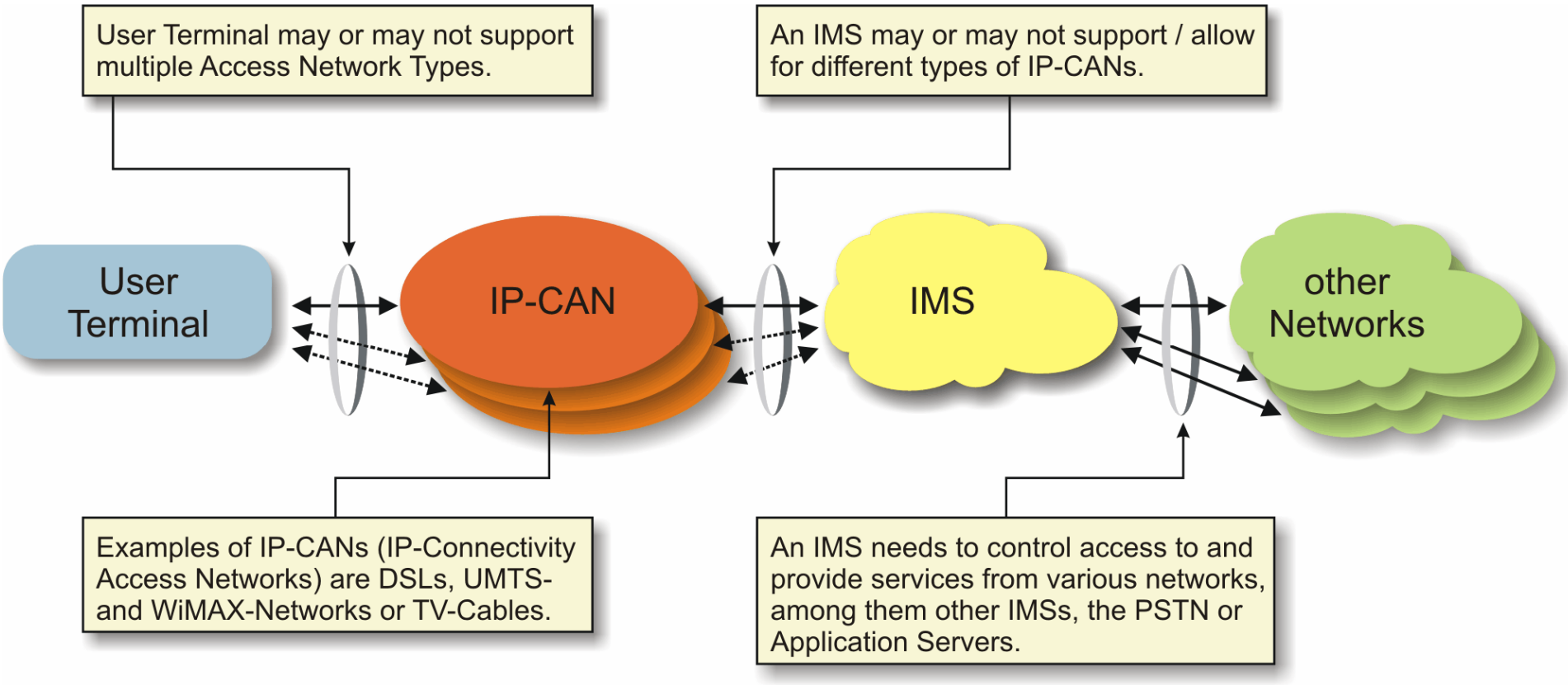
High Level View at the IMS and its Environment



High Level View at the IMS and its Environment

The figure illustrates how an IMS is embedded into the overall network architecture and infrastructure.

Mobility Issues

⇒ Two levels of mobility need to be distinguished: 1) The macro-mobility which allows a user to register from different IP-CAN's to the same IMS and 2) the micro-mobility which allows a user to roam within a given IP-CAN.

⇒ In that respect, micro-mobility is a function that resides in the IP-CAN and which is provided to the user by the IP-CAN. Obviously, the availability of micro-mobility depends on the type of access network. Inherently, only cable-free IP-CAN's will allow for micro-mobility.

⇒ On the other hand, macro-mobility may be provided by the IMS depending on operator preferences by allowing the user to access the IMS through different types of IP-CAN's.

Relationship to other Networks

⇒ The IMS takes on the role of a mediator between the user terminal on one side and the services on the other side. Services are provided by "other networks" which include heterogeneous networks like the PSTN, other IMS's or stand-alone application servers in an "Applications & Services" network cloud.

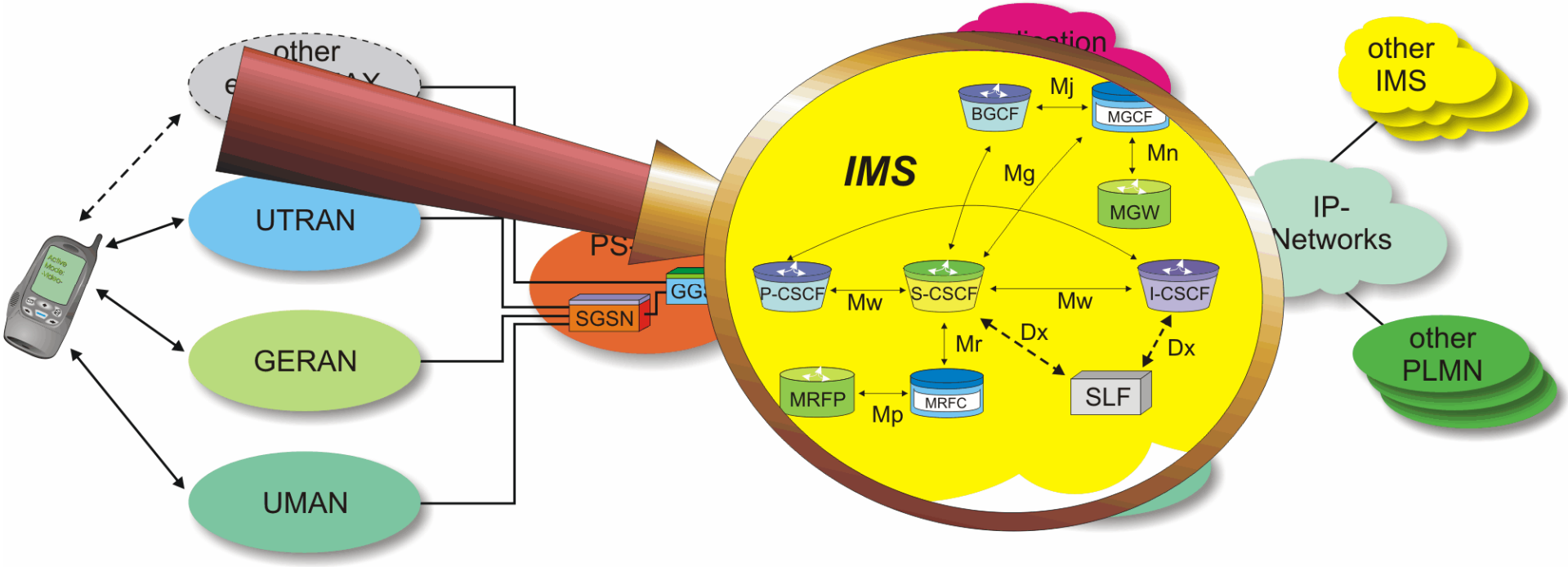
⇒ As illustrated in the figure, the IMS will mandatorily interconnect to various other networks but support for different access networks is optional.

User Terminals

⇒ The range of user terminals is wide and includes legacy analog telephones as well as multi-mode PDA's, supporting various different IP-CAN's. This depends on customer preferences and on the operator type (⇔ fixed line telephone companies need to slowly migrate their analog users to the new world and therefore they need to support these analog user terminals at least medium term).

⇒ In that respect, the type of user terminal also determines or limits the service types that a particular user is able to access. Obviously, an analog PSTN-terminal is unable to support video calls or instant messaging.

And what is inside the IMS?

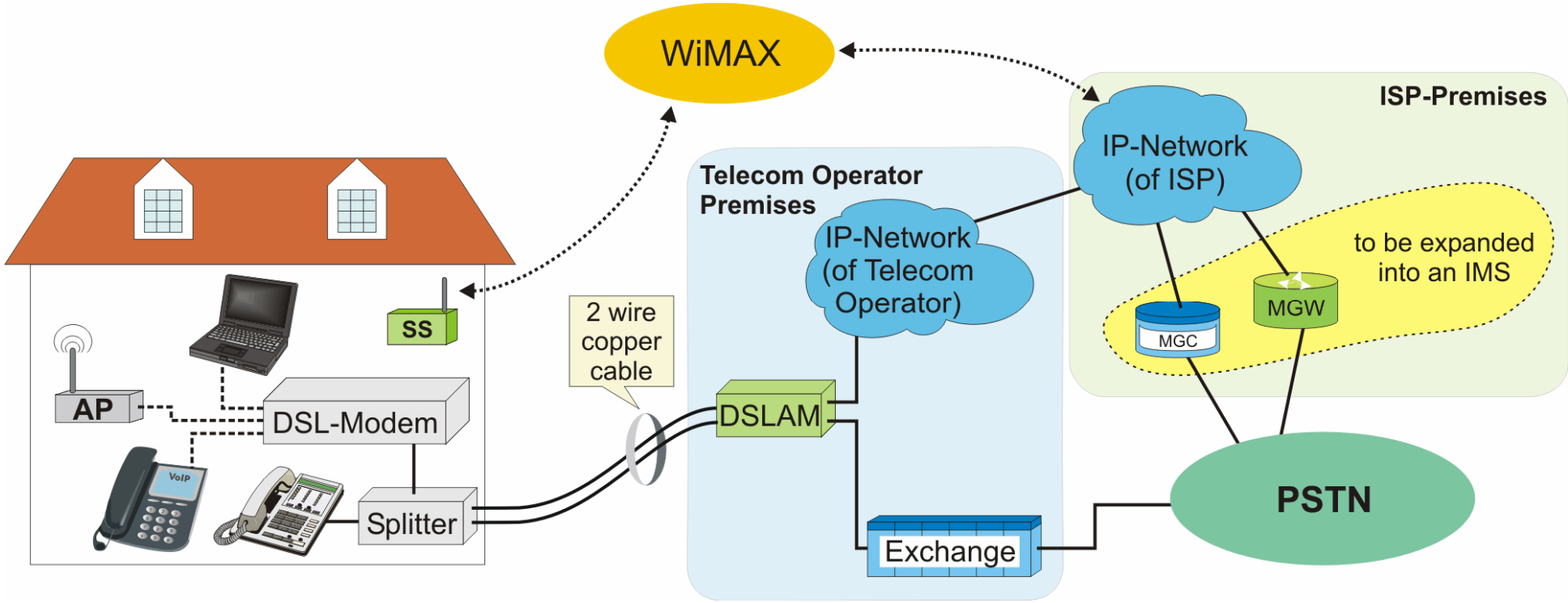


And what is inside the IMS?

The IMS is entirely based on IP as transport protocol. It therefore hosts IP-driven servers of which the majority uses SIP (Session Initiation Protocol) to communicate internally and externally. These servers are SIP-servers. Still, other protocols are also used within the IMS.
Other network elements within the IMS are media gateways.

[3GTS 23.228]

DSL Forum



DSL-Forum

We illustrate the DSL-forum's efforts mainly from the perspective of ISP's. Still, the DSL-forum's work also pertains to regular telecom operators who want to use their 2-wire copper cables for broadband service provision.

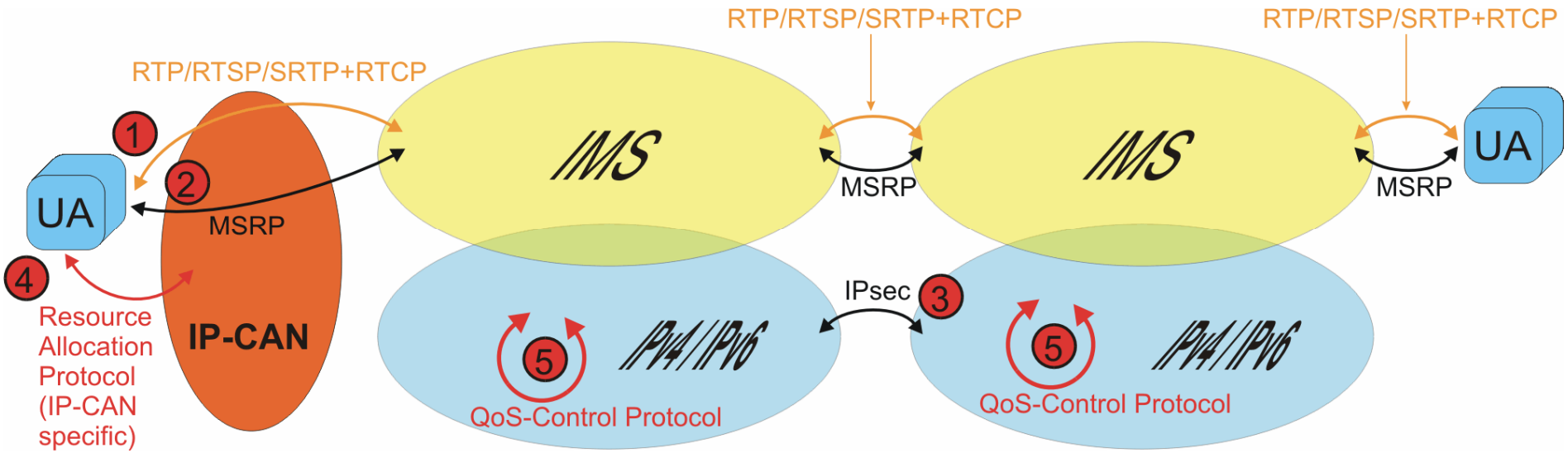
⇒ As the figure shows, the ISP's premises are usually behind the telecom operator's DSLAM and IP-network.

⇒ Similarly to cable TV operators, some ISP's already compete against the telecom operators for POTS and therefore these ISP's already have some soft switches available. These need to be expanded into an IMS within the next years, if the ISP wants to move on to become an integrated services network operator.

Note that we included an alternative WIMAX-based access network which is particularly interesting for the ISP as competitor of the telecom operator. Although apparently WIMAX has nothing to do with DSL it is frequently called "Wireless DSL" and it allows specifically the ISP to become independent from the telecom operator.

[http://www.dslforum.org/index.shtml]

Protocols within the IMS-User Plane



Protocols within the IMS-User Plane

RTP / RTSP / SRTP

All three protocols are there to convey user data. Neither protocol provides real-time capability by itself or by definition but they require some additional means like DiffServ, IntServ or MPLS to provide the necessary QoS. SRTP adds integrity protection and encryption to the capabilities of RTP.
[RFC 3550 (⇔ RTP), RFC 2326 ⇔ RTSP), RFC 3711 (⇔ SRTP)]

MSRP

The MSRP is used within the IMS for embedding IM-messages which can be differently encoded [draft-ietf-simple-message-sessions-XX]

Resource Allocation Protocols

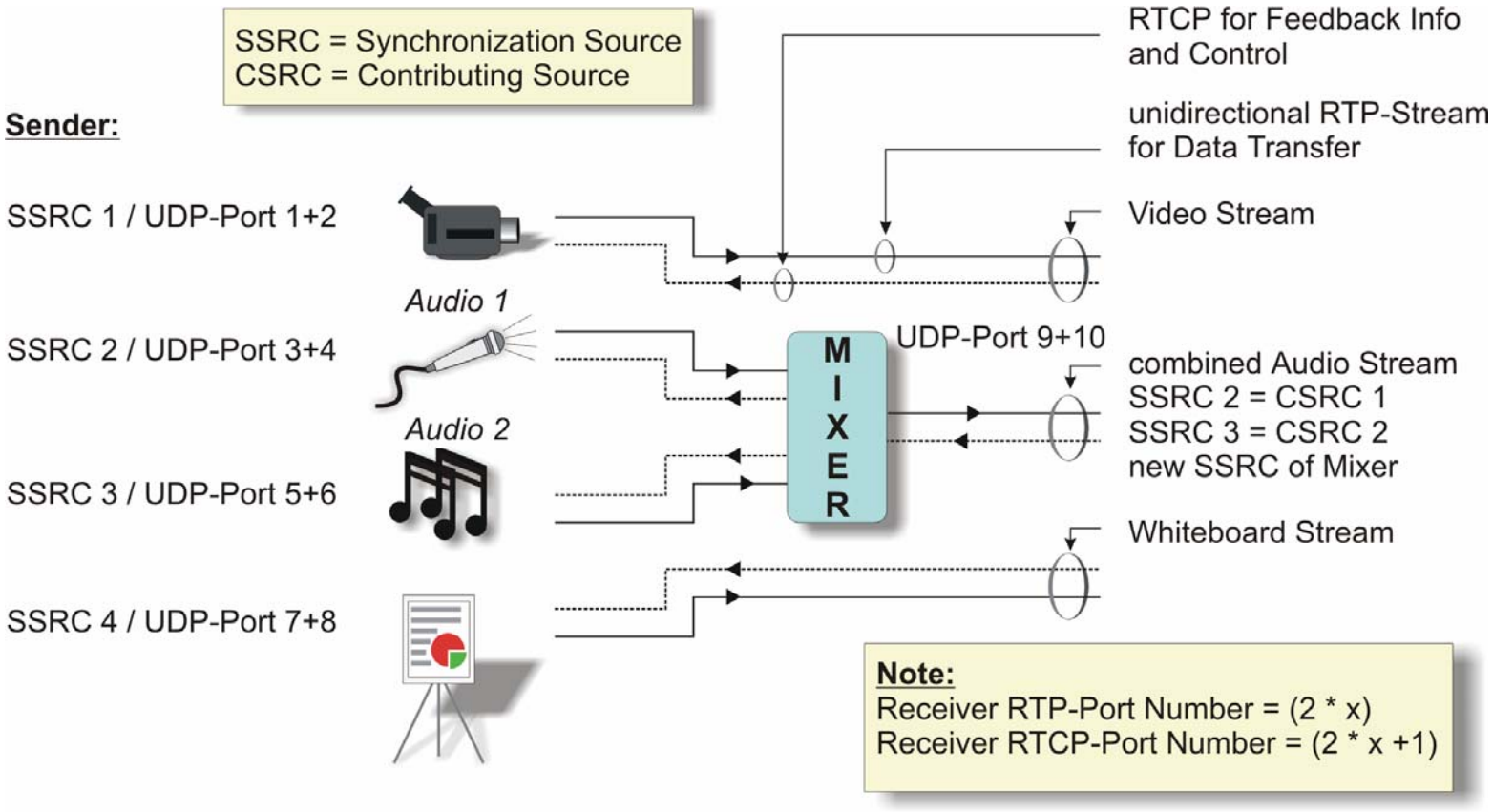
The bullet 4 relates to the IP-CAN specific resource allocation protocol which is used between UA and IP-CAN to obtain real-time QoS. In case of 3GPP-UTRAN / GERAN this would be session management.

QoS-Control Protocols

Different IP-transport networks may support different QoS-control protocols like IntServ, DiffServ or MPLS.

The Real Time Transport Protocol (RTP and RTCP)

- **Operation of RTP and RTCP**



The Real Time Transport Protocol (RTP and RTCP)

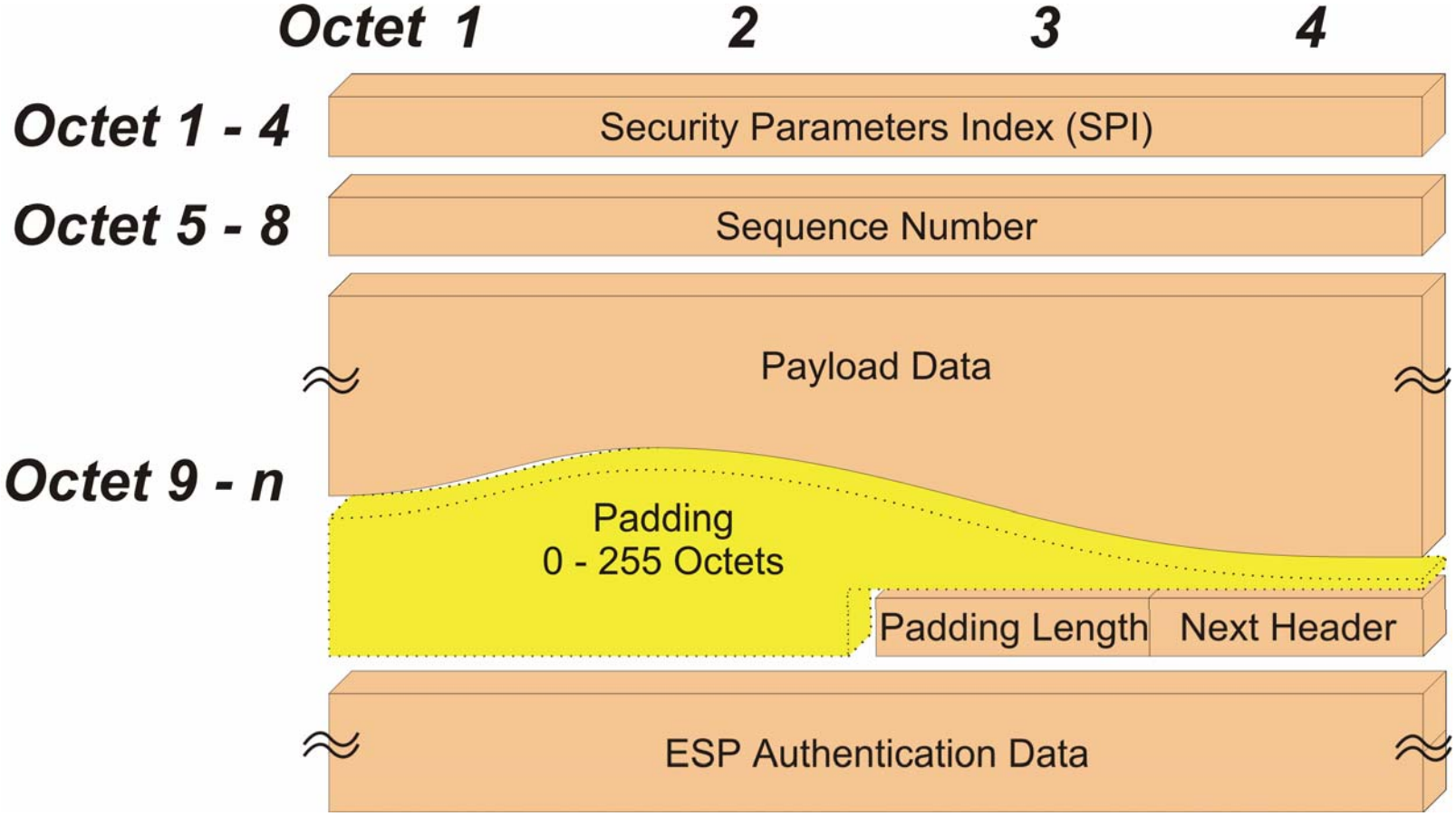
Operation of RTP and RTCP

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.

The IPsec Encapsulating Security Payload (ESP)



The IPsec Encapsulating Security Payload (ESP)

The ESP-header contains the following fields:

Security Parameters Index (SPI) (32 bit)

The Security Parameter Index is a pointer towards the Security Association (SA) that shall be used for that IP-frame. The SA has previously been negotiated upon setup of the IPsec-connection. Note that values 1 – 255_{dec} are reserved by ICANN and must not be used. SPI value '0' is only for local use and must also not be applied.

Sequence Number (32 bit)

The Sequence Number shall be incremented per IP-frame sent and shall disable any possibility for replaying. The Sequence Number is initialized to '00 00' upon IPsec-connection setup. Like for AH, the sequence number is not allowed to outrun its modulo.

Payload Data (n bit)

The original IP-frame

Padding (0 – 255 octets)

Padding may be applied for the following reasons: 1. to suit the requirements of a given encryption algorithm 2. to pad to a 4 octet boundary 3. to conceal the length of the actual payload. If not demanded else by the encryption algorithm, the padding octets shall be encoded with '1', '2', 3', '4', ….

Padding Length (8 bit)

Padding Length identifies, how many octets have been added for padding.

Next Header (8 bit)

Is actually the *Protocol*-field of the original IP-frame that was encrypted (and authenticated). Example: if a TCP-frame is included in the IP-frame, the Next Header = '6'.

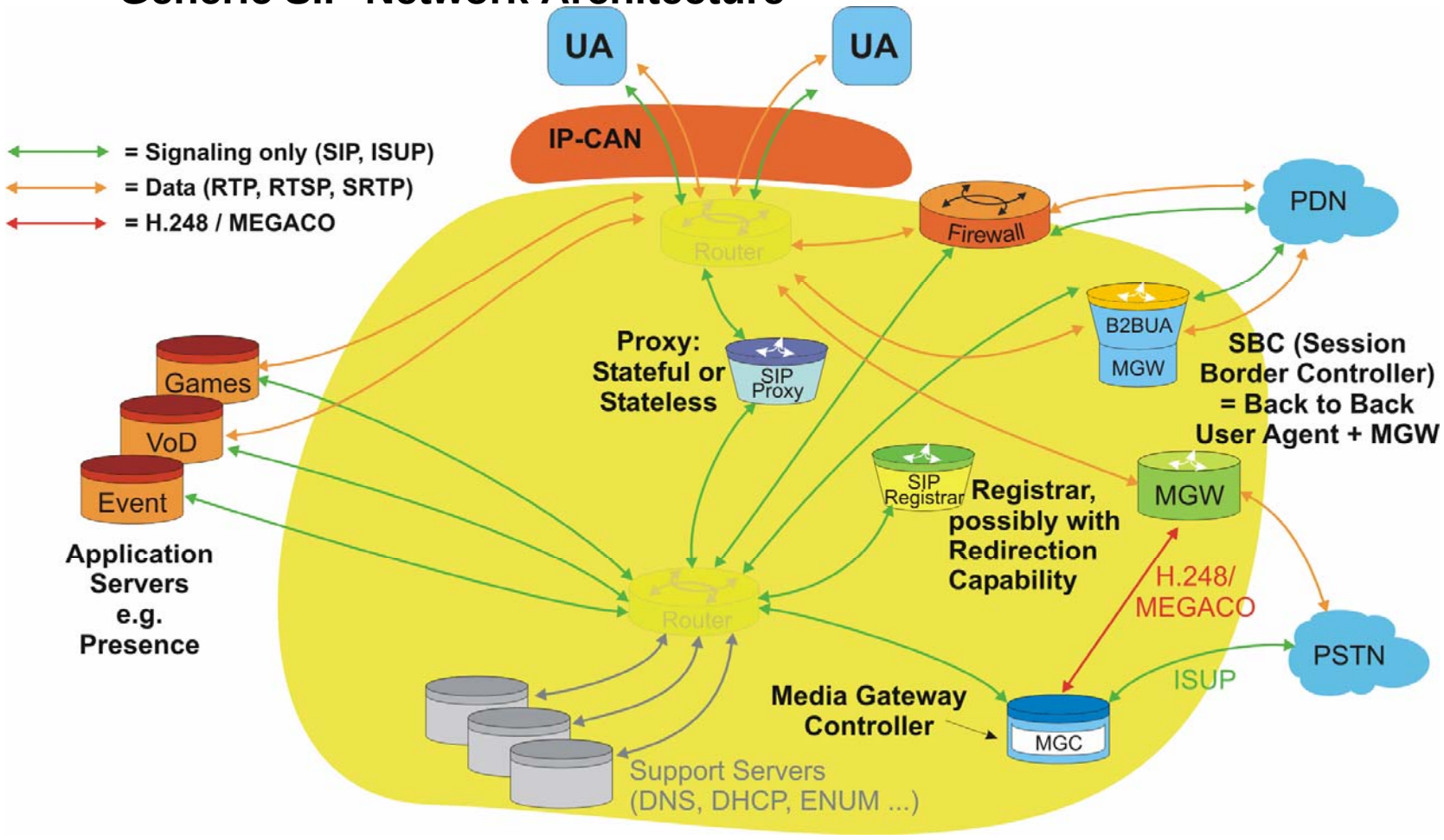
ESP Authentication Data (n bit)

The Authentication data will contain the Integrity Check Value (ICV) that validates and authenticates the IP-frame, using the authentication algorithm that has been negotiated for ESP. Padding may be applicable and shall be supported by all implementations.

[RFC 2406]

Basic Architectural Elements

- **Generic SIP-Network Architecture**



Basic Architectural Elements

Generic SIP-Network Architecture

The figure illustrates a typical network which uses SIP for session management functions. Please note the following information:

⇒ We specifically distinguished between the orange colored lines that the data take on one hand and the green colored lines for SIP-signaling messages. To be more precise: With the exception of an SBC (Session Border Controller), no SIP-proxy server or MGC deals with the data themselves.

⇒ The implicated *physical* network architecture with the two (SIP-independent) standard routers is an arbitrary possibility to interconnect the various network elements. We only added these routers to avoid irritations. Still, our focus is the logical network architecture. That's why we hid these routers behind shades.

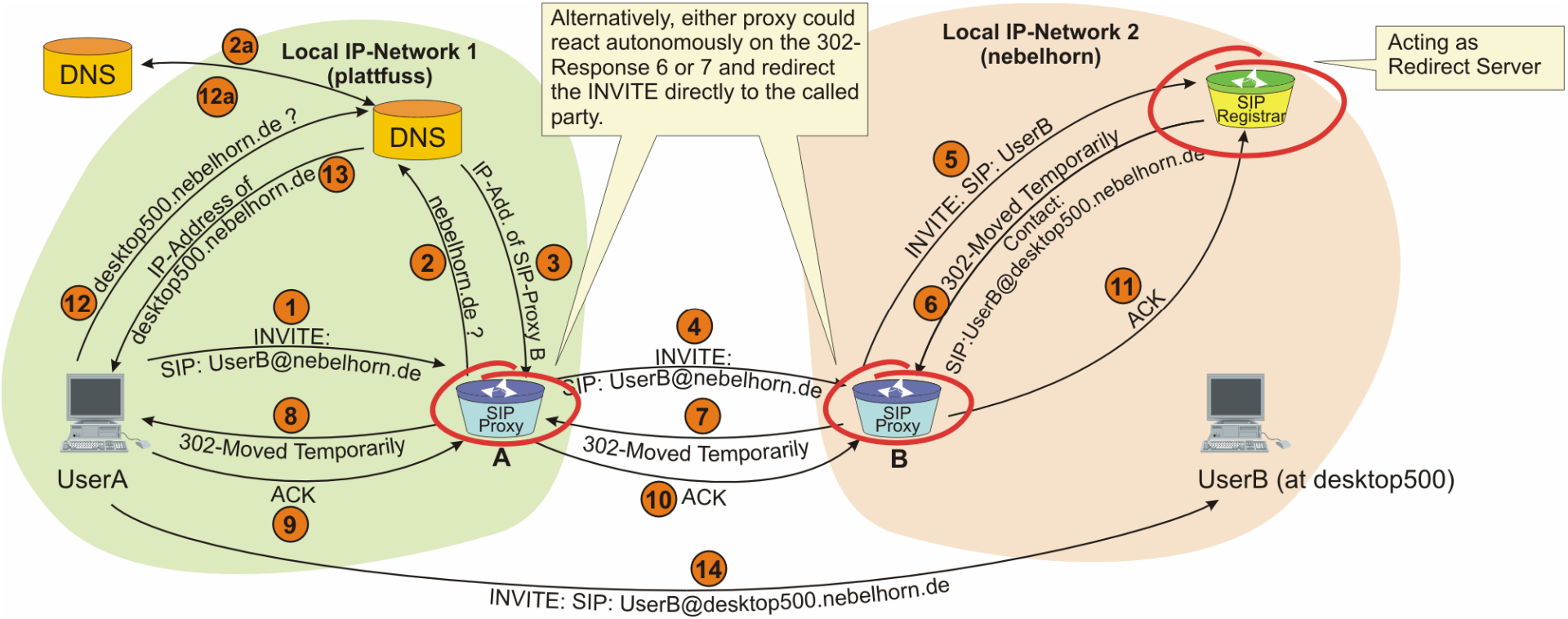
⇒ Sessions towards the PSTN require by default the interaction of soft switches (⇔ MGC + MGW).

⇒ Sessions towards external PDN's or to the internet will either traverse a firewall or they will traverse an SBC.

???? Question Section 3 ????

⇒ Why are there SIP-proxies to relay SIP-messages? Why is this task not taken care of by simple IP-routers?

Operation of Redirect Servers



Operation of Redirect Servers

Introduction

A redirect server is a SIP-proxy server or a SIP-registrar that responds an incoming Request: INVITE with a Response: 3XX (e.g. 302-"Moved Temporarily"). This response includes in the "Contact:"-header field the one or more current user device's addresses that shall be contacted instead or directly by the originating party.

Procedure Description

As illustrated in the figure, UserA sends an INVITE / sip: UserB@nebelhorn.de to its SIP-proxy server A (⇔ message 1). The proxy server invokes the help of one or more DNS-servers to resolve the IP-address of nebelhorn.de (⇔ message 2, 2a and 3). Consequently, proxy A relays the INVITE-message to SIP-proxy server B (⇔ message 4).

⇒ Proxy B sends the INVITE-message to the responsible SIP-registrar (⇔ message 5).

⇒ The SIP-registrar will issue a final Response: 302-"Moved Temporarily" which traverses all the way back to UserA (⇔ message 6, 7, 8) and which ultimately is the message that will trigger the redirection of the request.

Note the comment on the graphics slide: Either SIP-proxy server could react on the Response: 302-Moved Temporarily autonomously and redirect the Request: INVITE to its new destination directly.

⇒ As mentioned before, this response message type always carries in its "Contact:"-header field the current IP-address or FQDN where the requested part can be found. In our case, this is the fully qualified domain name desktop500.nebelhorn.de.

⇒ Before another INVITE to UserB at desktop500.nebelhorn.de can be sent, UserA needs to finish the previous INVITE-transaction by issuing a Request: ACK and sending it to the registrar in local IP-network 2 (⇔ message 9, 10, 11).

⇒ To be able to send an INVITE-message to sip: UserB@desktop500.nebelhorn.de, UserA invokes the support of one ore more DNS-servers to resolve the FQDN into an IP-address (⇔ messages 12, 12a and 13).

⇒ Finally, UserA can send a Request: INVITE / sip: UserB@desktop500.nebelhorn.de directly to UserB. *No SIP-proxy server is used* (⇔ message 14).

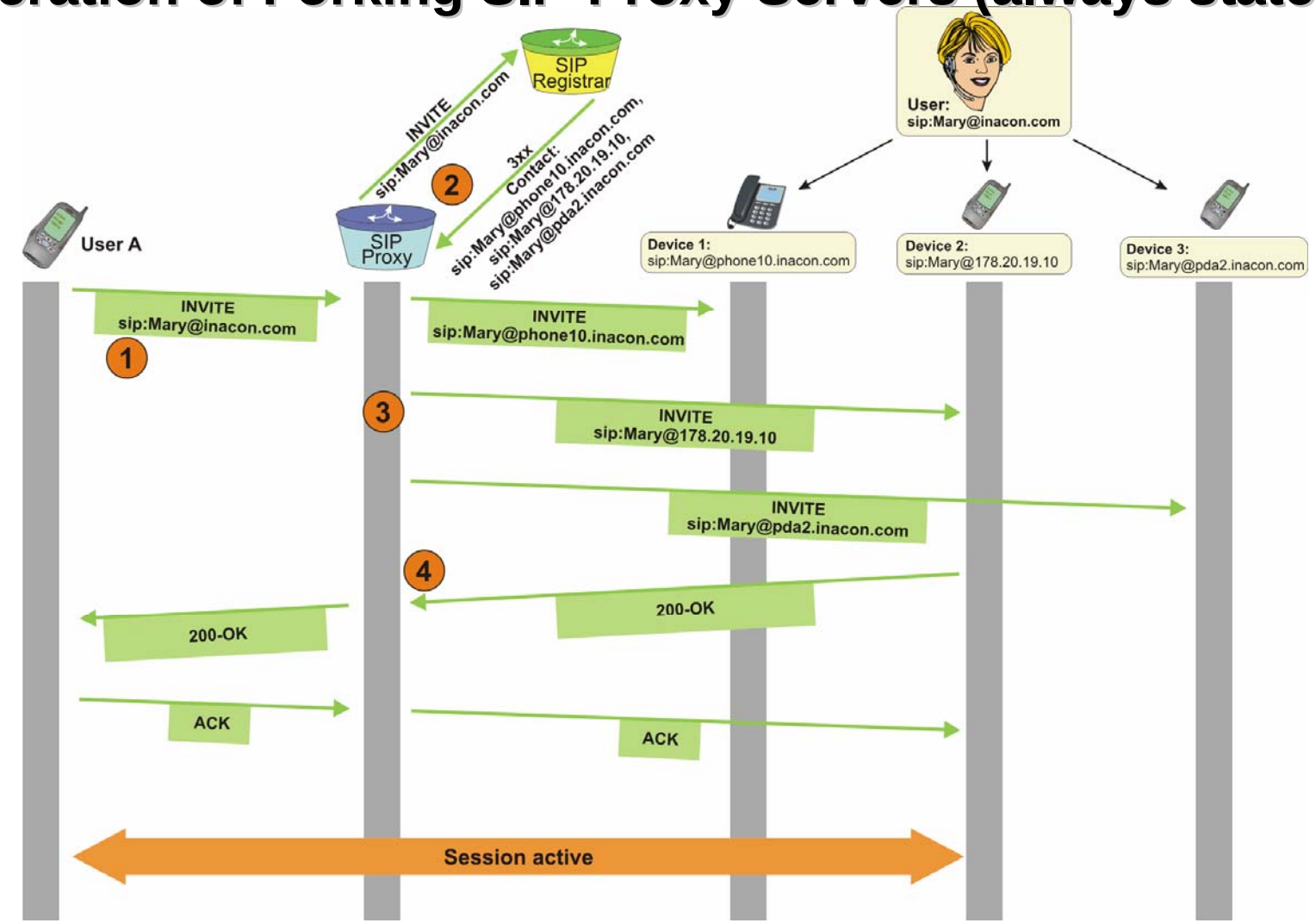
Redirection is well suited to reduce the load of SIP-proxies but it is not well suited for carrier based services which require at least a SIP-proxy server for charging purposes.

[RFC 3261 (8.3)]

??? Question Section 4 ???

⇒ If either SIP-proxy would autonomously redirect the INVITE to its new destination, would this be a stateful or a stateless proxy server or both?

(1) Operation of Forking SIP-Proxy Servers (always stateful)



(1) Operation of Forking SIP-Proxy Servers (always stateful)

⇒ The figure illustrates a case in which the forking proxy is not the registrar but it is an intermediate SIP-proxy server that receives a Response: 3XX for a Request: INVITE and autonomously redirects the Request: INVITE to all contact-addresses received within the Response: 3XX.

⇒ Of course, the registrar could have done the same in which case the registrar would be the forking proxy. This is an implementation issue.

- **Bullet 1:**
  As illustrated, the call starts with User A sending a Request: INVITE to its proxy server. The invitation is for Mary with the SIP-URI sip: Mary@inacon.com. This SIP-URI is the only information that User A has about Mary.

- **Bullet 2:**
  The SIP-proxy relays the Request: INVITE towards the registrar of Mary and receives back a Response: 3XX (e.g. Response 302-Moved Temporarilywhich includes a list of addresses to be contacted instead. This list of addresses is 1. sip: Mary@phone10.inacon.com, 2. sip: Mary@178.20.19.10 and 3. sip: Mary@pda2.inacon.com. Note that Mary's registrar operates as redirect server.

- **Bullet 3:**
  The proxy server uses the received contact information and sends out the invitation again but this time to all three received contact addresses / devices simultaneously. Consequentially, all three devices will start ringing more or less simultaneously.

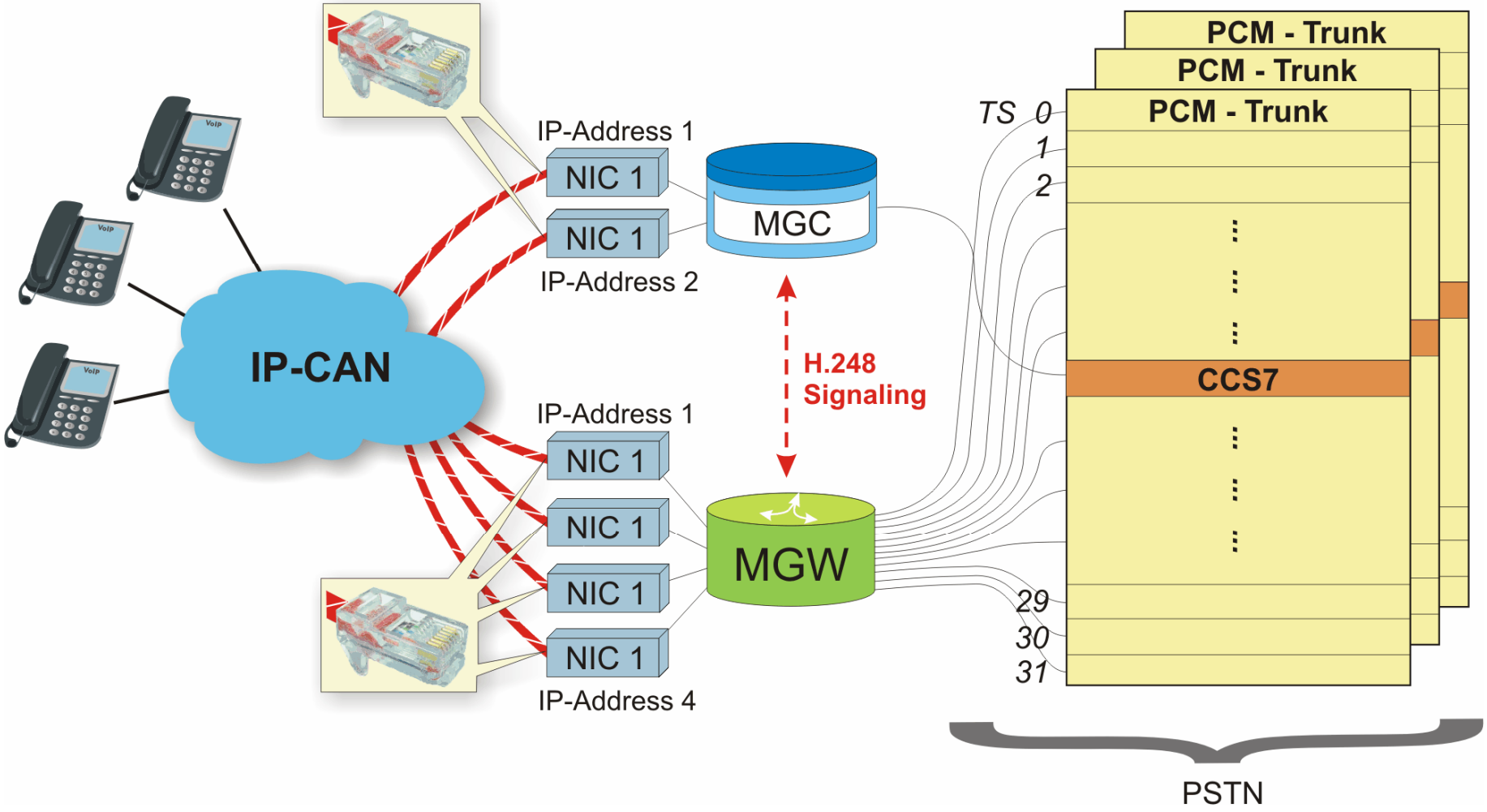
  This process of trying to reach the called user on different devices at the same time is called simultaneous forking.

- **Bullet 4:**
  Device 2 is the only one or the earliest one to send a Response: 200-OK to the forking proxy server. As illustrated, this successful response is relayed towards User A immediately. Accordingly, User A sends a Request: ACK through the forking proxy towards Mary's device 2.

  The session is active between User A and Mary's device 2 and media data are exchanged. However, since forking occurred, the still open INVITE-transactions towards device 1 and device 3 need to be handled. Please refer to the next page.

  [RFC 3261 (16.7.10)]

Soft Switches and their Controllers



Soft Switches and their Controllers

⇒ Definitely very important components of next generation networks are the media gateways which are frequently also called “soft switches”.

⇒ As the figure illustrates, media gateways are controlled by a media gateway controller (MGC). In the 3GPP-terminology this MGC becomes the MGCF (Media Gateway Control Function).

Media gateways and media gateway controllers are required to interface IP-based communication towards the legacy PSTN.

⇒ The control function between MGC and MGW is performed through a protocol called H.248 (⇔ ITU-T terminology) or MEGACO (⇔ IETF-terminology / RFC 3015).

⇒ The H.248 / MEGACO protocol allows for the seizure and release of resources that are controlled by the media gateway. This also relates to the control and conversion of codec types (AMR, PCM a-law, PCM µ-law, …).

⇒ Accordingly, the MGC terminates the call control signaling information from both sides: The SS7-signaling (ISUP) from the PSTN as well as the IP-based call or session control information (usually H.323 or SIP).

⇒ On the other hand, the MGW terminates all PCM-links (⇔ timeslots on the different PCM-trunks) and it is able to interconnect these PCM-links to packet-switched resources on the IP-network side (usually identified through the combination of Source IP-Address / Source UDP-Port Number and Destination IP-Address / Destination UDP-Port Number).

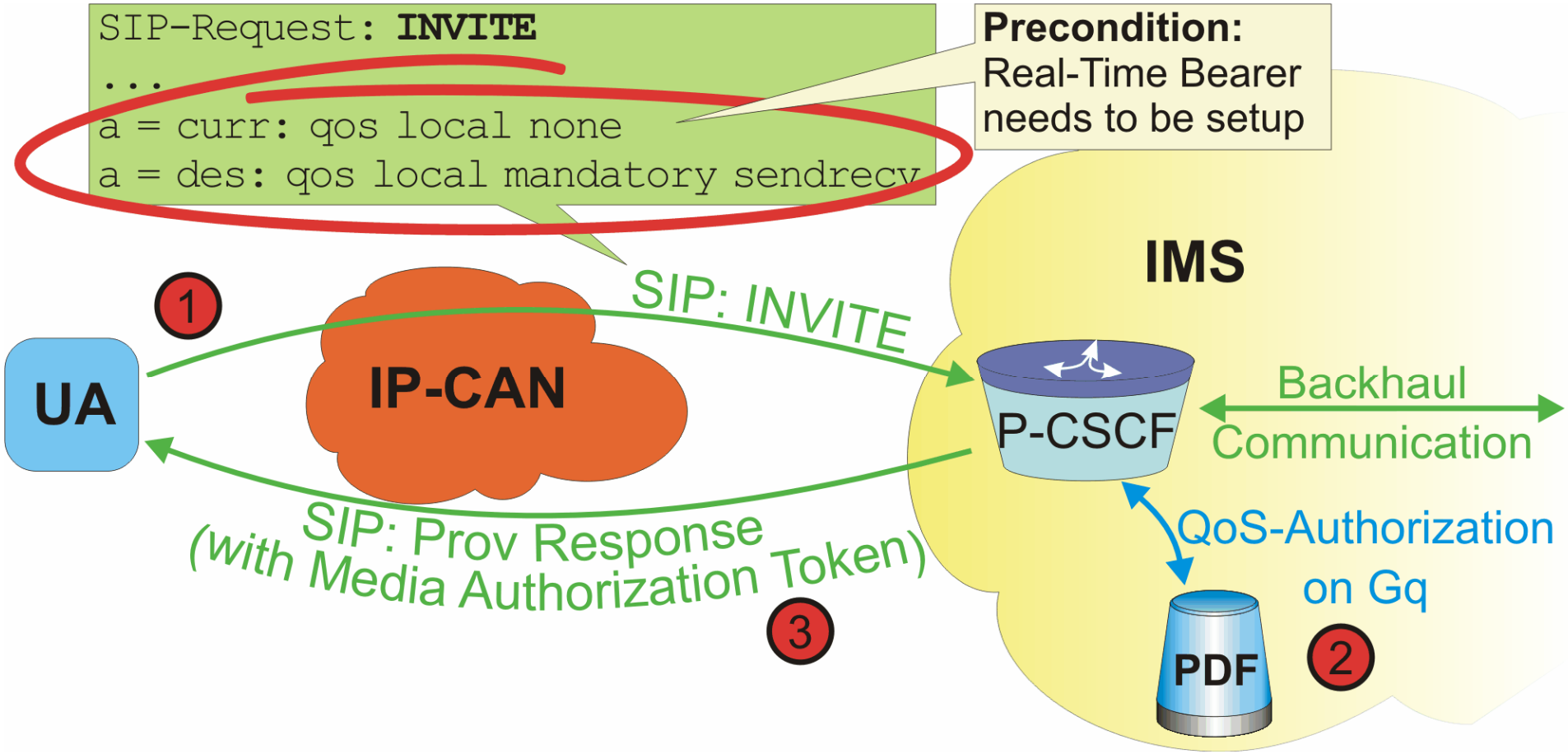
⇒ As the figure illustrates, media gateways and media gateway controllers usually are interconnected to the IP-network through more than one NIC (Network Interface Card) which means through more than one IP-address. This is done for load balancing and congestion control.

The figure tries to indicate what makes soft switches so appealing to network operators:

- They are connected to the IP-network simply through standard IP-network cables (e.g. RJ-45). No error-prone patch panel wiring is necessary.
- They usually do not require sophisticated configuration but they use some auto configuration to obtain IP-addresses etc..
- They usually have a smaller footprint than their predecessors, the public exchanges of the SS7-world.

P-CSCF Involvement for Session Setup and Policing

- **Step 1: Session Start**



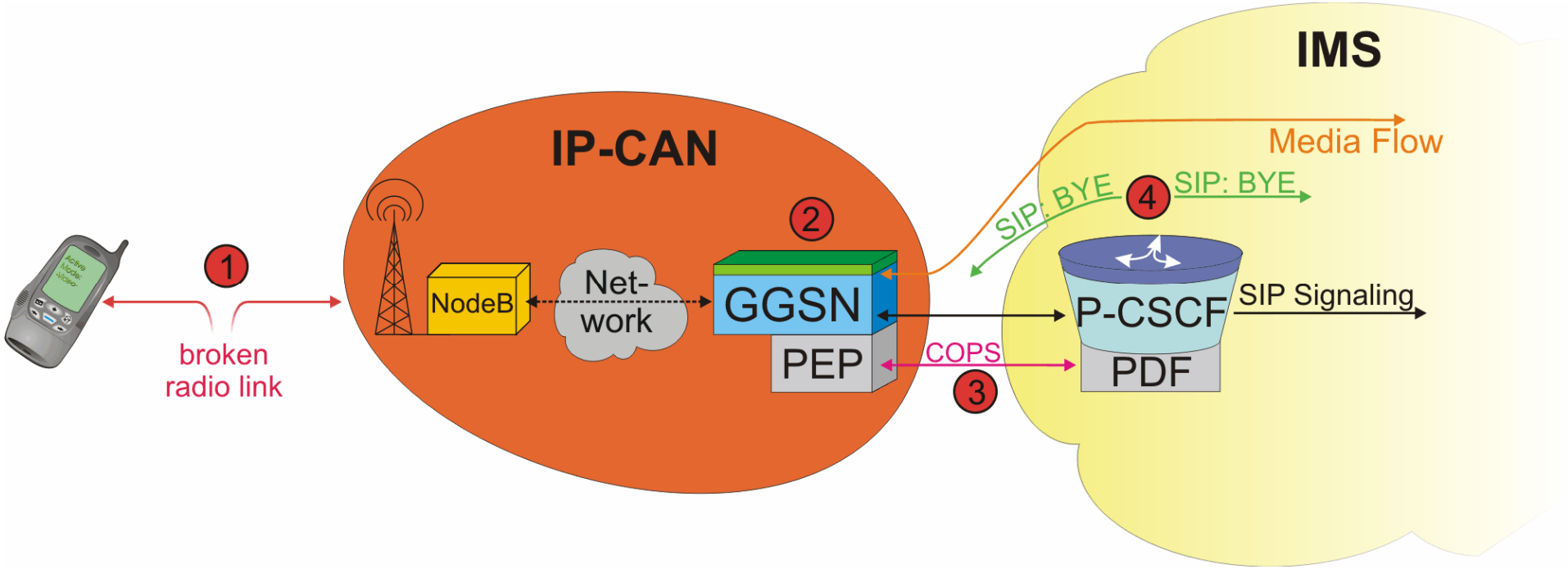
P-CSCF Involvement for Session Setup and Policing

Step 1: Session Start

It is the task of the P-CSCF to authorize the use of resources within an IP-CAN. That is, the P-CSCF will upon reception of a SIP: INVITE-message from the UA (bullet 1) communicate the session establishment request onwards (⇔ backhaul communication) but it will also communicate, possibly internally, with the PDF (Policy Decision Function).
The PDF will allocate a media authorization token (bullet 2) and provide it to the P-CSCF. This media authorization token is sent back (bullet 3) to the UA through the next possible SIP-message (most likely through a provisional response message).

[3GTS 29.209 (4, 5)]

P-CSCF Involvement during Ungraceful Session Release

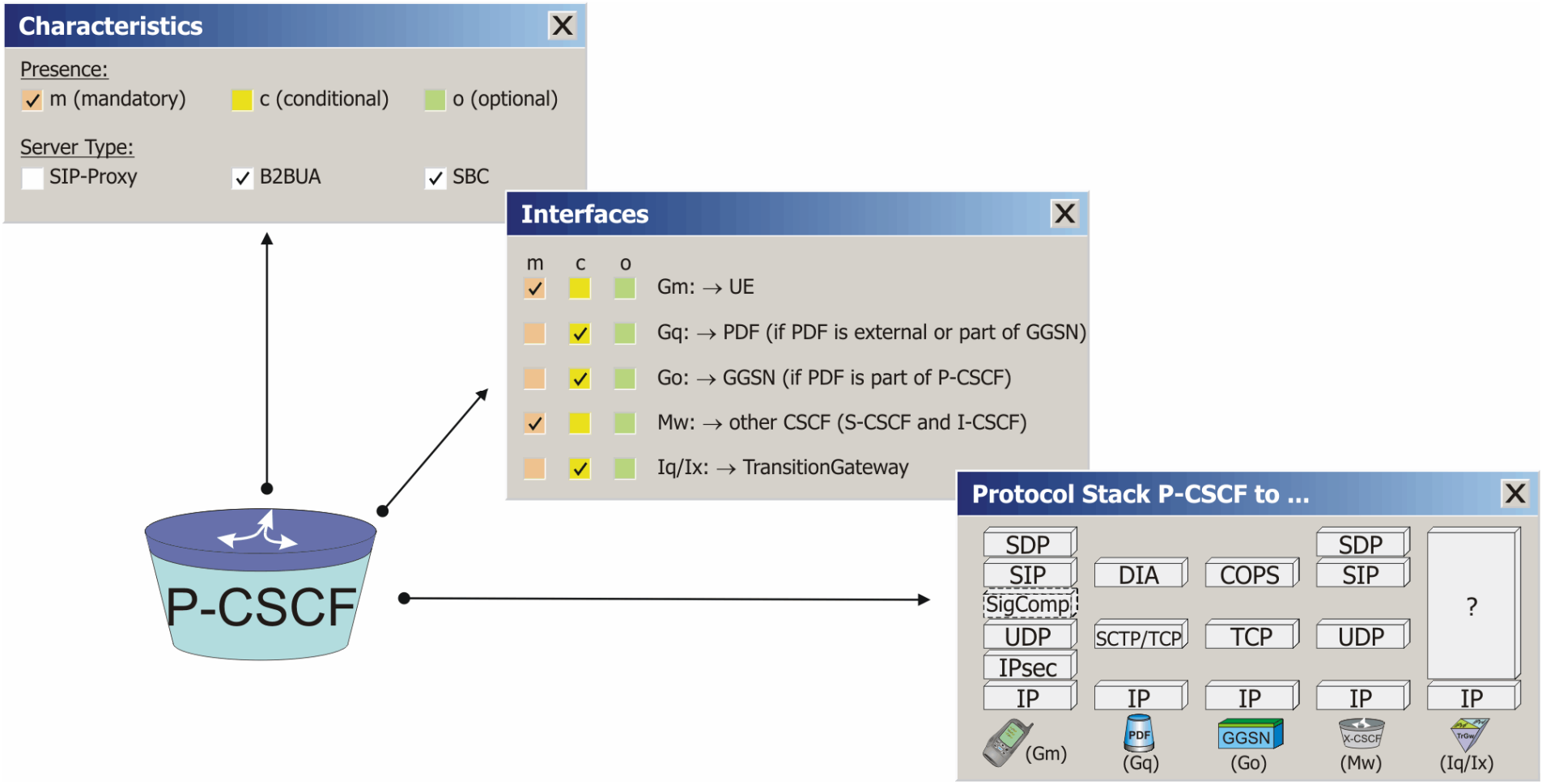


P-CSCF Involvement during Ungraceful Session Release

- ⇒ The P-CSCF plays an important role during ungraceful session release. Of course, there are multiple reasons that may lead to an ungraceful session release of which the P-CSCF only takes care of a few. They are mainly related to loss of radio coverage (call drop / broken radio link).
- ⇒ In such a case, the entire procedure is triggered by a broken radio link (1). The issue is that by definition the P-CSCF will not be able to recognize such a broken radio link because the media channels remain transparent to the P-CSCF.
- ⇒ However, the GGSN is well suited to recognize the lack of data packets (2). This function is based on some packet inspection function inside the GGSN which is initialized for certain IP-address / port number associations. It can be sophisticated enough to evaluate the quality of the data packets but this is not very likely.
- ⇒ When the GGSN detects that there are no more packets arriving in upstream direction for a certain period of time it may internally communicate towards the PEP that the radio link is interrupted and that there won't be any recovery.
- ⇒ Consequentially (3), the PEP will use COPS and the DRQ-message (Delete Request State) to tell the PDF that the media tunnel is broken. Accordingly, the PDF will internally or externally communicate with the P-CSCF that the media authorization has been withdrawn and …
- ⇒ Since the P-CSCF is a B2BUA, it will disconnect the session by sending a Request: BYE-message to both peers.

[3GTS 29.207, 3GTS 29.208, RFC 2748]

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.

Interfaces to other Network Elements

The Gm-interface to the UE 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

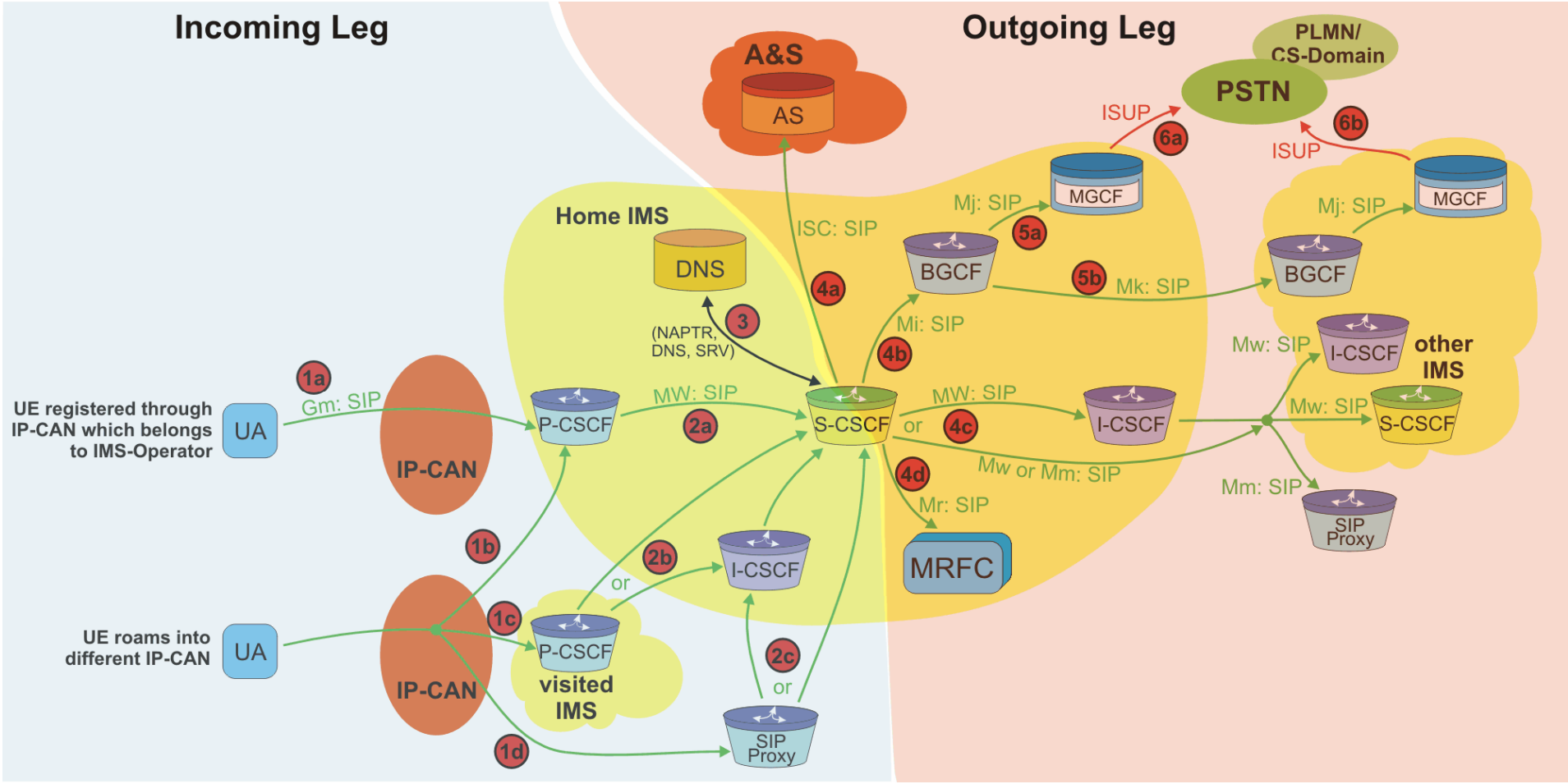
Please note that on the Gm-interface there is a new compression layer between UDP and SIP. All SIP-based interfaces only show UDP as transport protocol but specifically TCP is also possible.

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)]

Typical Use Cases of the S-CSCF

- **Involvement during IMS-Originating Transactions**



Typical Use Cases of the S-CSCF

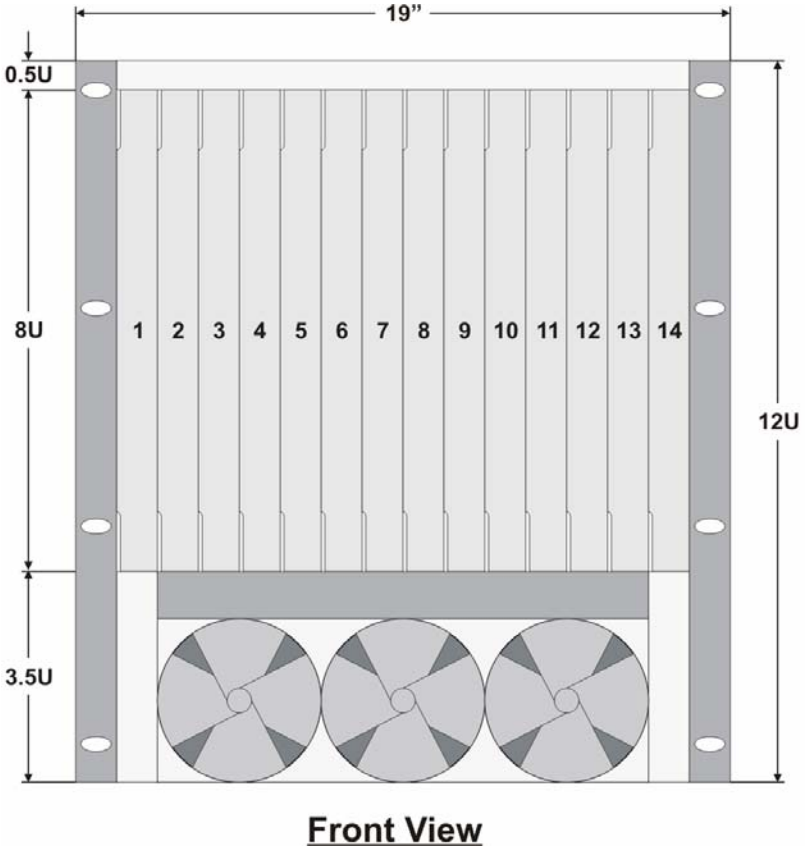
Involvement during IMS-Originating Transactions

How a UA-originating SIP-request message reaches the S-CSCF, obviously depends on the point of attachment of that UA. Note that for a better distinction we used two different colors to identify S-CSCF incoming and outgoing SIP-messages.

- ⇒ The UA may access the S-CSCF through a P-CSCF which is part of the same IMS to which the S-CSCF belongs to (bullet 1a). In this case, the IP-CAN may still belong to another operator (bullet 1b). Alternatively, the UA is using the P-CSCF of another IMS-operator which also means that the IP-CAN belongs to another operator (bullet 1c). And finally, the UA may use a generic SIP-proxy which belongs to or is related to the operator who owns the IP-CAN through which the UA registered previously (bullet 1d).
- ⇒ If the own P-CSCF is used then there will be no I-CSCF between the P-CSCF and the S-CSCF (bullet 2a). Note that at registration, the P-CSCF stores the address of the S-CSCF to be able to route UA-originating SIP-requests. However, an I-CSCF may be in the loop in case a foreign P-CSCF or generic SIP-Proxy is used (bullets 2b and 2c). This is an operator decision.
- ⇒ The S-CSCF may need to involve the help of a DNS-server (bullet 3) to be able to take the next hop decision to bring the SIP-request message closer to its destination. And definitely the S-CSCF will invoke service access control functions to check whether the UA is legitimate in the first place to send this SIP-request message.
- ⇒ Bullet 4a relates to the case that the UA wants to send a SIP-request message (e.g. SUBSCRIBE, PUBLISH, INVITE) to an Application Server.
- ⇒ Bullet 4b, 5a and b and 6a and b relate to the case of audio call establishment request towards a destination within the PSTN or within the circuit-switched domain of another PLMN (the MS-ISDN of a mobile phone was dialed and the DNS-query in bullet 3 resulted in a target within the PSTN / PLMN). The distinction between 5a / 5b and 6a / 6b is the geographic point of breakout into the PSTN / towards the second PLMN.
- ⇒ Bullet 4c relates to the case when the targeted SIP-URI resolves into another IMS or towards an outside generic SIP-proxy. It depends on operator policy (⇔ e.g. topology hiding y/n) whether the S-CSCF needs to route the SIP-request message through an I-CSCF or not.
- ⇒ Bullet 4d relates to the UA wanting to setup a conference session or accessing a chat room for instant messaging. In such case, the UA will send a Request: INVITE-message towards the MRFC.

Subrack Layout

- **Logical View:**



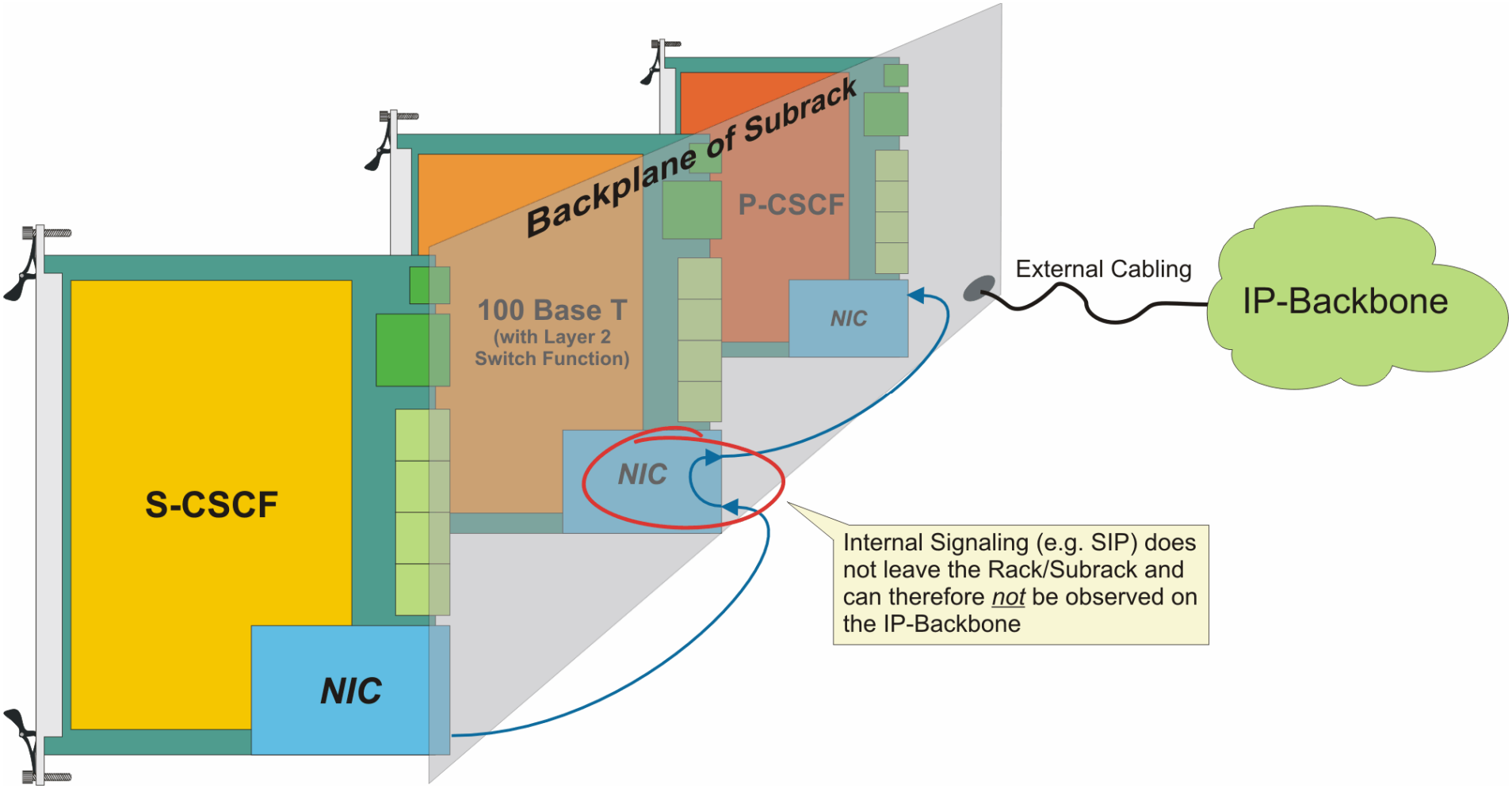
Front View

Physical View:

Subrack Layout

- ⇒ The figures illustrate the layout of an ATCA subrack. Note that this layout is standardized and defines, among many things, the mechanical characteristics of pins and cards.
- ⇒ Different options for the number of cards per subrack exist, in the presented one 14 cards can be inserted in one subrack. Underneath the cards, there are fans to provide for air conditioning.
- ⇒ Note that subracks can be mounted on top of each other to form larger nodes.

How does the Communication between Cards occur?

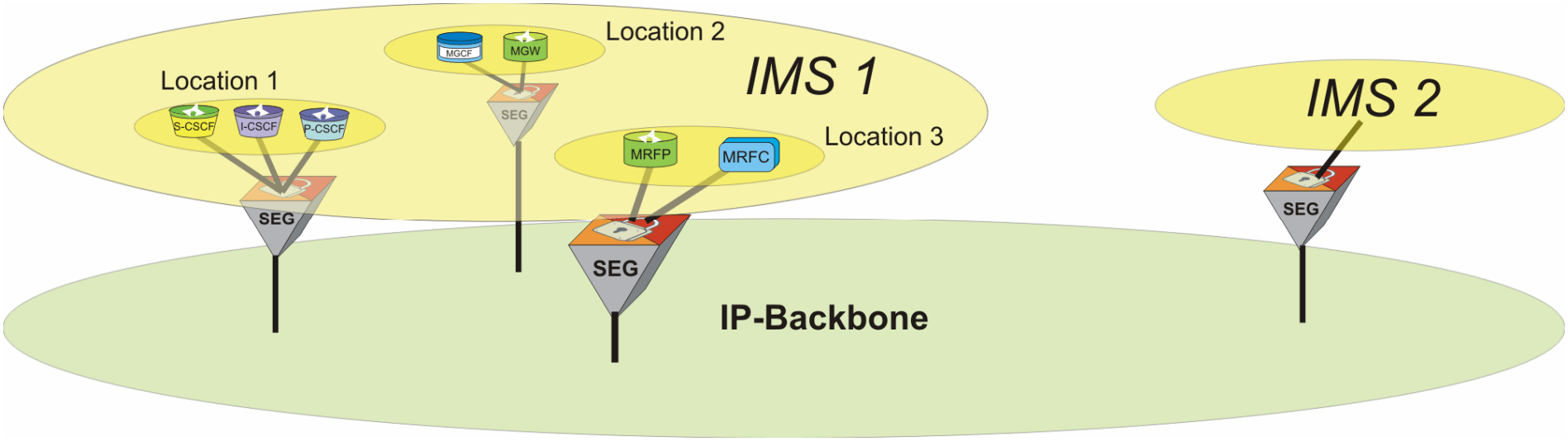


How does the Communication between Cards occur?

Since a single subrack may contain multiple logical network nodes, it would be a waste of resources to route each IP-frame between two boards of the same subrack through the IP-backbone network.
Accordingly, signaling between the different boards within a single rack or subrack is identified at the line card (e.g 100-BaseT) and sent to its destination without involving the backbone.

Consequentially, these IP-frames remain invisible on the IP-backbone

The General Security Architecture

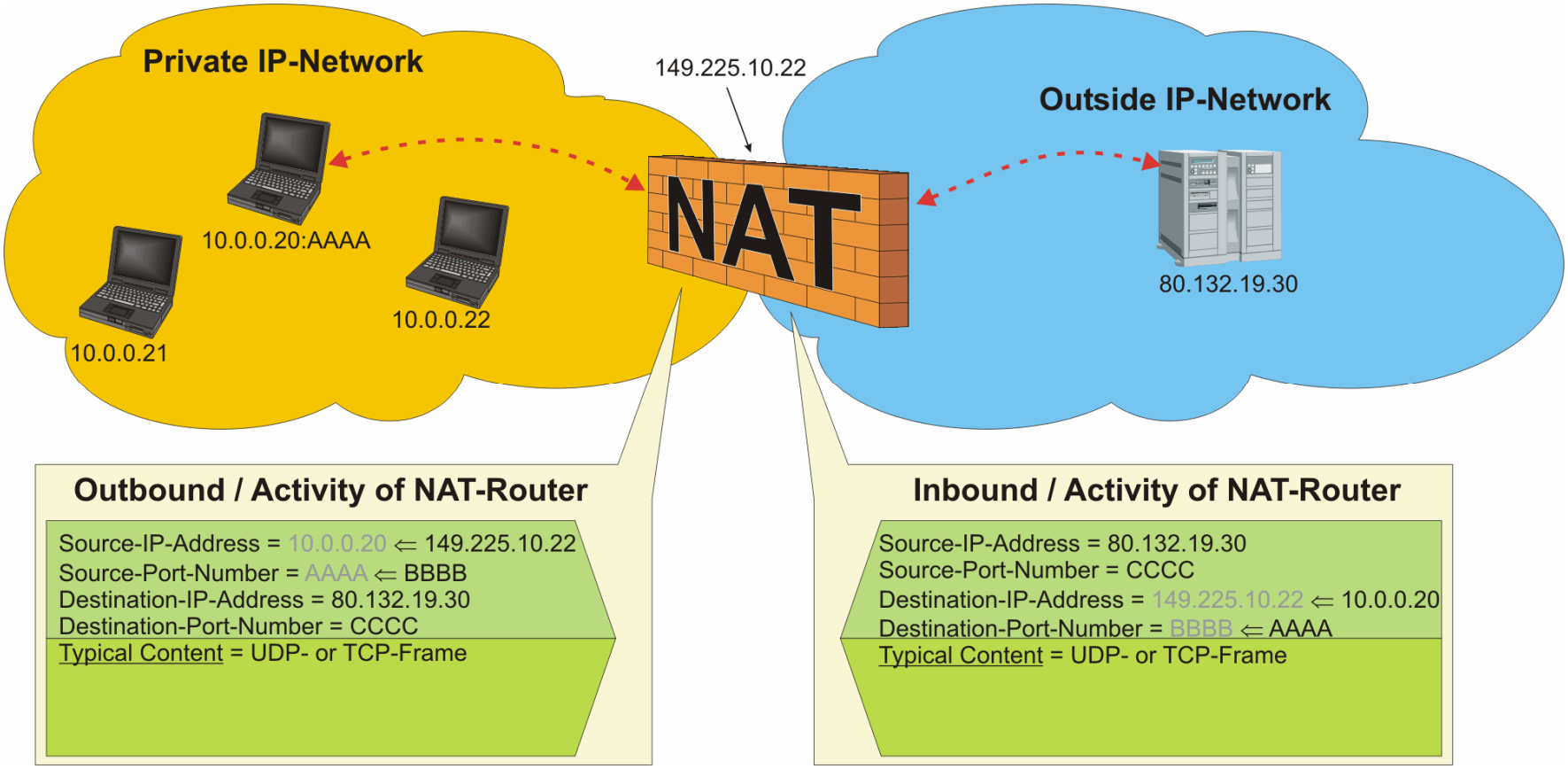


The General Security Architecture

- ⇒ With respect to security it is important understanding that typically a secure transmission is required only between sites, between different locations. Security is inherent within one facility that is controlled by the operator.
- ⇒ It is therefore not very likely to invoke special security techniques like IPsec or TLS for information exchange between network nodes which are situated within the same building or facility or even between boards in the same subrack.
- ⇒ These technologies are required, if information is exchanged e.g. between two BGCF's or between two I-CSCF's.
- ⇒ This approach is further outlined on the graphics page. It illustrates two different IMS's, both belonging to two different network operators. In that respect, IMS 1 is spread over different locations while IMS 2 uses only a single location.
- ⇒ As illustrated, each location of IMS 1 is protected by a dedicated SEG (security gateway) which may be based on TLS or IPsec. All IP-traffic between the different locations of IMS 1 is protected by these SEG's.
- ⇒ The interface between two SEG's is referred to as Za-interface and the interface between IMS-network elements and an SEG is called Zb-interface.
- ⇒ Literally the same applies for the traffic between IMS 1 and IMS 2. However, in this case the two network operators need to have negotiated their security architecture and they need to allow the exchange of security material (e.g. keys).

Tackling NAT-Issues within the IMS

- **Introducing NAT and NAPT**



Tackling NAT-Issues within the IMS

Introducing NAT and NAPT

The figure illustrates the operation of a typical NAT/NAPT-system.

⇒ Internally in the private IP-network, all devices have been assigned private IP-addresses; in the example class A-addresses 10.0.0.0 have been used.

Note that behind a NAT/NAPT, any range of IP-addresses (not only dedicated private ones) can be used and assigned to booting devices.

⇒ In the example, the laptop with address 10.0.0.20 sends an IP-frame from source port number AAAA (TCP or UDP) to the server in the outside IP-network which has an IP-address 80.132.19.30. The destination port number on the server is CCCC (TCP or UDP).

⇒ As illustrated, the NAT/NAPT-router is also the edge router of the private IP-network. Every outgoing IP-frame needs to be routed through the NAT/NAPT-router.

⇒ Its specific operation becomes clear when looking at the left message box: The NAT/NAPT-router replaces the source IP-address 10.0.0.20 with its own IP-address 149.225.10.22 which is also valid on the outside IP-network. The NAT/NAPT-router also replaces the source port number AAAA with a randomly selected but very important port number, say BBBB. The NAT/NAPT-router also buffers the association between its own source port number BBBB and the internal IP-address 10.0.0.20 / port number AAAA. The NAT/NAPT-router does not tamper with the content of the IP-frame.

The buffering of this association is time limited and the actual buffering time is implementation dependent. Recommended values are 5 minutes for UDP-associations and 24 hours for TCP-associations [RFC 4008].

⇒ Therefore, the NAT/NAPT-router really pretends to be the origin of the IP-frame and its content when it sends the IP-frame towards the server 80.132.19.30.

For fully understanding the operation of NAT/NAPT it is very important to understand the following step:

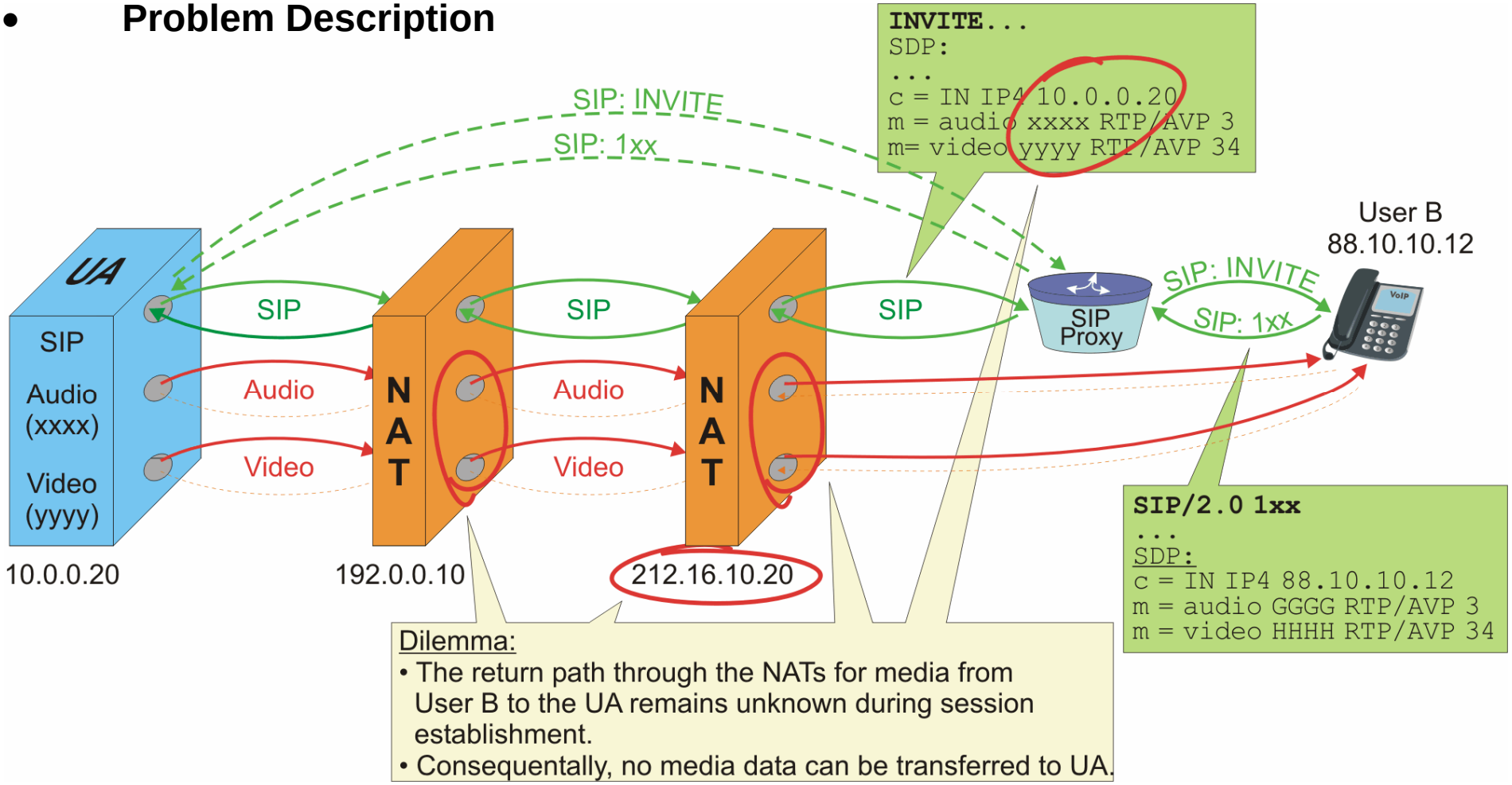
⇒ When the server 80.132.19.30 sends its reply for the previously received IP-frame, it sends it to 149.225.10.22 / port number BBBB. Since the NAT/NAPT-router has in the meantime probably sent many IP-frames to different destinations and possibly several ones also to 80.132.19.30, it is this port number BBBB that allows the NAT/NAPT-router to relate the incoming IP-frame to the previously stored association.

⇒ Therefore, the NAT/NAPT-router is enabled to replace the destination IP-address and port number with the internal values 10.0.0.20 and AAAA.

The whole meaning of NAT/NAPT lies in the sudden availability of a rather unlimited number of IP-addresses. This makes NAT so appealing. However, NAT also provides an inherent firewall function because no IP-frame may enter a private IP-network through a NAT/NAPT-router unless some process from the inside requested it.

Step 2: UA Originating Session Establishment behind 2 NAT's

- **Problem Description**



Step 2: UA Originating Session Establishment behind 2 NAT's

Problem Description

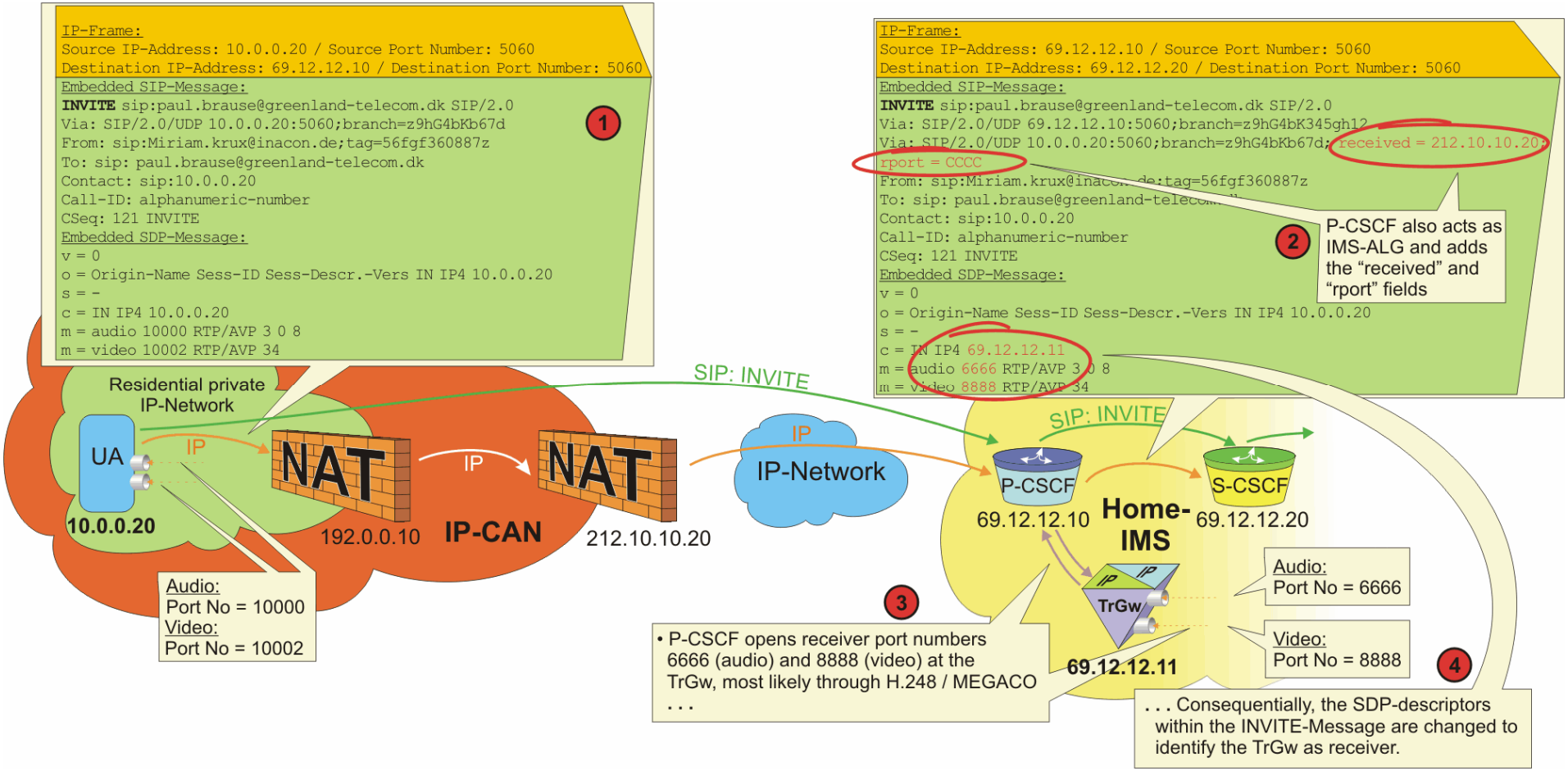
SIP-messages can flow back and forth through the NAT's, provided that symmetric port numbering is applied. Again, symmetric port numbering means that the SIP-proxy and the UA both send and receive SIP-message on/from the same port number, e.g. 5060.

The figure and the description underneath illustrate the reason why there is still a problem if media transfer shall occur, even when the aforementioned measures are taken during registration and even when the P-CSCF incorporates an IMS-ALG:

⇒ The UA conveys in its Request: INVITE-message its own connection address = 10.0.0.20 and a receiver port number XXXX for audio and YYYY for video.
⇒ The called User B conveys his/her connection address 88.10.10.12 and receiver port number GGGG for audio and HHHH for video. With this information available at the calling UA, the UA can send data to User B through the NAT's.
⇒ However, User B is unable to send media data anywhere. The media path from User B to the calling UA remains silent.

SIP-proxies can only operate on SIP-signaling messages and therefore the IMS-ALG function is not suited to address this issue. The amendment of a media gateway called TrGw in 3GPP becomes necessary.

SIP: INVITE-Message



SIP: INVITE-Message

The figure illustrates how the Request: INVITE-message which is sent by the UA behind the two NAT's reaches the P-CSCF, is processed by the P-CSCF and finally gets relayed to the S-CSCF before it is forwarded to the next hop (and finally it will reach its destination).
Note that we did not include the Response: 100-Trying message.

- **Bullet 1:**
  The entire IP-frame is shown together with the included SIP-message. Please note that the UA uses the IP-address of the P-CSCF as destination IP-address. And of course, the UA indicates its own IP-address 10.0.0.20 as connection address within the SDP-portion. The UA has reserved the two port numbers 10000 and 10002 to receive audio and video data from the called user "Paul Brause".

- **Bullet 2:**
  When the P-CSCF receives the Request: INVITE, it detects the NAT-interworking based on the difference between the source IP-address (within the IP-frame header) and the IP-address indicated in the topmost "Via:"-header field. As done during registration, the P-CSCF acts as IMS-ALG and adds to the "Via:"-header field of the UA the "received "- and "rport"- attributes. Both will allow the P-CSCF to forward upcoming response message to the UA through the closest NAT (⇔ received=212.10.10.20 / rport=CCCC).

- **Bullet 3:**
  However, most important for our current considerations is the fact that the P-CSCF also inspects the included SDP-portion. As illustrated, the P-CSCF will communicate with the TrGw (IP-address 69.12.12.11) and it will trigger the TrGw to open two port numbers to act as receiver port numbers for the upcoming audio and video streams from "Paul Brause".

- **Bullet 4:**
  Finally, the P-CSCF receives indication from the TrGw that port numbers 6666 and 8888 have been opened. Accordingly, the P-CSCF will alter the SDP-portion of the Request: INVITE-message and replace the UA's connection address and port numbers through the ones received from the TrGw.
  Then the P-CSCF forwards the Request: INVITE-message to the S-CSCF.

By replacing the UA's connection address and port numbers through the IP-address and port numbers of the TrGw, the invited party gets a valid sink for their media data.