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

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

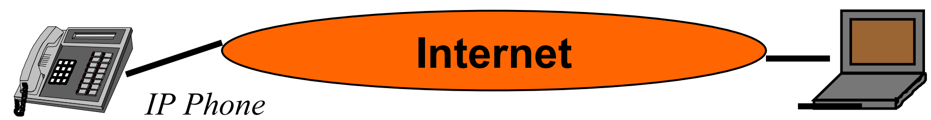
Other Related Bodies (cont.)

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

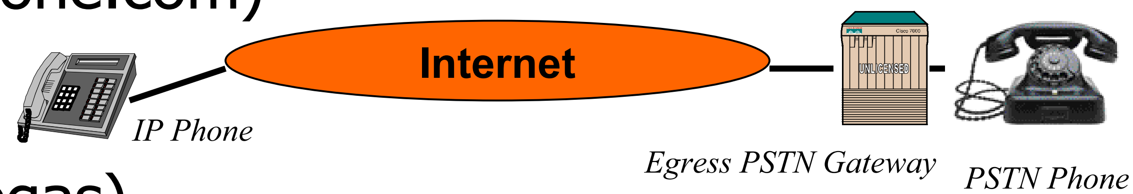
Protocol ZOO

Internet Telephony

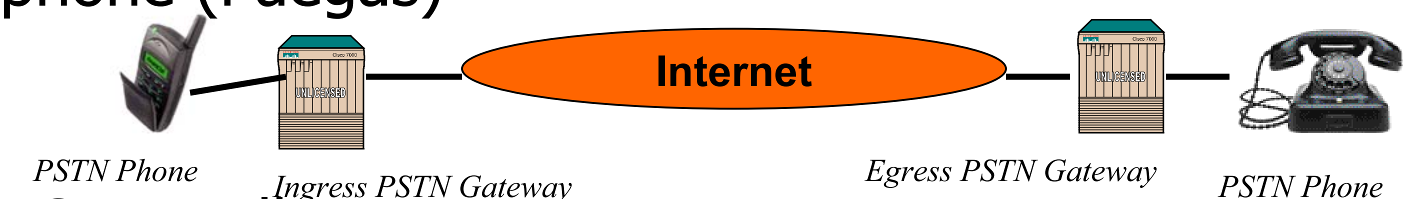
- Routing a call over the Internet
  - PC-to-PC (MS NetMeeting, appliances)



  - PC-to-phone (net2phone.com)



  - phone-to-phone (Paegas)



  - phone-to-PC as well

What Protocols Are Needed?

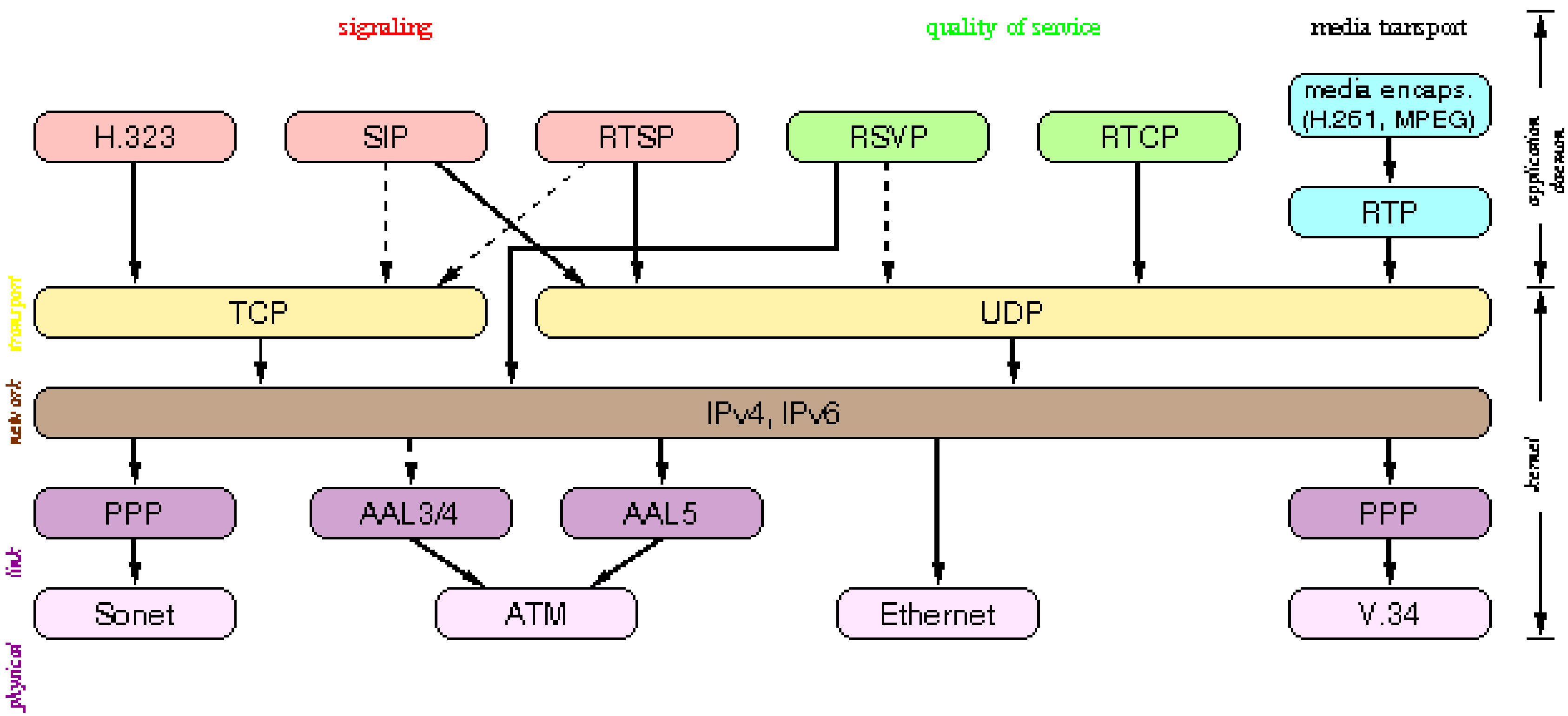
- Signaling protocol to establish presence, locate users, set up, modify and tear down sessions
- Media Transport Protocols for transmission of packetized audio/video
- Supporting Protocols
  - Gateway Location, QoS, interdomain AAA*, address translation, IP, etc.

** 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

SIP End-devices

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

SIP Workhorses

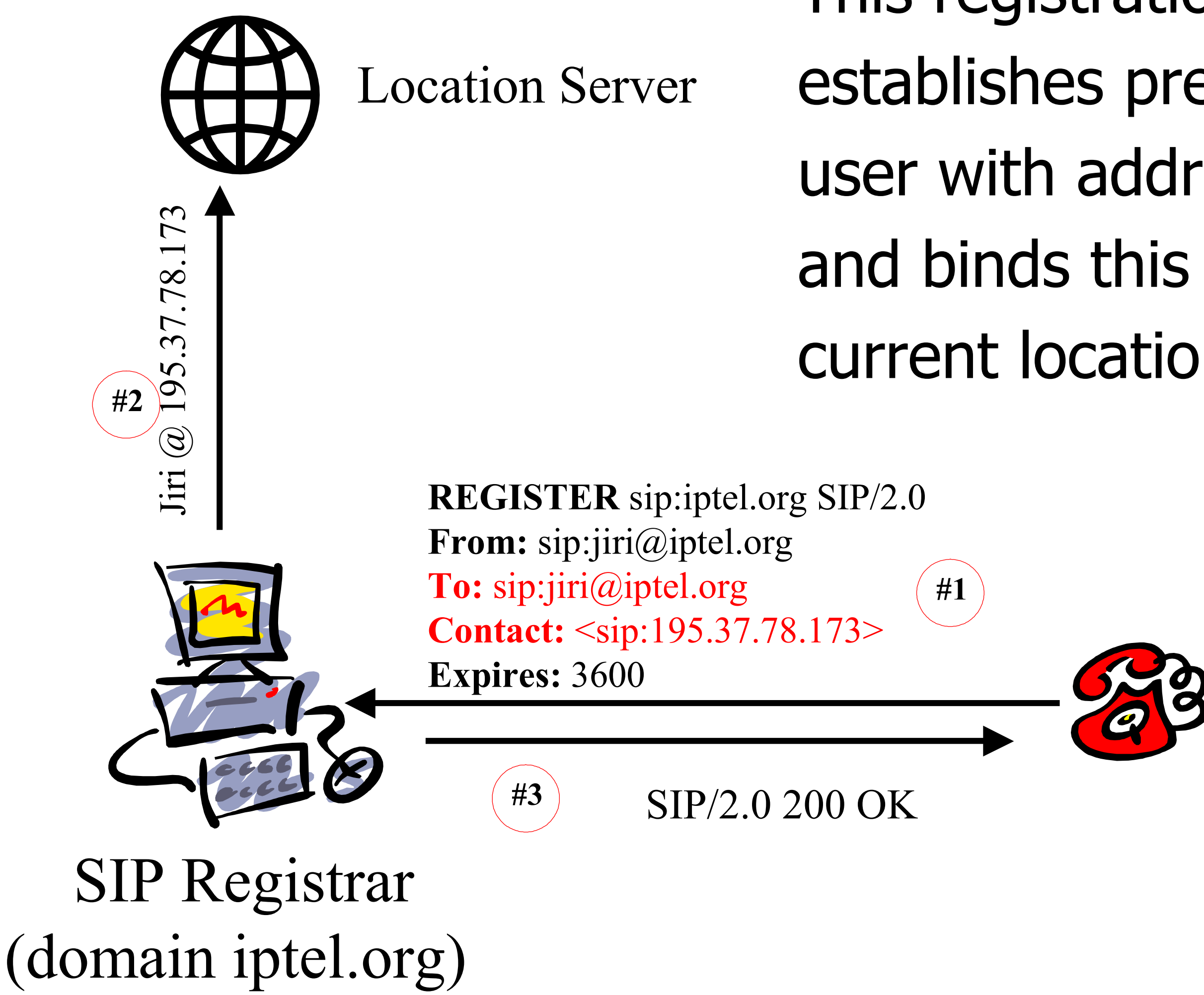
- SIP Proxy Server
  - relays call signaling, i.e. acts as both client and server
  - operates in a transactional manner, i.e., it keeps no session state
- SIP Redirect Server
  - redirects callers to other servers
- SIP Registrar
  - accept registration requests from users
  - maintains user's whereabouts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

SIP Addresses

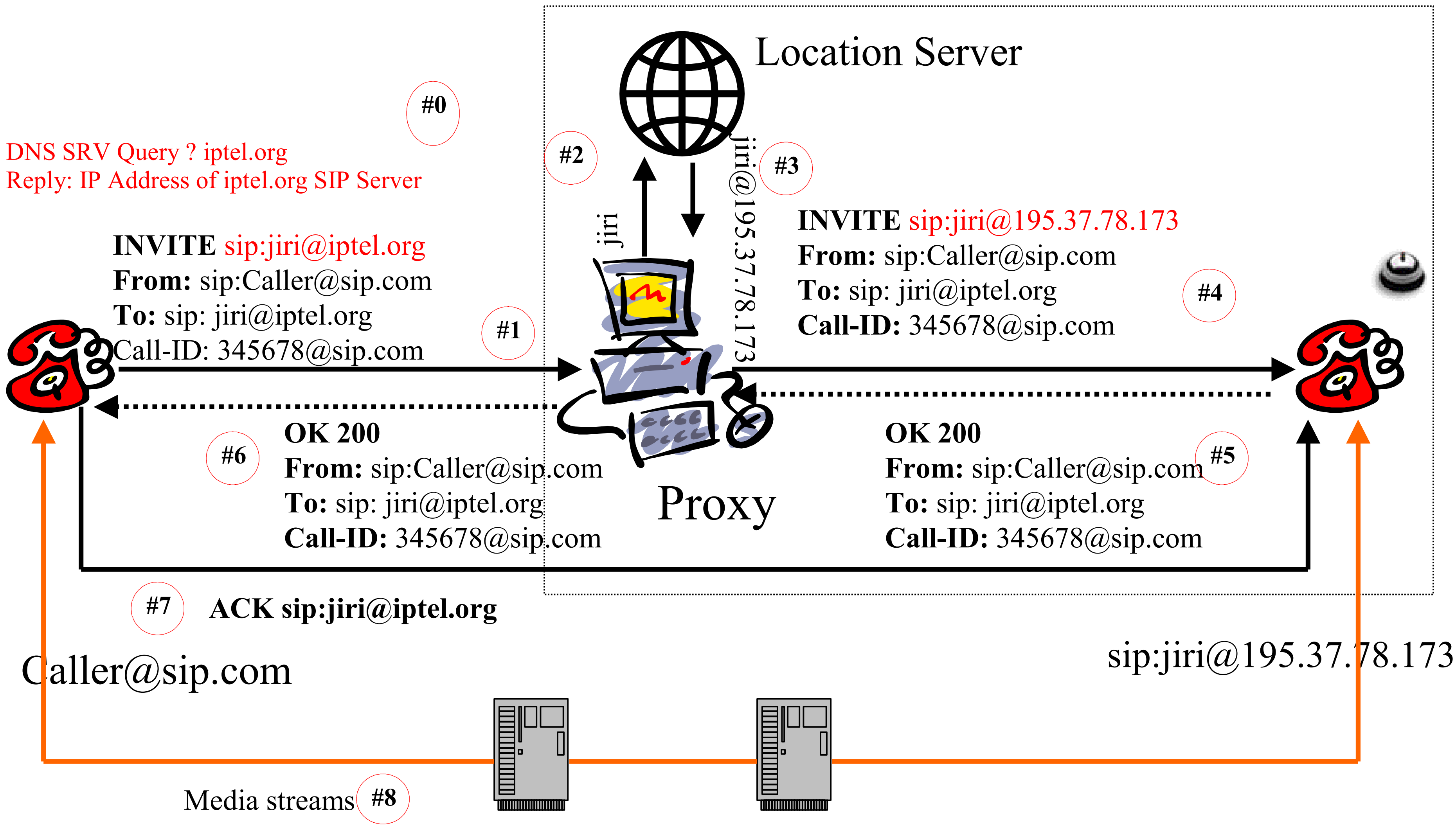
- SIP gives you a globally reachable address.
  - Callees bind to this address using SIP REGISTER method.
  - Callers use this address to establish real-time communication with callees.
- URLs used as address data format; examples:
  - sip:jiri@iptel.org
  - sip:voicemail@iptel.org?subject=callme
  - 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



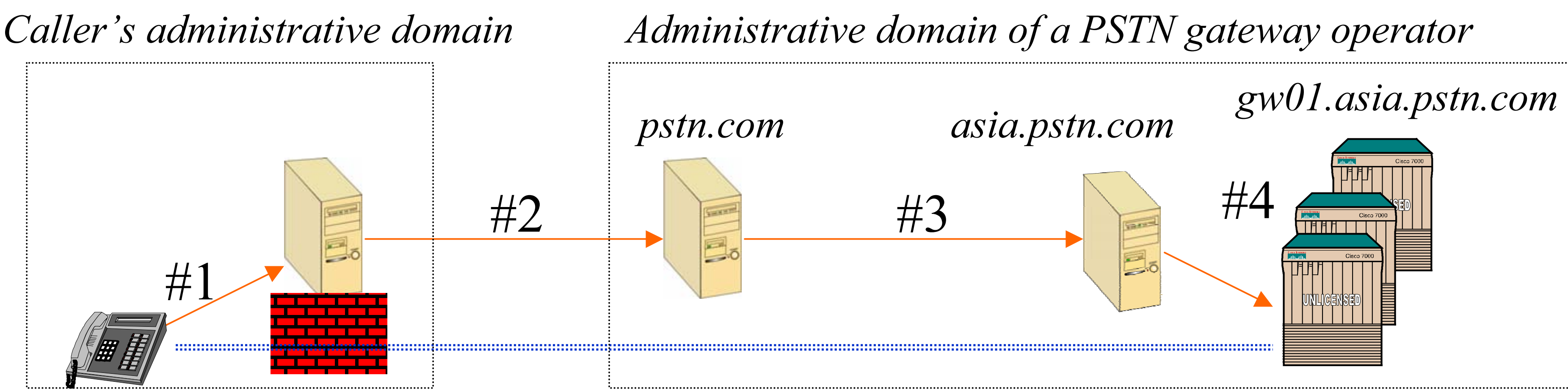
Proxy Server Functionality

- 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



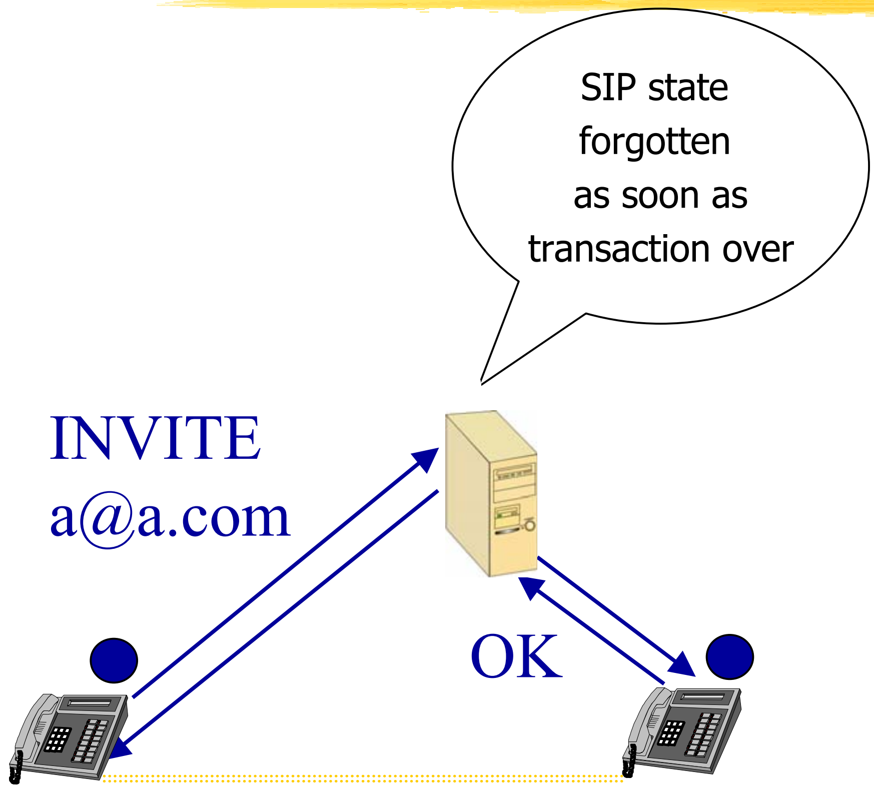
Caller's outbound proxy accomplishes firewall traversal.

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



Legend

SIP signaling

SIP state

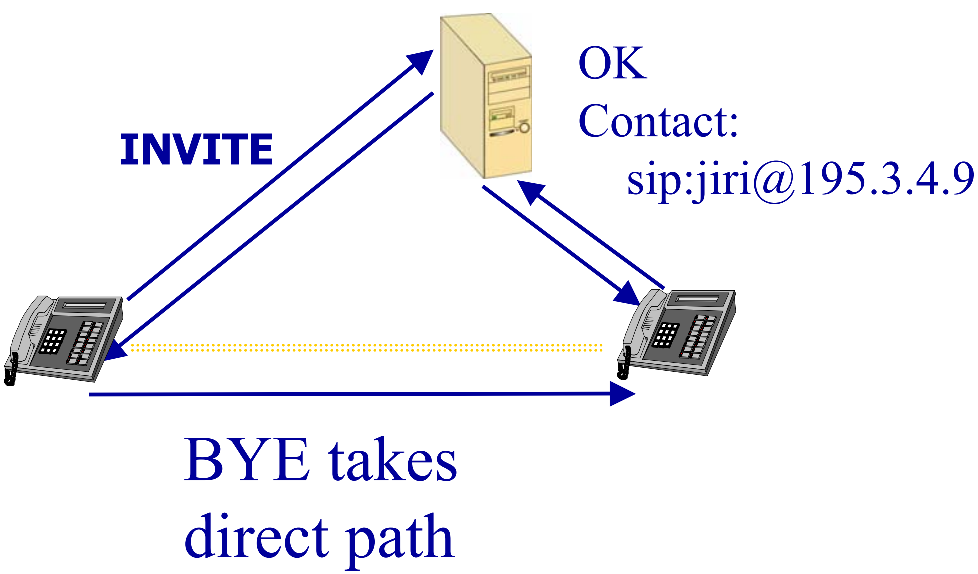
media

- 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

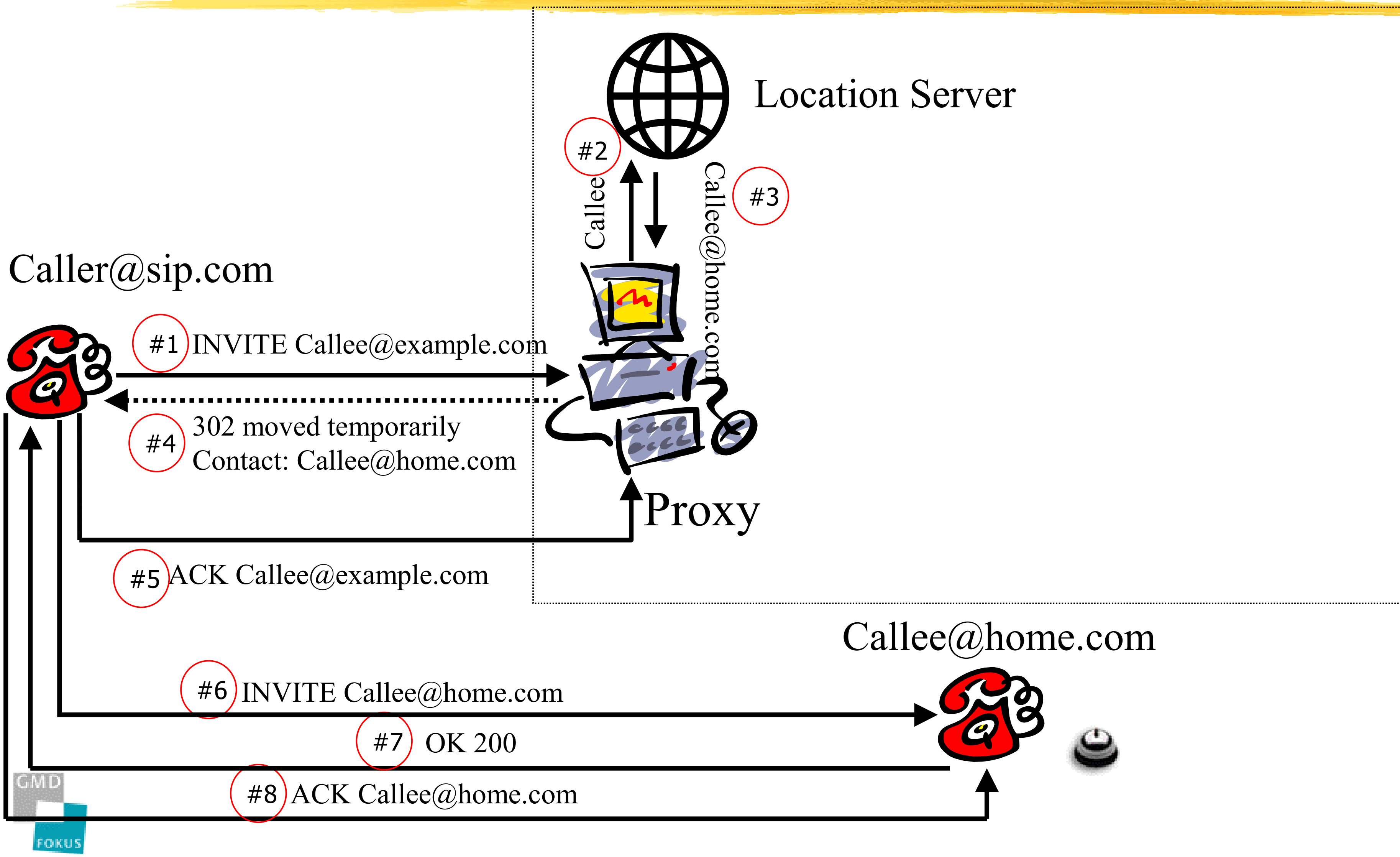
Subsequent Transactions Bypass Proxy

- 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

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

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

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.

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)