Why SIP?

Henning Schulzrinne
Dept. of Computer Science
Columbia University
New York, New York
(sip:)schulzrinne@cs.columbia.edu

SIP Services and Applications – Washington, D.C.

April 20th, 2001

Overview

- A brief history
- Service models
- SIP design principles
- Extensions in progress
- Potential hazards

Whence SIP?

Feb. 1996: earliest Internet drafts

Feb. 1999: *Proposed Standard*

March 1999: RFC 2543

April 1999: first SIP bake-off

November 2000: SIP accepted as 3GPP signaling protocol

December 2001: 6th bake-off, 200+ participants

March 2001: 7th bake-off, first time outside U.S.

SIP years

Year	development	trade rags
1996-1998	R&D	“academic exercise’, “distraction from H.323”
1999	standard & skunk works	“what does SIP stand for again?”
2000	product development	“SIP cures common cold!”
2001	pioneer deployment	“Where are the SIP URLs?”
2002	kmart.com/sip	SIP product comparisons

VoIP signaling architectures

- master-slave ➠ MGCP, Megaco
- (mostly) single administrative domain ➠ H.323
- peer-to-peer, cross domain ➠ SIP

Master-Slave Architecture

- master-slave: MGC controls one or more gateways
- allows splitting of signaling and media functionality
- “please send audio from circuit 42 to 10.1.2.3”
- uses MGCP (implemented) or Megaco/H.248 (standardized, but just beginning to be implemented)
- gateway can be residential
- basis of PacketCable NCS (network control system) architecture
- service creation similar to digital PBX or switch
- end system has no semantic knowledge of what’s happening
- $\longrightarrow$ can charge for caller id, call waiting

VoIP architectures

	SIP	H.323	Megaco/MGCP
multiple domains	x	?	–
Third-party control	x	–	single-domain
multimedia	x	fixed set	not likely
end system control	x	x	–
extensible	x	?	limited
generic events	x	–	–
cgi scripting	x	–	–
servlets	x	–	–
CPL	x	x	–

SIP inheritance

- URLs:
  - general references (“forward to email”)
  - recursive embeddding
- HTTP:
  - basic request/response format, status codes, . . .
  - proxies (but no caching)
  - cgi programming interface
- email/SMTP:
  - addressing
  - MX $\longrightarrow$ SRV records for load balancing, redundancy
  - header/body separation, MIME

SIP design choices

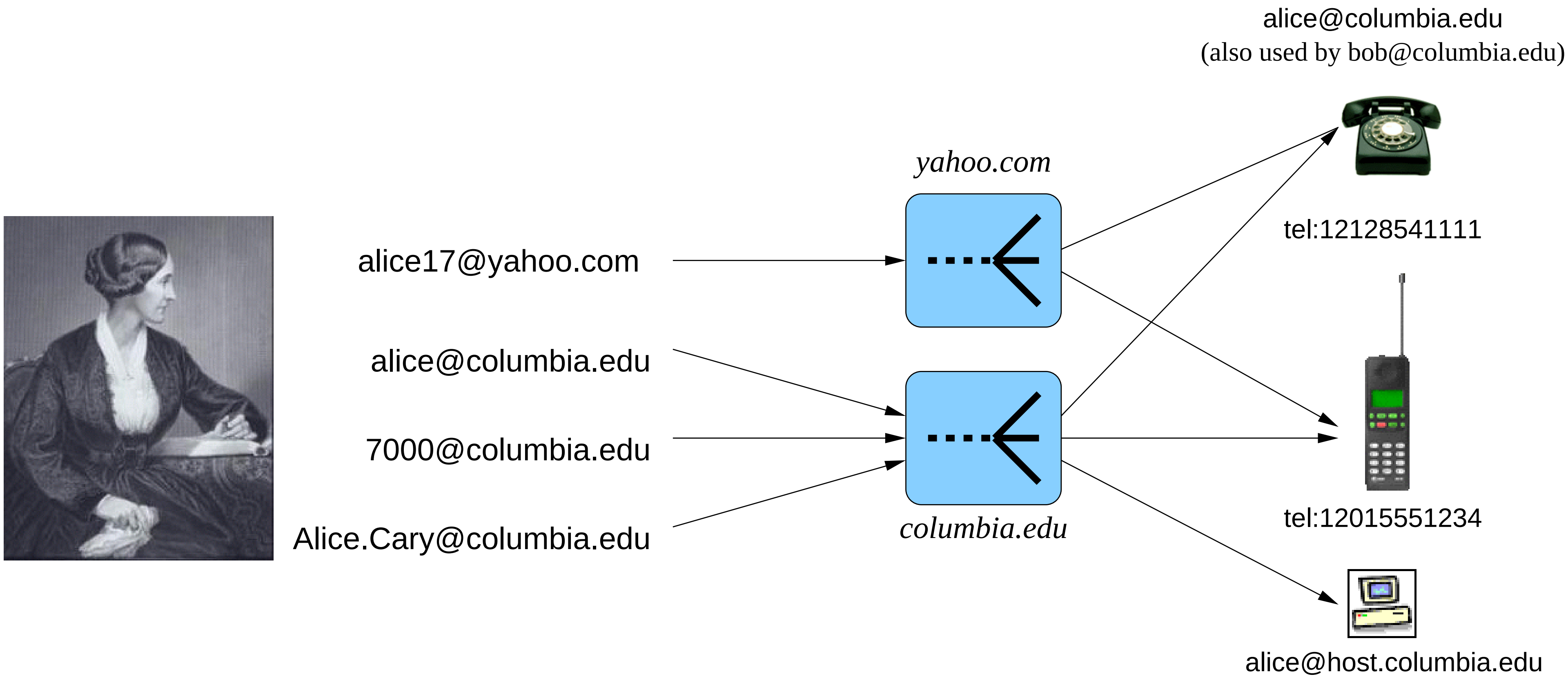
Transport protocol neutrality: run over reliable (TCP, SCTP) and unreliable (UDP) channels, with minimal assumptions

Request routing: direct (performance) or proxy-routed (control)

Separation signaling vs. media description: can add new applications or media types, SDP $\longrightarrow$ SDPng

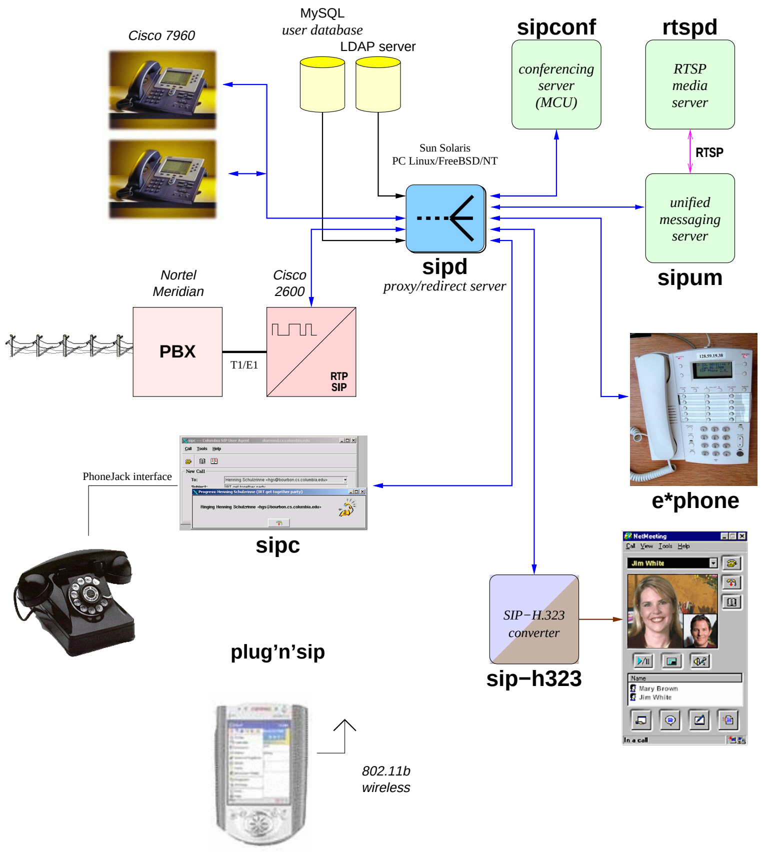
Extensibility: indicate and require proxy and UA capabilities

Personal mobility



Example: Columbia CS phone system

Expand existing PBX via IP phones, with transparent connectivity



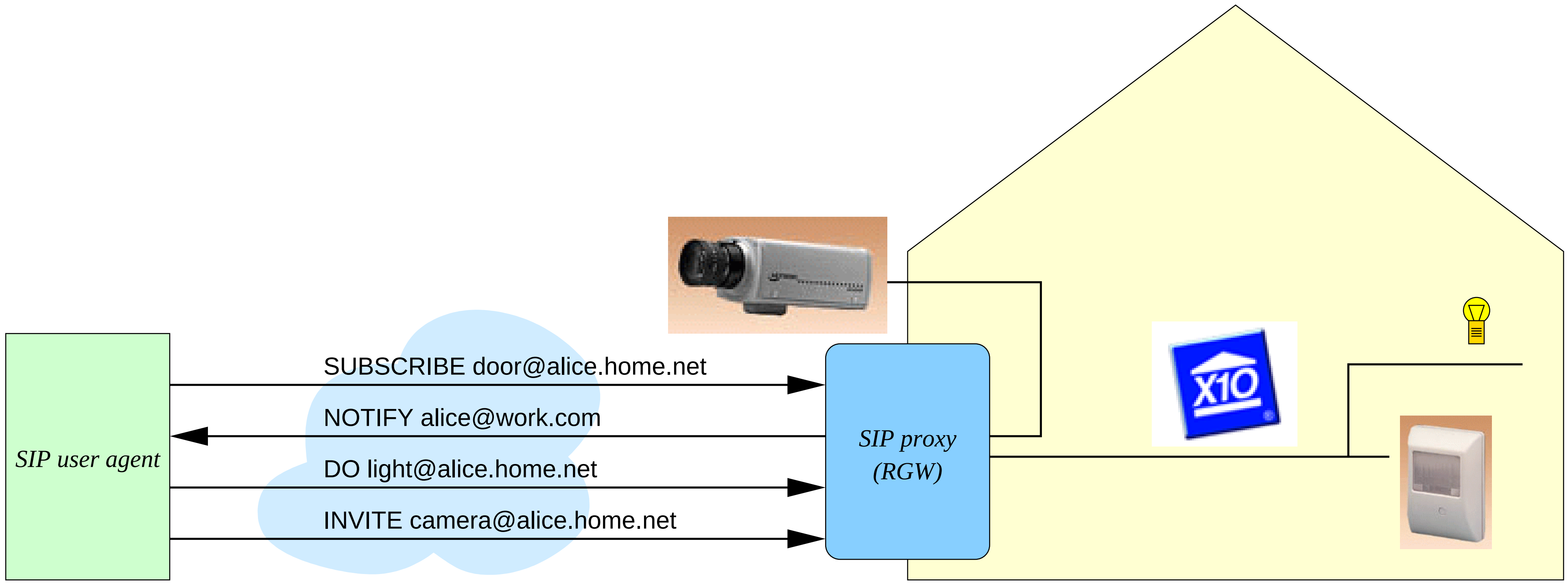
Events as universal glue

- currently, don't have general event notification in the Internet
- email is too slow: pull on the last hop (server to user)
- generic problem:
  - "voicemail has arrived"
  - "called party is reachable"
  - "new configuration data available"
  - "IR sensor has detected movement"
  - "boiler temperature above threshold"
  - ...
- same delivery (SIP), different data (XML DTDs)

SIP as a presence & event platform

- minimal SIP extension: SUBSCRIBE to request notifcations, NOTIFY when event occurs
- also, MESSAGE for IM, sessions for multi-party chats
- transition to true “chat” (and video)
- services such as reaching mobile phone while in meeting

Events: SIP for appliances



SIP service architectures

classical: Media and signaling in one box

distributed: request routing and coordination, with service components (storage, IVR, location, . . .)

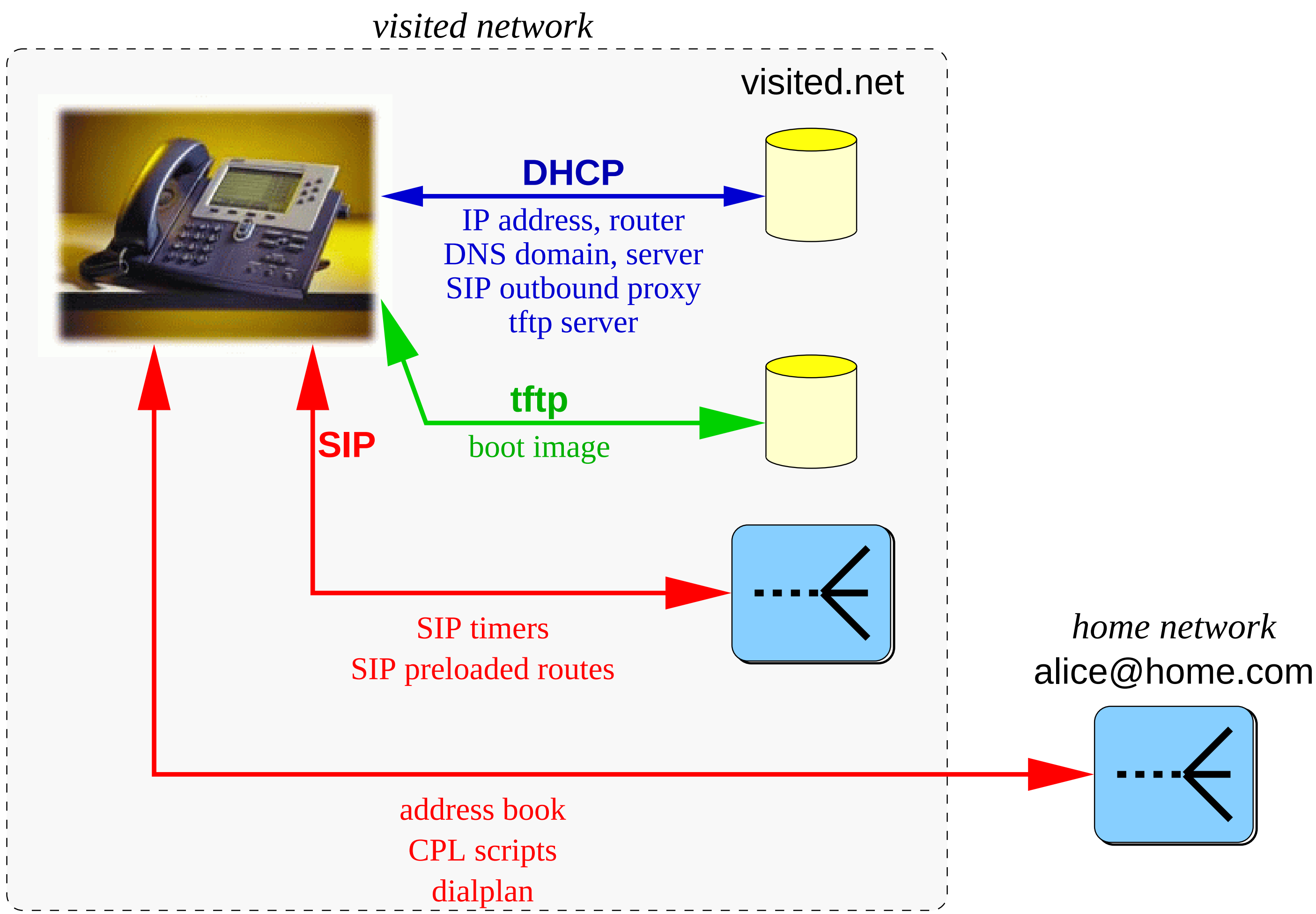
Challenges and obstacles

- scalable device configuration
- PSTNv3
- “walled garden”
- service infrastructure
- standardization
- invisible Internet telephony

Device configuration

- need to plug in store-bought phone, without more than personalization
- limited user interface
- configuration from local (visited) network and from home network
- don't want current PBX single-vendor tie-ins
- cannot rely on California-style upgrades
- notifications of new configurations ➠ SUBSCRIBE/NOTIFY

Device configuration



Potential obstacles

- SIP as transport – for legacy signaling
  - due to proxies, UDP not designed for volume data
  - doesn’t add significant value
- NATs and firewalls – can engineer around them, but ugly
  - leads to IP-over-HTTP solutions, defeating firewall
  - proxy boxes outside NATs

PSTN legacies to avoid

- E.164 numbers – might as well wear bar codes
- overlap dialing
- tones and announcements
- in-band signaling for features (DTMF)
- systems with user-interface knowledge (12 keys, voice)
- voice-only orientation (BICC, MGCP/Megaco)
- integration of bit transport and services
- service-specific billing ➠ separate signaling & billing
- trusted networks without crypto

➠ confine PSTN knowledge to edge of network

“Walled garden” model

- 3G wireless carriers adopting SIP, but used to closed services
- SIP users should be able to use any proxy for services, not just carrier service
- typical users have many identities (and, thus, servers):

work	hgs@cs.columbia.edu
travel	schulzrinne@yahoo.com
home	henning@schulzrinne.leonia.nj.us
professional	h.g.schulzrinne@ieee.org

- hard to prevent: SIP can use any port number
- if not, requires draconian restrictions on IP packets, not just filtering port 5060 (SIP port)
- also, services may be split across servers

So I want to build a SIP network. . .

Ready for trials, but probably not quite for shrink-wrap status:

- installation and operation still requires fair amount of expertise
- lots of web and email experts, few SIP experts
- needs some external infrastructure: DHCP and SRV, possibly AAA
- inconsistent configuration for Ethernet phones (being worked on)
- SIP phones still more expensive than analog phones ➠ hard to justify PBX replacement (incremental cost)
- no just-download or ship-with-OS “soft” clients

Need for service infrastructure

- need carriers that offer SIP gateways
- without having to provide SS7 connectivity
- with *outbound* PSTN calling
- with *inbound* calls and *number portability* – need to be able to keep old PSTN numbers
- either IP Centrex model or in-house servers – like ISP services for email or web
- for commercial-grade conferences, need nailed-up Internet connectivity, orderable (at least) by web page – across providers!
- PBX revenue already decreasing

Why aren't we junking switches right now?

What made other services successful?

email: available within self-contained community (CS, EE)

web: initially used for local information

IM: instantly available for all of AOL

All of these ...

- work with bare-bones connectivity ($\geq$ 14.4 kb/s)
- had few problems with firewalls and NATs
- don't require a reliable network

Why aren't we junking switches right now?

Telephone services are different:

- reliability expectation 99.9% $\nearrow$ 99.999%
- PC not well suited for making/receiving calls – most residential handsets are cordless or mobile
- business sets: price incentive minor for non-800 businesses
- services, multimedia limited by PSTN interconnection
- initial incentive of access charge bypass fading (0.5c/min.)
- international calls only outside Western Europe and U.S.

Standardization

- SIP working group is one of the most active in IETF
- located in “transport” area, but really an application
- about 80 active Internet drafts related to SIP
- typically, 400 attend WG meetings at IETF
- but few drafts are working group items
- 80-20% – 80% of the technical work takes 20% of the time

Invisible Internet telephony

“VoIP” technology will appear in

- Internet appliances
- home security cameras, web cams
- 3G mobile terminals
- fire alarms and building sensors
- chat/IM tools
- interactive multiplayer games
- 3D worlds: proximity triggers call

Conclusion

- SIP maturing – base stable, extension in progress
- avoid creating PSTN replica
- leverage, not inhibit, Internet flexibility
- significant deployment challenges remain

For more information. . .

SIP: http://www.cs.columbia.edu/sip

RTP: http://www.cs.columbia.edu/~hgs/rtp

Papers: http://www.cs.columbia.edu/IRT