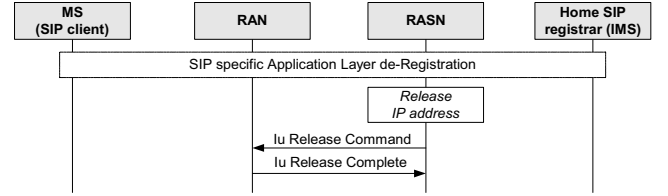

Fig. 5. GPRS detach.

The same messages are also exchanged for the GPRS Detach (Fig. 5) with the only difference to be that the expiration time is set to zero in both REGISTER requests.

B. Routing Area Update

After the MS attaches to the network, and while being in IDLE state, it needs to inform the latter of its location whenever crossing the boundaries of a routing area (RA). The RA Update (RAU) procedure is used for this purpose and, in the case of the SAILOR network, is discriminated into the intra-RASN and inter-RASN RAU, depending on whether the old and the new RAs the MS is located at, belong to the same RASN. The RASN –equivalent to the SGSN in UMTS networks– is configured with the binding <Routing Area ID – set of RNCs> for being able to page the MS within a routing area, when data addressed to the latter arrives.

The intra-RASN RAU (Fig. 6) is performed by means of a SIP REGISTER request. The Routing Area Identifier (RAI) is included in the message, allowing the RASN to determine the type of RAU, i.e. intra or inter-RASN. In the former case, the request is intercepted by the RASN SIP$_{RAN}$ protocol entity, and since the MS remains attached under the same IP subnet, there is no need for the registrar to be updated. The inter-RASN RAU is performed in the framework of an inter-RASN handoff as described in the next section.
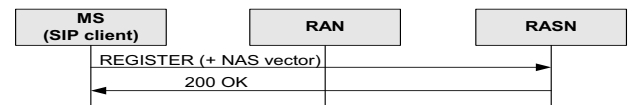

Fig. 6. Intra-RASN Routing Area Update.

C. S-RNC Relocation

In case of intra-RASN S-RNC relocation (Fig. 7), the source and the target serving RNC belong to the same RASN. The fact that the RASN remains unchanged results in no change at the IP address of the MS as well as the IP user data path. The REGISTER request is received by the RASN, which takes care of the RAN aspects by interworking with the corresponding RNCs by means of the RANAP protocol interface. In parallel, it communicates the message to the IMS (after removing the NAS vector) and receives the reply, which

subsequently is relayed to the MS. Since in intra-RASN mobility events, the IP address of the MS remains unchanged, a future optimization would be that the RASN automatically generates the 200 OK response towards the client, instead of relaying the REGISTER to the IMS. However, this has been considered to be a minor overhead to the system, since the SIP operation imposes that periodic registrations to the IMS are sent anyway.
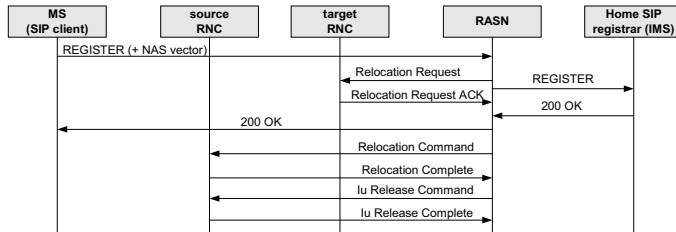

Fig. 7. Intra-RASN Serving – RNC relocation.

The inter-RASN handoff causes the relocation of the MS's serving RASN. The scope of the procedure is twofold; on the one hand, it takes care of the RAN aspects, such as the RAU and on the other, it updates the MS's IP path towards the target RASN, by communicating its new IP address to the home registrar. The inter-RASN handoff may occur either when data transfer towards the MS is in progress or when no packet exchange is taking place. For these two cases, the terms *Continuous* and *Discontinuous* RASN Relocation will be used, respectively.

In the discontinuous RASN relocation (Fig. 8), the RAU is performed by means of a SIP REGISTER message, indicating the MS's old RAI among other information. The RASN identifies that this is an inter-RASN handoff since the indicated RA does not belong to its serving area. It consequently deduces the source RASN and sends to the latter a Context Transfer Request message (implemented by means of the SIP CORE method, a new method defined in the framework of SAILOR). Note that the UMTS-specified Gn interface (inter-GSN interface) has been eliminated in the SAILOR network, while an IP-based protocol – SIP_{RAN} – is used instead. The source RASN responds with a Context Transfer OK message (implemented by means of a SIP 200 OK), which carries MM/SM context for the MS as well as SIP-based context, if any. After establishing the user contexts, the target RASN reserves a new IP address for the MS and forwards the SIP REGISTER message to the IMS, including the MS's IP address and having removed the NAS-related information. The target RASN immediately replies with a 200 OK response, carrying the MS's new IP address used by the latter to reconfigure its IP stack. Moreover, it updates the established user contexts with the new address and continues with the removal of the MS's context from the source RASN. To this aim, it communicates a REGISTER request to the source RASN, with the expiration time set to zero, so that the latter will deregister the MS.

In the continuous RASN relocation (Fig. 9), the aim is to update the MS's IP path while the UMTS L2 handoff is in progress, i.e. while data are still forwarded from the source RASN to the MS (through the target RNC). In this way, as soon as S-RNC relocation is performed (i.e. L2 connectivity with the target RASN is established), the data already follow the new IP path, allowing for a seamless handoff. Therefore, any IP handoff mechanism, such as IP tunneling, is refrained, leading to reduction of packet loss and latency.
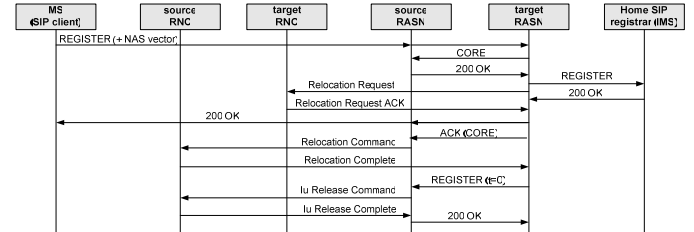

Fig. 8. Inter-RASN Serving – RNC relocation / no active session.

Upon receiving the REGISTER, the target RASN transmits a CORE message to the source RASN, in order to obtain the MS's MM/SM contexts. The procedure continues with the establishment of RABs on the new UMTS path and the temporary forwarding of data from the source to target RNC (by means of the Iur interface). The target RASN reserves a new IP address for the MS and relays the REGISTER to the IMS. The allocated IP address is included within this message so that the latter to be updated with the MS's new IP address. The IMS replies with a 200 OK (carrying the new IP address), which is intercepted by the RASN and forwarded to the MS. Subsequently, the MS re-INVITEs the Correspondent Hosts (CHs), indicating its new IP address. As before, a SIP REGISTER with expiration time set to zero is transmitted from the target to the source RASN so that resources (MS's context and RABs) are finally released on the old path.
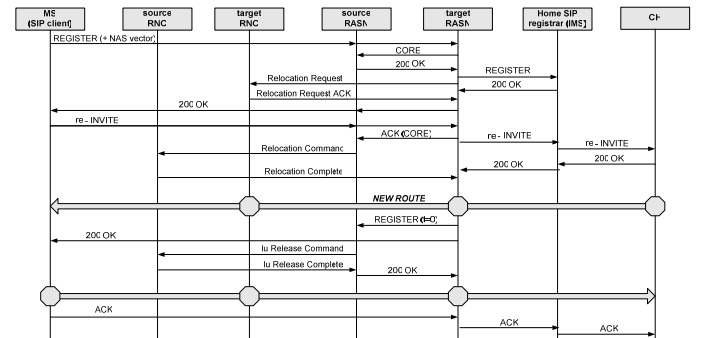

Fig. 9. Inter-RASN Serving – RNC relocation / session(s) active.

D. PDP Context Activation/Deactivation

The PDP context activation in UMTS networks is executed, so that the user data path is established. The CSCF discovery and the SIP-based application layer registration with the IMS must follow, so that the user can access the service provided by the MSP. The process in the SAILOR approach is performed by means of the SIP INVITE procedure, including the corresponding response codes and acknowledgement mechanisms. Fig. 10 illustrates the procedure for MS originated session initiation, while the MS terminated one is

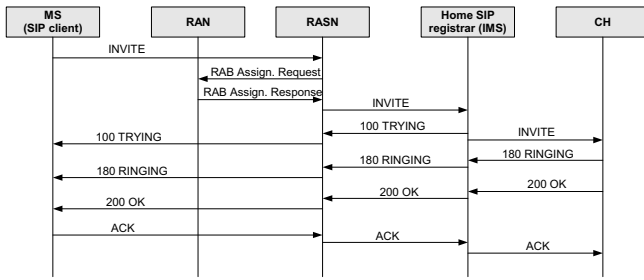analogous. It is noted that the IP address allocation for the MS has been performed at registration time.

Fig. 10. Mobile originated session initiation.

The SIP BYE method is used for session termination. Fig. 11 depicts the procedure for the MS originated session termination, while the MS terminated one is symmetric. The release of the RABs is accomplished via standard RANAP signalling.
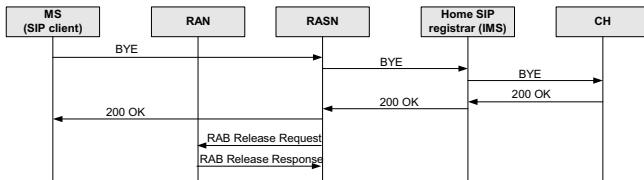

Fig. 11. Mobile originated session termination.

IV. DESIGN DESCRIPTION OF RASN

The RASN software has been developed in the context of SAILOR. The most important of its design blocks are the following:

- *RANAP interface*: Since the RASN plays the role of the standard SGSN from the AN point of view, it is required to support RANAP communication with the RNCs. The RANAP handler of the RASN is responsible for transmitting and evaluating RANAP messages emanated from the AN.
- SIP_{RAN}/*SIP signalling relay*: This is one of the most important building blocks of the RASN. It assumes the responsibility of intercepting the SIP_{RAN} messages from the AN and communicates them to the IMS, after excluding the NAS vector and vice versa. In a sense, it functions as a limited operational SIP proxy, since it receives/transmits SIP signalling from/towards both directions and maintains the state of each transaction. In certain cases, it is able to generate rejection messages towards the MSs or the IMS (i.e. when a request is rejected from the AN, it is not forwarded to the IMS; instead, a SIP rejection message is generated and sent back from the RASN).
- *User Data Handler*: The RASN operates as the first hop IP router for the mobile hosts situated within its serving area. The relaying/routing functionality is under responsibility of the user data handler. It retains a binding of the form <RAI, Serving - RNC, MS IP address> so that it is enabled to deliver incoming IP traffic to the MSs registered to the RASN.
- *IP address management*: A handler for managing the IP address pool of the RASN. It is responsible for the

assignment/releasing of IP addresses of the MSs entering/leaving its realm.

For more details concerning the design specification of the RASN, please refer to [10].

V. CONCLUSION

A UMTS-IP converged architecture, aiming at the IP penetration in the CN of the cellular infrastructure has been presented in this document. The benefits of the proposed architecture are multiple; SM/MM at IP level results in eliminating anchoring points in the user's data path, the replacement of GTP by simpler IP-based protocols minimizes the complexity of MM/SM procedures in the CN, the elimination of the GTP header introduced to IP packets reduces the overhead to the network and the use of simpler IP routers instead of GSNs facilitates the network deployment. Moreover, a SIP-based approach is adopted for carrying out SM/MM functions, while replacing the legacy UMTS NAS protocols. Hence, in the proposed architecture, SIP covers MM/SM procedures at both UMTS and IP level. The performance gains are significant; for example, while the registration of a mobile host to its registrar would necessitate a) its attachment to the UMTS CN, b) the activation of a PDP context, and c) the SIP registration to take place, the whole procedure now requires only the third step to be performed. Most importantly, the consumption of radio resources is significantly reduced. Seamless handoff can be achieved by relying on the well-known S-RNC relocation mechanism, while trying at the same time to synchronize the update of the IP path with the re-establishment of L2 connectivity.

REFERENCES

[1] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.
[2] SAILOR (Satellite Integrated UMTS Emulator-IST-2001-37266) web site, URL: http://www.newapplication.it/sailor.
[3] 3GPP TS 29.060 v6.7.0, "General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface".
[4] 3GPP TS 23.060 v6.0.0, "General Packet Radio Service (GPRS); Service description; Stage 2," March 2003.
[5] Henning Schulzrinne, Elin Wedlund, "Application-Layer Mobility using SIP", ACM SIGMOBILE Mobile Computing and Communications Review, Vol. 4, No 3, July 2000.
[6] C. Pollak, S. H. Oh, A. Giralda, V. Nikas, G. Lioudakis, J. M. Sanchez, "A SIP based concept for NAS signaling in combination with an all-IP core network architecture", Mobile Summit 2004.
[7] 3GPP TS 23.221 v.6.1.0, "Technical Specification Group Services and System Aspects; General UMTS architecture".
[8] 3GPP TS 23.101 v.4.0.0, "Technical Specification Group Services and System Aspects; Architectural Requirements".3GPP TS 23.002 v.5.8.0, "Technical Specification Group Services and System Aspects; Network Architecture".
[9] C. Pollak et al., "Requirements, Design and Test Specification for CN entities and UE", IST-2001-37266 SAILOR deliverable D0401, 2003.
[10] C. Pollak et al., "Test Report on RASN, IMS and UE", IST-2001-37266 SAILOR deliverable D0402, 2004.