Understanding SIP

Dorgham Sisalem
Jiri Kuthan

Mobile Integrated Services
GMD Fokus
Sisalem,kuthan@fokus.gmd.de

Attention!

- **Update Notice**

Authors are committed to ongoing improvement of this tutorial. Thus, this version may include updates and differ slightly from printed version. You can get the updated version at the following address:

http://www.fokus.gmd.de/mobis/siptutorial/

- **Frequent Misunderstandings**

There are numerous issues that turned out to be difficult to understand. Such issues are labeled with the symbol bellow. Please, pay special attention to them.

Frequently
Misunderstood
Issue

Outline

- It's IP Telephony
- Who is who
- IP Telephony Basics
  - Protocol ZOO
  - SIP Signaling
  - Multimedia Communication
- Advanced Signaling
  - Programmability
  - QoS Preconditions
- Mobility and 3gpp
- SIP vs H.323
- Robustness
- Security
- Legacy
- Political Issues
- Status Update
- Conclusions
- References

The Big FAQ

- Q: You are too IP-centric, aren't you?
- A: Of course, we are.

- Internet telephony (which has Internet in its name) is about IP.
  - IP telephony runs on top of IP and utilizes the IP service model.
  - It is not about re-engineering PSTN -- PSTN is good enough.

- SIP is much more similar to HTTP rather than to legacy signaling both in terms of service model and protocol design.

Appeals of IP Telephony

- Saving, but ...
  - lower QoS
  - Telcos lower prices (**1998**: Berlin-Prague, 99 Pf/Min, **1999**: 39 Pf/Min, **2000**: 32 Pf/Min call-by-call, 23 Pf/Min preselection)
- **Internet Service integration**
  - IP is the first true Integrated Services Digital Network
  - Major argument: **convenience**
- **In IP, you are your own master**
  - Open service market: access providers located across the globe; even you can be a provider.
  - Programmability: programs by user as well as third parties.

Integrated Applications

- Distributed games
  - SIP Quake sighted!
- Virtual reality
- Web-pages and applets
- Links in e-mails
- Web-IVRs
- Click-to-dial
- Directory Services
- Video conferencing
- Instant Messaging
  - voicemail notifications
  - stock notifications
  - callback notification
- Calendars
  - pre-setup conference calls
- Unified Messaging
  - voicemail2email

etc.

IP Service Model

- Split of Transport and Application Services
  - these are different businesses run on top of different technologies
  - service promiscuity: anyone can access services brought by any providers
  - anyone with IP connectivity can become a provider
  - setting up a signaling service as easy setting up a web server
  - **service market is completely open**
- Applications Are Split As Well
  - Example:
    - IP operated by UUNET
    - SIP signaling by WCOM
    - PSTN call termination by mypstn.com and another-pstn.xy
    - least-cost PSTN termination routing by yet another company

Example: iptel.org Trial Site

- Provides just signaling services
  - gives users a unique globally reachable address
  - resembles Web-hosting in IP world or NetCentrex in PSTN world
  - no media transport -- only signaling relayed, media does not hit the server at all
- To set it up, we needed
  - PC
  - Freely available software
  - IP access
  - one part-time undergraduate student
- Users need
  - IP phone (either in SW or HW)
- Complimentary services may be easily provided by other parties, users just need to set up their signaling preferences:
  - bridging to PSTN, voicemail--2-email, etc.

IP Design Concepts

- **Distributed end-2-end design**
- Intelligence and states resides in end-devices
- Network maintains almost zero intelligence (except routing) and state (except routing tables).
- End-devices speak to each other using whatever applications they have. There is almost no logic in the network affecting this behavior.
- Result:
  - Flexibility. Introducing new applications is easy.
  - Failure recovery. No state, no problem on failure.
  - Scalability. No state, no memory scalability issues.

Who is Who

Who Engineers the Internet

- Internet Engineering Task Force (www.ietf.org)
- "large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual."
- IETF's business:
  - Design and standardization of interoperable protocols
  - Almost anything else out of scope: deployment, promotion, API specification, etc.

IETF - Standardization Procedure (RFC 2026)

- Much of the work is handled via mailing lists. The IETF holds meetings three times per year
- Proposals submitted for discussion as Internet Drafts. If approved they are published as RFCs.
- No formal voting -- rough consensus
- RFC
  - Most of them are NOT standards - informational, experimental, historic, funny (Check April 1st ones (RFC 1149)).
  - Published RFCs never change.
  - multiple instances of running code required before standardizing
- New topic ➔ BOF

Concepts of the Internet Design (RFC 1958, 2775)

- Single inter-networking protocol deployed **end2end**
- State stored only in end-devices, no single point of failure, scalable core, higher message overhead
  - example: TCP cb stored only in end-devices; no TCP state in routers (per-link reliability would not solve the e2e problem)
- Keep it simple and stupid (avoid options and parameters)
- Be conservative when sending and liberal when receiving.
- Performance and cost subject to consideration
- **Modularity** is good. (Puzzle/LEGO concept)
- **Distributed** design
- Some of current technical triggers: IPv4 scaling limits, gigabit speeds, QoS, security

GMD FOKUS

Advantages of the IETF Standardization Process

- Anyone can join both actively and passively and contribute to quality of standards.
- Standards available for free.
- Long years of Internet engineering practice.

Related IETF Working Groups

- SIP: Session Initiation Protocol
- IPTEL: Internet Telephony
- AVT: Audio Video Transport
- MIDCOM: Firewall/NAT Traversal
- SIMPLE: SIP for Instant Messaging and Presence Leveraging
- MMUSIC: Multiparty Multimedia Session Control
- QoS Related: DiffServ, IntServ, RSVP
- PSTN legacy: SigTran, Megaco
- interaction of PSTN and IP services: PINT,SPIRITS

Other Related Bodies

- Third Generation Partnership Project (3gpp)
  - creation of technical specifications for 3rd generation mobile systems
  - uses SIP as call signaling in IP networks
- ITU-T SG 16
  - H.323 V1-V4 umbrella standard
  - H.248 (Megaco)
- ETSI Tiphon
  - concerned with IP/PSTN interoperability
  - analysis of security threats, Open Settlement Protocol

Other Related Bodies (cont.)

- SIP Forum for promotion of SIP technology
- IMTC concerned with interoperability
- PacketCable established by CableLabs to look at cable technologies
- Telecommunications Industry Association (TIA) involved in layers bellow IP
- Softswitch promoting IN replicas in IP

Other Related Bodies (cont.)

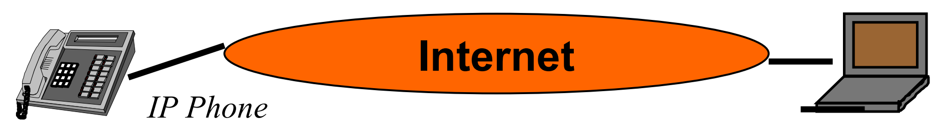
- *The list still goes on...*
- JAIN developing abstract APIs for developing service creations across PSTN, ATM, IP, etc.
- TIPIA
- TTL
- VoiceXML Forum

Protocol ZOO

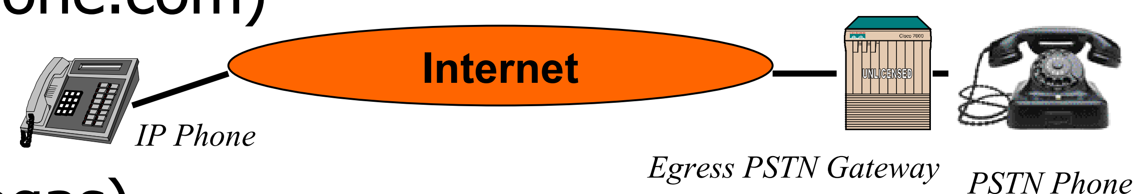
Internet Telephony

⌘Routing a call over the Internet

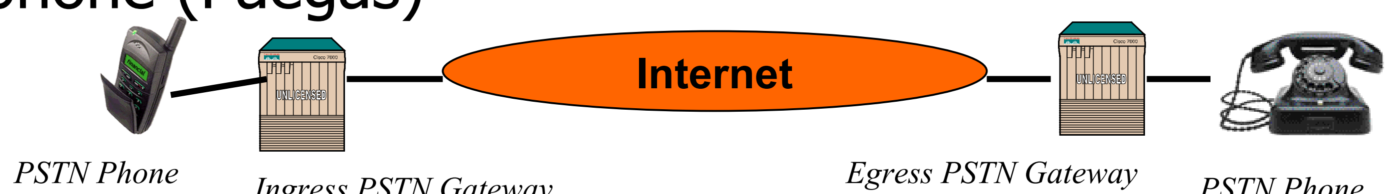
- PC-to-PC (MS NetMeeting, appliances)



- PC-to-phone (net2phone.com)



- phone-to-phone (Paegas)



- phone-to-PC as well

What Protocols Are Needed?

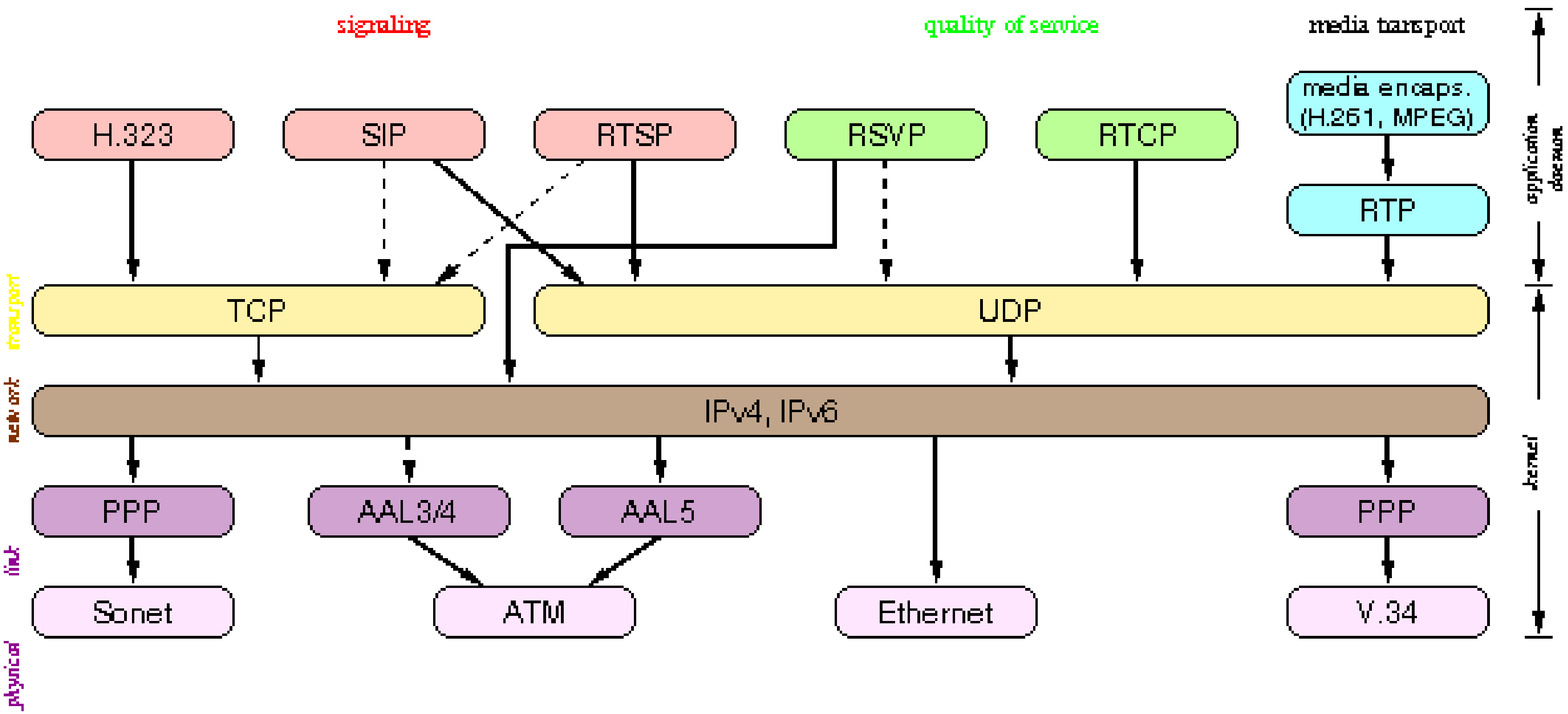
- Signaling protocol to establish presence, locate users, set up, modify and tear down sessions
- Media Transport Protocols for transmission of packetized audio/video
- Supporting Protocols
  - Gateway Location, QoS, interdomain AAA*, address translation, IP, etc.

** AAA = Authentication, Authorization, Accounting*

What Protocols Are There

- Signaling: SIP/SDP (IETF), H.323 (ITU-T)
  - Note: SIP adopted by 3gpp; lower production and operation costs reported
- Media: RTP (IETF's, adopted by ITU-T)
- Transport: UDP, TCP, (Stream Control Transmission Protocol - RFC 2960)
- Supporting protocols:
  - DNS
  - TRIP - Telephony Routing over IP - discovery and exchange of IP telephony gateway routing tables between providers
  - RSVP - Resource Reservation Setup Protocol
  - COPS - Common Open Policy Service - protocol for for supporting policy control over QoS
  - Diameter - Authentication, Accounting, Authorization

Protocol ZOO



Source: Henning Schulzrinne,
http://www.cs.columbia.edu/~hgs/internet/

SIP Signaling

Session Initiation Protocol

- SIP is end-to-end, client-server session signaling protocol
  - SIP's primarily provides presence and mobility
  - Protocol primitives: Session setup, termination, changes
- Arbitrary services built on top of SIP, e.g.:
  - Redirect calls from unknown callers to secretary
  - Reply with a webpage if unavailable
  - Send a JPEG on invitation
- Features:
  - Textual encoding (telnet, tcpdump compatible)
  - Programmability

SIP - General Purpose Presence Protocol

- SIP is not limited to Internet telephony
  - SIP establishes user presence
  - SIP messages can convey arbitrary signaling payload: session description, instant messages, JPEGs, any MIME types
- Suitable for applications having a notion of session
  - distributed virtual reality systems,
  - network games (Quake II/III implementations),
  - video conferencing, etc.
- Applications may leverage SIP infrastructure (Call Processing, User Location, Authentication)
  - Instant Messaging and Presence
  - SIP for Appliances

SIP Is Not

- Transport Protocol
- QoS Reservation Protocol
- Gateway Control Protocol

- Some argue it may be used for accessing IP-enabled appliances ...
- It does NOT dictate ...
  - Product features and services (color of your phone and distinctive ringing melodies, number of simultaneous calls your phone can handle, don't disturb feature, ...)
  - network configuration

SIP History

- Work began in 1995 in IETF mmusic WG
- 02/1996: draft-ietf-mmusic-sip-00: 15 ASCII pages, one request type
- 12/1996: -01 30 ASCII pages, 2 request types
- 01/1999: -12 149 ASCII pages, 6 methods
- 03/1999: RFC 2543, 153 ASCII pages, 6 methods
- 11/1999: SIP WG formed
- 11/2000: draft-ietf-sip-rfc2543bis-02, 171 ASCII pages, 6 methods
- 12/2000: it was recognized that amount of work at SIP WG was becoming unmanageable; 1 RFC; 18 I-Ds on WG's agenda; numerous individual submissions
- 04/2001: proposal for splitting SIP WG into SIP and SIPPING announced

- 2001: SIP implementations widely available
  - http://www.cs.columbia.edu/~hgs/sip/implementations.html
  - http://www.pulver.com/sip/products.html

GMD FOKUS

SIP End-devices

- User Agent (user application)
  - UA Client (originates calls)
  - UA Server (listens for incoming calls)
  - both SW and HW available

SIP Workhorses

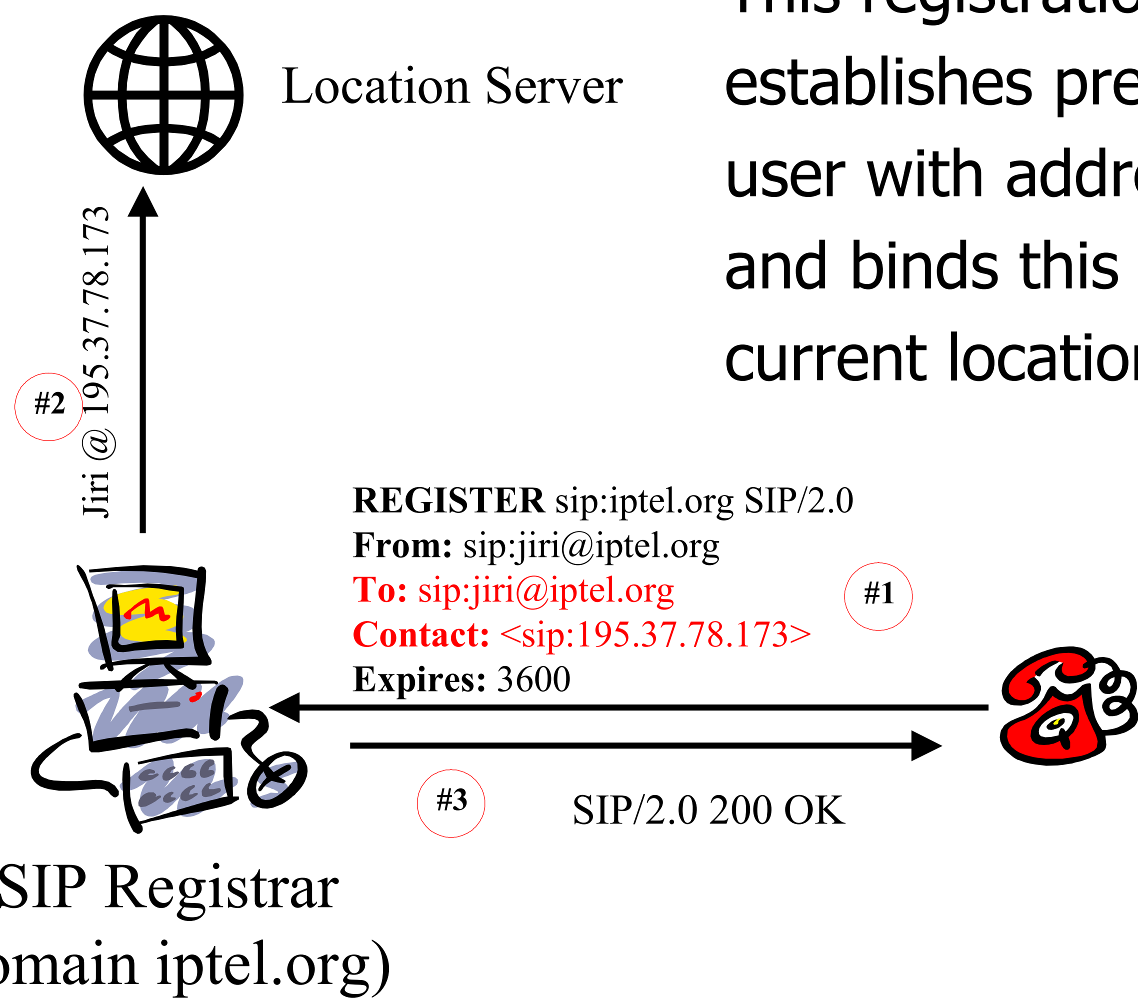
- SIP Proxy Server
  - relays call signaling, i.e. acts as both client and server
  - operates in a transactional manner, i.e., it keeps no session state
- SIP Redirect Server
  - redirects callers to other servers
- SIP Registrar
  - accept registration requests from users
  - maintains user's whereabouts at a Location Server (like GSM HLR)

SIP Addresses

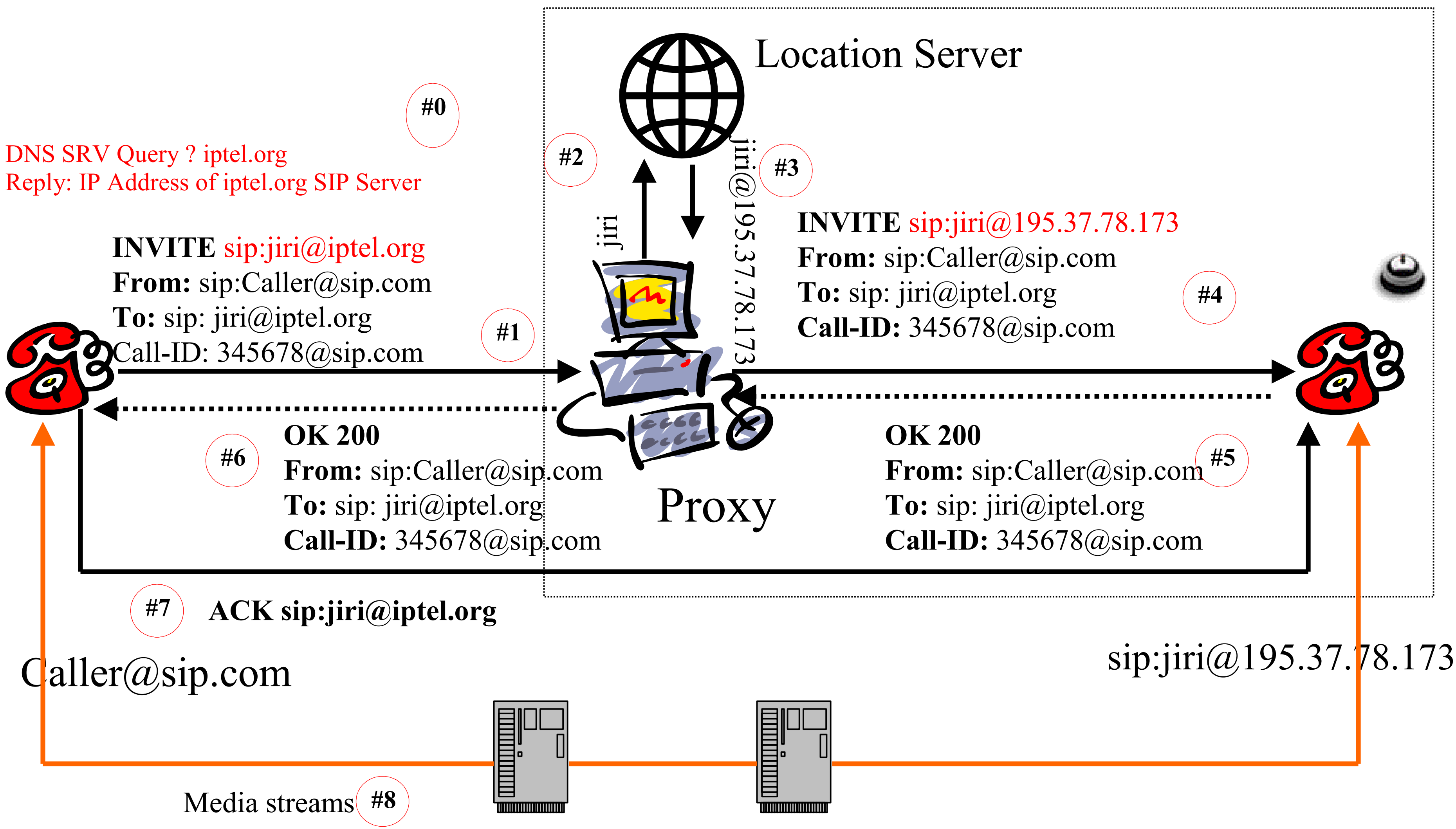
- SIP gives you a globally reachable address.
  - Callees bind to this address using SIP REGISTER method.
  - Callers use this address to establish real-time communication with callees.
- URLs used as address data format; examples:
  - sip:jiri@iptel.org
  - sip:voicemail@iptel.org?subject=callme
  - sip:sales@hotel.xy; geo.position:=48.54_-123.84_120
- must include host, may include user name, port number, parameters (e.g., transport), etc.
- may be embedded in Webpages, email signatures, printed on your business card, etc.
- address space unlimited
- non-SIP URLs can be used as well (mailto:, http:, ...)

SIP Registration

This registration example establishes presence of user with address jiri@iptel.org and binds this address to user's current location 195.37.78.173.



SIP Operation in Proxy Mode



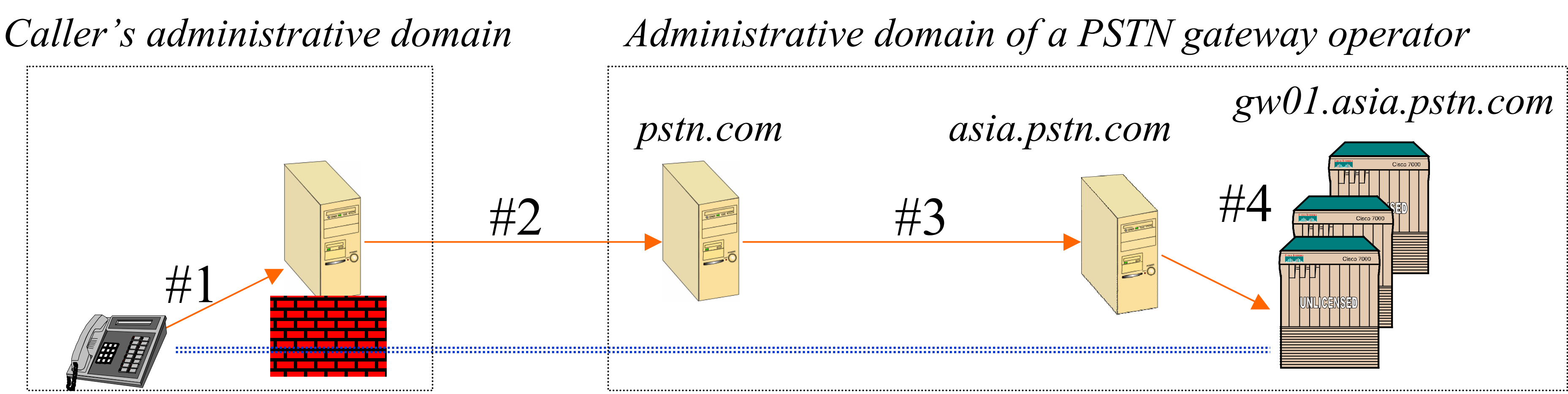
Proxy Server Functionality

- Serve as rendezvous point at which callees are globally reachable
- Perform routing function, i.e., determine to which hop (UA/proxy/redirect) signaling should be relayed
- Allow the routing function to be programmable. Arbitrary logic may be built on top of the protocol
  - user’s signaling preferences
  - AAA
  - firewall control
  - etc.
- Forking: Several destinations may be tried for a request sequentially or in parallel.

Proxy Chaining

- There may be also cases when a local outbound proxy may be involved
  - provides locally important call processing logic (e.g., identifying nearest 911)
  - manages firewall
  - provides least-gateway-cost routing service
  - IP phones must know address of the proxy:may be configured manually or with a configuration protocol (DHCP, TFTP, ...)
- In general, servers may be arbitrarily chained
  - a central company's server may distribute signaling to departmental servers
  - a user may want to forward incoming calls to her cell phone
- Servers have to avoid loops and recognize spirals

Proxy Chaining - an Example



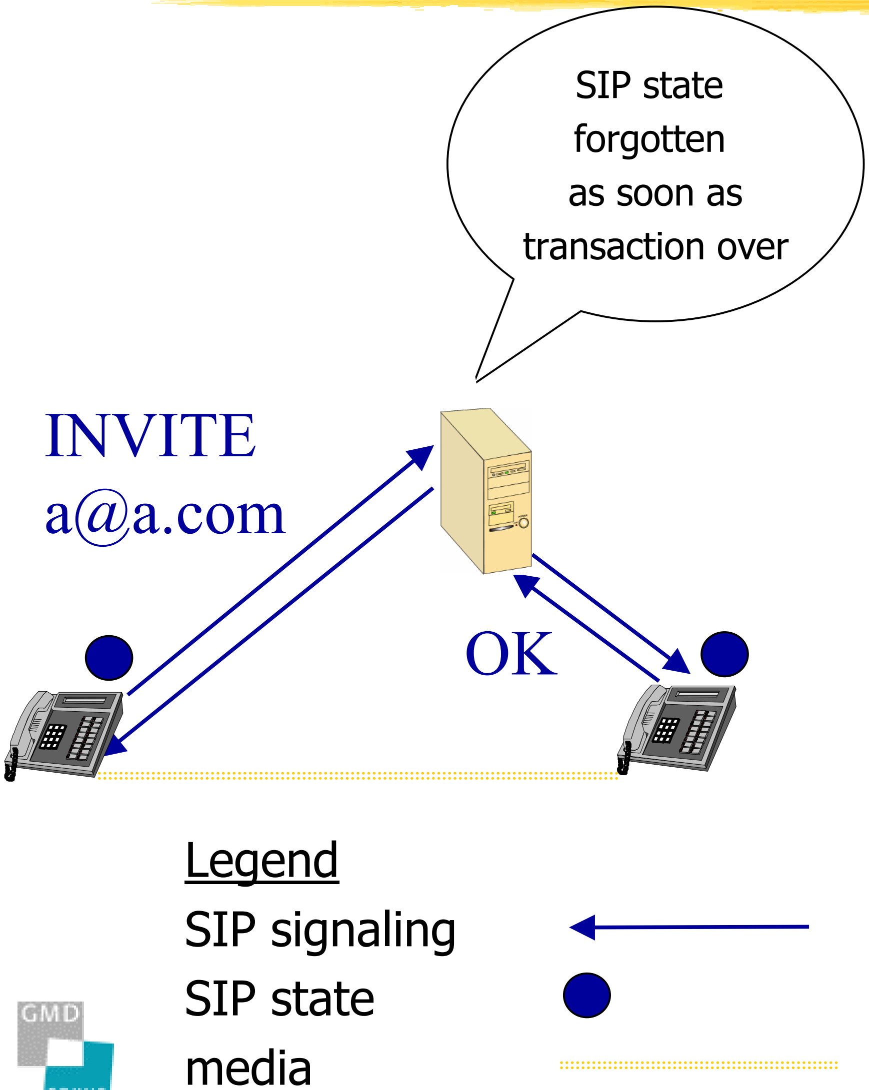
Caller's outbound proxy accomplishes firewall traversal.

Destination's "first-hit proxy" identifies a proxy serving dialed area.

Proxy in the target area distributes load in a gateway farm.

Note: signaling (in red) may take a completely different path from media (in blue).

"Stateful" Proxy Refers to Transactions

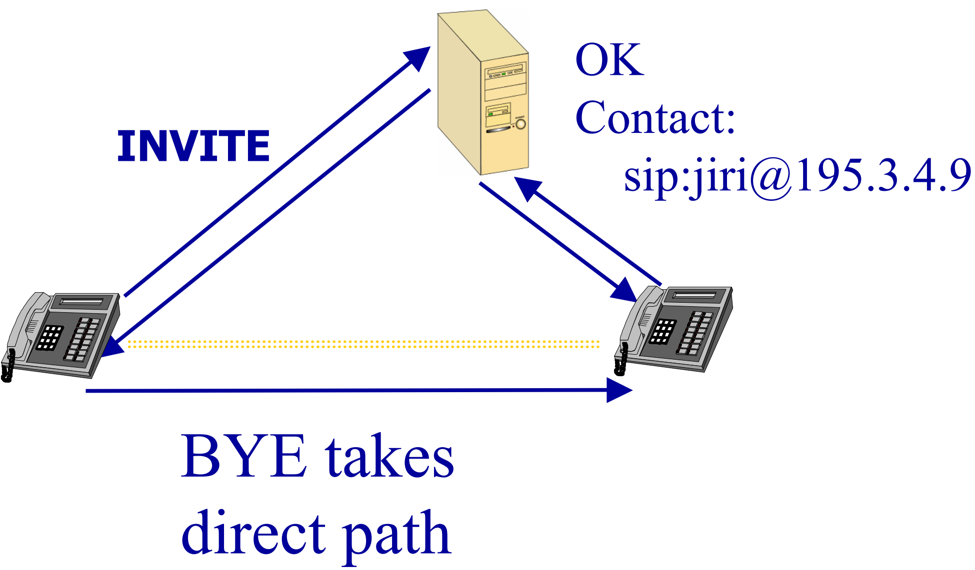


- If a proxy is stateful it keeps state during a SIP transaction and completely forgets it afterwards.
- A SIP proxy is not aware of existing calls
- Unless route recording is used, BYE may take a completely different path (I.e., cannot be expected to terminate the state.)
- Theoretically, there may be session state as well. Unless there is a well defined use of it, it indicates unscalable implementation.

Frequently Misunderstood Issue

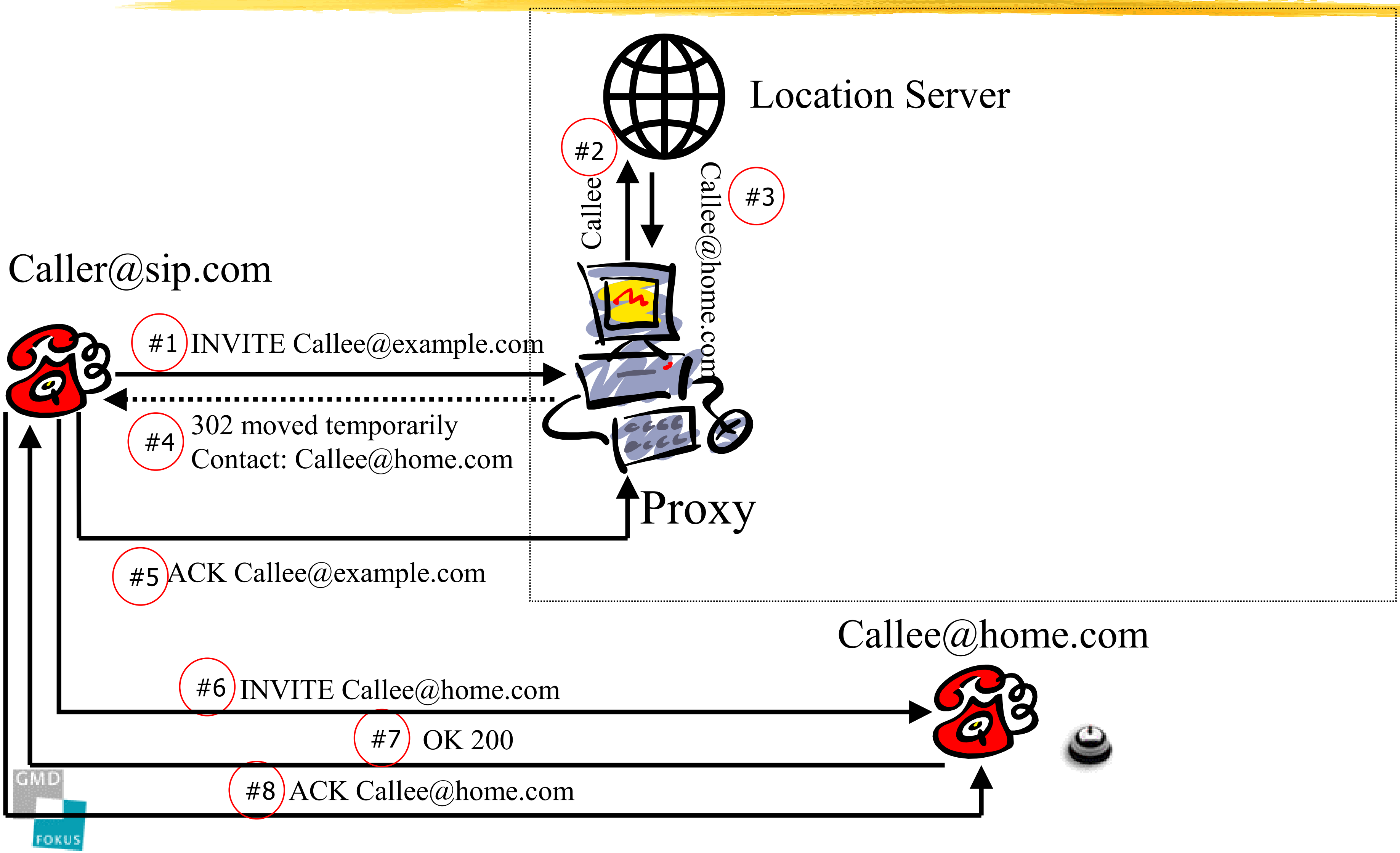
Subsequent Transactions Bypass Proxy

- Unless route recording is used, BYE may take a completely different path to destination indicated in **Contact:** header field.



Frequently Misunderstood Issue

SIP Operation in Redirect Mode



SIP Server -- Proxy versus Redirection

- A SIP server may either **proxy** or **redirect** a request
- Which of the two method applies is a configuration issue. It may be statically configured or dynamically determined (CPL).
- Redirection useful if a user moves or changes her provider (PSTN: “The number you have dialed is not available.”) -- caller does not need to try the original server next time. Stateless.
- Proxy useful if forking, AAA, firewall control needed. In general, proxying grants more control to the server.

SIP RFC2543 Methods

- **INVITE** initiates sessions
  - session description included in message body
  - re-INVITEs used to change session state
- **ACK** confirms session establishment
  - can only be used with INVITE
- **BYE** terminates sessions
- **CANCEL** cancels a pending INVITE
- **OPTIONS** capability inquiry
- **REGISTER** binds a permanent address to current location; may convey user data (CPL scripts)

SIP Extension Methods

⌘ INFO	mid-call signaling (RFC 2976)
⌘ COMET	precondition met (draft-ietf-sip-manyfolks-resource)
⌘ PRACK	provisional reliable responses acknowledgement (draft-ietf-sip-100rel)
⌘ SUBSCRIBE/ NOTIFY/ MESSAGE	instant messaging (draft-rosenberg-impp-*)

SIP Response Codes

- Borrowed from HTTP: xyz explanatory text
- Receivers need to understand x
- x80 and higher codes avoid conflicts with future HTTP response codes
- 1yz Informational
  - 100 Trying
  - 180 Ringing (processed locally)
  - 181 Call is Being Forwarded
- 2yz Success
  - 200 ok
- 3yz Redirection
  - 300 Multiple Choices
  - 301 Moved Permanently
  - 302 Moved Temporarily

SIP Response Codes (cont.)

- 4yz Client error
  - 400 Bad Request
  - 401 Unauthorized
  - 482 Loop Detected
  - 486 Busy Here
- 5yz Server failure
  - 500 Server Internal Error
- 6yz Global Failure
  - 600 Busy Everywhere

SIP Message Structure

Request Method

INVITE sip:UserB@there.com SIP/2.0

Via: SIP/2.0/UDP here.com:5060
From: BigGuy <sip:UserA@here.com>
To: LittleGuy <sip:UserB@there.com>
Call-ID: 12345600@here.com
CSeq: 1 INVITE
Subject: Happy Christmas
Contact: BigGuy <sip:UserA@here.com>
Content-Type: application/sdp
Content-Length: 147

Message Header Fields

v=0
o=UserA 2890844526 2890844526 IN IP4 here.com
s=Session SDP
c=IN IP4 100.101.102.103
t=0 0
m=audio 49172 RTP/AVP 0
a=rtpmap:0 PCMU/8000

Payload

Response Status

SIP/2.0 200 OK

Via: SIP/2.0/UDP here.com:5060
From: BigGuy <sip:UserA@here.com>
To: LittleGuy <sip:UserB@there.com>;tag=65a35
Call-ID: 12345601@here.com
CSeq: 1 INVITE
Subject: Happy Christmas
Contact: LittleGuy <sip:UserB@there.com>
Content-Type: application/sdp
Content-Length: 134

v=0
o=UserB 2890844527 2890844527 IN IP4 there.com
s=Session SDP
c=IN IP4 110.111.112.113
t=0 0
m=audio 3456 RTP/AVP 0
a=rtpmap:0 PCMU/8000

"receive RTP G.711-encoded audio at 100.101.102.103:49172"

Session Description Protocol (SDP)

- Convey sufficient information to enable participation in a multimedia session
- SDP includes description of:
  - Media to use (codec, sampling rate)
  - Media destination (IP address and port number)
  - Session name and purpose
  - Times the session is active
  - Contact information
- Note: indeed SDP is a data format rather than a protocol.

Session Description Protocol (SDP)

```
v=0
o=sisalem 28908044538 289080890 IN IP4 193.175.132.118
s=SIP Tutorial
e=sisalem@fokus.gmd.de
c=IN IP4 126.16.69.4
t=28908044900 28908045000
m=audio 49170 RTP/AVP 0 98
a=rtpmap:98 L16/11025/2
```

Address Header Fields

- **From:** message originator
- **To:** final recipient
- **Request-URI:** current destination; may change along signaling path
- **Contact:** appears in INVITE / OPTIONS / ACK / REGISTER requests and in responses. It indicates direct response address to which subsequent transactions are sent.
  - A UA may send subsequent BYE or ACK to Contact: address (unless configured to use an outbound proxy).
  - It includes redirection address in 3xx and 485 responses.
  - It includes additional error information in 4xx, 5xx, and 6xx responses.
  - It may include preference weights.
  - It includes current location in REGISTER requests.
  - Multiple Contact: header fields may be included.

GMD FOKUS

SIP Protocol Design

- Infrastructure follows IP state model
  - Most intelligence and state in the end-devices
  - Network core maintains at most transactional state
  - Network edge may maintain session state
  - Benefits: memory and CPU consumption low in servers, reliability and scalability high (no single point of failure)
- UDP Support
  - faster set-up, less state
- Idempotent INVITEs (no collection of data spanning multiple requests)

Extensibility

- Range of future services unknown -> make signaling service-independent.
- History lesson: HTTP is not about hypertext transport any more.
  - It also provides e-mails, e-commerce, pc-banking, movies, etc.
  - Programmability adds numerous applications, the protocol remains almost the same.
- SIP designers took lesson from HTTP
  - Self-identifying Attribute-Value-Pairs (AVPs) followed by separators (EoL)
  - best-effort: receivers ignore unknown AVPs and skip to next separator
  - SDP support compulsory, arbitrary MIME payloads may be included (JPEG, ISUP, charging info, Multipart, ...)
  - transparent proxying

Extensibility (cont.)

- SIP designers took lesson from HTTP (cont.)
  - Require, Proxy-require, Supported Header Fields
  - classes of status codes (1xx in-progress, 2xx success, 3xx forwarding, ...)
  - guidance on designing new extensions provided (draft-ietf-sip-guidelines)
  - capability inquiry with OPTION -- returns supported methods (Allow), media types (Accept), compression methods (Accepted-Encoding), Supported (supported features)

GMD FOKUS

Multimedia Communication

IP Based Multimedia Communication

- SIP mainly establishes the IP addresses and port numbers at which the end systems can send and receive data
- SIP does not transport data and does not depend on a certain compression
- Data packets most probably do not follow the same path as the SIP packets

IP Based Multimedia Communication (cont.)

- Audio/Video samples are digitized, compressed and sent in UDP packets
- Compression schemes use limitations of human ears/eyes to reduce bandwidth
- Reduce audio bandwidth using silence suppression
- Reduce video bandwidth using motion detection

Compression Codecs

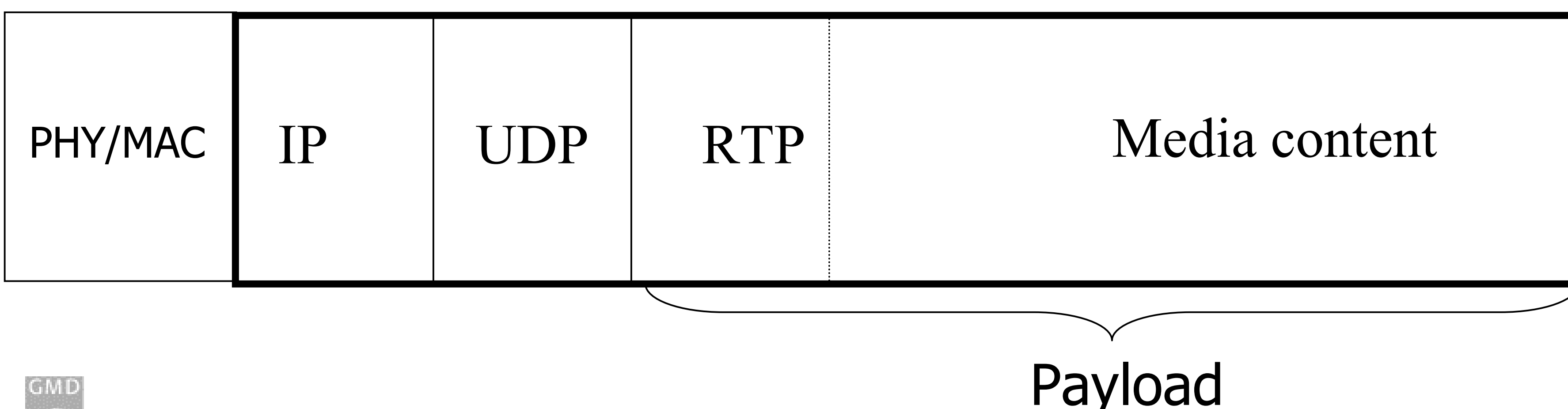
Codec	Unidirectional Bandwidth (kb/s)
G.723	5.3/6.3
GSM	13.0
G.711	64 (telephone)
MPEG L3	56-128
Video	depends on content, frame rate compression and motion

more http://www.cs.columbia.edu/~hgs/(audio/video)

GMD FOKUS

Real Time Transport Protocol (RTP)

- Standardized by the IETF and used by ITU-T as well
- Designed to be scalable, flexible and separate data and control mechansms



RTP: Functions

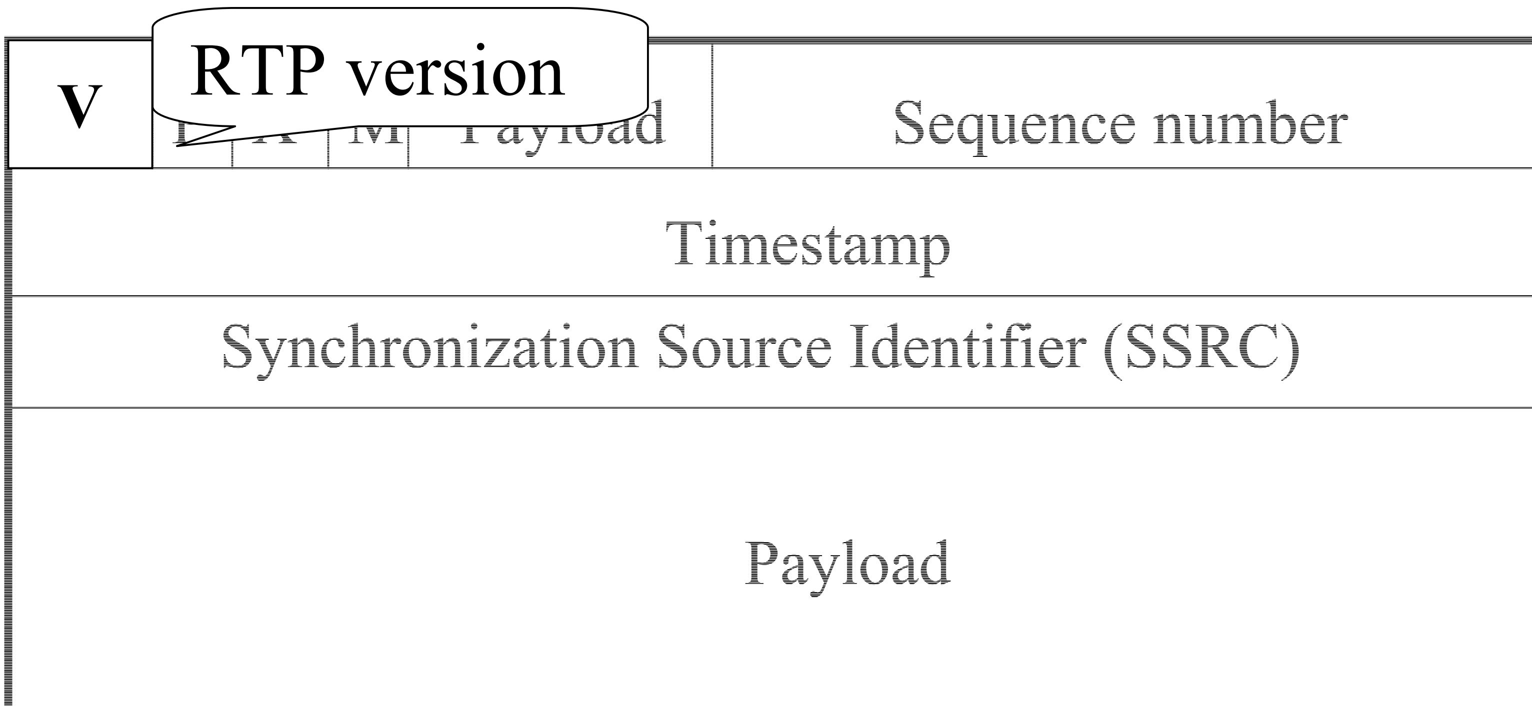
- Provides information for:
  - media content type
  - talk spurts
  - sender identification
  - synchronization
  - loss detection
  - segmentation and reassembly
  - security (encryption)

RTP: Header

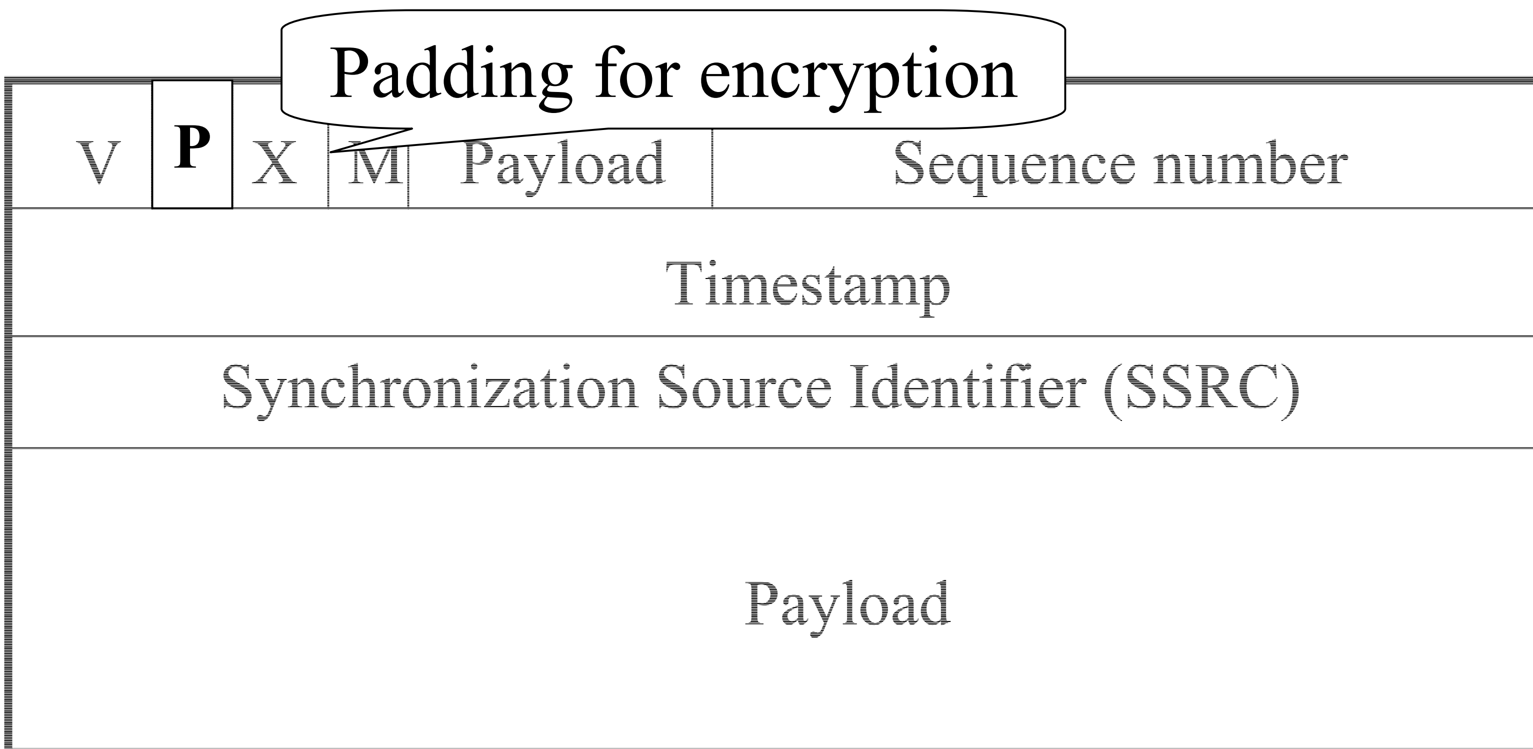
<table>
<tr><td>V</td><td>P</td><td>X</td><td>M</td><td>Payload</td><td>Sequence number</td></tr>
<tr><td colspan="6">Timestamp</td></tr>
<tr><td colspan="6">Synchronization Source Identifier (SSRC)</td></tr>
<tr><td colspan="6">Payload</td></tr>
</table>

GMD
FOKUS

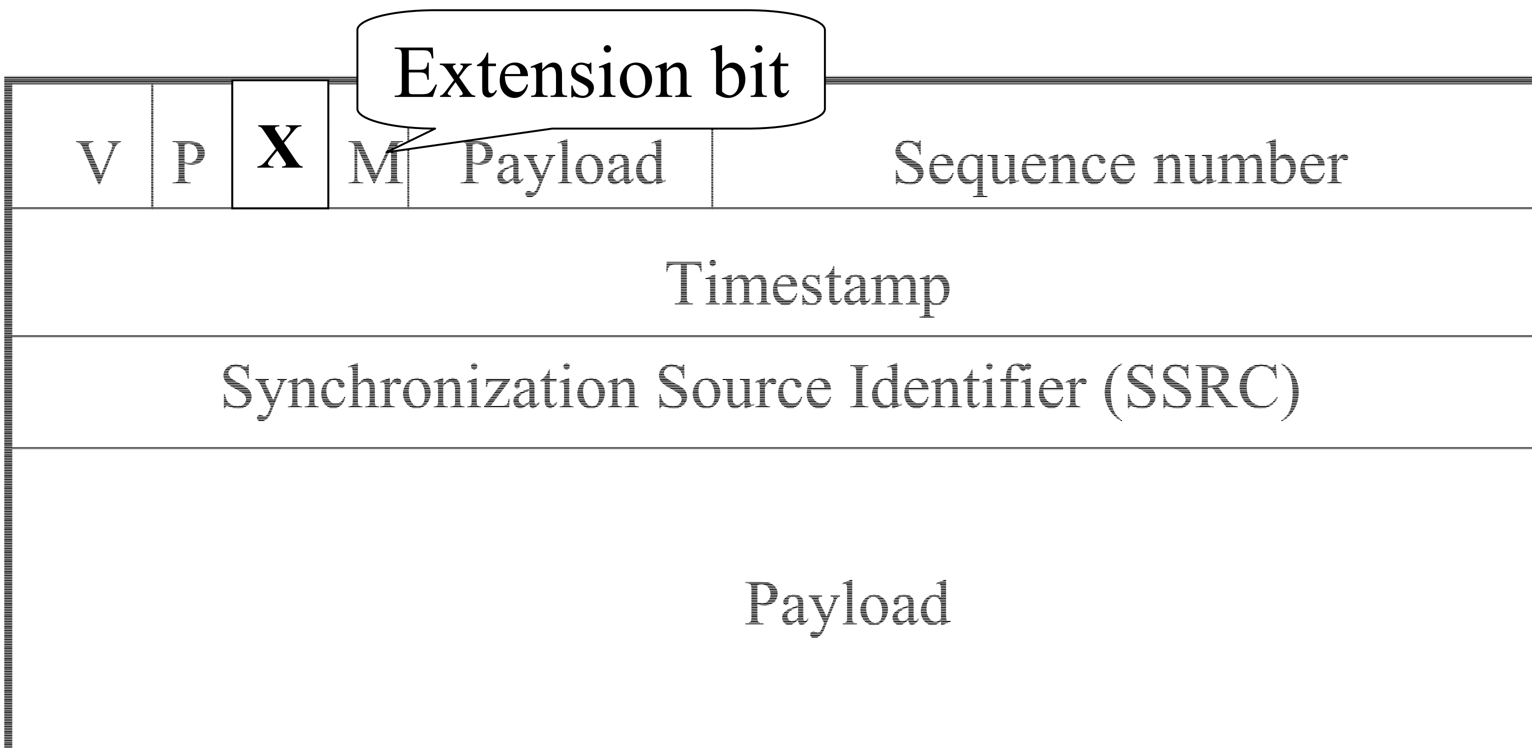
RTP: Header


V
RTP version
Sequence number
Timestamp
Synchronization Source Identifier (SSRC)
Payload

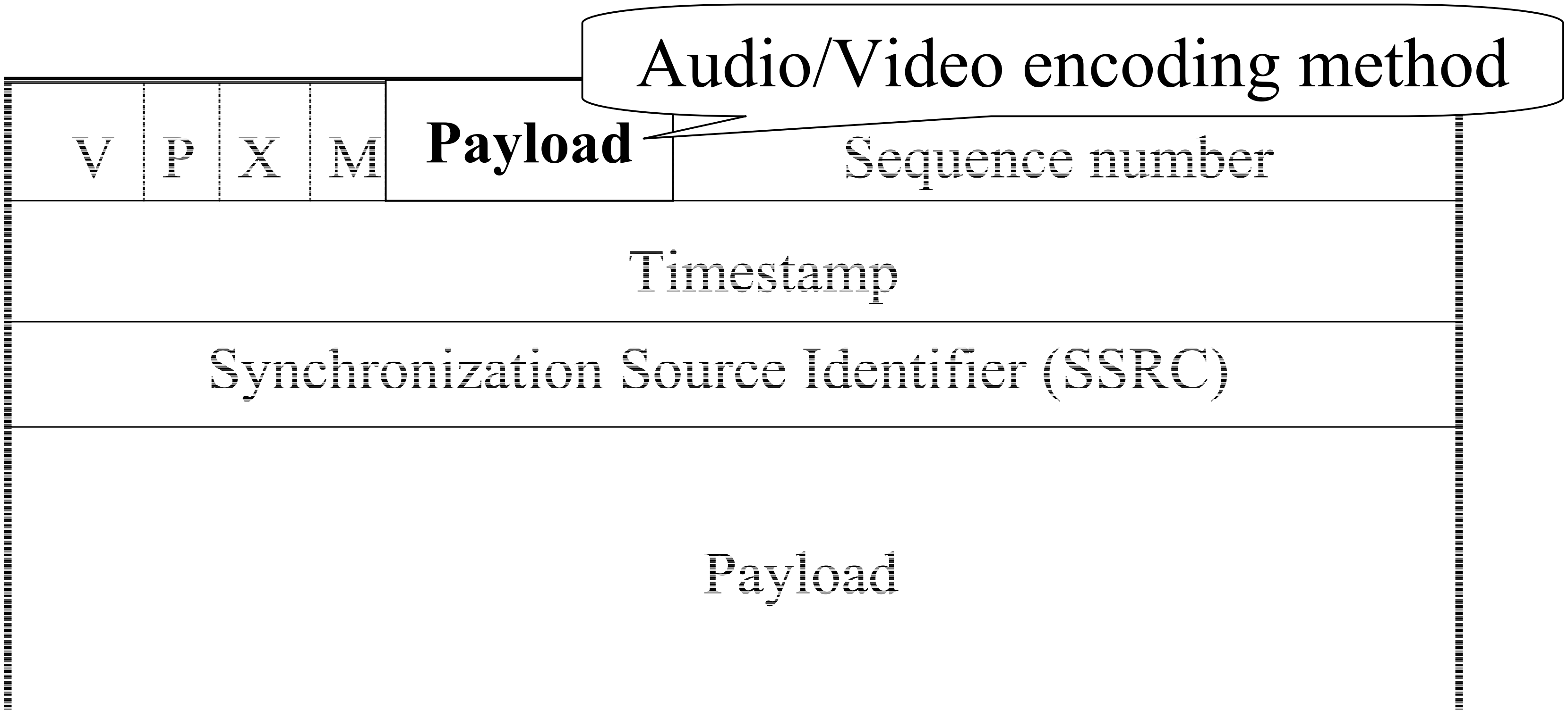
RTP: Header


Padding for encryption
V
P
X
M
Payload
Sequence number
Timestamp
Synchronization Source Identifier (SSRC)
Payload

RTP: Header


Extension bit
V
P
X
M
Payload
Sequence number
Timestamp
Synchronization Source Identifier (SSRC)
Payload

RTP: Header



Audio/Video encoding method
V
P
X
M
Payload
Sequence number
Timestamp
Synchronization Source Identifier (SSRC)
Payload

RTP: Header

Number of packet increased by one for each new packet

V	P	X	M	Payload	**Sequence number**
Timestamp					
Synchronization Source Identifier (SSRC)					
Payload					

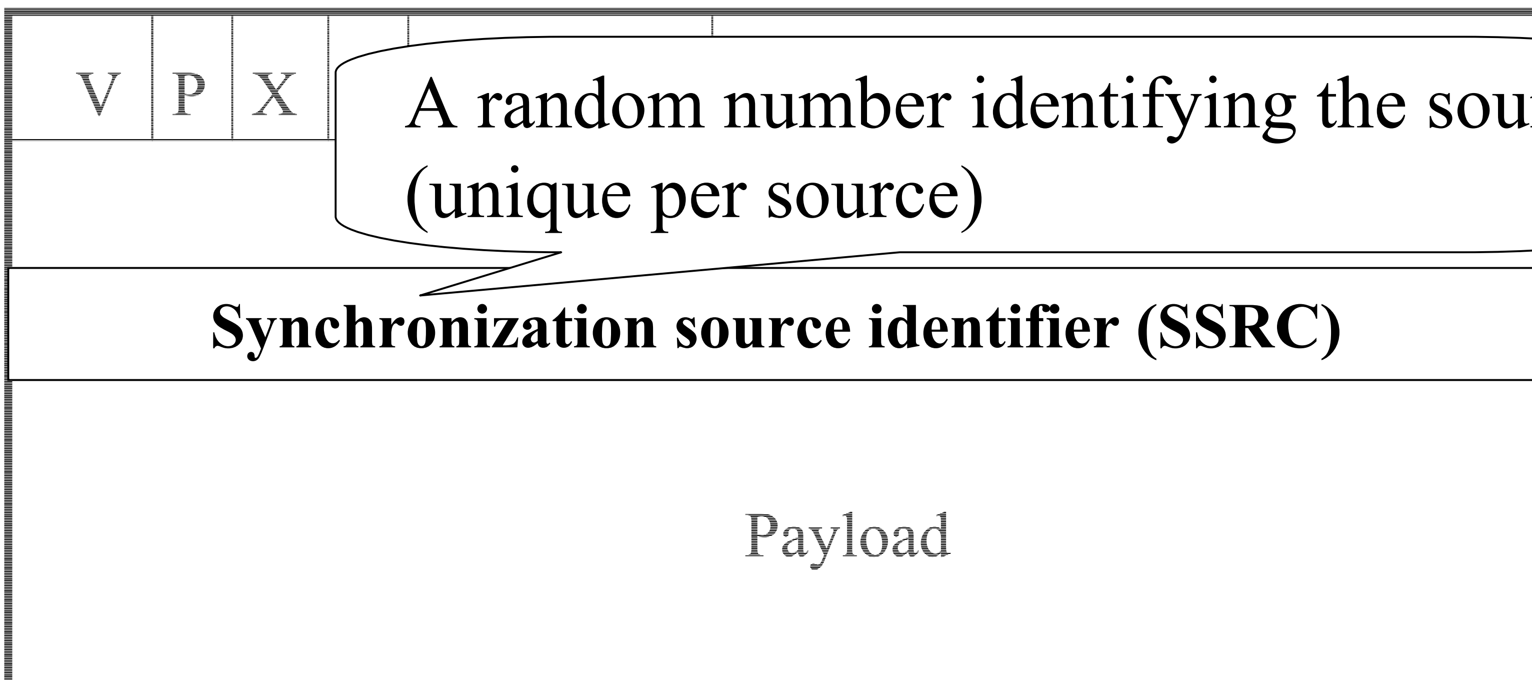
RTP: Header

Different fixed value for each compression type
(160 for 20 ms at 8000 Hz)

V	P	X	M	Payload	…quence number
Timestamp					
Synchronization Source Identifier (SSRC)					
Payload					

GMD FOKUS

RTP: Header


V
P
X
A random number identifying the source (unique per source)
Synchronization source identifier (SSRC)
Payload

Real time Transport Control Protocol (RTCP)

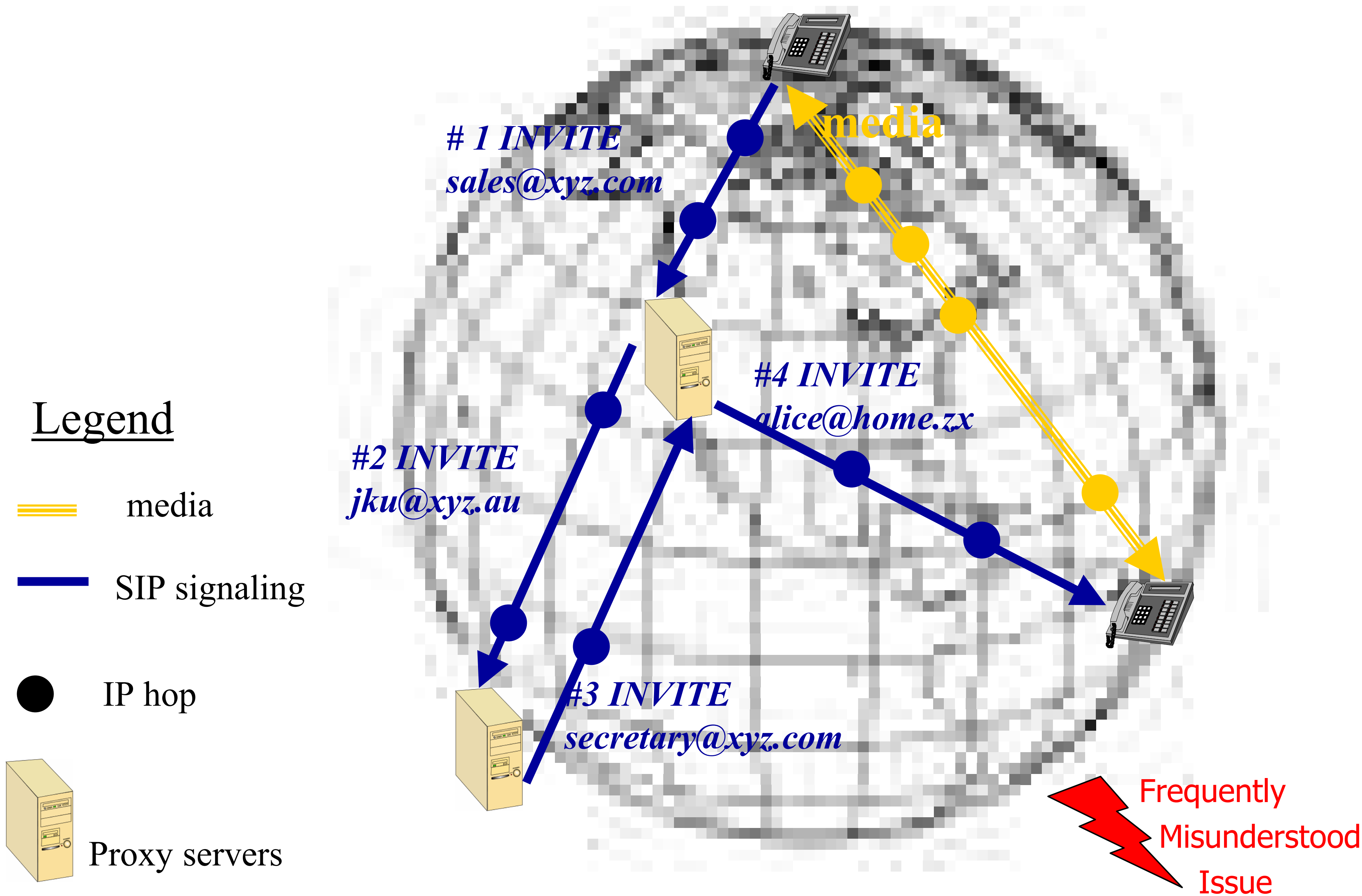
- Separate packets sent on a different port number
- Exchange information about losses and delays between the end systems
- Packets sent in intervals determined based on number of end systms and available bandwidth

GMD FOKUS

Real time Transport Control Protocol (RTCP)

- **Sender Reports**: Information about sent data, synchronization timestamp
- **Receiver Reports**: Information about received data, losses, jitter and delay
- **Source Description**:Name, Email, Phone, Identification
- **Bye**: Explicit leave indication
- **Application defined parts**: Parts for experimental functions

Media Path !=Signaling Path



SIP Proxies Have <u>NO</u> Notion of Media Path...

- SIP proxies **can not usually** control media path as there is **split between signaling and media**.
  - IP, DiffServ, and RSVP are the protocols for communication between end-devices and the network.
  - Attempts to manipulate media flows in the middle of path will tend to fail:
    - A proxy does not know all IP hops along an end-to-end media path
    - Hops may belong to foreign administrative domains.
  - Signaling and media transport (possibly w/QoS) are two different businesses.
  - A SIP proxy may be located far apart from media path.

Frequently Misunderstood Issue

GMD FOKUS

... and Attempts to Do So Would Be Difficult to Deploy

- For generality, extensibility and performance purposes, proxies do not parse SDP.
- Even if they did, their operation might result in failure as new extensions (e.g., new codecs) or entire payload types are introduced by end-devices.
- Even with SDP knowledge, proxies do not know entire media flow selectors -- SDP indicates only destination address of media streams.
- SDP may be encrypted.
- Unless route recording used, subsequent SIP requests (including ACK w/SDP) may take completely different path.

- Exception to the rule: firewall control
  - better than embedded ALGs
  - firewalls located in the same administrative domain as a call party and its SIP proxy
  - the construct still suffers from shortcomings listed previously

Frequently Misunderstood Issue

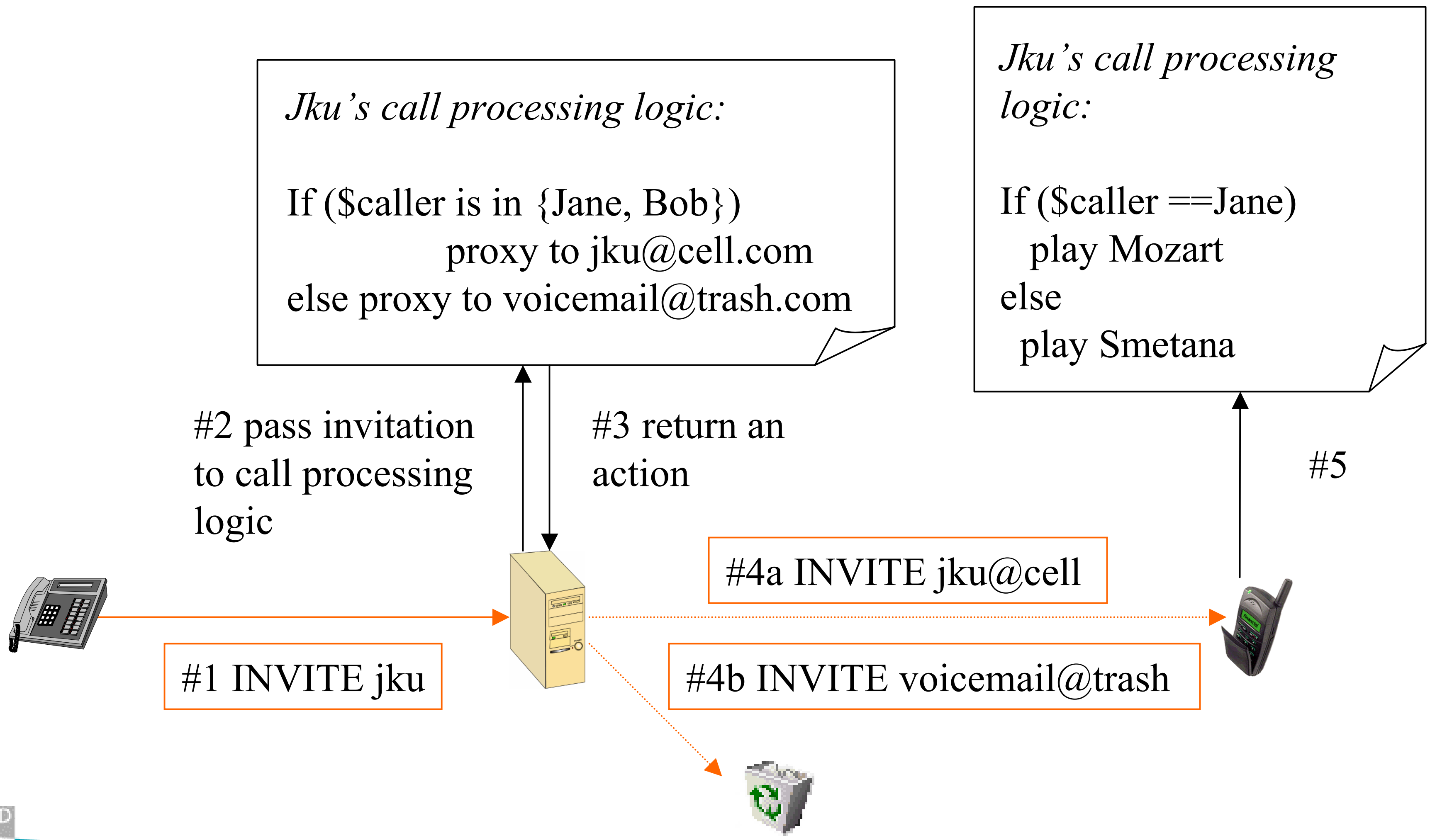
GMD FOKUS

Programming SIP

Programming SIP

- Examples
  - "discard all calls from Monica during my business hours"
  - "redirect authenticated friends to my cell phone, anyone else to my secretary"
  - "if busy, return my homepage and redirect to recorder"
- Users and third parties may program
- SIP follows HTTP programming model
- Mechanisms suggested in IETF: CGI, Call Processing Language (CPL), Servlets

Call Processing Logic Example



Where May Signaling Services Live?

- Some services have to live in the network:
  - call distribution
  - services for dial-up users without always-on IP connectivity
- Some services can be implemented in both places:
  - forward on busy
- Some services work best in end-devices:
  - distinctive ringing

Service Location Examples

Feature	End-device	Proxy	Network w/media
Distinctive Ringing	Yes	Can assist	Can assist
Visual call id	Yes	Can assist	Can assist
Call Waiting	Yes	No	Yes
CF Busy	Yes	Yes	Yes
CF No Answer	Yes	Yes	Yes
CF No Device	No	Yes	Yes
Location hiding	No	Yes	Yes
Transfer	Yes	No	Yes
Conference Bridge	Yes	No	Yes
Gateway to PSTN	Yes	No	Yes
Firewall Control	No	No	Yes
Voicemail	Yes	No	Yes

Source: H. Schulzrinne: "Industrial Strength IP Telephony"

CGI

- Follows Web-CGI. Unlike Web-CGI, SIP-CGI supports proxying and processes responses as well.
- Language-indpendent (Perl, C, ...)
- Communicates through input/output and environment variables.
- CGI programs unlimited in their power. Drawback: Buggy scripts may affect server easily.
- Token is passed between SIP server and CGI to keep state across requests and related responses.

Call Processing Language

- Special-purpose call processing language.
- May be used by both SIP and H.323 servers.
- Target scenario: users determine call processing logic executed at a server.
- Limited languages scope makes sure server's security will not get compromised.
- Portability allows users to move CPL scripts across servers.
- Scripts may be manually written, generated using convenient GUI tools, supplied by 3rd parties, ...

CPL Example

```
<incoming>
    <address-switch field="origin" subfield="host">
        <address subdomain-of="example.com">
        <location url="sip:jones@example.com">
        <proxy timeout="10">
                <busy> <sub ref="voicemail" /> </busy>
                <noanswer> <sub ref="voicemail" /> </noanswer>
                <failure> <sub ref="voicemail" /> </failure>
        </proxy>
        </location>
     </address>
     <otherwise>
        <sub ref="voicemail" />
     </otherwise>
    </address-switch>
 </incoming>
```

- Actions may include redirection, proxy, rejection

Java Servlets

- Compromise between security and power: still a powerful generic language but security provided by Java “sand-box”.
- Well-defined API is needed. As APIs are not IETF’s business this work moved to JAIN.
- JAIN thought to be a generic API applicable to almost any signaling (SIP, H.323, PSTN, etc.)
- http://java.sun.com/products/jain/index.html

Call Processing Tradeoffs

- Generality versus security
  - multipurpose programming languages provide a huge service space
  - but also a huge vulnerability space
- Performance versus portability
  - portable languages (CPL) need to be interpreted
    - higher processing delay
  - portability needed if services deployed at multiple servers or end-devices (e.g. if stored at USIMs)
- Recommendation
  - choice of appropriate service creation mechanism depends on deployment scenario, i.e. where the service is executed and by whom the service is maintained

Call Processing - Generality versus Security

Generality

- CGI Highest. Any binaries may be executed.
- Servlets Medium. All commands known to Java Virtual Machine may be executed.
- CPL Lowest. Only CPL commands may be executed.

Security by

	language	RT code verification	admin. policy
CGI	✗	✗	✓
Servlets	✗	✓	✓
CPL	✓	✓	✓

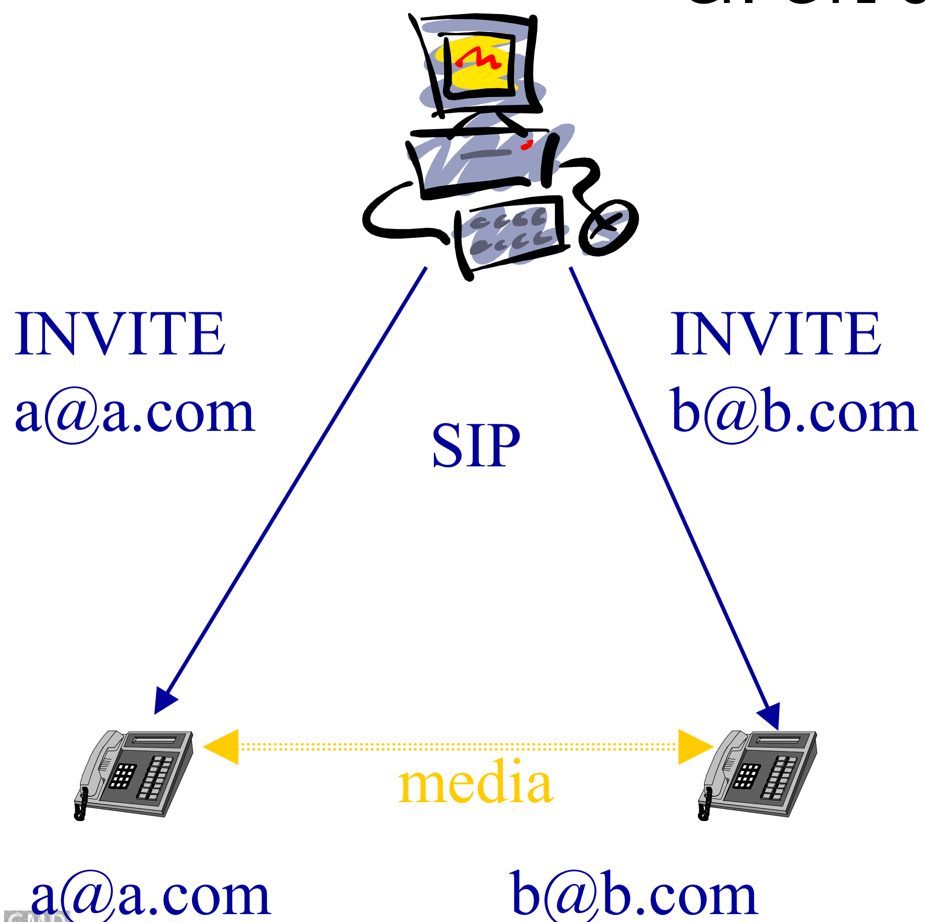
Other Work

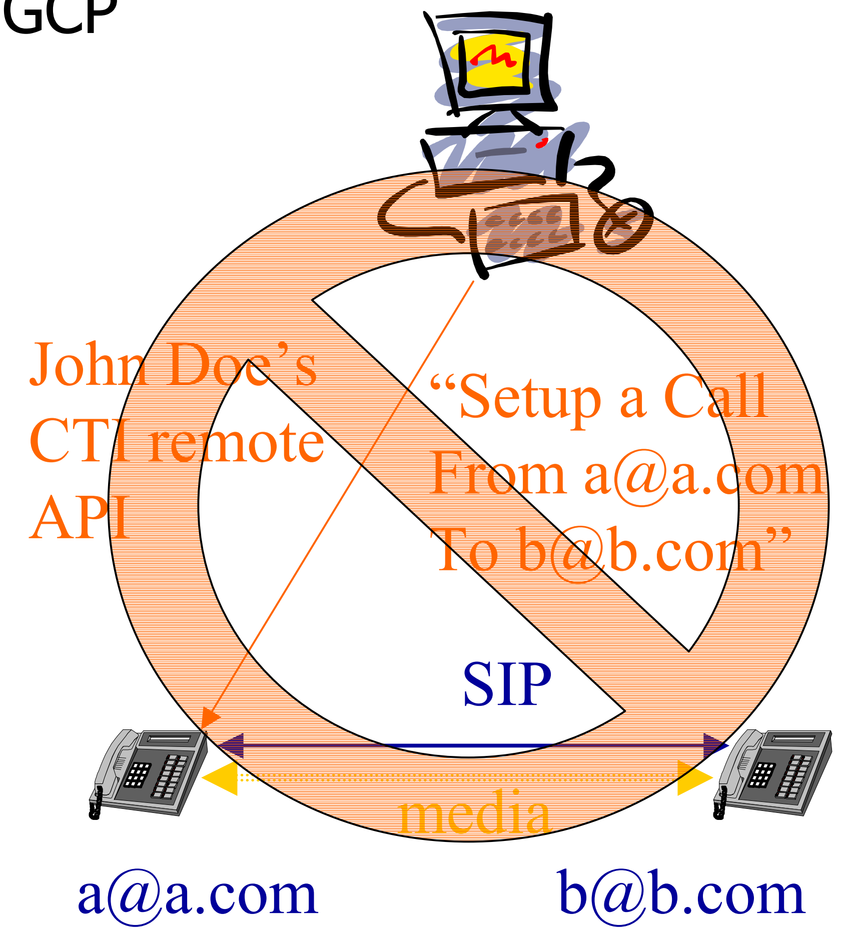
- There seems to be a huge interest in creating call control APIs. Other efforts include:
  - Parlay
  - TAPI
  - JTAPI
  - CTI
  - ...

SIP Can Be Easily Used as "Control Protocol"

Frequently Misunderstood Issue

Cf. CTI or GCP



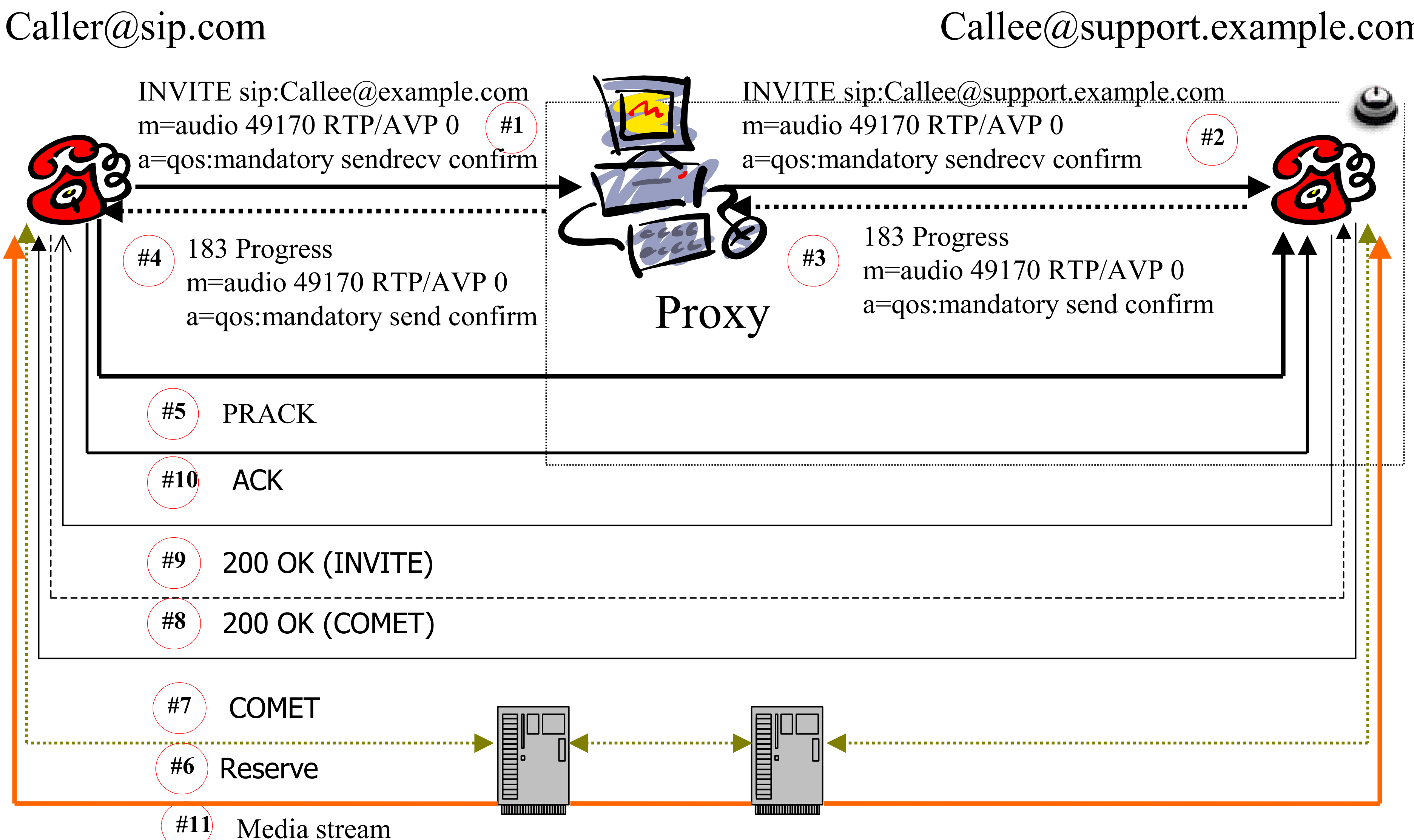


SIP & QoS

QoS: SIP and QoS Control

- SIP DOES NOT provide QoS support.
- QoS is coupled with SIP through the notion of preconditions.
- Objective is to ensure that resources are made available before the phone rings.
- Invitations might indicate in SDP that QoS assurance is mandatory.
  - Call setup should only proceed after satisfying the preconditions
- SIP extended method (COMET) indicates the success or failure of the preconditions.

SIP and QoS Control



SIP and Mobility

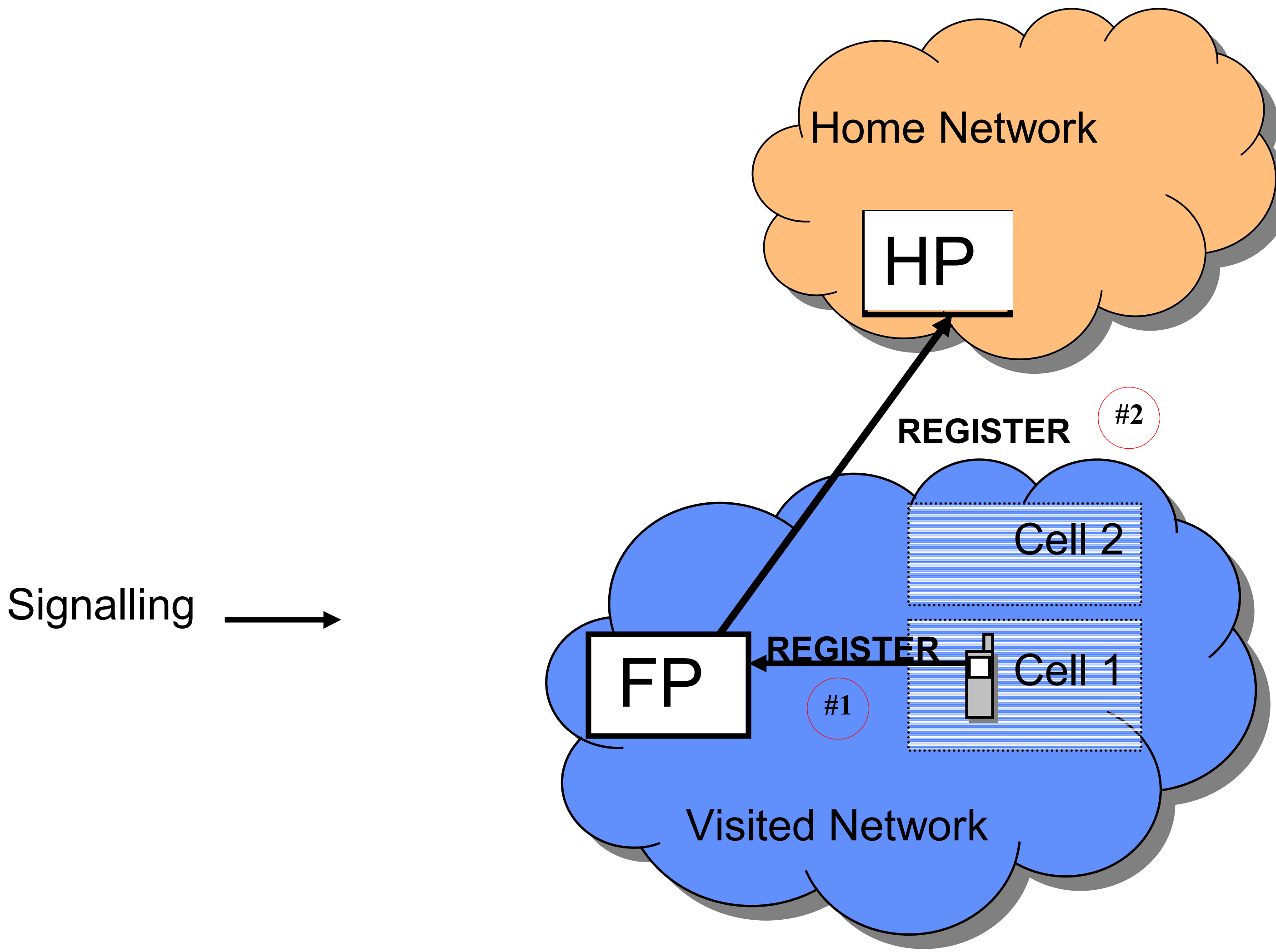
SIP and Mobility

- SIP-based mobility support
- SIP and Mobile-IP
- SIP in 3G Networks

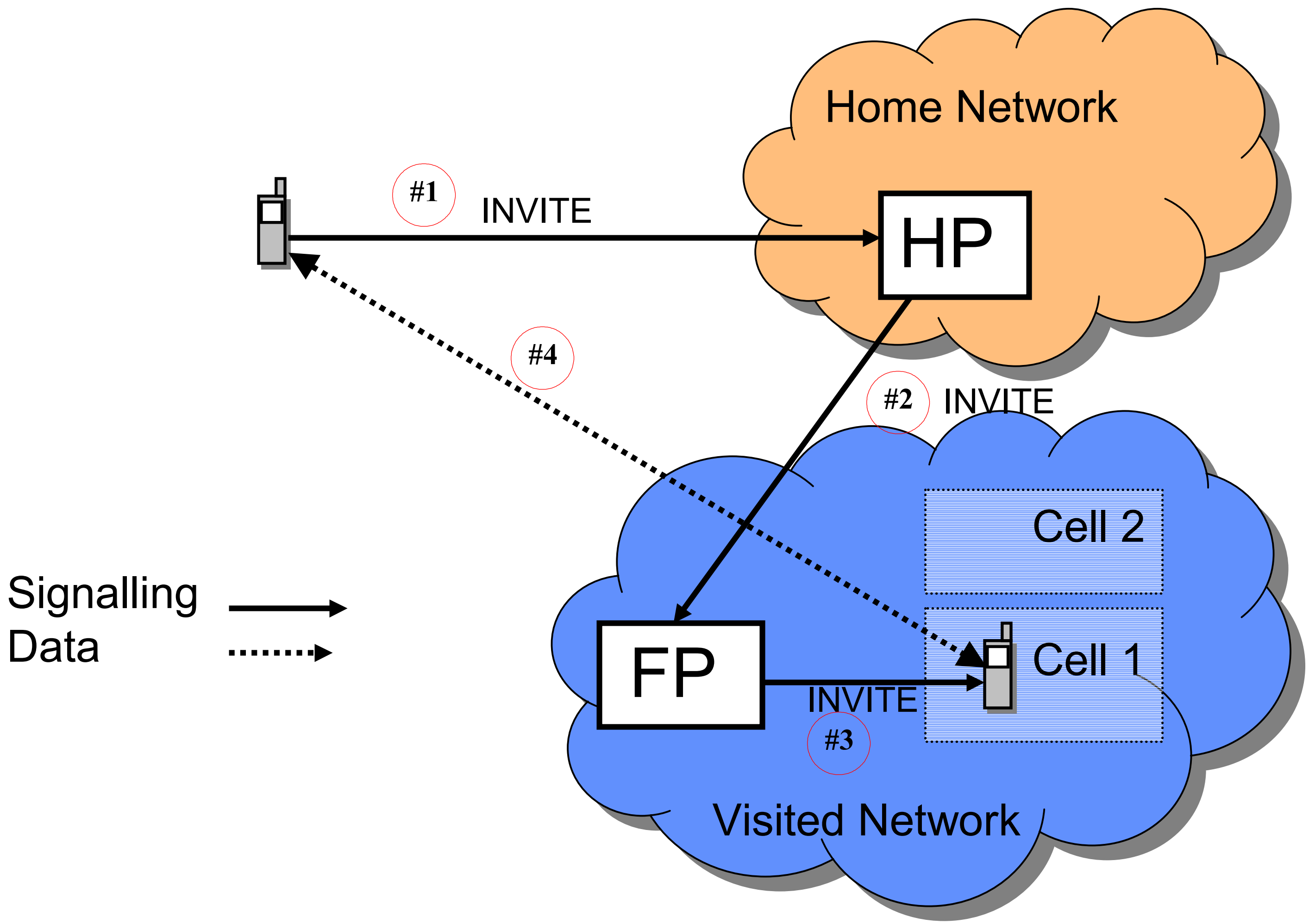
SIP and Terminal Mobility

- Terminal can move between subnetworks
- Realised today with GSM and wireless LAN
- Issues to consider:
  - Handoff performance
  - Redirection authentication
- Mobile hosts (MH) inform their home proxy about their new locations using REGISTER
- Mid-call mobility (*Session mobility*) is dealt with using reINVITE

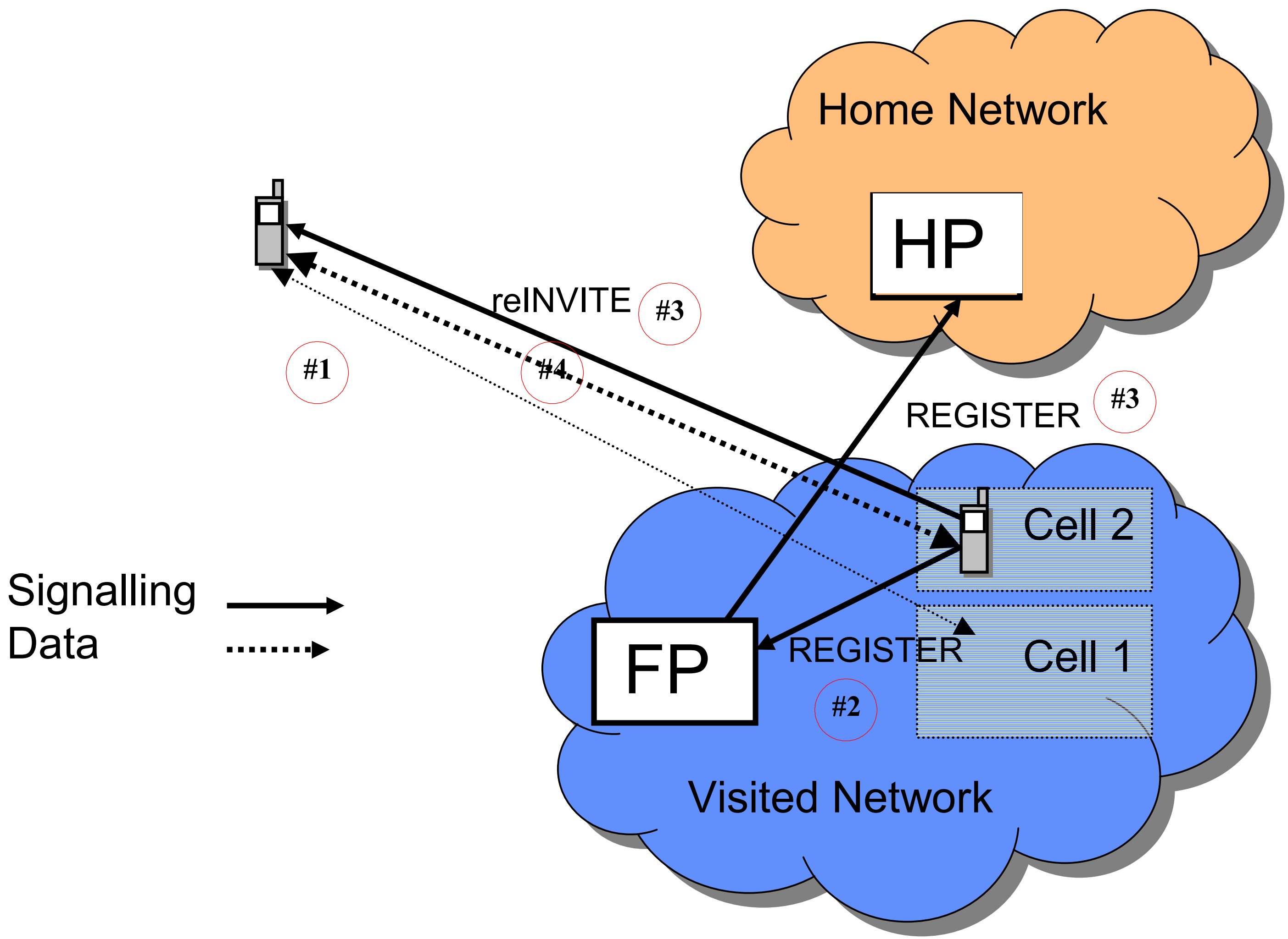
SIP and Terminal Mobility



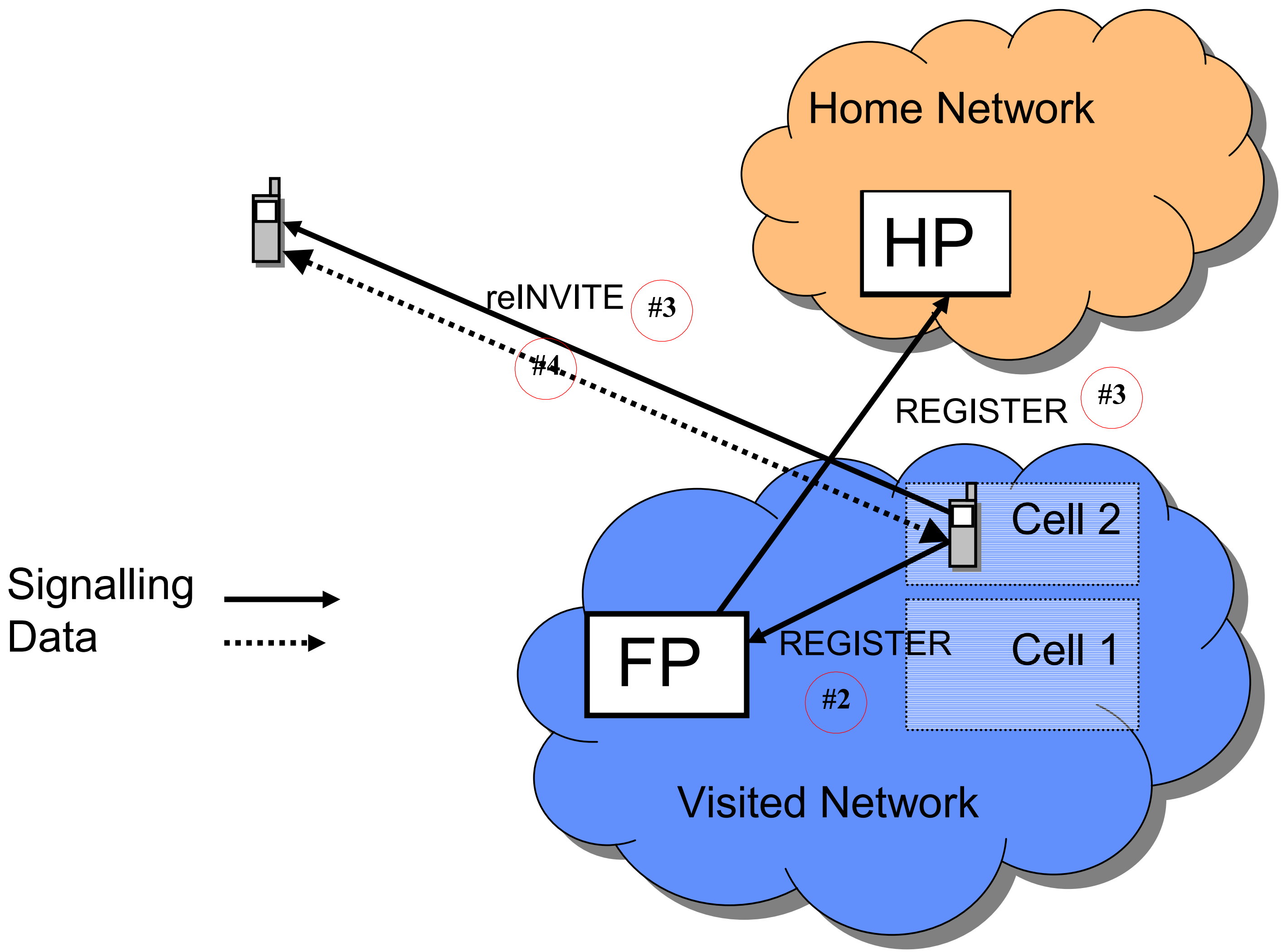
SIP and Terminal Mobility



SIP and Terminal Mobility



SIP and Terminal Mobility



SIP and Personal Mobility

- Person uses different Devices and possibly address
- REGISTER binds a person to a device
- Proxy and redirect translate address to location and device
- Issues to consider:
  - Authentication: finger print, IR, password ..
  - Binding different addresses to single person: LDAP ..

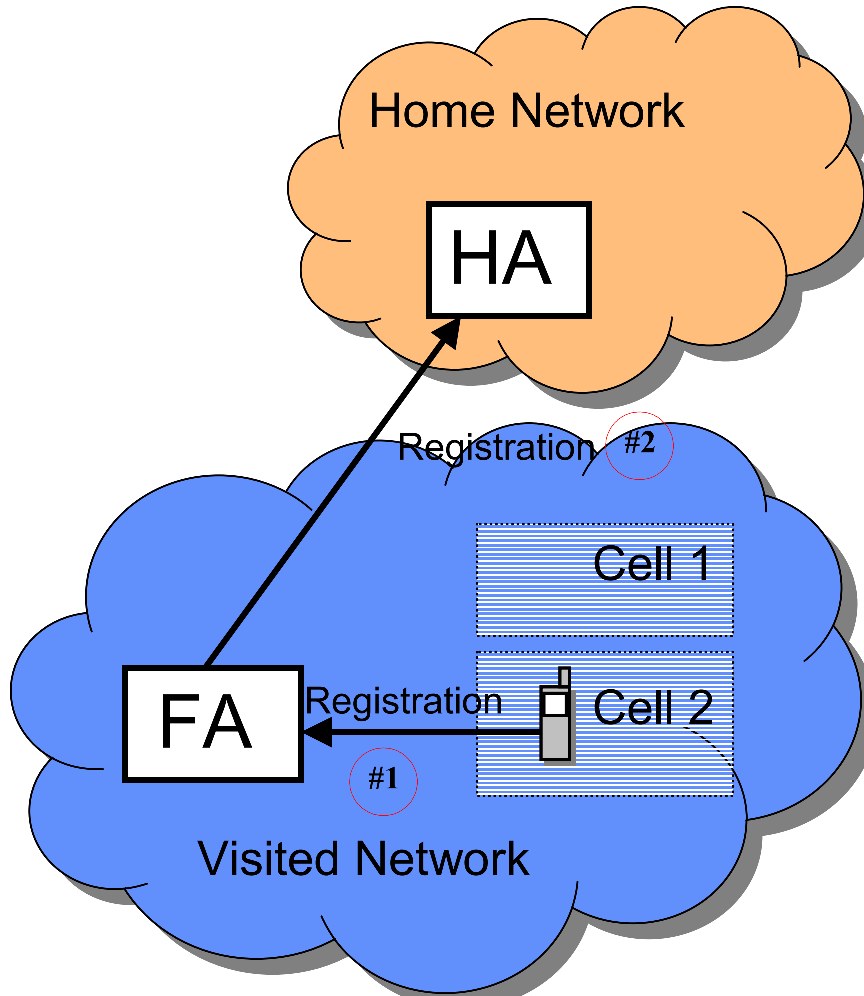
SIP and Service Mobility

- Use same services from different locations and devices
  - Speed dial, address book, media preferences, call handling
- Services located at home server
  - RECORD-ROUTE home proxy to force calls to be processed by home servers
  - Services located at end systems
  - retrieve with REGISTER
- Issues to consider
  - Services need to be device independent: standardised service description (CPL) ..
  - User recognition and authentication

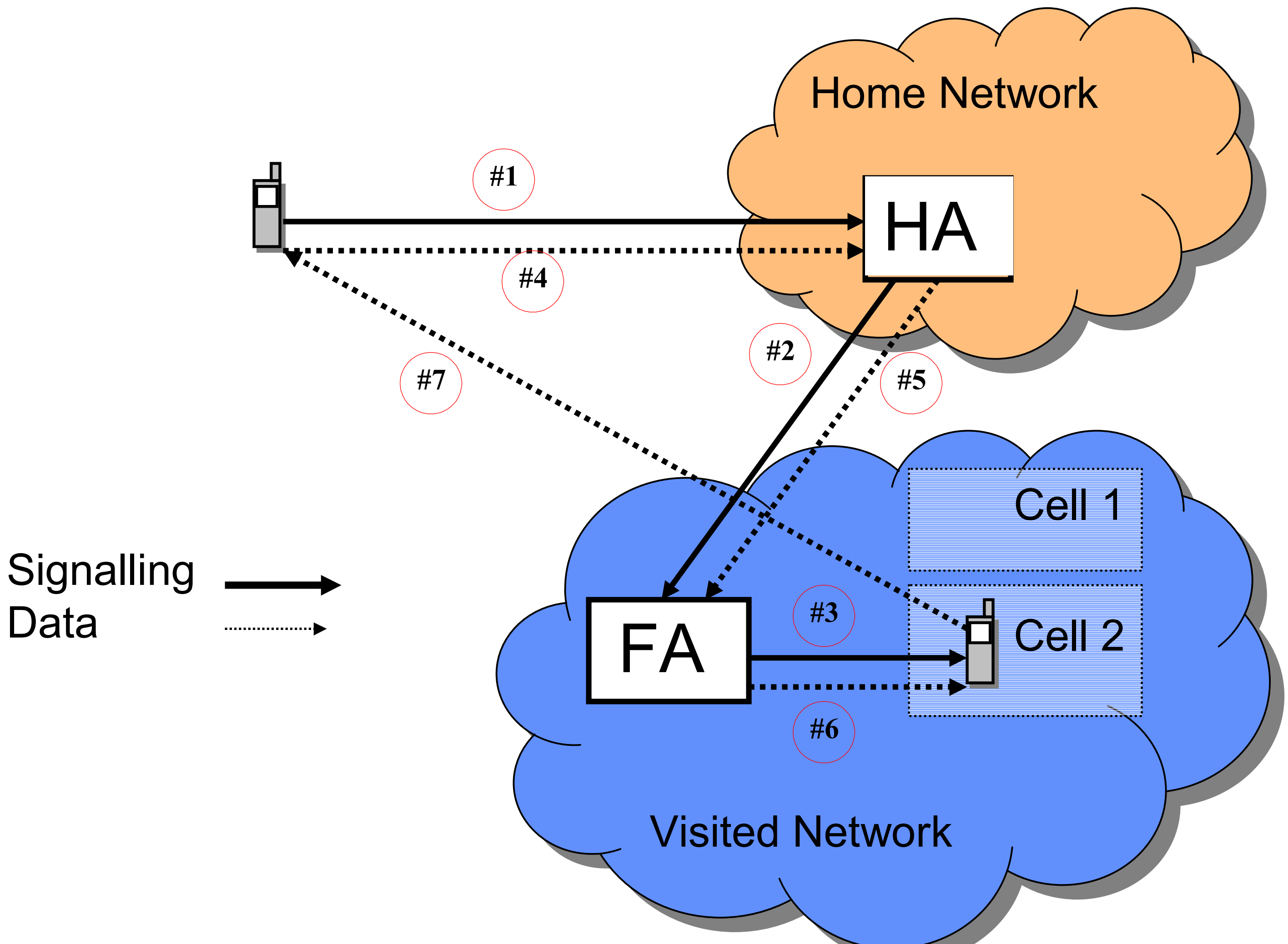
SIP and Mobile-IP

- Mobile-IP is a well established standard for mobile communication in the Internet
- Allow hosts to be reached under the same address regardless of location
- Mobile hosts register a care-of-address with home agent
- Correspondent nodes (CN) send data to home agent
- Home agent tunnels traffic to care-of-address
- MH sends traffic directly to CN
- Triangular routing increases delay
- Tunnelling increases bandwidth consumption

Mobile-IP (Registration)



Mobile-IP (Communication)



SIP and Mobile-IPv6

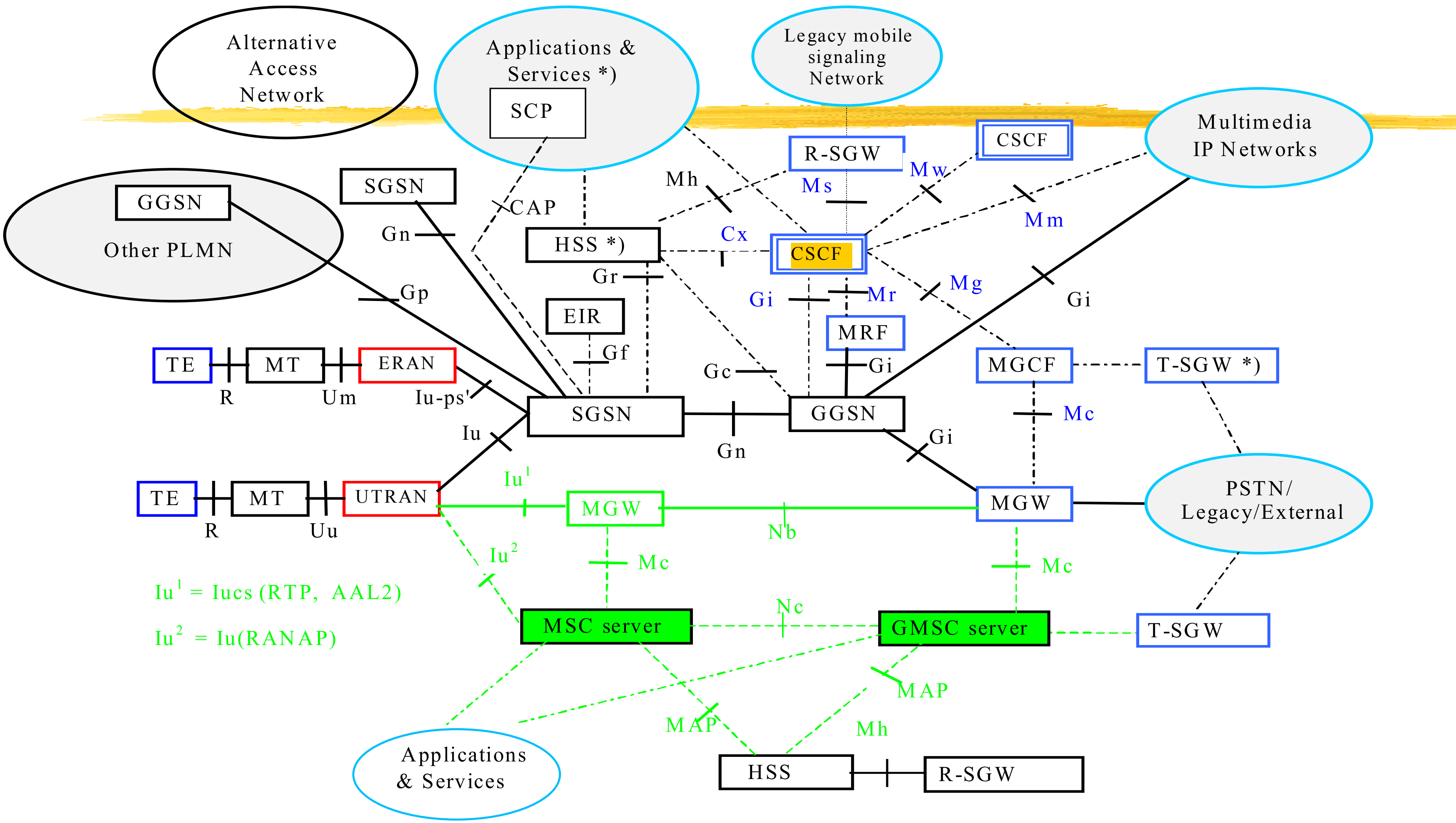
- IPv6 is especially interesting for mobile Internet
- Mobile-IPv6 uses Binding updates similar to SIP registration and reinvitations to avoids triangular routing
- Use routing header option to avoid tunnelling
- Could be a solution for providing a unified protocol for mobile data and voice communication??

3GPP Networks: Introduction

- 3GPP consortium consists of ETSI, ARIB, TTA, T1 and CWTS
- UMTS R00 is an All-IP architecture with support for CS terminals
- Architecture based on GPRS with multimedia enhancements
- SIP is used for establishing and terminating IP-telephony calls
- H.248 is used for gateway control
- Support for integration of intelligent services

GMD FOKUS

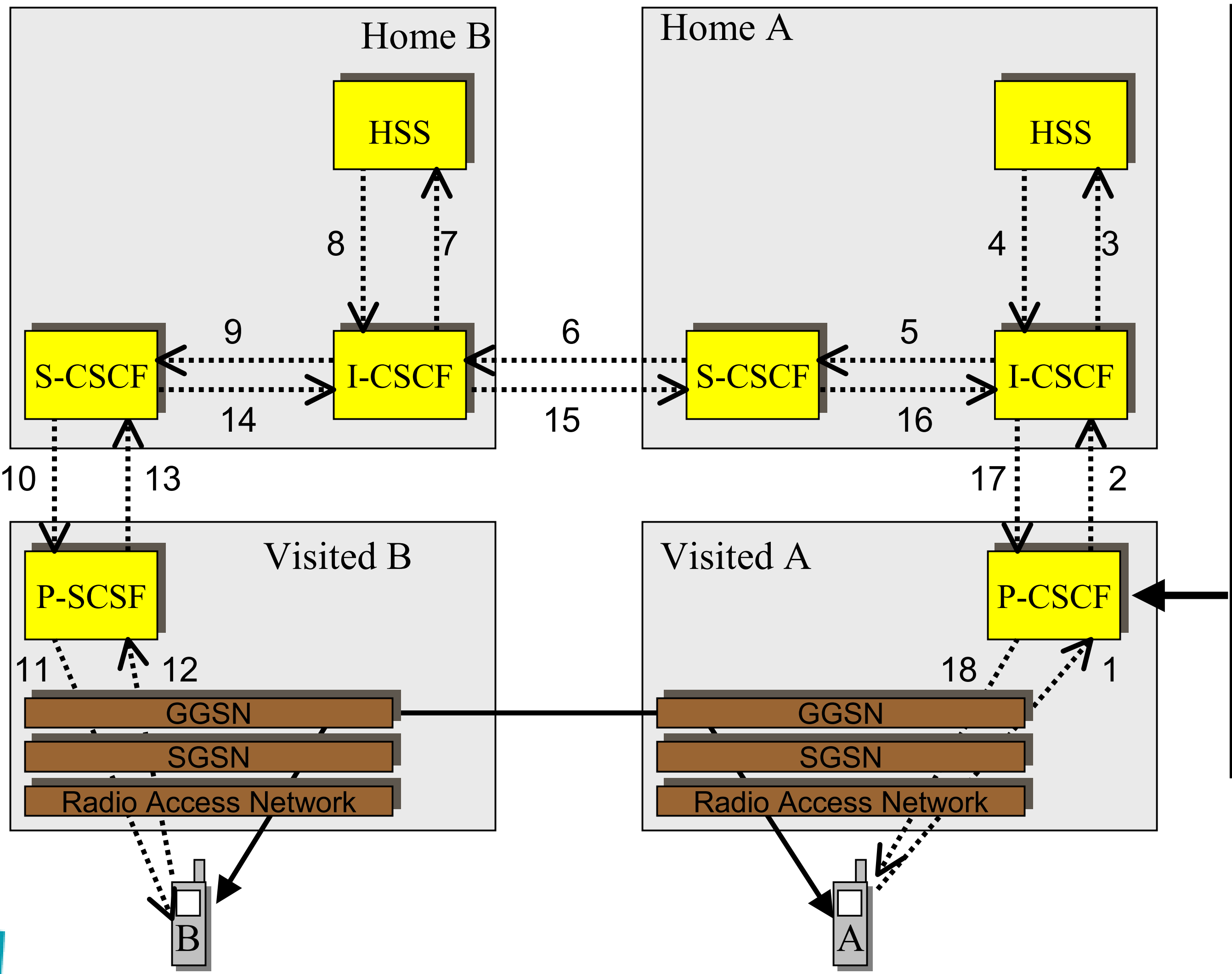
3GPP: Architecture



Signalling Interface

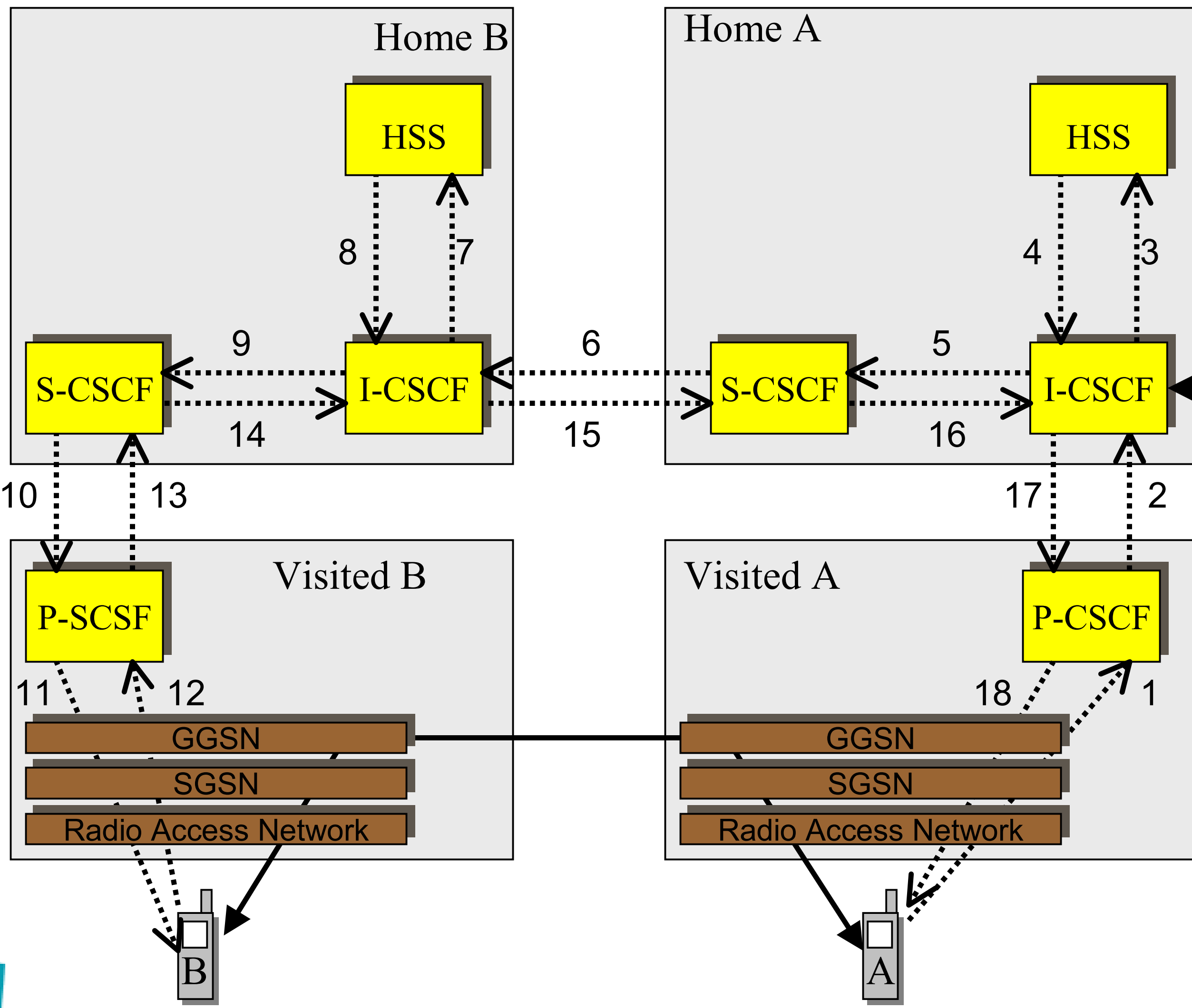
Signalling and Data Transfer Interface

3GPP: Proxy CSCF



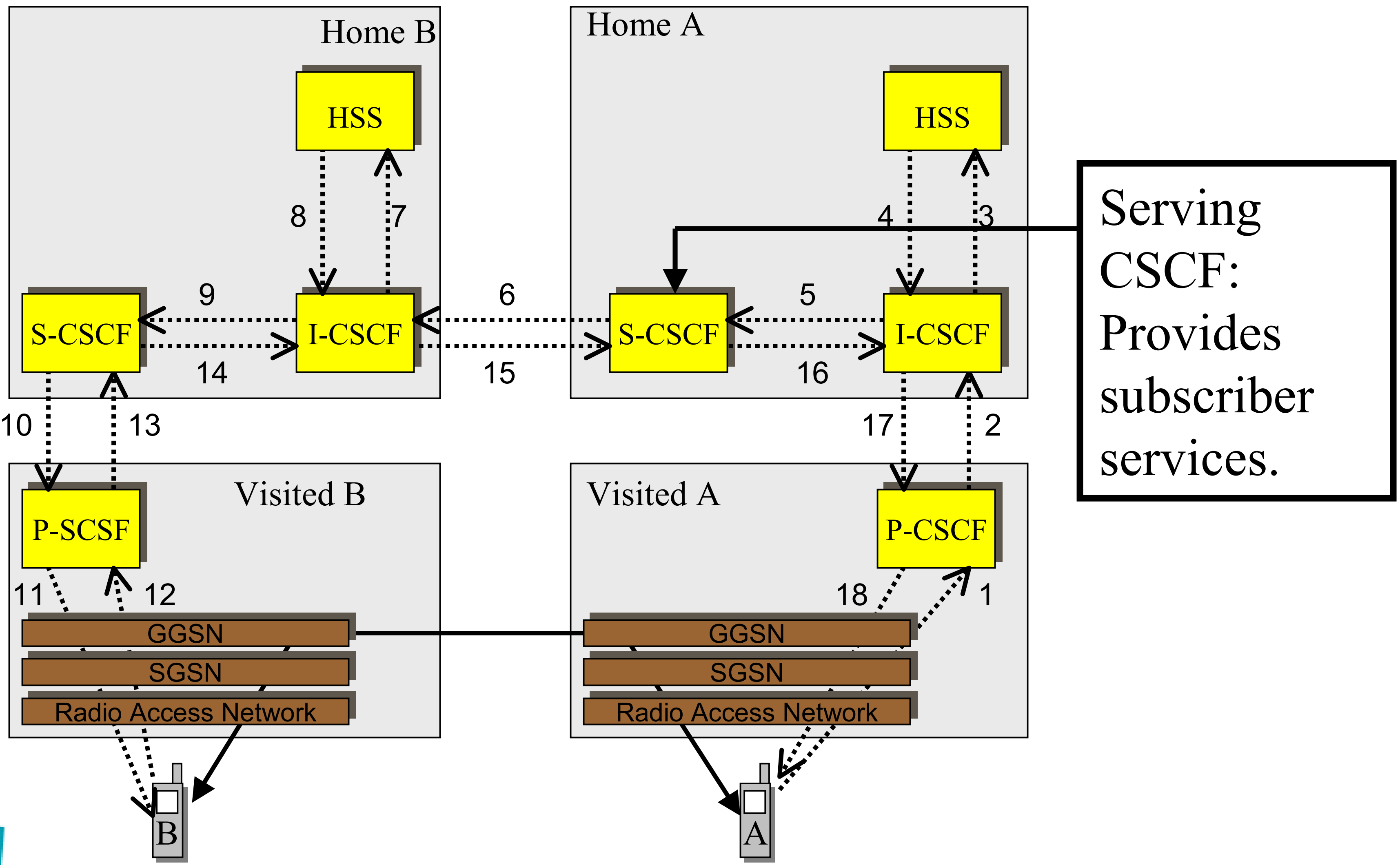
Proxy CSCF: Provides emergency service breakout, triggers for locally-provided services, and number normalizing (per local dialing plan)

3GPP: Interrogating CSCF



Interrogating CSCF: Queries the HSS to find the correct S-CSCF. First point of contact for incoming call signalling.

3GPP: Serving CSCFs



SIP vs H.323

Outline

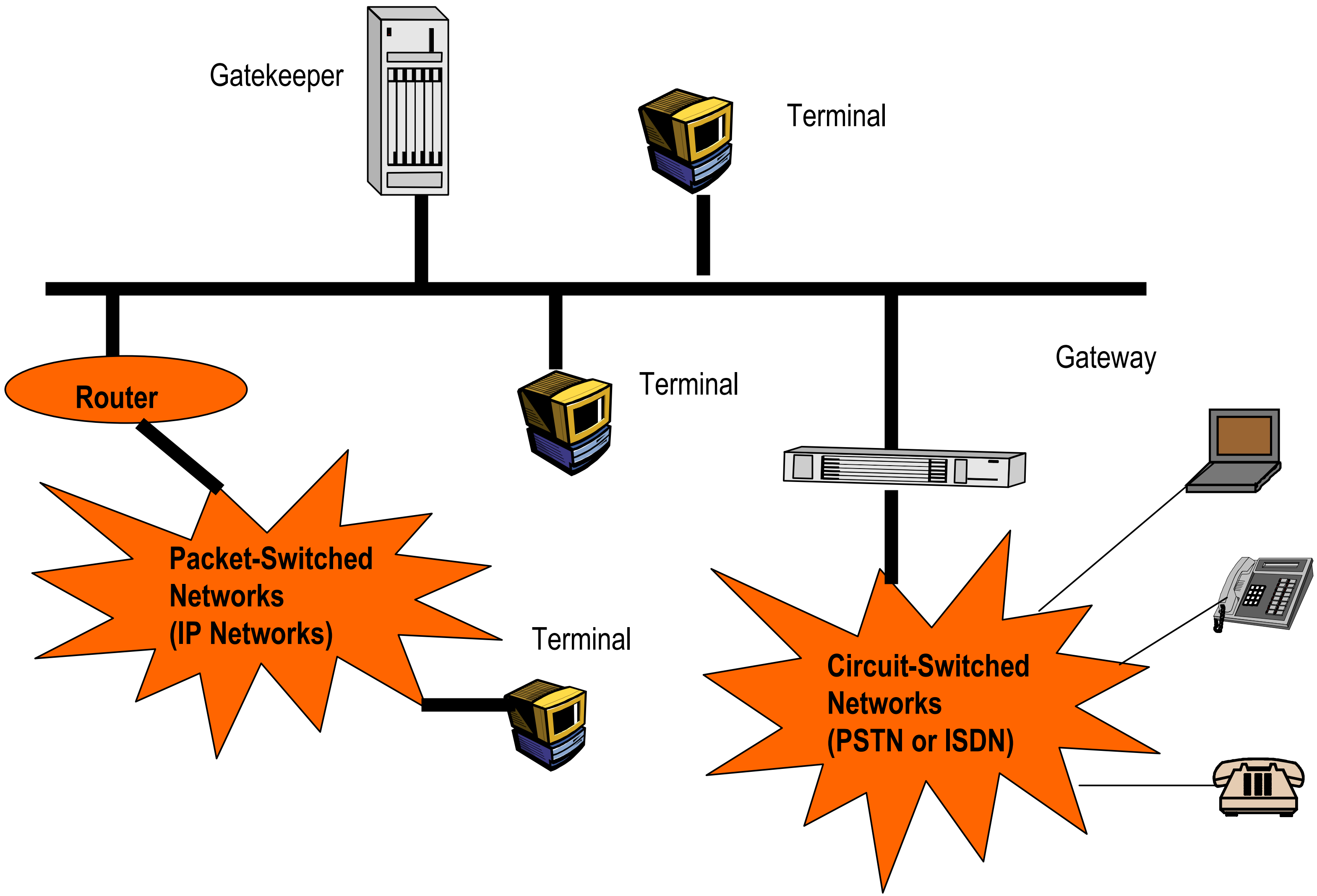
- H.323 overview
- H.323/SIP comparision
  - Functionality
  - Quality of Service
  - Scalability
  - Flexibility / Extensibility
  - Implementation
- Summary

GMD FOKUS

H.323 overview

Name	Description of protocols
H.323	Specification of the hole system
H.225.0	Call Control, Call Setup
H.235	Security protocol for authentication etc.
H.245	Capability exchange and mode switching
H.450	Supplementary services
H.246	Interoperability with circuit-switched networks
H.332	For large size conferences
H.26x	Video codecs
H.7xx	Audio codecs

H.323 Endpoint types



H.323/SIP Comparison

	H.323	SIP
Architecture	Stack	Element
Origin	ITU	IETF
Transport	Mostly TCP	Mostly UDP
Encoding	ASN.1	HTTP-like
Emphasis	Telephony	Multimedia, multicast
Address	Aliases	SIP URLs

H.323 vs. SIP: Basic Call Control

Service	H.323v1	H.323v2	H.323v3	SIP
Call hold	No	Yes	Yes	Yes
Call transfer	No	Yes	Yes	Yes
Call forward	No	Yes	Yes	Yes
Call waiting	No	Yes	Yes	Yes

H.323 vs. SIP: Advanced features

Service	H.323v1	H.323v2	H.323v3	SIP
Third party call	No	No	No	Yes
Conference	No	Yes	Yes	Yes
Click-to-dial	No	Yes	Yes	Yes
Capability exchange	Yes+	Yes+	Yes+	Yes

GMD FOKUS

H.323 vs. SIP QoS

Service	H.323v1	H.323v2	H.323v3	SIP
Call setup delay	6-7 RT	3-4 RT	2.5 RT*	1.5 RT
Packet Loss recovery	TCP	TCP	Yes+	Yes+
Loop detection	No	No	PathValue	Via, hops
Fault tolerance	No	No	backup	Yes

GMD FOKUS

* *mixed-mode transport may gain an advantage compared to SIP's UDP-to-TCP fallback*

H.323 vs. SIP Scalability

H.323

- Interaction between many sub-protocols make it very complex
- Stateful servers in Version 1+2
- H.323v3 more complex

SIP

- SIP and SDP are less complicated
- Servers can be stateless

H.323 vs. SIP Extensibility of functionality

H.323

- Only *NonStandardParm* field useful (consists of vendor codes)
- New features could be supported using H.450.1 generic functions

SIP

- Hierarchical namespace of features
- Hierarchical error codes
- New features can be registered with IANA
- Transparent proxying
- Arbitrary MIME Types
- SUPPORTED, REQUIRED, OPTIONS protocol elements

H.323 vs. SIP Ease of customization

H.323

- Interaction between protocols makes customization complicated
- Full compatibility with all version must be guaranteed (more code)

SIP

- Handled by simple header field
- Unknown header fields can be ignored

H.323 vs. SIP Transport Protocol neutral

H.323

- Not before Version3
- Support for TCP/UDP in H.323v3

SIP

- Can use any transport protocol

H.323 vs. SIP Ease of Implementation

H.323

- H.323 messages are binary
- Encoded using ASN.1
- Special parsers needed to map into readable form and vice versa
- Implementation and debugging complicated

SIP

- SIP messages are text-based (unicode supported)
- Easy implemented in Perl, Tcl, Java
- Easy debugging: tcpdump, ngrep, netcat, ...

Summary: SIP versus H.323

H.323

- Deployment started earlier
- Shorter messages

SIP

- Scalability
- Extensibility
- Less Complexity
- Ease of Implementation
- Customization
- Call forking
- Third-party call control

Note: implementations of SIP-H.323 signaling gateways available! Transition to SIP while preserving investments in existing infrastructure possible.

SIP Robustness

Robust Protocol Design

- Robustness determined by state maintenance model
- Amount of state in SIP Servers minimized
  - servers may be stateless (SL) or maintain transaction state (TS) or session state (SS)
  - less state the more robustness; failure of a SL or TS proxy does not affect existing sessions
  - transactional state is needed to enable services such as forking/forward-on-busy or if SIP runs over TCP
  - session state may be needed for maintaining firewalls or generating failure-resistant CDRs; keep-alive possible using re-INVITEs and session timer
- SIP INVITEs convey full signaling state
- Subsequent messages may take different path

DNS for Failure Recovery & Load Balancing

- Unavailable SIP servers can be dealt with using DNS in the same way as mail servers are:
  - DNS servers maintain multiple prioritized SRV entries
  - callers initiate calls to high-priority server; if unavailable, they proceed to lower-priority server
- Load balancing can be accomplished similarly
  - DNS servers maintain multiple SRV entries with equal priority
  - a random pick is chosen out of the server list
- Notes on DNS
  - it's good do have multiple DNS servers for each zone of authority;
  - DNS may be a pain ...

Other Load Balancing Methods

- A front-end proxy may dispatch calls to a proxy farm
- Load-balancing NAT may be used
- Call processing logic may be off-loaded to end-devices

Interoperability

- Interoperability events “SIP Bake-offs” three times a year
- 6th bake-off took place in December 2000
  - 57 companies, 202 attendees
  - complex test scenarios demonstrated
  - “torture tests” conducted
- *http://www.cs.columbia.edu/~hgs/sip/bakeoff/*

Trademark War

- Pillsbury, through the law offices of Fulbright & Jaworski, has demanded that Columbia and other users of the term "bake-off" cease doing so, claiming that it infringes on their trademark.
- *http://www.cs.columbia.edu/~hgs/sip/bakeoff/pillsbury.html*

SIP Security

Internet Security

- Internet is open
  - anyone with Internet access may try to attack anyone else
  - increasing complexity and programmability results in lots of easily exploitable bugs
  - packets can be dumped anywhere in the middle of packet path
- Security of both users and providers inherently suboptimal

Security Services

- Availability
  - subject to Denial of Service Attacks: burdening servers with enormous load, uploading hostile applications, physical violence
  - difficult to beat: self vs. non-self problem
- Privacy
  - prevents unauthorized persons from inspection of both signaling and media
  - can be solved using encryption
  - problems: encryption computationally expensive; key exchange protocols needed; no PKI available

Security Services

- Message Integrity
  - prevents unauthorized users from changing packets
  - can be solved using Message Authentication Checks
- User Authenticity
  - prevents unauthorized users from using someone's else identity to fool other users or accounting & charging systems
- Anonymity
  - prevents other call parties from knowing who is calling

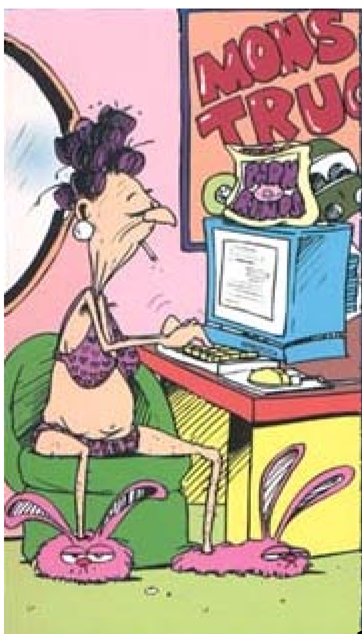
Disclaimers & Problems

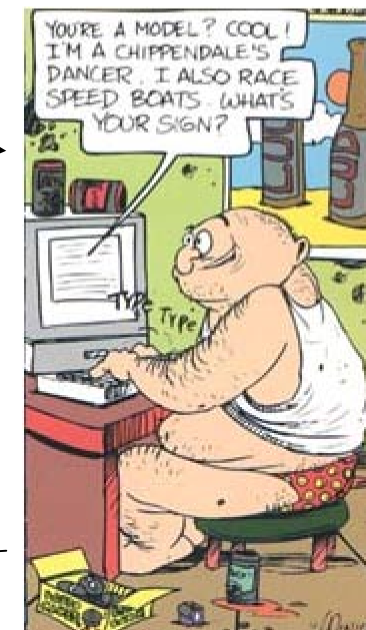
- Disclaimer #1: Protocol security is only a piece of the big picture; security of a system may always be compromised by naïve implementation or administration.
- Disclaimer #2: Security of a single protocol does not help; all participating protocols have to be made secure.
- Disclaimer #3: Physical security counts as well!!!
- Disclaimer #4: Security protocols cannot solve social-layer issues.

Disclaimer #4

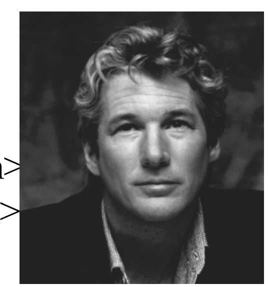
SIP INVITE w/JPEG

INVITE sip:UserB@there.com SIP/2.0
Via: SIP/2.0/UDP here.com:5060
From: BigGuy <sip:UserA@here.com>
To: LittleGuy <sip:UserB@there.com>
Call-ID: 12345600@here.com
...





200 OK w/JPEG



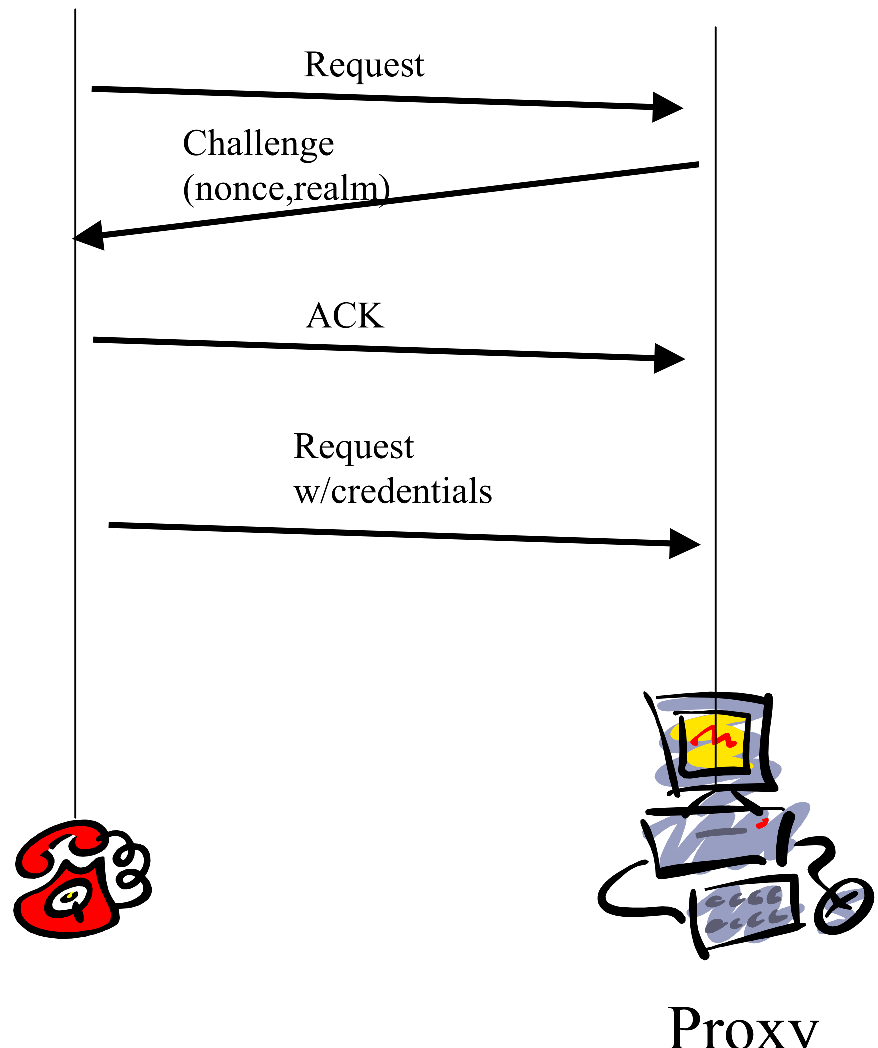
SIP/2.0 200 OK
Via: SIP/2.0/UDP here.com:5060
From: BigGuy <sip:UserA@here.com>
To: LittleGuy <sip:UserB@there.com>
Call-ID: 12345601@here.com...

Signaling Security

- End-2-End Security
  - cannot cover entire signaling -- fields needed for routing have to be visible
  - no intermediate proxies can corrupt security
  - mechanisms: basic and digest authentication, PGP
- Hop-by-Hop Signaling Security
  - requires belief in transitive trust
  - immense computational stress on servers if public-key used
  - can deal with firewalls/NATs
  - may cover entire signaling
  - mechanisms: ipsec, TLS
- Combination of both may be used
- Keying: no established solution

SIP Authentication

- The most needed part of the security picture.
- Protocols: Basic, Digest, PGP
- All of them challenge-response, Basic & Digest use shared secret



Media Security

- Encryption of media content
- May take place either at IP or RTP layer
- Performance overhead considerable
- No established solutions for keying

Firewall Traversal

Outline

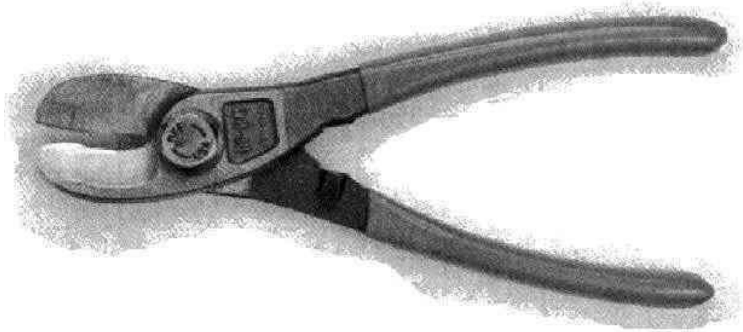
- Where firewalls cause problems to Internet telephony
- Where NATs cause problems to Internet telephony
- Mapping solution space
- Our proposal: link SIP proxies to firewalls/NATS
- Report on IETF efforts
- Conclusions

Firewall Traversal

- Firewalls static
  - protect networks by enforcing a restrictive packet filtering policy
  - frequently deployed in corporate networks
  - policy permits flows from and to trusted addresses
- Internet telephony dynamic
  - signaling conveys dynamic addresses and port numbers
  - 3-rd party call control
  - user mobility
- Problem
  - signaling static and easy
  - static firewalls do not know and permit dynamic media packet flows
  - changing policy to default-permit-explicit deny seriously changes security model -- not a valid solution
  - trade-off between sufficiently restrictive policy and accommodating applications needs sought

Ultimately Secure Firewall

Installation Instructions: For best effect install the firewall between the CPU unit and the wall outlet. Place the jaws of the firewall across the power cord, and bear down firmly. Be sure to wear rubber gloves while installing the firewall or assign the task to a junior system manager. If the firewall is installed properly, all the lights on the CPU will turn dark and the fans will grow quiet. This indicates that the system has entered a secure state. For Internet use install the firewall between the demarc of the T1 to the Internet. Place the jaws of the firewall across the T1 line lead, and bear down firmly. When your Internet service provider's network operations center calls to inform you that they have lost connectivity to your site, the firewall is correctly installed. (© *Marcus Ranum*)



NAT Traversal

- NATs
  - conserve IP space by transparent IP address sharing
  - NAT-PT can be deployed on boundaries between IPv4 and IPv6
  - various flavors (NAPT) and applications (load balancing, renumbering avoidance)
  - problems: session addresses indicated in signaling (SDP, Contact:, Route:, Record-Route:) do not match NAT-ed addresses; sessions fail to get established
- Solution:
  - Eliminating the need for NATs by mass introduction of IPv6 unlikely to happen in near future
  - RSIP experimental
  - Use application patchwork, Application Level Gateways, to resynchronize applications with IP/transport

Where FWs/NATs affect SIP

INVITE sip:UserB@there.com SIP/2.0
Via: SIP/2.0/UDP 192.168.99.1:5060
From: BigGuy <sip:UserA@here.com>
To: LittleGuy <sip:UserB@there.com>
Call-ID: 12345600@here.com
CSeq: 1 INVITE
Subject: Happy Christmas
Contact: BigGuy <sip:UserA@192.168.99.1>
Content-Type: application/sdp
Content-Length: 147

v=0
o=UserA 2890844526 2890844526 IN IP4 here.com
s=Session SDP
c=IN IP4 100.101.102.103
t=0 0
m=audio 49172 RTP/AVP 0
a=rtpmap:0 PCMU/8000

- Contact, From, To address header fields
- Via header fields (received tag)
- Route and Record-route
- SDP payload

GMD FOKUS

The FW/NAT Problem Summary

- Implications: No users behind firewalls/NAT can interoperate with other Internet users
- Problem Size: Unknown, probably huge
  - nobody knows how many users are behind FWs/NATs
  - IP addresses shared by hosts, hosts shared by users
  - hugely deployed by enterprises, some ISPs deploy NATs as well
  - Brian Carpenter (January 2001): "*My hand waving estimate is that 40% (160M) of users are behind a firewall and/or NAT, 50% (200M) on dial-up, and 10% (40M) have direct always-on access. But there is no way I can justify these numbers.*"
- Solution Status: very few products have VoIP ALGs
- ALGs are no "Wunderwaffe" (all-disease-cure)
  - Firewall ALGs fail to operate if data encrypted
  - NAT ALGs fail to operate if data encrypted or authenticated
  - embedded ALGs suffer from dependency on vendor, lower performance, higher development costs
  - problems with multiple FW/NATs

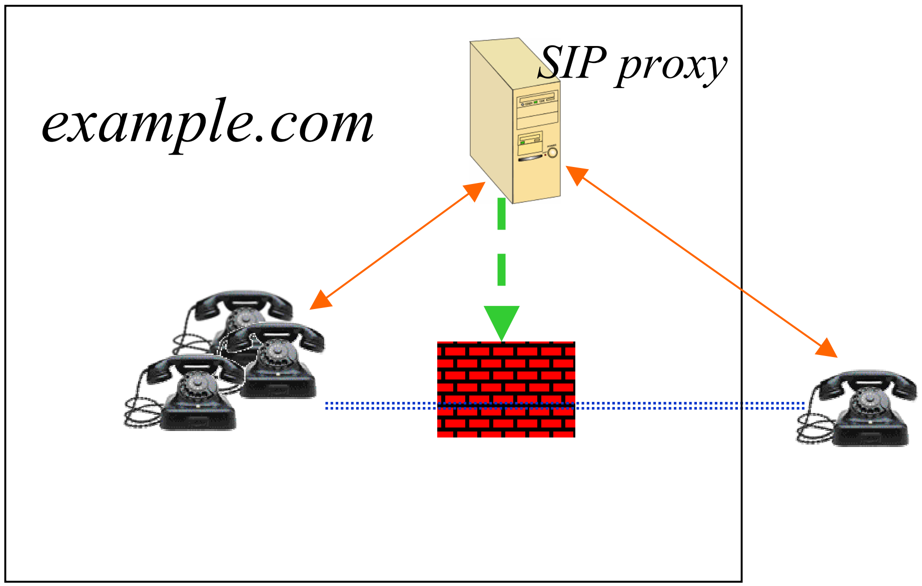
Solution Space

- Junk FW/NATs ... Unlikely to happen in near future.
- Subvert FW policy ... Not sure your admin will like it.
- Build ALGs into your FWs/NATs.
- Use external application-awareness
  - in end-devices (SOCKS/RSIP) ... Protocol stack in your appliances needs to be changed
  - in proxies

Make VoIP ALGs Easier to Live With

- Idea: split ALGs from NATs/FWs and reuse application awareness residing in SIP proxies
- Benefits:
  - intermediate network devices need to speak a single control protocol; ALG may be supplied by third parties easily; no more vendor dependency
  - existing application-awareness (e.g., SIP proxies) may be reused (as opposed to duplicating it in network devices)
  - hop-by-hop security works
- Wanted: Protocol for Reconnection of the split pieces: **Firewall Control Protocol**
  - addressed by MidCom WG

FCP Controlled Firewall/NAT



Protocol functionality

- Open and close firewall pinholes
- Allocate and release NAT translations

Legend

SIP

FCP

media streams

FCP Benefits

- Reduction of development costs
- Relieves from vendor dependency
- Hop-by-hop signaling security supported
- Likely to improve performance
- Easy to deploy

FCP Design Concepts

- Objective: easy-to-deploy
- Scope: pinhole opener versus management tool
  - our recommendation: Keep It Simple and Stupid
  - do not add additional complexity unless a need for it is clearly documented
  - retain extensibility so that new applications will be able to use it: notion of attributes
- Control Model: end-devices versus proxies
  - end-devices: huge deployment overhead and security concerns
- Layering
  - FCP application-independent
  - the cutting line splits application from transport
  - particular wrong ideas include but are not limited to:
    - making FCP controllers maintain NAT pools
    - bringing “application clues” back again to controlled devices
- Application-aware soft-state

FCP Security

- Who may act as controller?
  - Anyone with valid permissions
  - Protocol specification does not dictate if it is end-devices, SIP proxy, human being, whatever
  - From deployment perspective, the scenario most likely with long-lasting relation between a few of controllers such as SIP proxies and network devices all of them trusted and belonging to the same administrative domain.
- Application-layer security applies as well: a proxy never opens pinholes until both parties agree to set up a call, proxy approves it; proxy's approval may require at least one party to authenticate
- Mutual authentication and message integrity desparately needed; can be accomplished at transport/IP layer
- Permissions defined by ACLs

FCP Status

- Three proprietary solutions reported
  - operational experience: support for route recording and session timers rare
- IETF: MidCom in a beginning phase since one year
- Discovery ruled out as an orthogonal issue
- Further issues to be dealt with:
  - extensibility (e.g., ability to add new pinhole attributes such as throughput constraints)
  - nightmare: multiple boxes
  - performance
  - failure recovery
  - mapping FCP to a real protocol
  - SIP issues: FCP timing wrt to session state, rule timers, "funnel rules"
  - etc.

Conclusion

- MidCom is a horrible, horrible hack.
- However, it is horribly needed.
- MidCom Firewalls are a NextGen technology; MidCom WG is in a very early stage.
- In the meantime, embedded ALGs likely to dominate.
  - Many customers have no SIP support in their firewalls.
- Other solutions unlikely to fly -- too tricky from deployment point of view
  - end-device driven middleboxes
  - junking firewalls/NATs
- NATs could be dealt by making end-devices and SIP proxies "little bit" NAT aware

GMD FOKUS

Information Resources

- Repository of related I-Ds available at
  http://www.fokus.gmd.de/glone/projects/ipt/players/ietf/firewall
- draft-rosenberg-sip-firewalls
- draft-biggs-sip-nat
- draft-kuthan-fcp
- draft-shore-h323-firewalls
- draft-rosenberg-sip-entfw-01.txt
- FCP Site:
  http://www.fokus.gmd.de/glone/employees/jiri.kuthan/private/fcp/

Interworking with Legacy Networks

About PSTN

- Long innovation cycle
- High costs
- Walled garden service model (see RFC 3002)
  - complete control over services
  - applications bundled with access
  - rigorous service definitions
  - security easier to accomplish
- Various national signaling dialects
- Huge customer base -- backwards compatibility needed

Interoperability Issues

- IP-PSTN Gateways make the conversion job
  - convert both signaling and media
  - may be split into media and signaling gateways (MGCP/Megaco)
  - many pains: DTMF, IVRs, overlapped dialing, national signaling dialects
  - gateways act as UAs from SIP perspective
- Convergent Services
  - PINT
    - allow Internet users to trigger PSTN services
    - e.g., click to PSTN-dial
  - SPIRITS
    - allow PSTN events to trigger Internet services
    - e.g., Internet Call Waiting
- Sigtran - Trunk replacement

GMD FOKUS

MGCP/Megaco

- Both protocols of Master-Slave nature
- Use of the protocol
  - The protocol to reconnect split signaling/media gateways
  - Architectures envisioned in which MGCP controllers control behavior of simple IP phones
    - costs of Megaco/MGCP devices comparable to full end-to-end devices (at least, TCP/IP and a signaling protocol must be present anyway)
    - only services mediated by the protocol supported
    - lack of user mobility
    - not end-to-end compatible (QoS)
- History
  - MGCP is a result of an "individual effort" whereas Megaco protocol is output of Megaco working group; Megaco adopted at ITU-T (H.248)
  - More MGCP implementations reported

Routing Calls between SIP and PSTN Devices

- Addressing PSTN destinations from SIP devices
  - PSTN phone number and destination domain known:
    - sip:+1-212-555-1212@gateway.com;user=phone
    - Address the gateway directly
  - only PSTN phone number known:
    - tel:+358-555-1234567 (RFC 2806)
    - Ask somebody (TRIP, a proxy ..)
- Addressing SIP destinations from PSTN devices
  - SIP devices use E.164 numbers
  - PSTN routes calls to a gateway
  - Translate phone number into a SIP URL using ENUM
  - Continue as usual

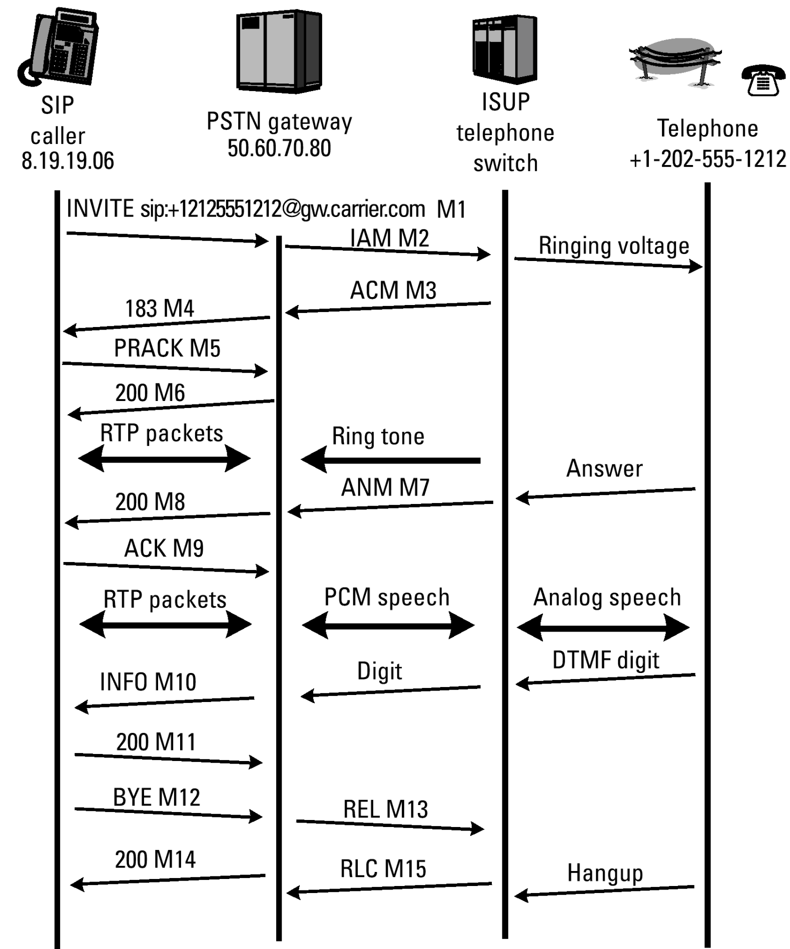
The Call Routing Protocol: TRIP (formerly gwloc)

- exchange of call routing information between cooperating providers
- routing services (e.g. ‘find cheapest gateway to China) may be provided by third parties
- Design
  - follows IP routing protocols (BGP4, IS-IS)
  - exploits scalable techniques: routing information is aggregated and redistributed, incremental updates, soft-state design
  - TRIP used to send, receive or send&receive
- References
  - RFC2871, draft-ietf-iptel-trip-03.txt

GMD FOKUS

Call Flow SIP to PSTN

- Request-URI in the **`INVITE`** contains a Telephone Number which is sent to PSTN Gateway.
- The Gateway maps the **`INVITE`** to a SS7 ISUP IAM (Initial Address Message)
- **`183 Session Progress`** establishes early media session so caller hears Ring Tone.
- Two way Speech path is established after ANM (Answer Message) and **`200 OK`**



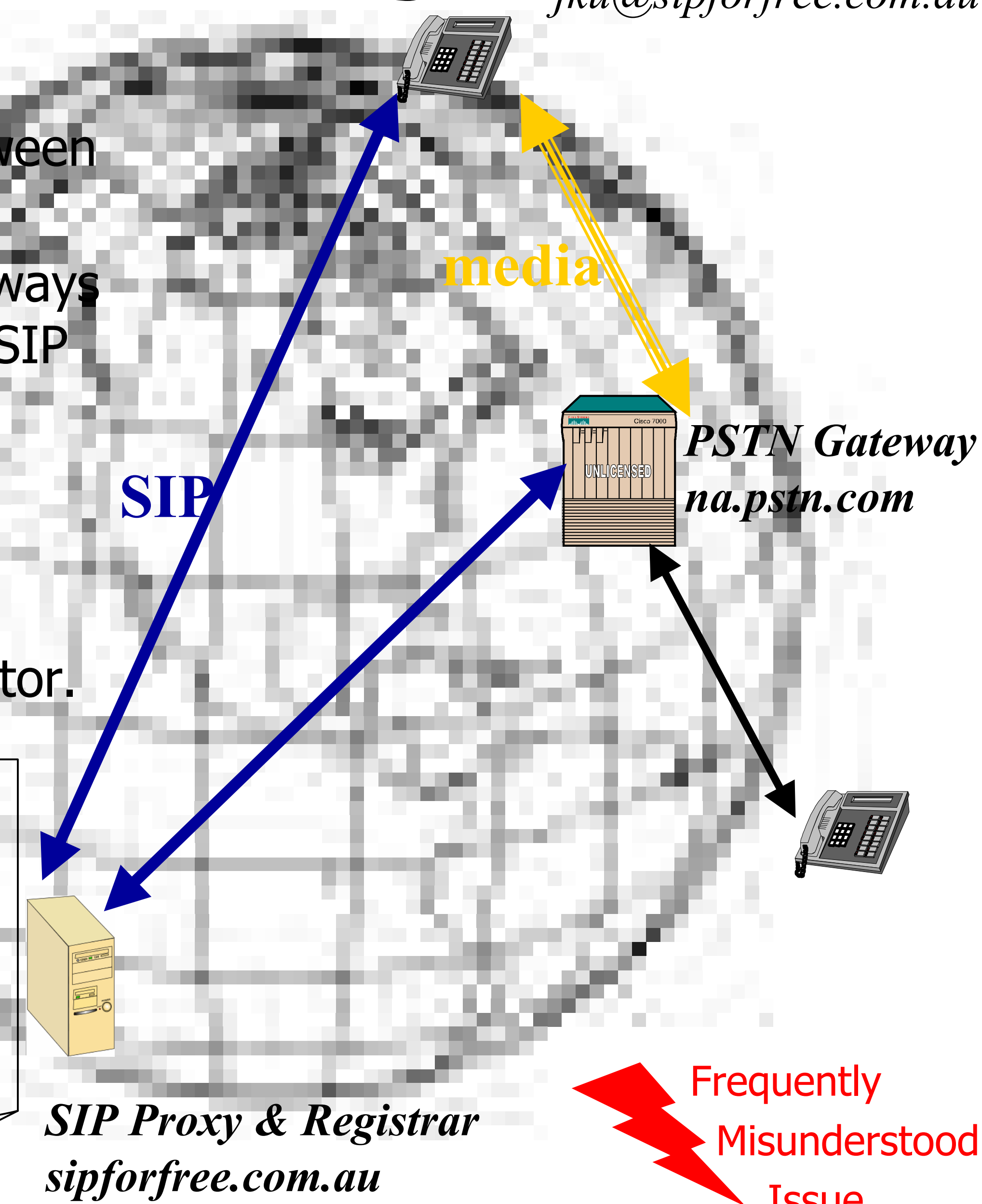
Slide courtesy of Alan Johnston, WorldCom. (See reference to Alan's SIP book.)

PSTN GW != SIP proxy

- PSTN gateways are adapters between two different technologies.
- From SIP perspective, PSTN gateways are SIP termination devices, i.e., SIP User Agents just like IP phones.
- **PSTN gateway functionality separate from call processing logic residing at a proxy.**
- Gateway operator != proxy operator.

call processing logic:

If *($destination* ***in*** *PSTN)* ***then***
route_to_least_cost_gateway();
elseif *local("sipforfree.com.au")* ***then***
lookup_registry;
else *proxy_to_foreign_domain();*



Political Issues

Political Issues - Wiretapping

- Wiretapping
  - RFC 2804: "The IETF has decided not to consider requirements for wiretapping as part of the process for creating and maintaining IETF standards."
  - IETF is international and cannot standardize protocols for enforcement of local laws
  - Eliminate security loopholes.
  - Source of complexity. Complexity inevitably jeopardizes security.
  - Etc.
- Telecommunication Industry Association and ETSI/Tiphon working on it
- Check VON Coalition at www.von.org

Political Issues - Regulations

- Some government agencies have sought to ban VoIP (Czech Republic) and even PC-based VoIP (Pakistan, India)
- Most have taken no action
  - EU Commission, 1998:"Status of Voice Communications on Internet under *Community Law and in particular, under directive 90/388/EEC": "These services cannot for the time being be considered as "voice telephony" in the sense of this Directive and they therefore fall already within the liberalized area, before the deadlines set for the implementation of full competition."*
  - US FCC, 1998: "Report to Congress,#96-45": *"We continue to believe that alternative calling mechanisms are an important pro-competitive force in the international services market ... it may not be appropriate to apply the international accounting rate regime to IP telephony."*
- Hungary (1999): most detailed regulation policy in the world; QoS must be poor
- ITU circulates a new draft for the Policy Forum (March 2001) and "continues to think that old regulations should be imposed on new technologies" (Pulver Report, Nov 6th, 2000)
- Further links: ITU: http://www.itu.int/iptel

GMD FOKUS

Case Study: People's Republic of China (ITU-T, Dr. Lovelock)

- 1998: the Chen brothers began offering IP phone services; police detained the brothers, seized their equipment, the Chens filed a suit against China Telecom "Computer Services were not listed in the Arrangement for Approval and Regulation of Decentralized Telecommunication Services; accepted at the court
- until 1998, Ministry of Information Industry (MII) via China Telecom resisted proliferation of IP Telephony Services
- then, new licensing framework limited to government-affiliated operators (China Telecom, China Unicom, Jitong)
- May 1999: At Jitong's offices 2e3+ people queued some of them at as early as 2am to get prepaid cards
- June-August 1999: revenue $35 millions
- Population 1.3 billion (1e9), less IP addresses than Stanford University (18.0.0.0/1800, 6951 graduates, 7553 graduate students, 1595 tenured faculty)

GMD FOKUS

Current Status

Current Status

- SIP moving to Draft Standard
- Adopted for 3G mobile networks
- Products available:
  - software phones
  - appliances
  - proxies
  - application servers
  - PSTN gateways: carrier grade, enterprise, single line adaptors
  - etc.
- Services available:
  - Telia, MCI WorldCom, Level3

Still on the Agenda

- Mobility
- Firewall Traversal
- Inter-domain Aspects
- Advanced Services (transfer, conferencing, instant messaging)
- Compatibility with existing H.323 base
- Emergency services

Conclusions

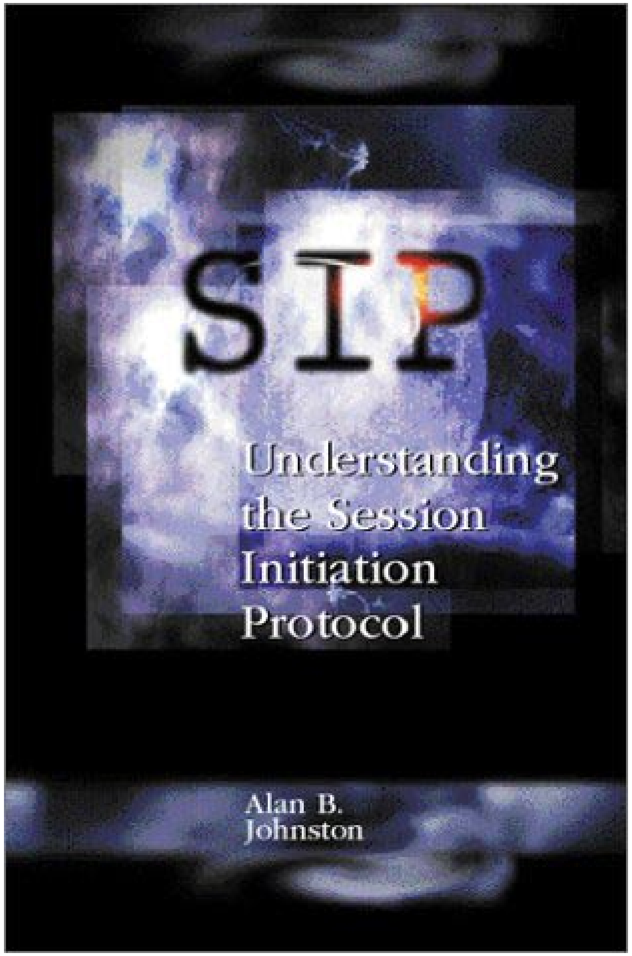
- Internet telephony
  - opens telephony to unlimited competition
  - integrates voice with arbitrary services (cf. ISDN)
- The core tool: Session Initiation Protocol
  - explores Internet legacy: textual protocol, stateless design, extensibility
  - enables end-2-end services (service portfolio not limited by a control protocol)
- Commercial SIP products and services available

References

References

- Internet Telephony -- Exhaustive collection of Links and References
  - http://www.fokus.gmd.de/glone/projects/ipt/
- Session Initiation Protocol
  - http://www.cs.columbia.edu/~hgs/sip
- Internet Resources
  - http://www.cs.columbia.edu/~hgs/internet/
- Requests for Comments and Internet Drafts
  - http://www.normos.org

There’s a SIP Book!



- Alan B. Johnston: “SIP: Understanding the Session Initiation Protocol”
- Artech House 2001
- ISBN 1-58053-168-7

RFCs and Internet Drafts

- SIP: RFC 2543, draft-ietf-sip-rfc2543bis
- SDP: RFC 2327
- SIP call flows: draft-ietf-sip-call-flows
- SIP services call flows: draft-ietf-sip-service-examples
- SIP-CGI: RFC 3050
- CPL: draft-iptel-cpl
- preconditions: draft-ietf-sip-manyfolks-resource-00.txt
- RTP: RFC 1889, draft-ietf-avt-rtp-new
- PIM: draft-rosenberg-impp-*

Information Resources

- Dorgham Sisalem, sisalem@fokus.gmd.de
- Jiri Kuthan, kuthan@fokus.gmd.de

- The End -

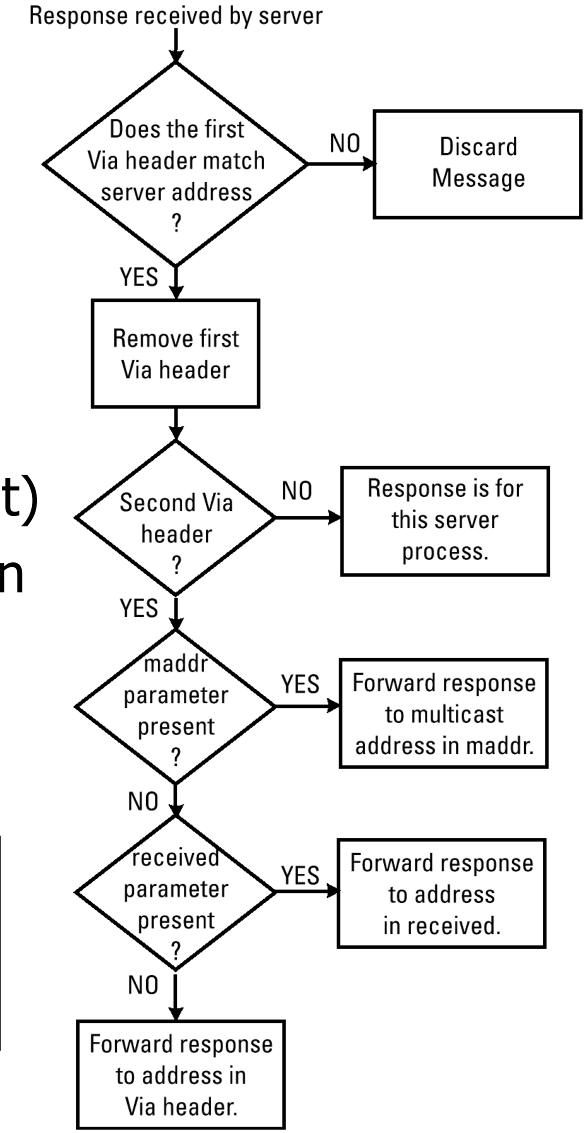
Backup Slides

This section contains unordered slides that turned out to be less useful than we expected.

Signaling

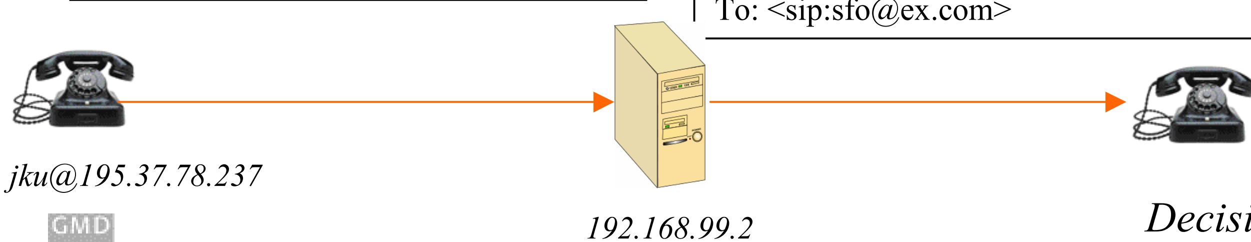
Via Header Field

- Every proxy adds a Via header with its address to requests to make sure responses within a transaction will take the same path (e.g., to avoid loops or make sure the same SIP firewall will be hit)
- All Via headers copied from Request to Response in order.
- Response is sent to the address in top Via header; decision tree shows next hop processing.




```
INVITE sip:sfo@ex.com:5060 SIP/2.0
Via: SIP/2.0/UDP 195.37.78.237:5060
From: "Jiri Kuthan" <sip:jku@ex.com>
To: <sip:sfo@ex.com>
```

```
INVITE sip:sfo@ex.com:5060 SIP/2.0
Via: SIP/2.0/UDP 192.168.99.2:5060
Via: SIP/2.0/UDP 195.37.78.237:5060
From: "Jiri Kuthan" <sip:jku@ex.com>
To: <sip:sfo@ex.com>
```



Decision tree diagram courtesy of Alan Johnston, MCI WorldCom. (See reference to Alan's SIP book.)

Frequent Misconceptions

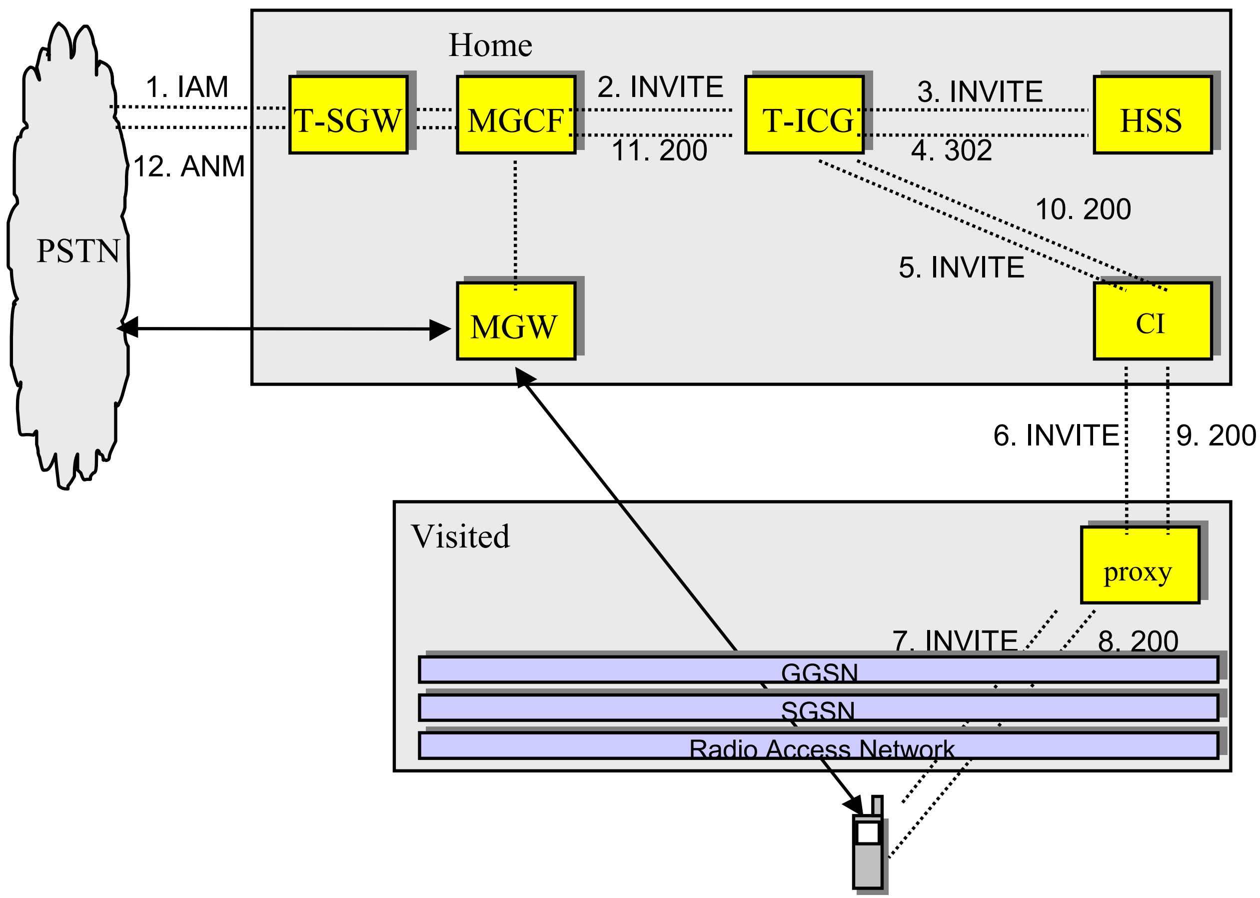
Avoiding SIP Duplication

- Most attempts to build protocols that do what one can already do with SIP are a waste of time.
- Quick-check: If your new protocol conveys pieces of information conveyed in SIP, it indicates SIP could have been used without having to build a new protocol.
- Advise: route your signaling directly through the place of your logic.
- Examples:
  - 3rd party call control
  - inspecting SIP messages by an anti-spam site

GMD FOKUS

3gpp

3GPP-PSTN Interaction



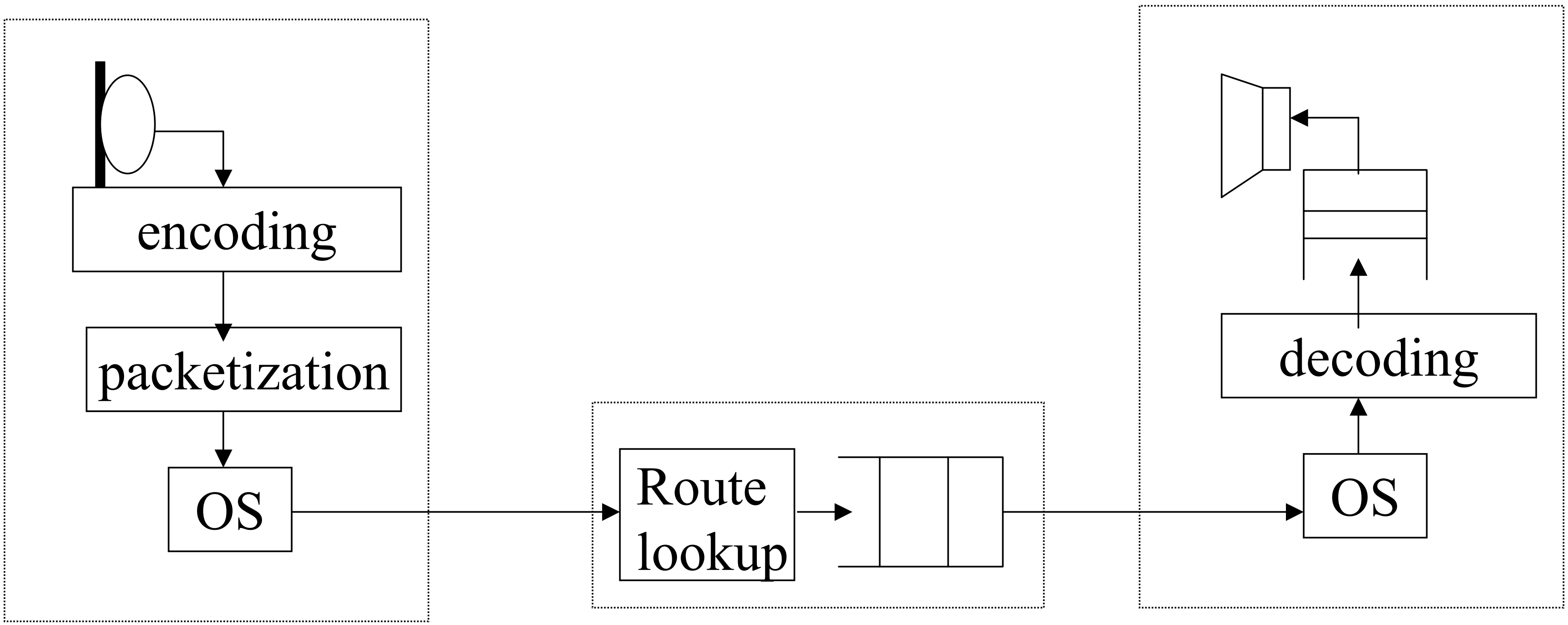
QoS

QoS: Issues to Consider

- Bit and Packet Losses
  - media distortion
  - congestion collapse
  - application-layer retransmission (DNS)
- Delay
  - non-interactive communication
- Jitter
  - higher delays and losses

GMD FOKUS

QoS: The Problem



QoS: End-System Solutions

- Use Real-Time OS or dedicated hardware
  - more complexity
- Deploy FEC and concealment schemes
  - higher bandwidth consumption
- Adaptive Playout buffers
  - higher delays
- Congestion control
  - no fixed quality ensured

QoS: Traffic Engineering

- Estimate the required resources
- Provide more than estimated resources (over-provision)
- Does not require changes to network structure
- Estimation complexity increases with increased network size
- No absolute guarantee

GMD FOKUS

QoS: Integrated Services

- Network supports different QoS classes
- End systems signal their required resources
- Routers decide to accept or reject reservation requests
- Routers classify and schedule packets based on the reserved flow resources
- RSVP proposed for QoS signaling

QoS: Integrated Services

- Signaling increases load and processing overhead
- Per-flow handling causes scalability problems
- Classification and scheduling increases complexity
- No clear billing is defined

QoS: Differentiated Sevices

- Services are negotiated between ISPs and customers (SLA)
- At the edge packets are marked, dropped or shaped based on the SLA
- Within the core packets are treated based on the marks
- Marks are mapped to PHB
- Two PHBs standardized: Expedited and Assured Forwarding

QoS: Differentiated Sevices

- Dynamic SLAs are difficult
- Network engineering and bandwidth provisioning not clear
- Achieved quality might vary (not predictable)

GMD FOKUS