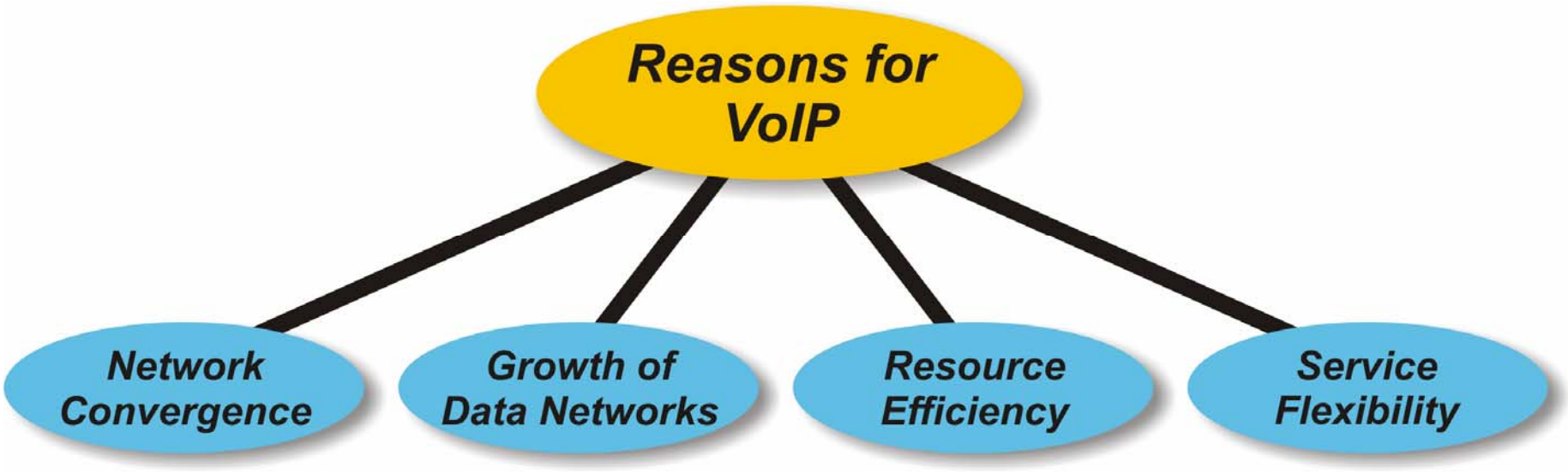
Reasons for VoIP



Reasons for VoIP

Network Convergence
For corporate customers it is quite expensive to administrate and operate two different network types: The first one for voice and the second one for data. Why not combining both networks and offering all applications through one network?

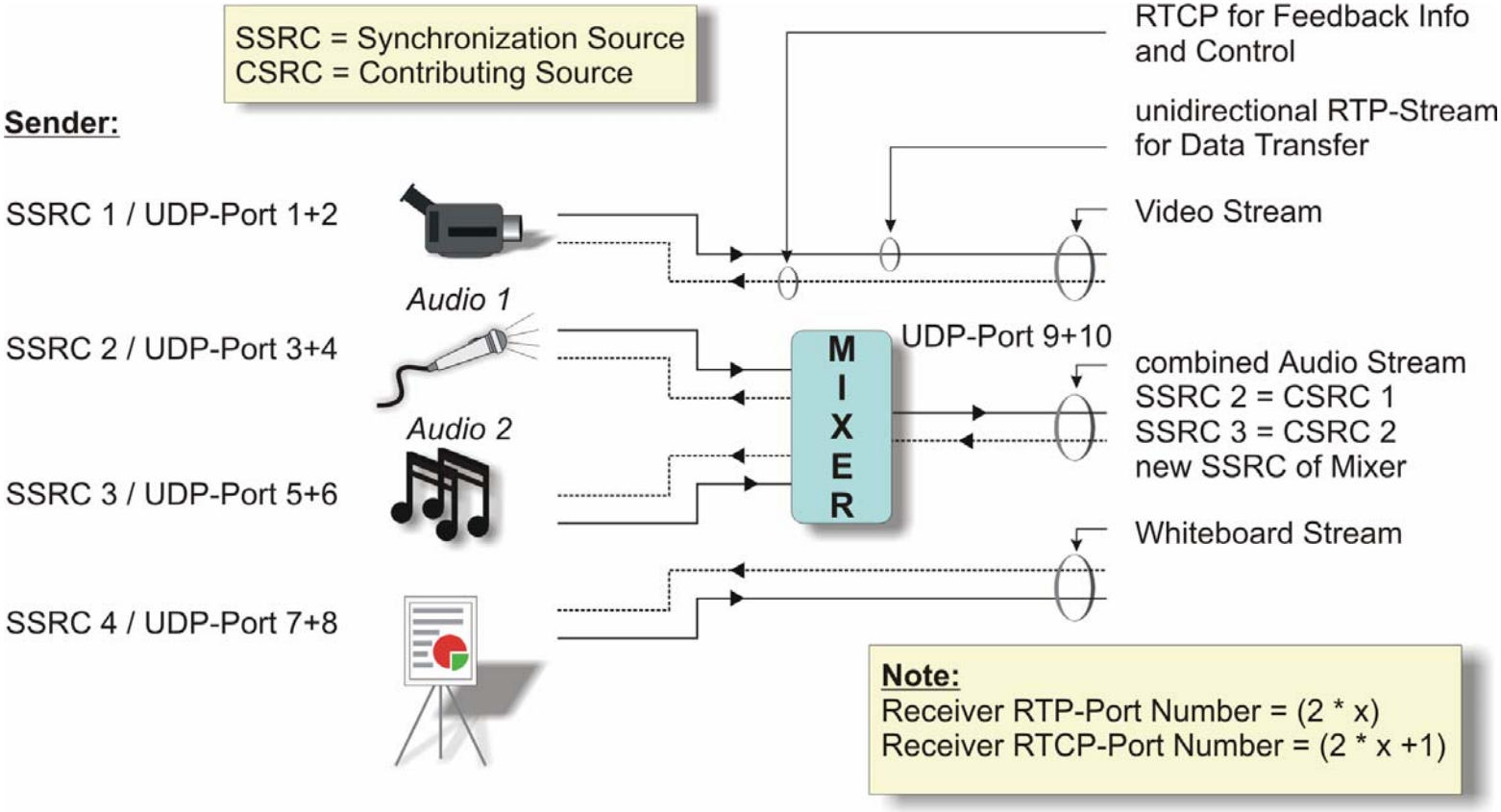
Growth of Data Networks
Nowadays, the growth of data traffic is by far bigger than the growth of voice traffic. This is also the reason why there are specialized data networks in the first place. When data traffic was still low, this traffic was routed through the existing voice network resources. Since this relationship has turned around, the question arises why voice traffic should not be routed through data networks which are more and more IP-based.

Resource Efficiency
Even when only POTS is required and provided, the PSTN uses the G.711-voice coder (⇔ PCM) which requires 64 kbit/s in each direction to support a single voice call. Nowadays, more advanced voice coders are available like for instance the G.729-voice coder that provide a good voice quality but only require 8 kbit/s of bandwidth. Another advantage of new voice coders is their capability to seize transmission during silent phases (⇔ DTX) which further helps to save resources. The regular PSTN-network is not well suited for processing this variable rate information since it is tailored to route 64 kbit/s-channels. Obviously, the PSTN *could* process this information, but there is no point in doing so because no resources would be saved. Therefore, the advantages of speech compression cannot be used for cost reduction within the PSTN.
It can be expected that speech compression algorithms will be even more advanced in the future. Therefore, it makes sense to remain flexible in the way how resources are provided. This approach asks for using a packet-switched network to transmit voice.

Service Flexibility
Although regular voice traffic can be expected to remain the dominant service also in the future, other services will arise when the preconditions are fulfilled. Those other services are for instance video transmission, text messaging or a whiteboard function during a call. Another option is any combination of these services with voice. Therefore, it makes sense to use very flexible protocols for service selection (⇔ call control) and resource reservation. The respective VoIP-protocols like H.323 and SIP provide this flexibility.

The Real Time Transport Protocol (RTP and RTCP)

- **Operation of RTP and RTCP**



The Real Time Transport Protocol (RTP and RTCP)

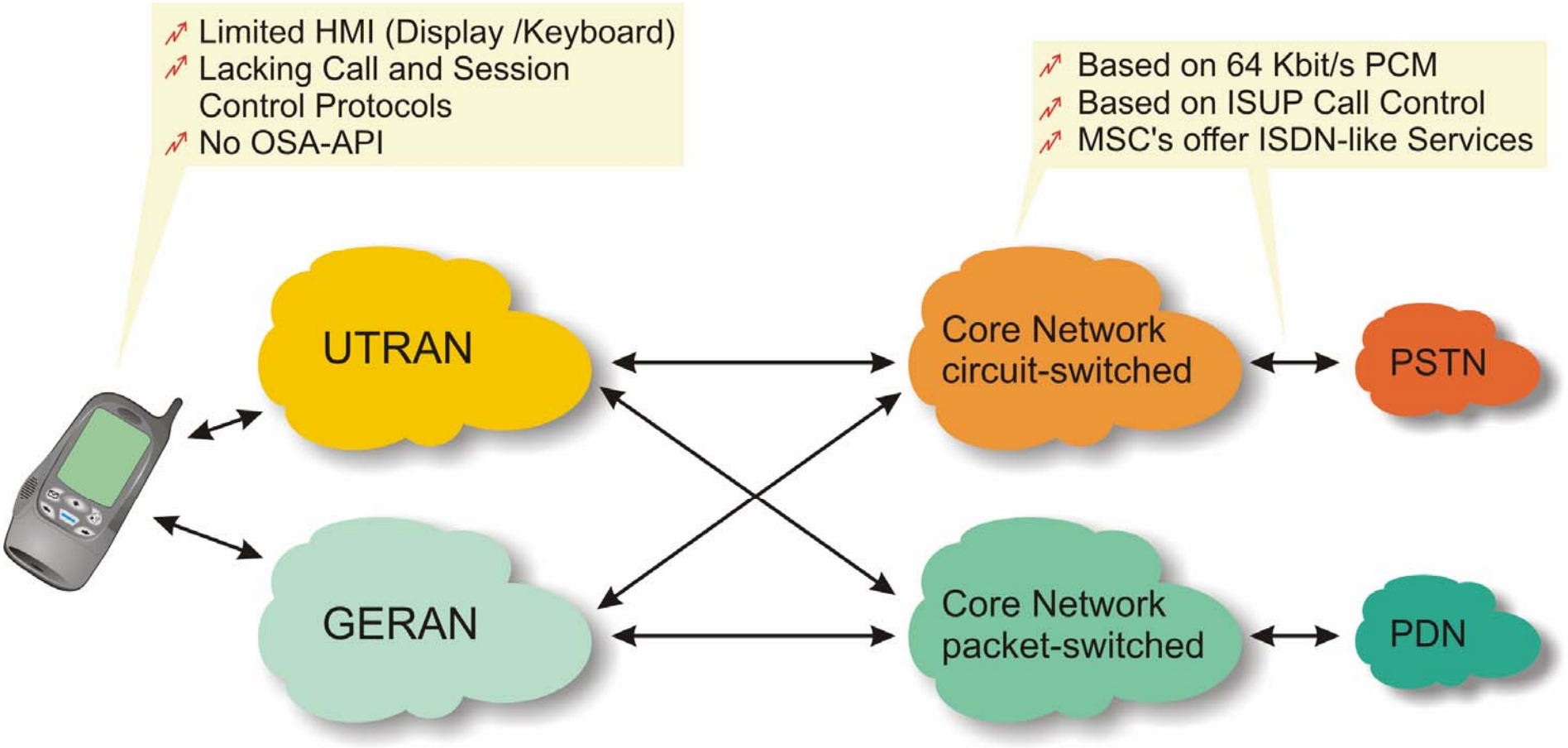
Operation of RTP and RTCP

The operation of RTP and RTCP is illustrated in the figure, using the example of three parallel data streams which are transmitted by the sender on the left side to an invisible receiver. The three parallel data streams carry video information (⇔ on UDP-port 1), combined audio information (⇔ on UDP-port 9 and 10) and whiteboard information.

Note:
- Without previously invoking resource reservation through RSVP, RTP is not capable of providing real-time service.
- RTP shall always use an even-numbered destination port (⇔ 2 * x) while the related RTCP-signaling shall occur on the following port (⇔ 2 * x + 1).
- The actual port numbers to be used are negotiated between the peers through the very session control protocol (e.g. SIP, H.323) before RSVP is reserving the related resources for RTP-streams.

- **SSRC (Synchronization Source / 32 bit)**
  Each sender is uniquely identified in an RTP-stream through its SSRC which is randomly selected by the sender and relates for instance to a camera.
- **Payload Type / Media Type**
  The media type which is conveyed within an RTP-stream is uniquely identified through the Payload Type-field which is part of the header of each RTP-frame.
- **CSRC (Contributing Source / 32 bit)**
  In our example, the two separate audio streams are de-multiplexed at the sender through a “mixer”. To still identify the modules which contributed to the combined audio stream, the two SSRC’s of the two audio separate audio devices is inserted in the header of the related RTP-frames as CSRC-field.
- **Timestamp Information**
  To provide for jitter calculations at the receiver side, each RTP-frame carries a 32 bit long timestamp which identifies the very time when the first data octet within that RTP-frame was sampled.

Limitations of the Release 99 Network & Software Architecture



Limitations of the Release 99 Network & Software Architecture

With a Release 99 compliant implementation of the network architecture, there are only a few differences to our current perception and experience of mobile communication. In fact, the major changes are the advent of UMTS and EDGE in the access network domains. The major consequence of this is the resolution of the bandwidth bottleneck on the air interface. However, many other issues remain unresolved:

Which new services become realistic with Rel. 99?

The new UTRAN is definitely much better suited for new services with stringent and variable QoS-requirements on possibly many different data streams than the GSM-based access network. However, the major question is *which* new services can be offered by mobile network operators that will generate new revenue streams. In addition to the business and marketing aspects of this question there are technical issues. These technical issues relate in particular to the lacking call or session control capabilities of the core network and the terminals. Another issues is the end-to-end negotiation capabilities of specific QoS-requirements. In that respect, the circuit-switched core network should play a major role since it can inherently provide real-time QoS. However, the circuit-switched core network is based on 20 years old ISDN equipment that is not well suited for the communication of the 21st century.

How do the narrow-band MSC's handle broadband service requests?

Nowadays, the circuit-switched core network is primarily based on the MSC's which are basically ISDN-switches, tailored for the use in mobile networks. These ISDN-switches and the accompanying equipment like modem banks can perfectly handle telephone calls with 64 kbit/s. However, their update into broadband switches that can digest and process new services on multiple simultaneous data streams doesn't make much sense, considering the alternative of flexible media gateways with smaller footprint which are already available.

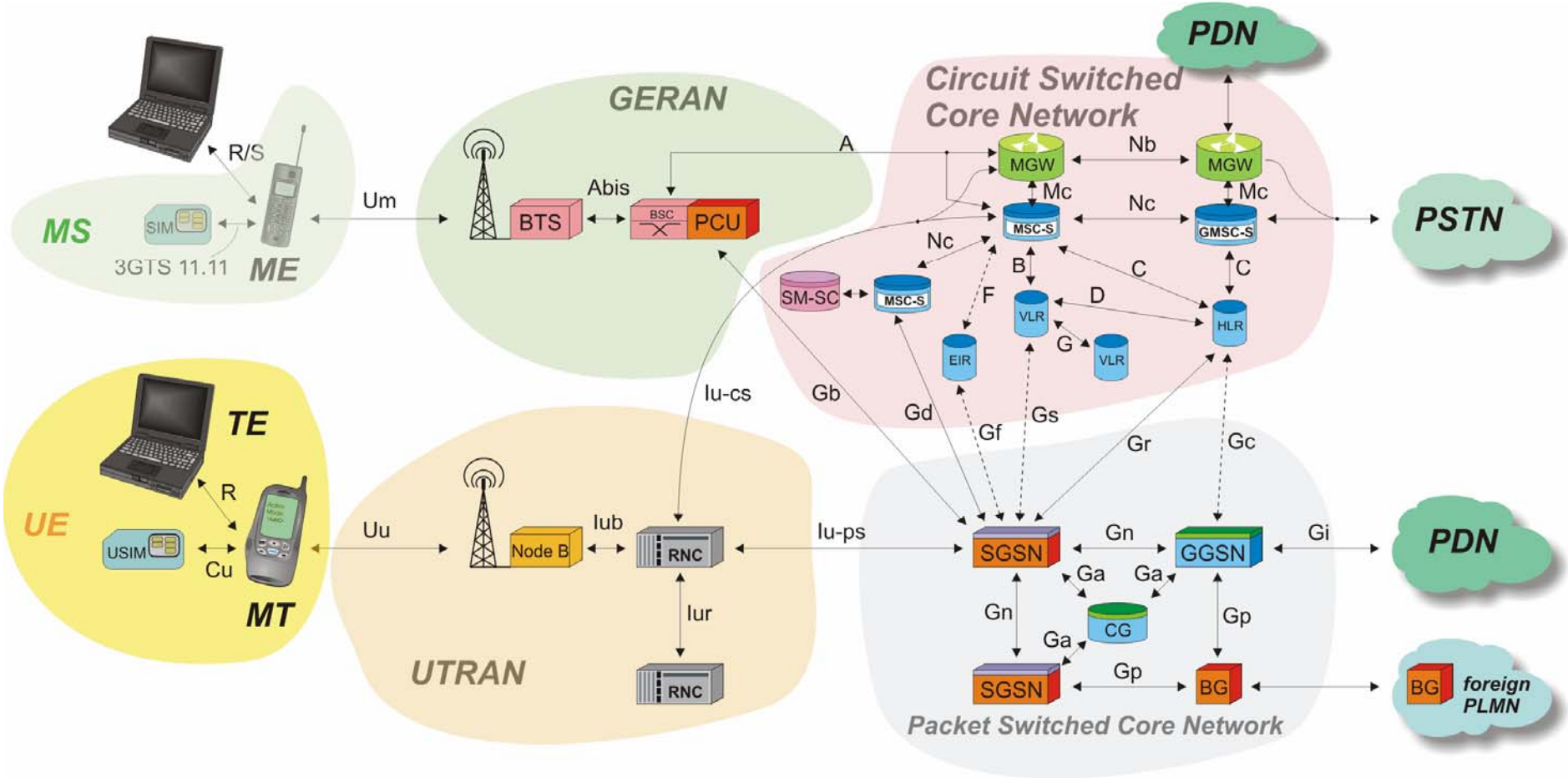
How can the user gain access to these new services?

This appears to be a minor issue. However, how do you select a video telephony service on your device? Which number do you dial from which application? Finally, the available terminals today are still rather plain telephones than multi-function communicator devices. This issue cannot be resolved with Rel. 99. New call and session control software within these terminals is required to make new applications happen. The advent of SIP and the approach of Open Services Access (OSA), inherited from the IT-world, may provide alternatives in the future. However, neither SIP nor OSA are applicable with Rel. 99.

Conclusion:

- With Rel. 99, new services primarily relate to the packet-switched core network domain. These new services are enabled by a more sophisticated QoS-concept and by the higher bandwidths on the air interface which are enabled by UTRA and EDGE.
- The circuit-switched core network domain with Rel. 99 remains identical to previous releases and becomes a bottleneck for new services.
- Both Rel. 99 and Rel. 4 do not support new multimedia call control concepts like SIP in neither the core network nor the mobile station.

Access and Core Network Architecture with Release 4



Access and Core Network Architecture with Release 4

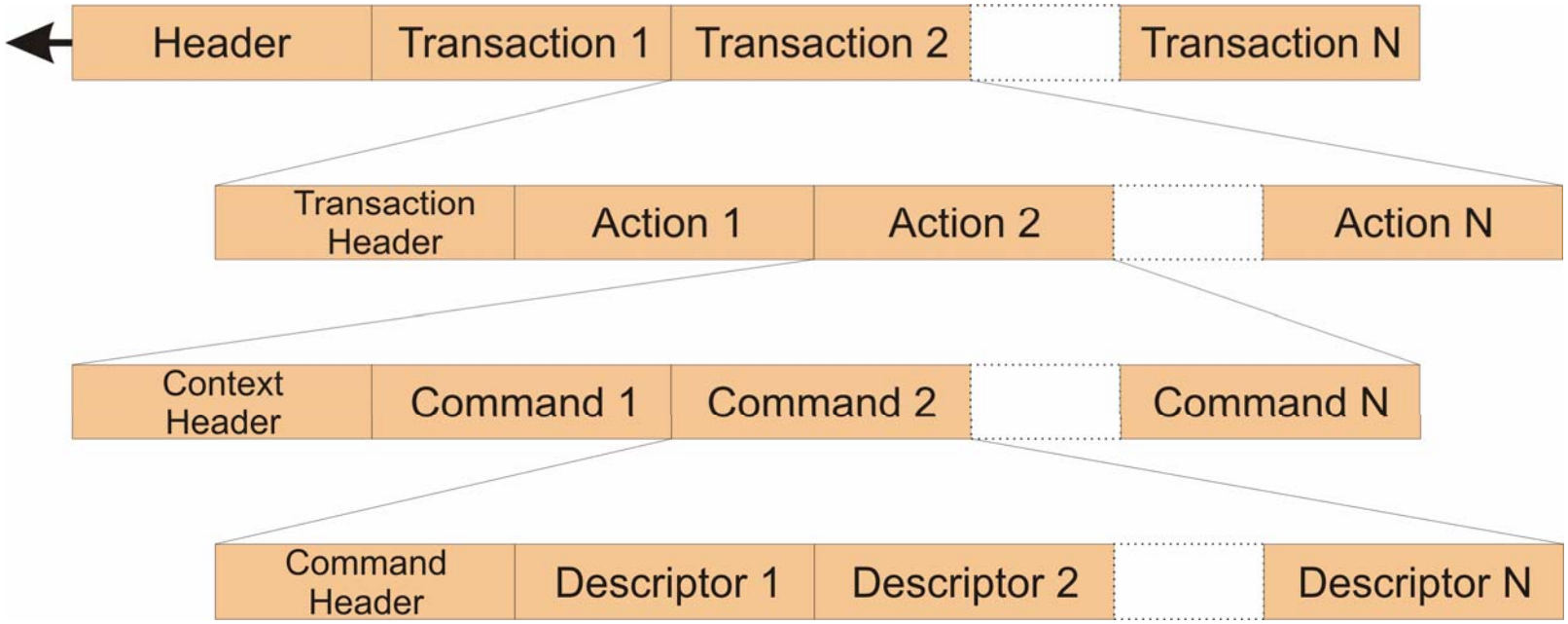
An overview of the entire network architecture with Rel. 4 is provided in the figure. Please note the following specifics:

⇒ Both the A-interface and the Iu-cs-interface are divided into a data part and a control part. The data part terminates at the MGW while the control part terminates at the MSC-S.
⇒ The figure implies a monolithic architecture which is intentional with Rel. 4, considering the fact that there is a one-on-one relationship between RNC / BSC on one hand and MSC-S / MGW on the other hand (as mentioned and explained before).
⇒ Obviously, there may and will be mixed configurations of MSC's (< Rel. 4) and MSC-S / MGW (Rel. 4). In such a case, interworking between ISUP and BICC is required and needs to be provided through the MSC-Server.
⇒ Please consider that many Rel. 4 implementations will go for an all-IP core network which translates into an IP-cloud interconnecting the various network elements.

[3GTS 23.002]

The H.248 Message Structure

- **Overview**

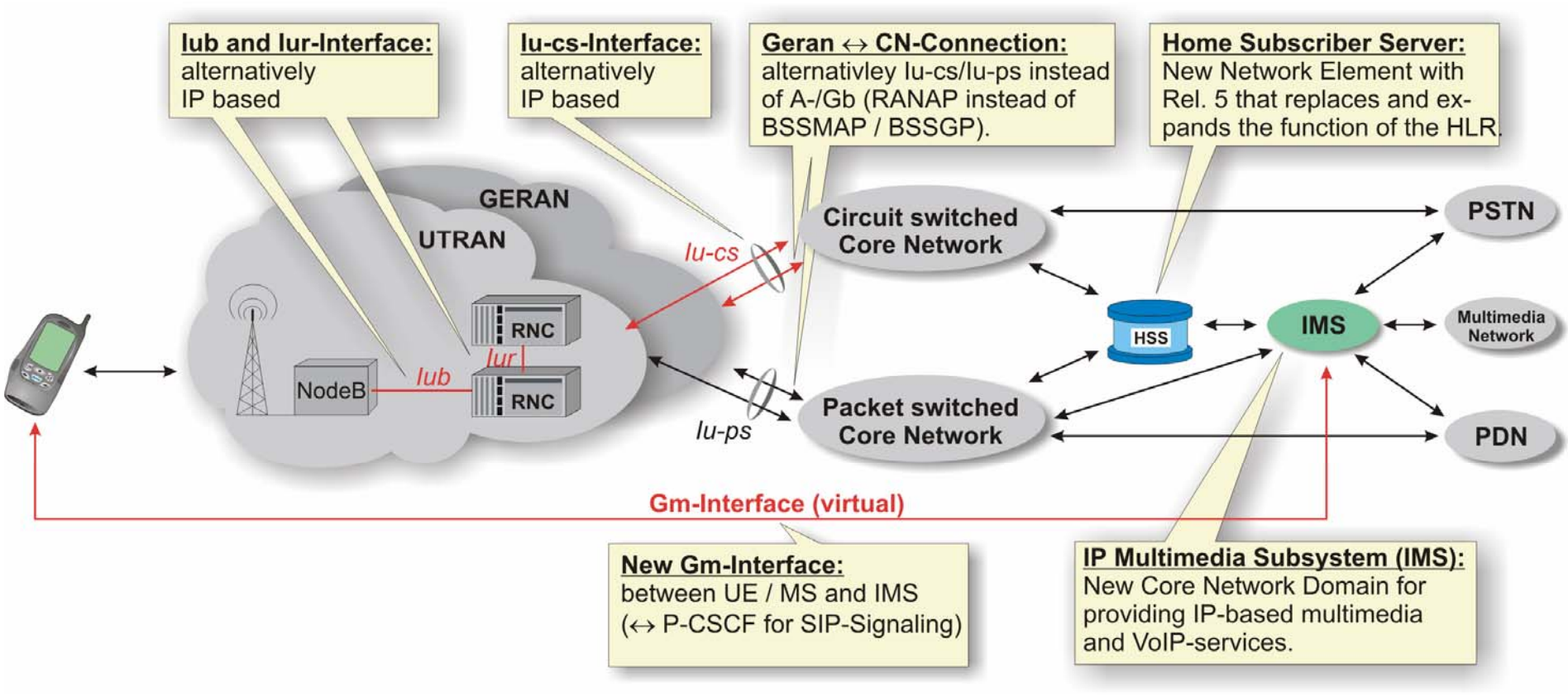


The H.248 Message Structure

Overview
The principles of H.248-message formatting are:

- **Hierarchical structure**
  As can be seen in the figure, each information element type has its fixed position in the structure. The header is basically only there to identify the message originator (the destination is identified through lower layers). In addition, the header may contain an IPsec-authentication header.
  ⇒ Behind the header, there is the hierarchical level of the “transactions”. Each information exchange between MSC-S and MGW is called a transaction that comes with a transaction ID. Upcoming H.248-message will use the same transaction ID to relate responses to commands.
  ⇒ The next level is the hierarchical level of “actions”. Actions are requests or replies and they are related to a single context.
  ⇒ For each action and therefore for a single context, there is then a sequence of commands to be imposed on that context like adding a termination or applying a certain notification signal.
- **Concatenation of elements of the same type into a single message.**
  As can be seen in the figure, an H.248-message allows for the concatenation of transactions, actions, commands and descriptors. In this way, each H.248-message may contain signaling information which is related to various different contexts and terminations.

Important Architectural Changes with Release 5



Important Architectural Changes with Release 5

The figure illustrates the most important architectural changes with 3GPP's Rel. 5 recommendations:

IP-Multimedia Subsystem (IMS)

The IMS represents a new core network domain consisting of different network elements like the P-CSCF (Proxy Call Session Control Function) or the MGCF (Media Gateway Control Function). The IMS provides new IP-based call and session control functionality to the PLMN (⇔ SIP in particular). The IMS is connected to the PLMN exclusively through the packet-switched core network domain (⇔ GGSN).

Home Subscriber Server (HSS)

With Rel. 5, the HLR is replaced by the more powerful HSS. Opposed to the HLR, the HSS is capable to communicate with the new IMS core network domain and the HSS can provide subscription services for multimedia services.

New Gm-Interface

Between the UE or the MS and the P-CSCF within the IMS, there is a new virtual interface, the Gm-interface for the exchange of SIP-related signaling messages. The virtual Gm-interface is established via the access network (GERAN or UTRAN) and the packet-switched core network domain.

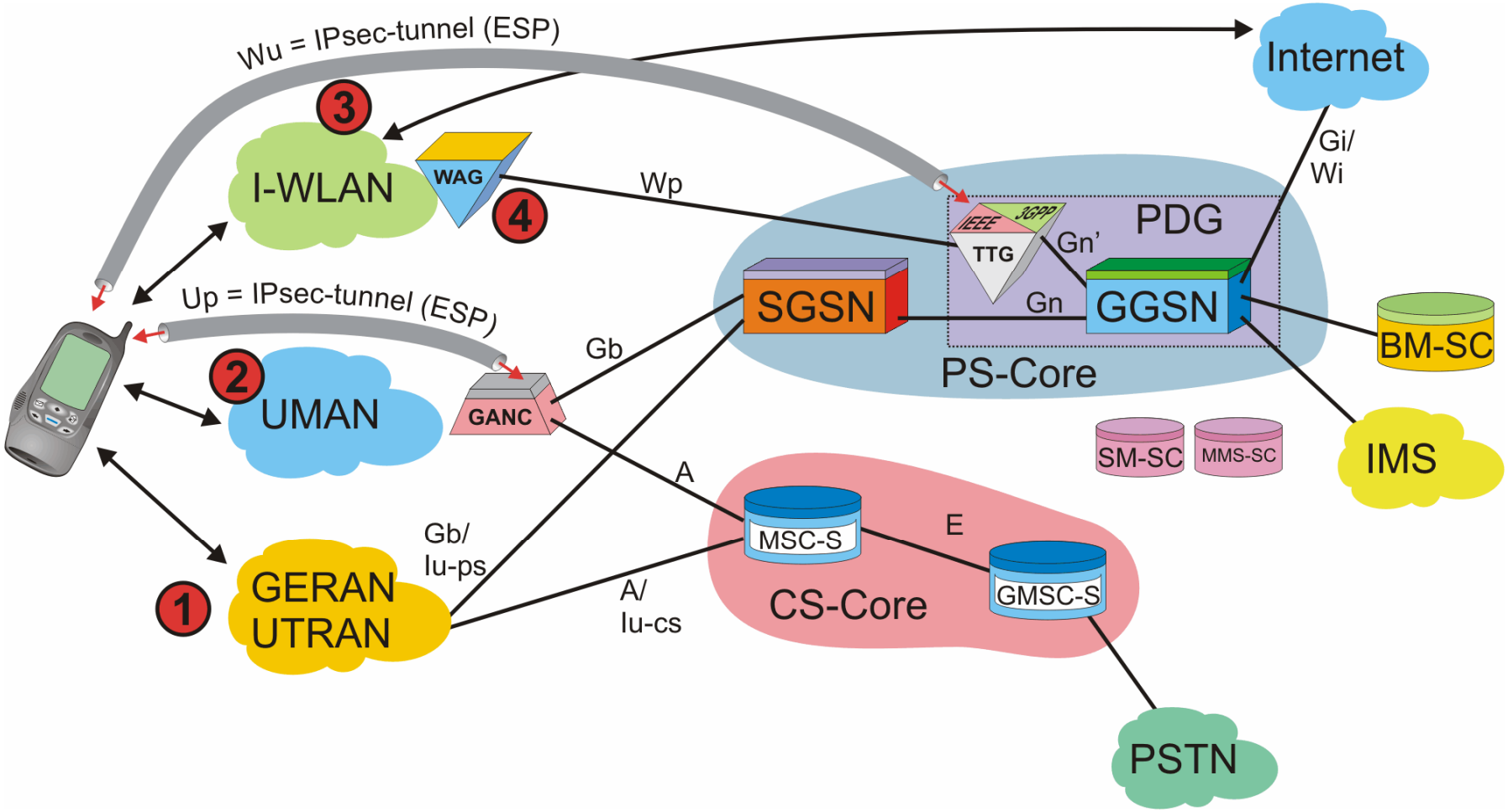
GERAN ⇔ Core Network Connection as Iu-Interface

With Rel. 5, the A- and Gb-interfaces can be replaced by Iu-cs- and Iu-ps-interfaces. Consequently, the GERAN can be connected to the core network domains the same way as the UTRAN. This reduces the number of protocols to be supported and simplifies the network architecture.

Iub-, Iu-cs- and Iur-Interface alternatively IP-based

With UMTS Rel. 4, the circuit-switched core network can be built on an IP-network. With Rel. 5, this strategy is expanded down to the Iub-, Iu-cs- and the Iur-interface. Therefore, NodeB's can be interconnected to their C-RNC using IP. The same applies to the interconnection between RNC's on one hand and MSC-S and MGW on the other hand. Obviously, the Iur-interface can also be based on IP.

Access and Core Network Architecture with Release 6



Access and Core Network Architecture with Release 6

Overview

The figure illustrates the entire network architecture with Release 6 no more in full detail but focusing on the changes which are applicable with R6. The most important changes with R6 are the alternative radio access technologies and the advent of an MBMS-server, called BM-SC.

Interconnection of Alternative RAT's

Bullet 1: Access through GERAN/UTRAN
Through GERAN/UTRAN the mobile station gets access to all circuit-switched and packet-switched services that the PLMN offers, namely to the PSTN, to the internet (incl. IP-address allocation), to MBMS (through BM-SC), SMS and MMS and to the IMS.

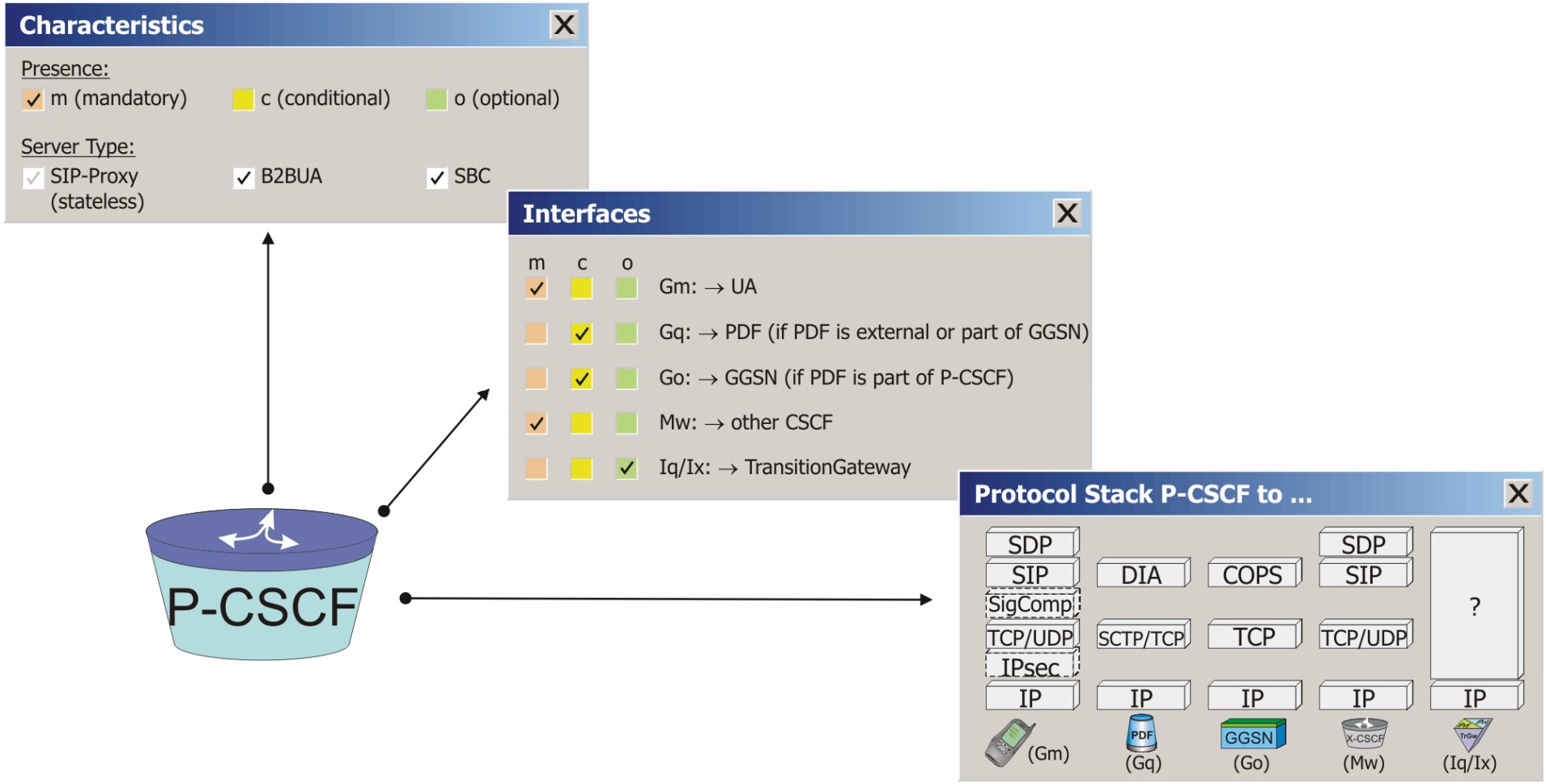
Bullet 2: Access through GAN/UMAN
GAN/UMAN offers the same access service to the mobile station. The GANC interconnects to both, the circuit-switched and the packet-switched core network domain and therefore to all services that these offer.

Bullet 3: I-WLAN Direct IP-Access
In this case, the mobile station only uses the PLMN for access authentication and authorization. The access to internet based services is done directly without involvement of the PLMN. However, the mobile station can access PLMN-based packet-switched services like MBMS or SMS only through GERAN/UTRAN.

Bullet 4: I-WLAN 3GPP IP-Access
In this case, the mobile station and the packet-switched core network domain establish an IPsec-tunnel between themselves and the mobile station gets access to all PLMN-based packet-switched services, namely IMS, MBMS, MMS and SMS. It is a configuration issue whether direct internet traffic is routed through the PLMN or is handled as in case of "I-WLAN Direct IP-Access".

[3GTS 23.002]

Facts Sheet



Facts Sheet

The figure illustrates some genuine characteristics of the P-CSCF.

Characteristics

⇒ The presence of the P-CSCF is a must in an IMS, considering its specific tasks of e.g. SIP-compression or establishment of security associations towards the UE.

⇒ The P-CSCF needs to be at least a B2BUA but it may even be an SBC, if the IMS-Access Gateway or TrGw becomes integral part of the P-CSCF. In such a case, the Iq/Ix-interface is no longer an open interface.

⇒ Most interestingly, the P-CSCF may behave as a stateless SIP-proxy in a very important situation: On port number 5060, all received SIP-messages should be handled stateless to allow the IMS coping with DoS-attacks.

Interfaces to other Network Elements

The Gm-interface to the UA is obviously mandatory; the Gq-interface is only there when the PDF is not integrated into the P-CSCF; the Go-interface is only there when the PDF is integrated into the P-CSCF and this integrated PDF communicates through the Go-interface with the PEP (Policy Enforcement Point). In case of 3GPP, the PEP is integrated into the GGSN; the Mw-interface is mandatory as it allows the P-CSCF to communicate with the other CSCF's; the Iq/Ix-interface is optional and only there, if NAT- and/or IP-version Interworking with the IP-CAN are necessary or if the TrGw is external to the P-CSCF.

Protocol Stacks of the P-CSCF

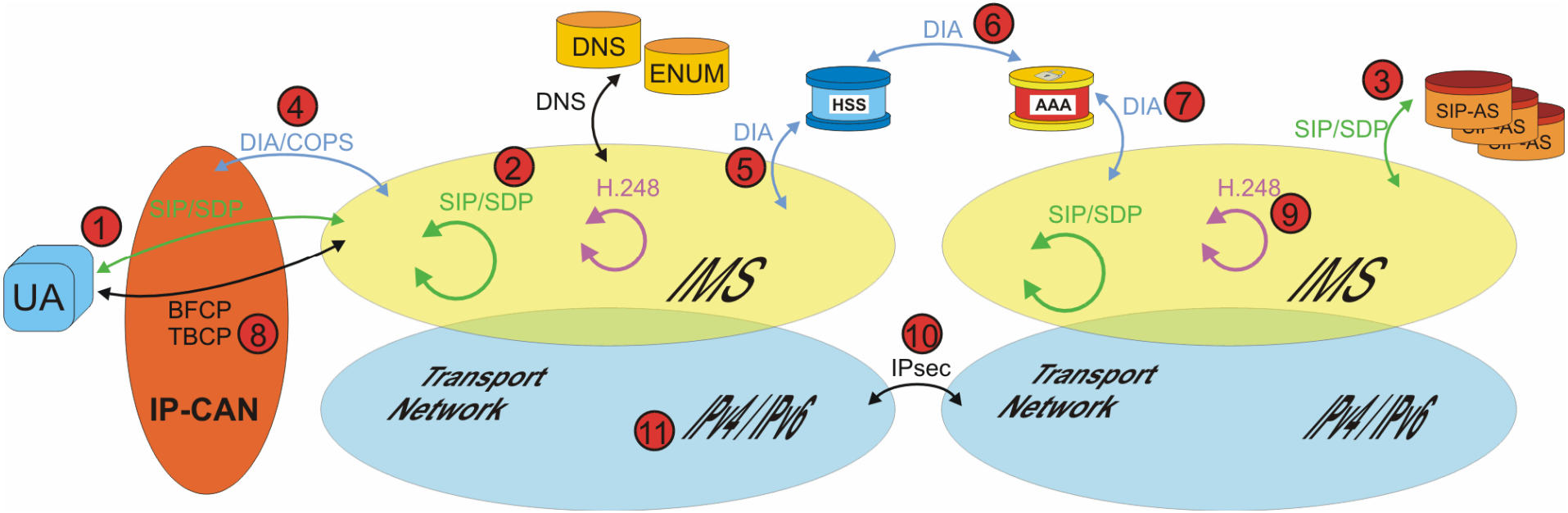
Please note that on the Gm-interface there is a new compression layer between UDP/TCP and SIP. All SIP-based interfaces show UDP and TCP as transport protocol but SCTP may also be supported.

Note that each P-CSCF is identified by its FQDN/IP-address and a SIP-URI (e.g. sip: P-CSCF_No1@operator.com).

[3GTS 23.228 (4.6.1)]

Overview

- **Protocols within the IMS-Control Plane**



Overview

Protocols within the IMS-Control Plane

SIP and SDP

It is noticeable that SIP is used for end-user signaling (⇔ UNI) (bullet 1) as well as among SIP-proxies (⇔ NNI) (bullet 2) and towards application servers (bullet 3). However, SIP requires SDP to describe the media of a session and therefore, the previous statement also applies to SDP.
[RFC 3261 (⇔ SIP), RFC 2327 (⇔ SDP) draft-ietf-mmusic-sdp-new-26.txt (⇔ SDP), RFC 3266, RFC 3264]

DIA (DIAMETER))

The DIAMETER Protocol (bullet 4 and 5) is very important as it allows the SIP-proxy servers to interrogate and interact with databases like the HSS (Home Subscriber Server). DIAMETER in general is there to exchange subscriber related information like:
Bullet 5, 6 7: Is a subscriber authorized to use a service and provision of the authentication information for that subscriber etc.
Bullet 4: QoS-Authorization over the Gq-interface.
DIAMETER is also used to for session offline and online charging. In many aspects, DIAMETER can be compared with the MAP-protocol of 3GPP's 3GTS 29.002. DIAMETER is frequently considered the successor of RADIUS. This is also the explanation of the strange term "Diameter" in the protocol name: A diameter is twice the radius and in terms of protocols, DIAMETER shall be considered as successor of RADIUS. DIAMETER is defined as a base protocol (⇔ DBP) to be supported by all implementations and application specific amendments. 3GPP for instance defined various amendments to the DBP for use on the Cx-, Dx-, Sh- and charging interfaces. [RFC 3588, RFC 3589, http://www.diameter.org/, 3GTS 29.229, 3GTS 29.329]

COPS

The Common Open Policy Service Protocol is used for policing between the PDF and the PEP (bullet 4) [RFC 2748].

H.248 / MEGACO

H.248 / MEGACO (bullet 9) allow the MGC to control one ore more media gateways. Control relates particularly to the seizure and release of resources for user data transfer [ITU-T H.248, RFC 3015].

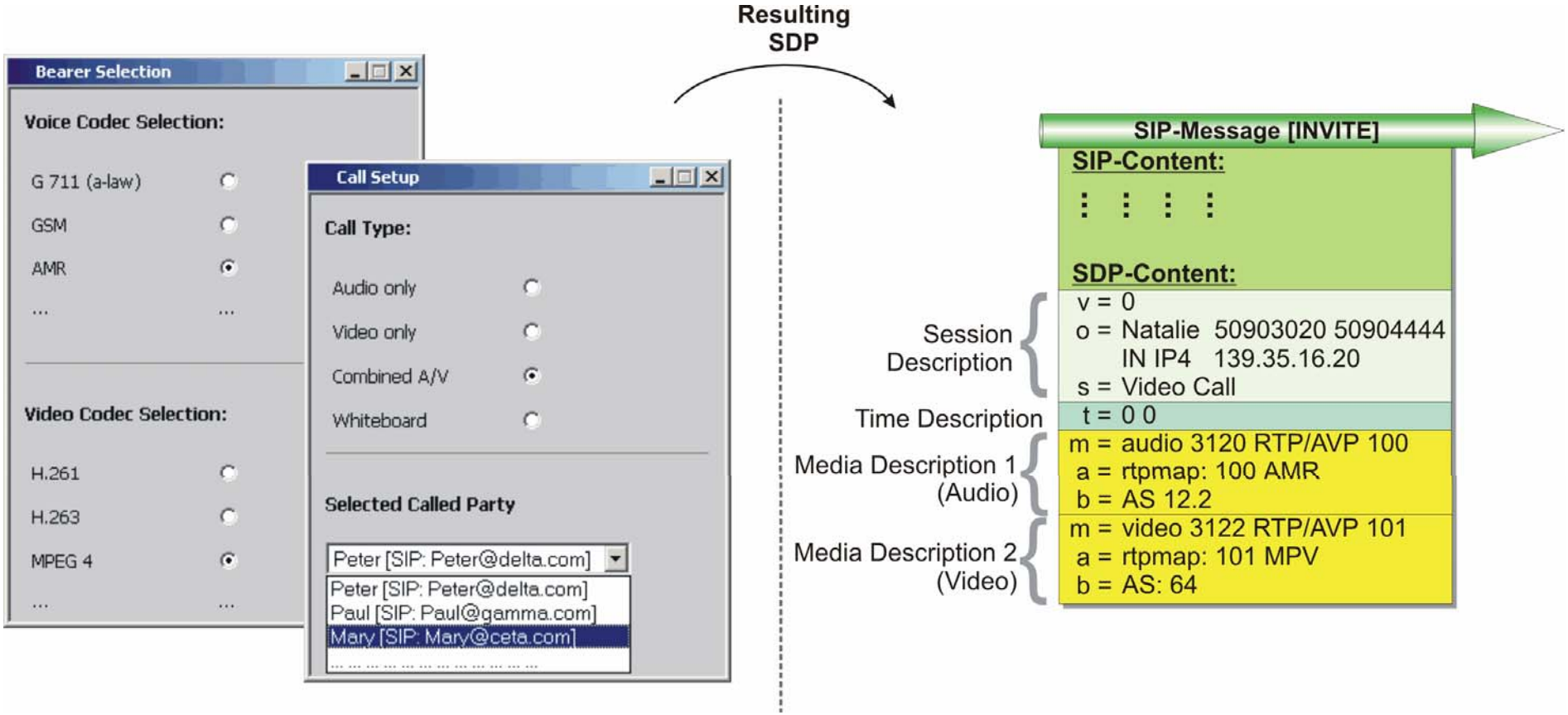
BFCP / TBCP

Binary Floor Control Protocol (bullet 8) and Talk Burst Control Protocol are used for floor control within the IMS. TBCP is restricted to use within the PoC-service [draft-ietf-xcon-bfcp-05].

IP, IPsec and TLS

The IMS uses IPv4 or IPv6 in the transport network and IPsec or TLS to provide for secure links between IMS-facilities.

Interworking between Application and SDP

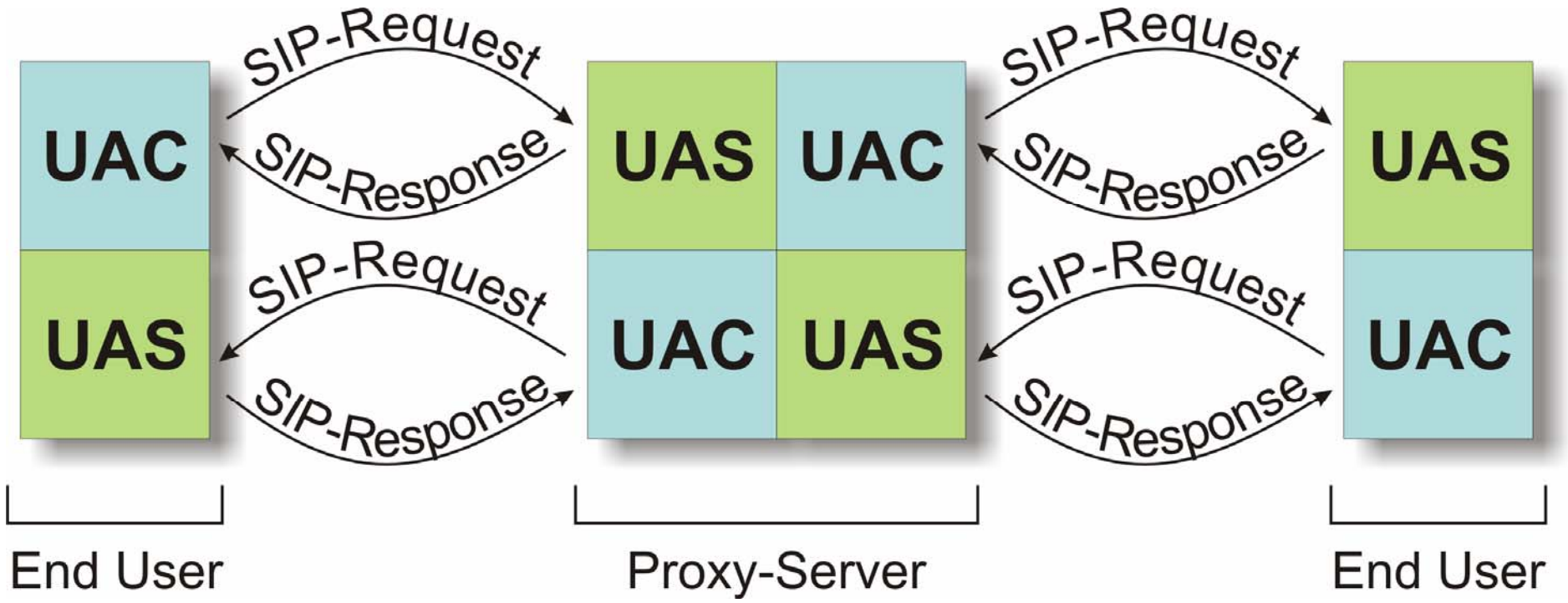


Interworking between Application and SDP

- **Session Description Items**
  - ⇒ The user's name is apparently Natalie (o = Natalie …) Natalie is on an IPv4 network and her current host has the IP-address 139.35.16.20.
  - ⇒ Natalie has somewhere else (not shown) named this call type "video call" (s = Video Call)
- **Time Description Items**
  - ⇒ The call shall start immediately <start time = 0> and shall last infinitely <stop time> = 0. Obviously, the call will be terminated by other means when desired by either user.
- **Media Description Items**

  Two different media descriptors follow. Each is indicated by an m-item.
  - ⇒ The first media description refers to the audio portion of the call. Natalie's host computer intends to use RTP/UDP/IP as transport protocol and the UDP-port number on the called party's computer where audio data will be sent is '3120'$_{dec}$. Note that port number '3121'$_{dec}$ is reserved for the related RTCP-signaling.
  - ⇒ Dynamic Payload type indication for RTP is selected. The RTP-Payload Type shall be '100'. Through the attribute line "a = rtpmap: 100 AMR", this dynamic payload type is assigned to the AMR-coder which has also been selected by Natalie on the user interface.
  - ⇒ The b = AS: 12.2 tells the receiver that 12.2 kbit/s are required for the transfer of the AMR-data. The term 'AS' means that this bandwidth only relates to this media stream.
  - ⇒ The second media description is related to the video portion of the call. As for the audio part, RTP/UDP/IP shall serve as transport protocol. The destination UDP-port number where the video data shall be sent is '3122'$_{dec}$.
  - ⇒ The dynamic RTP-Payload Type 101 is used. This Payload Type is assigned to MPV (MPEG-Video) in the line "a = rtpmap: 101 MPV.

User Agent Client and User Agent Server



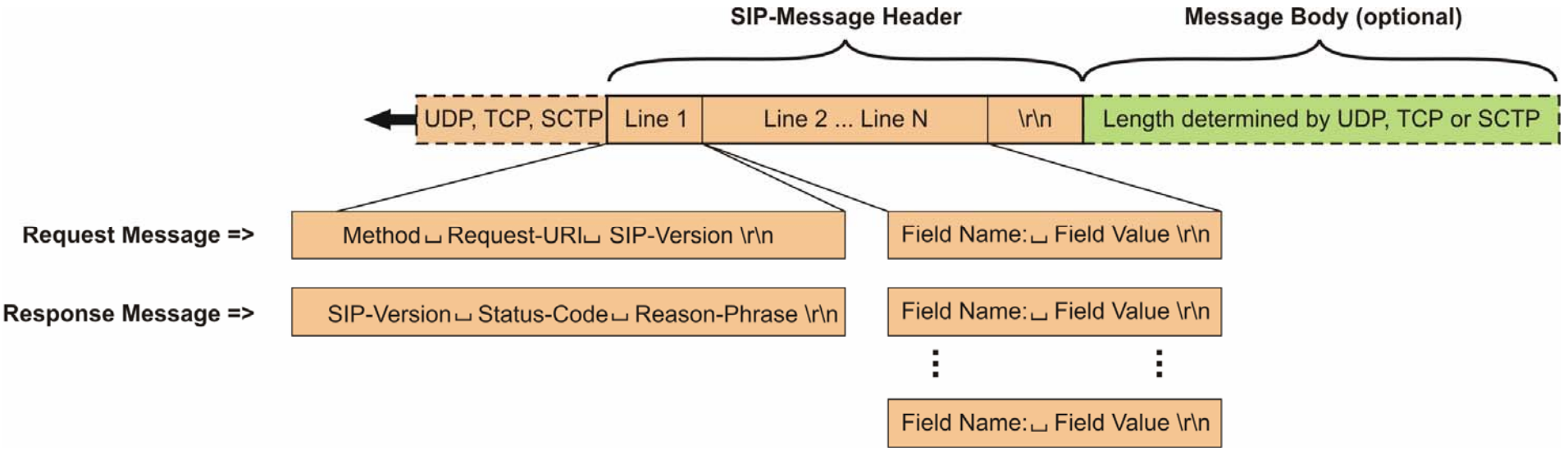
User Agent Client and User Agent Server

All SIP-entities are considered to be UA's (User Agents). User agents are again differentiated into

- **UAC's (User Agent Client)**
  A UAC is a logical entity which is generating SIP-requests.
- **UAS's (User Agent Server)**
  A UAS is a logical entity which receives SIP-requests, processes them and responds to them with SIP-responses.

The definition of UAC and UAS is on a per-transaction basis. Intermediate proxies have to act as both UAC and UAS during a transaction.

SIP-Message Format



SIP-Message Format

General Information

SIP-messages are either requests or responses. Both requests and responses consist of a message header and an optional message body which contains for instance an SDP-description.
Each SIP-message is embedded in exactly one TCP- or UDP-frame. Note that by default, SIP will use UDP as transport protocol. The default destination port number for SIP-messages in TCP, SCTP and UDP is '5060'$_{dec}$. If a secure transport layer (⇔ TLS) is used or if SIPS (secure SIP) is used then the default port number shall be '5061'$_{dec}$.

Request Messages

Request messages are sent by a UAC (User Agent Client) to a UAS (User Agent Server). Each request messages has a Request-Line as the first header line. This first line contains the:

⇒ SIP-message type, called "method". Examples for methods are INVITE, ACK or REGISTER. We will later get back to the different methods.
⇒ Request URI (sip-URI, sips-URI or tel-URI) which specifies the destination of that SIP-message.
⇒ SIP-Version. This version shall be "SIP/2.0" if RFC 3261 is used.

The request line shall terminate with a <CRLF>. Various additional header fields follow of which some are mandatory while others are optional. Each header field line shall terminate with <CRLF>. The end of the header is indicated through an empty line which only consists of <CRLF>.

Response Messages

Response messages are sent by a UAS (User Agent Server) to a UAC (User Agent Client). Each request messages has a Status-Line as the first header line. This first line contains the:

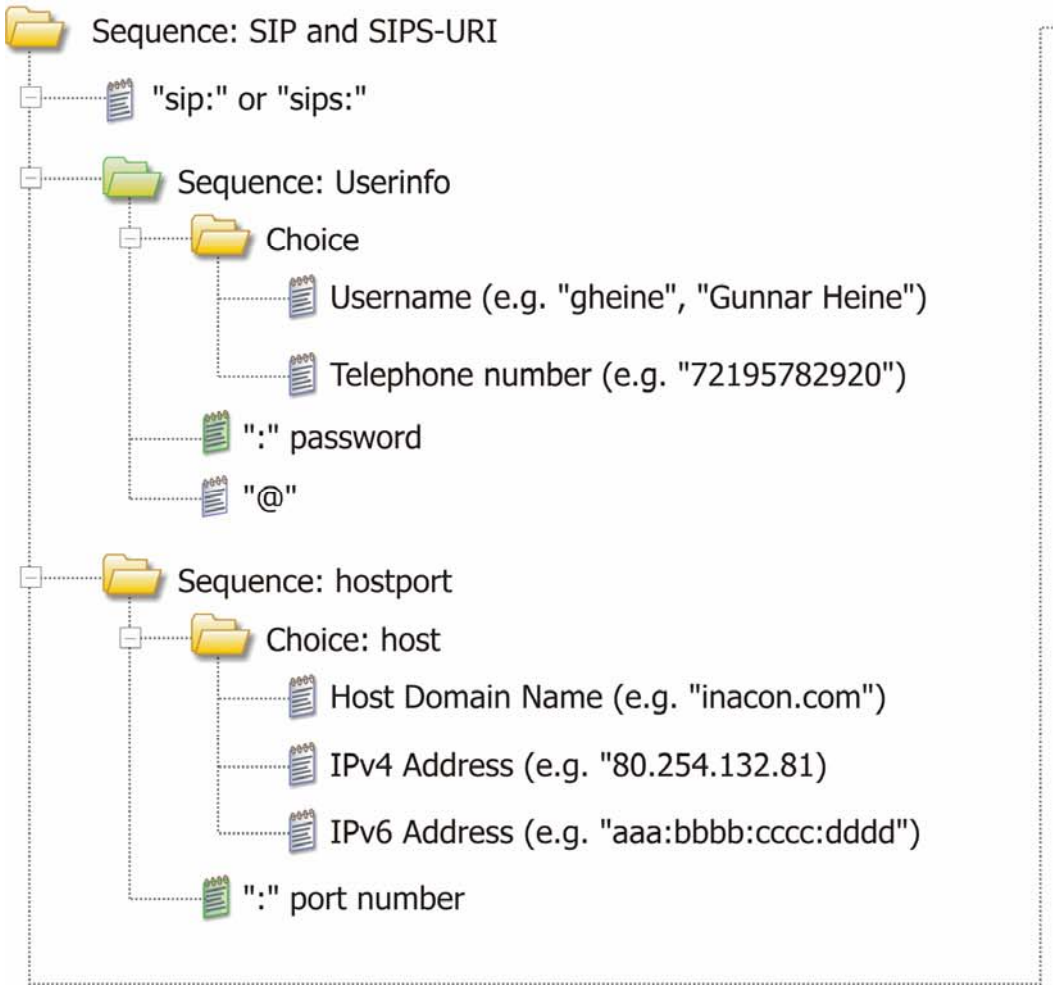
⇒ SIP-Version. This version shall be "SIP/2.0" if RFC 3261 is used.
⇒ Numeric status code (e.g. '200'dec).
⇒ Human readable interpretation of the status code which is called a reason-phrase.

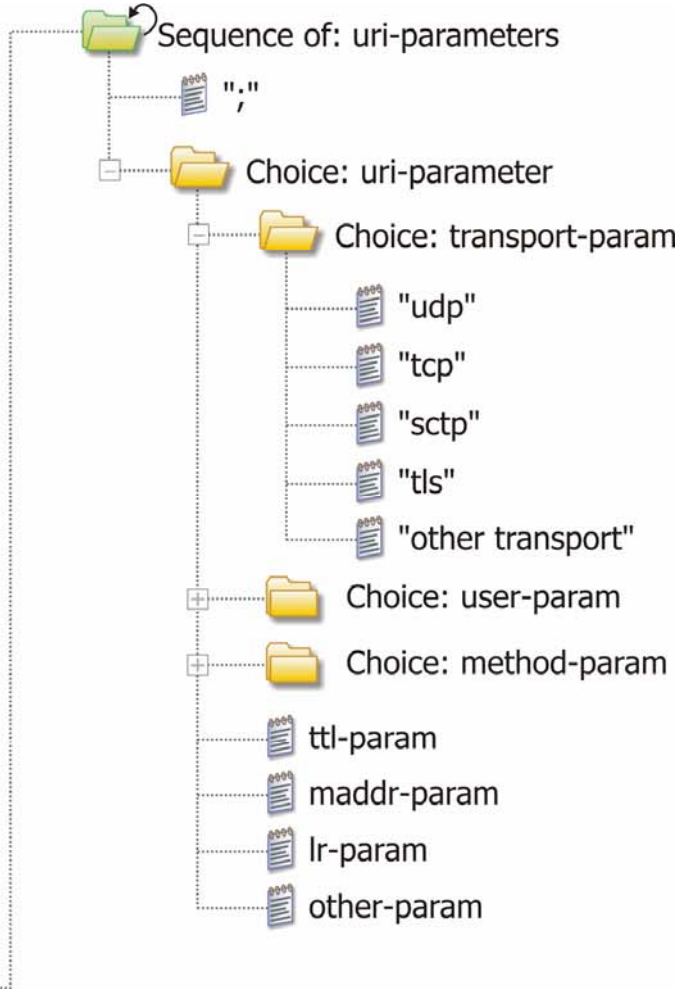
Like the request line in request messages, the status line in response messages shall terminate with a <CRLF>. Various additional header fields follow of which some are mandatory while others are optional. Each header field line shall terminate with <CRLF>. The end of the header is indicated through an empty line which only consists of <CRLF>.

Note: The format of SIP-messages is inherited from HTTP.

[RFC 3261 (7)]

The SIP(S)-URI



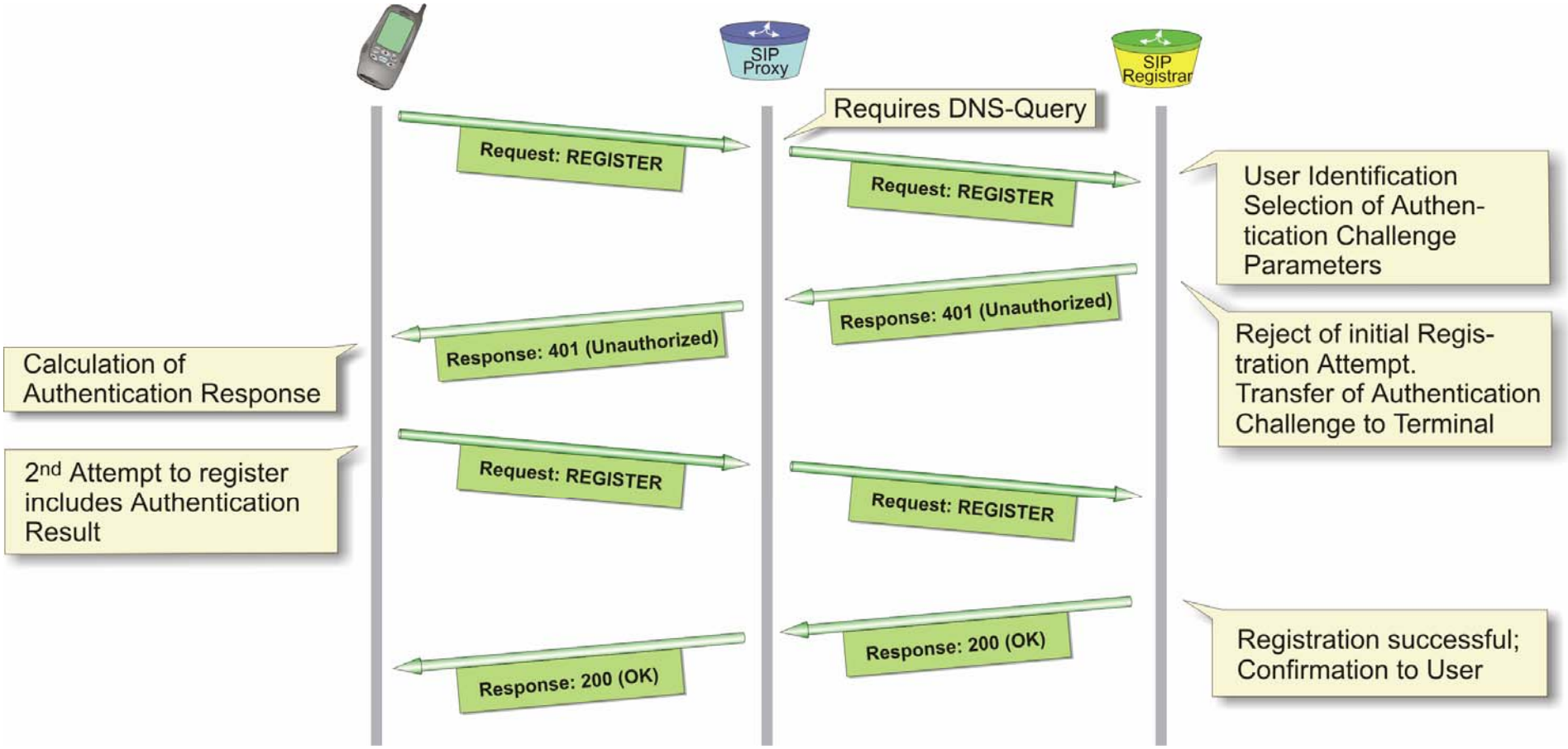


The SIP(S)-URI

- **SIP or SIPS**
  The SIP(s)-URI starts with the text string “SIP:” to indicate that the following identifiers relate to a SIP-URI or with the text string “SIPS:” to indicate that a SIPS-URI follows.
- **Userinfo**
  The presence of the element “userinfo” is optional but highly recommended to identify a user using his / her fully qualified domain name (e.g. baulbrause@cakao.net). The userinfo consists of either a username or a telephone number. Either element may be followed by an optional password which is separated from the userinfo through the character “:”. The use of a password in the Request-URI is not recommended by the specifications. In either case, the userinfo ends with the character “@”.
- **Hostport**
  The presence of the element “hostport” is mandatory. It consists of:
  ⇒ the element “host” which includes either an IP-address (in which case there would be no userinfo) or a host domain name (e.g. cakao.net). Note that the use of IP-addresses as host-ID is not recommended.
  ⇒ optionally a port number where this request shall be sent to at the receiver side (the default port number for SIP is ‘5060’$_{dec}$.
- **uri-parameters**
  The presence of uri-parameters is optional. If one or more uri-parameters are present, their listing is separated from the element “hostport” through the character “;” which is also used to separate different parameters from each other. Different parameters depend on each other. For instance, only when the maddr-parameter (server address of that user) identifies a multicast IP-address, then the “ttl”-parameter (time to live) shall be present.
  The “transport” parameter specifies the transport protocol to be used for SIP-messages relating to this request. The “lr”-parameter indicates that the sending UA uses loose source routing.

  [RFC 3261 (19.1)]

Registration: Message Flow



Registration: Message Flow

Initial Conditions

The SIP-UAC has been powered on or needs to re-register because the respective timer (set through “expires”-parameter in “Contact:”-header field has expired. The SIP-UAC knows the IP-address of the SIP-proxy (in 3GPP provided during PDP-context activation or through a SIP-DNS-operation). The SIP-UAC is unaware of whether it is currently logged into the home or a foreign network.

Applicability of this Procedure

This procedure is applicable for all successful SIP-registrations that include authentication.

Description

⇒ The SIP-UAC sends a REGISTER-request message to the IP-address of the SIP-proxy server. Among other elements, the REGISTER-request message identifies the pre-configured SIP-registrar of that UAC.

⇒ The SIP-proxy server needs to access a DNS-server to identify the IP-address of the SIP-registrar. After a successful response, the SIP-proxy server will forward the REGISTER-request message to the registrar.

⇒ The SIP-registrar rejects the registration attempt because of lacking authentication (⇔ Response: 401 (Unauthorized)). This response message will also include the challenge parameter for the SIP-UAC to respond to. In case of UMTS, this relates in particular to the inclusion of the RAND- and AUTN-parameters (base64-encoded).

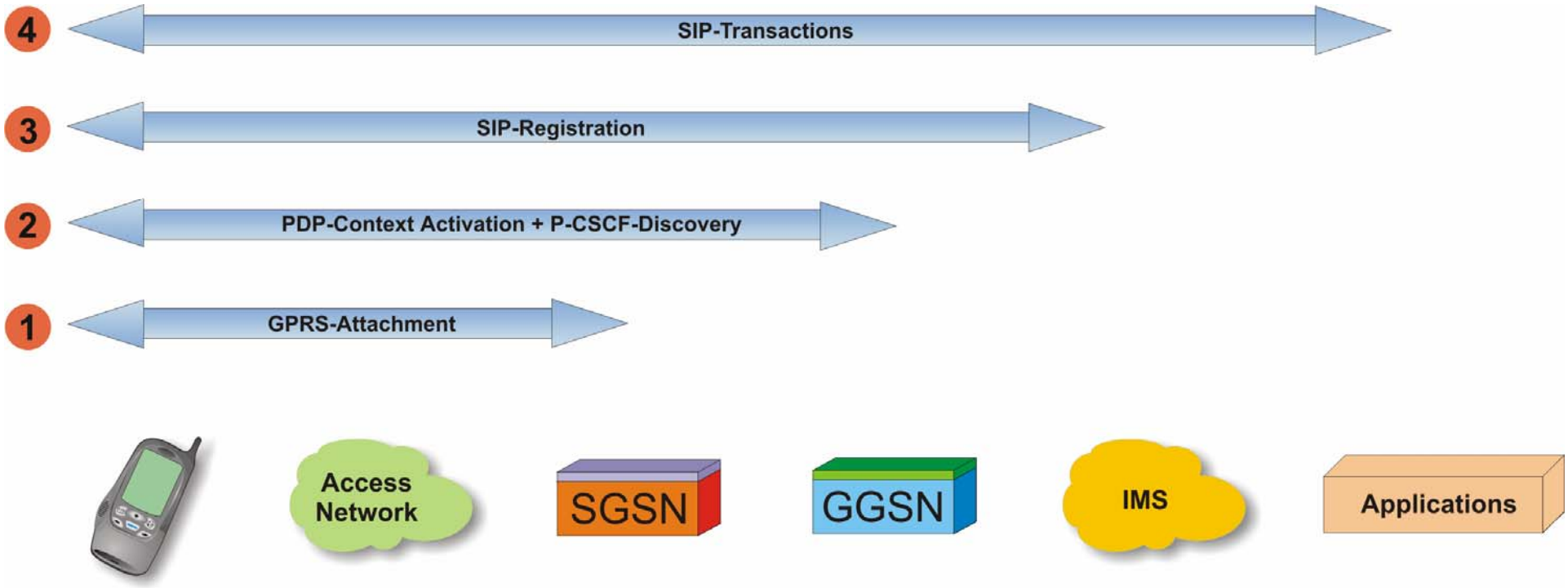
⇒ The “401 (Unauthorized)”-response message is forwarded to the SIP-UAC.

⇒ The SIP-UAC will use the challenge parameter to calculate the authentication response. In UMTS, this relates to the calculation of RES.

⇒ Consecutively, the SIP-UAC will send another REGISTER-request message to the SIP-proxy which includes the authentication result. In case of UMTS, this relates to the RES-parameter. The SIP-proxy will forward the second REGISTER-request message to the SIP-registrar which will check whether the authentication has been successful.

⇒ If yes, the SIP-registrar will send a “200 (Successful)”-response message via the SIP-proxy to the SIP-UAC which is now registered.

SIP-Procedure Preparation in 3GPP-Networks



SIP-Procedure Preparation in 3GPP-Networks

In 3GPP-networks, the IMS is interconnected to the UE / MS exclusively through the packet-switched domain. Consequently, the user needs to perform the respective internal registration procedure towards that domain before SIP-procedures can be performed:

1. The MS / UE shall attach to the GPRS using the regular GPRS-attachment procedure.
2. The MS / UE shall establish a PDP-context with best-effort QoS to be able to register to the IMS. Either during the PDP-context activation or immediately afterwards the UE / MS receives the address of the P-CSCF (Proxy Call Session Control Function) to send its SIP-message to.
3. The next step will be the registration towards the IMS-domain (or to be more accurate to the SIP-registrar which is called S-CSCF in 3GPP-networks).
4. Finally, SIP-transactions can be established using the packet-switched domain.

Note that SIP-sessions will usually require the activation of a secondary PDP-context to obtain the necessary QoS.