Scaling the Internet

Henning Schulzrinne
Dept. of Computer Science
Columbia University
schulzrinne@cs.columbia.edu

IFIP WG 7.3 Workshop
Rosario, Washington

June 19, 1997

Overview

- scaling = numbers + applications
- applications: telephone, TV, radio, . . .
- Internet integrated services architecture ➠ RSVP alternatives
- real-time transport ➠ RTP scaling
- signaling: SIP
- stream control: RTSP
- reliability, charging

Applications

Internet Integrated Services Applications

integrate ➠ single delivery mechanism:

- data
- telephony, multimedia conferencing
- radio
- “cable television”

trench/conduit: typical residential, commercial (CATV, phone)

physical: SONET, ADSL

data link: ATM, ISDN

network: IP ➠ largest speed, media range (300 b/s to 2.4 Gb/s)

Internet Telephony Advantages

- functional enhancements
- need LAN/WAN anyway
- (currently) not encumbered by regulatory burdens ("the FCC bypass")
- second phone ➠ reliability not as crucial
- no 7c/min access charge
- flat-rate billing
- cost of TAT transatlantic line is only $0.03/hour
- compression (but: already done for overseas)
- AT&T: "loose $350m in 2001 to Internet telephony"
- lower call-setup delay (H.324 with V.34: 14 seconds)

Internet Telephony Advantages: Control

- separation call and connection ➠ avoids triangle routing
- clear who handles in-band signaling (# sign: local switch, provider, called voice-mail system)
- no difference for call between intra-PBX, inter-LATA and general
- separation of connectivity from resource availability
- datagram ➠ less bootstrapping
- multiple call presences ➠ call waiting easier
- end system handles many functions
- fewer call states

Problems with Internet Telephony

- total DEM/minute: $5.5 \cdot 10^6$ (local + long-distance)
- ➠ total use 350 Gb/s, interstate 50 Gb/s
- "more than half is data"?
- already: Japan, Australia $\leftrightarrow$ US: mostly Internet
- predictable sound quality
- not suitable for dial-up users: signaling by email, telephone
- no advanced services (yet...)
- 640 mio. phone lines $\leftrightarrow$ 180 mio. computers ➠ interworking

Switching costs

switching method	ports	capacity (Gb/s)	cents/kb/s	$/interface
10BaseT Ethernet hub	12	0.12	0.6	57
10BaseT Ethernet switch	24	0.24	1.2	121
100BaseTX Ethernet switch	8	0.80	0.5	474
router		2.1	16.0	
local ATM switch	16	2.48	1.0	1581
PBX	256	0.02	218.	140
Lucent 5ESS local (no AIN)	5,000	0.32	469.	300
Lucent 5ESS local (AIN)	20,000	1.28	273.	175
Lucent 4ESS toll (100k lines)	6.40	7.8		

Internet Radio

- 12,140 U.S. AM and FM radio stations, only 100 in Germany
- FM quality (56 kb/s) ➠ backbone capacity of 680 Mb/s
- New York City: 45 FM stations ➠ 2.5 Mb/s
- DirecTV: 31 audio channels ➠ 1.7 Mb/s
- easy time-shifting, content-labeling ➠ near media-on-demand
- niche markets ➠ “shortwave radio”

Daily Media Doses

media	minutes/day
long-distance calls, US	14
America Online (AOL)	32
local phone calls, US	38
total phone line usage	52
Internet	57
Television, Sweden	120
television, USA	240

Consumer Access

New service break $19.95 flat-rate model:

- acceptable RT bandwidth = acceptable web bandwidth
- 200 concurrent users / T1 ➠ 7.5 kb/s
- currently: 10 customers/modem (AOL: 30) ➠ will decrease
- ➠ only 75c/month for pipe ISP $\rightarrow$ backbone
- ➠ “Internet radio” costs $30/month for access

Internet Integrated Services Architecture

Internet Integrated Services Architecture

link-level: scheduling, link-layer-specific

IP multicast: trade-off state vs. tree optimality

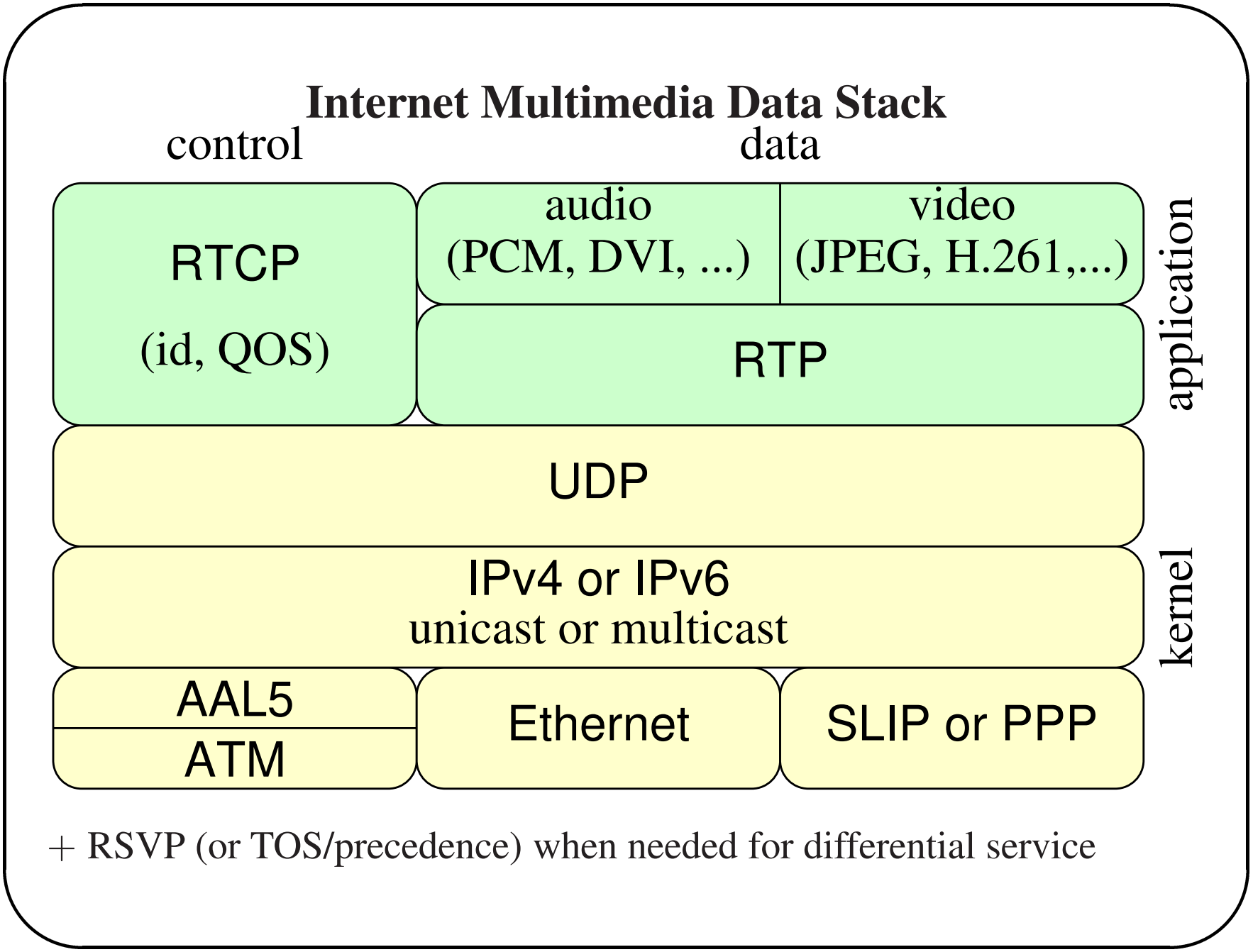
reservation: set up scheduling state ➠ RSVP

transport: timing, QOS feedback ➠ RTP

applications: playout delay adjustment, error concealment, bit-rate adaptation, VCR control (➠ RTSP), signaling (➠ SIP) . . .

Internet Integrated Services Architecture

- philosophy: work both with and without resource reservation ➠ robustness
- most traffic remains best-effort ➠ reservation efficiency not too important
- few guaranteed service classes:
  - guaranteed load = "lightly loaded network"
  - guaranteed delay = Parekh-Gallagher bound, WFQ
  - flow or aggregate?


Internet Multimedia Data Stack
control
data
RTCP
(id, QOS)
audio
(PCM, DVI, ...)
video
(JPEG, H.261,...)
RTP
application
UDP
IPv4 or IPv6
unicast or multicast
kernel
AAL5
ATM
Ethernet
SLIP or PPP
+ RSVP (or TOS/precedence) when needed for differential service

Resource Reservation Scaling Issues

- soft-state: state disappears unless periodically refreshed
- per-flow state?
- per-flow scheduling?
- variable aggregation in backbone and leaves – “Internet VPs”?
- make refresh implicit in data traffic?

RSVP Problems

- if reservation/tear down request lost, no immediate feedback
- can increase reservation latency or “phone off hook”
- ➠ exponential back-off
- large number of refreshes ➠ scaling problems
- hop-by-hop confirmation ➠ extend refresh interval
- receiver-oriented ➠ complexity (RESV, PATH), not always needed
- most applications don’t need full flow spec

Sender-Based Implicit Reservations

- appropriate for distribution applications, conferences
- use RTCP sender reports ➠ bandwidth
- soft-state: about 1% of data bandwidth
- IPv4, IPv6 router alert option ➠ forward packet to CPU
- requires no standardization
- error reports?

Real-Time Transport Protocol

RTP

Data:

- end-to-end transport of streams with timing information
- labeling: source (independent of network source), payload type
- encryption, sequencing, loss detection, ...

Control = RTCP ➠ periodic ($T = f(N)$) reports to multicast group:

- member names:
  - member list
  - loop detection
  - multiple media $\rightarrow$ single conference
- sender reports ➠ rate estimation, lip-sync
- receiver reports ➠ loss, jitter, delay feedback

RTP Scaling

RTCP Algorithm

- RTCP packets sent periodically
- in $[1.25\text{s}, 3.75\text{s}]$ after joining
- media bandwidth b, RTCP packet size p, session size N
- period

$$T = \max(5\text{s}, \frac{pN}{0.05b \cdot 0.75})$$

- compute $\rightarrow$ set timer $\xrightarrow{T}$ send
- randomized $[0.5, 1.5] \leftrightarrow$ packet synchronization

RTCP Scaling Problems

works fine for high-speed networks, smaller groups (< 1000)

initial transient: lots join at the same time: "movie at ten", channel surfing ➠ flooding, announcements lost

state storage: each receiver stores whole receiver list

slow estimate: learn only $\leq 1/\mathrm{s}$ for 28.8 kb/s

QOS transients: can't distinguish 50% loss, 1 minute vs. 5%, 10 minutes

multicast: DVMRP prune timeout ➠ RTCP packets flooded everywhere

BYE: everybody leaves at once

RTCP Transient Suppression

Conditional reconsideration: if N at send instant $> N$ at timer computation, recompute T and check interval

Unconditional reconsideration: recompute even if N hasn't increased

SendMax: accelerate convergence for new members ➠

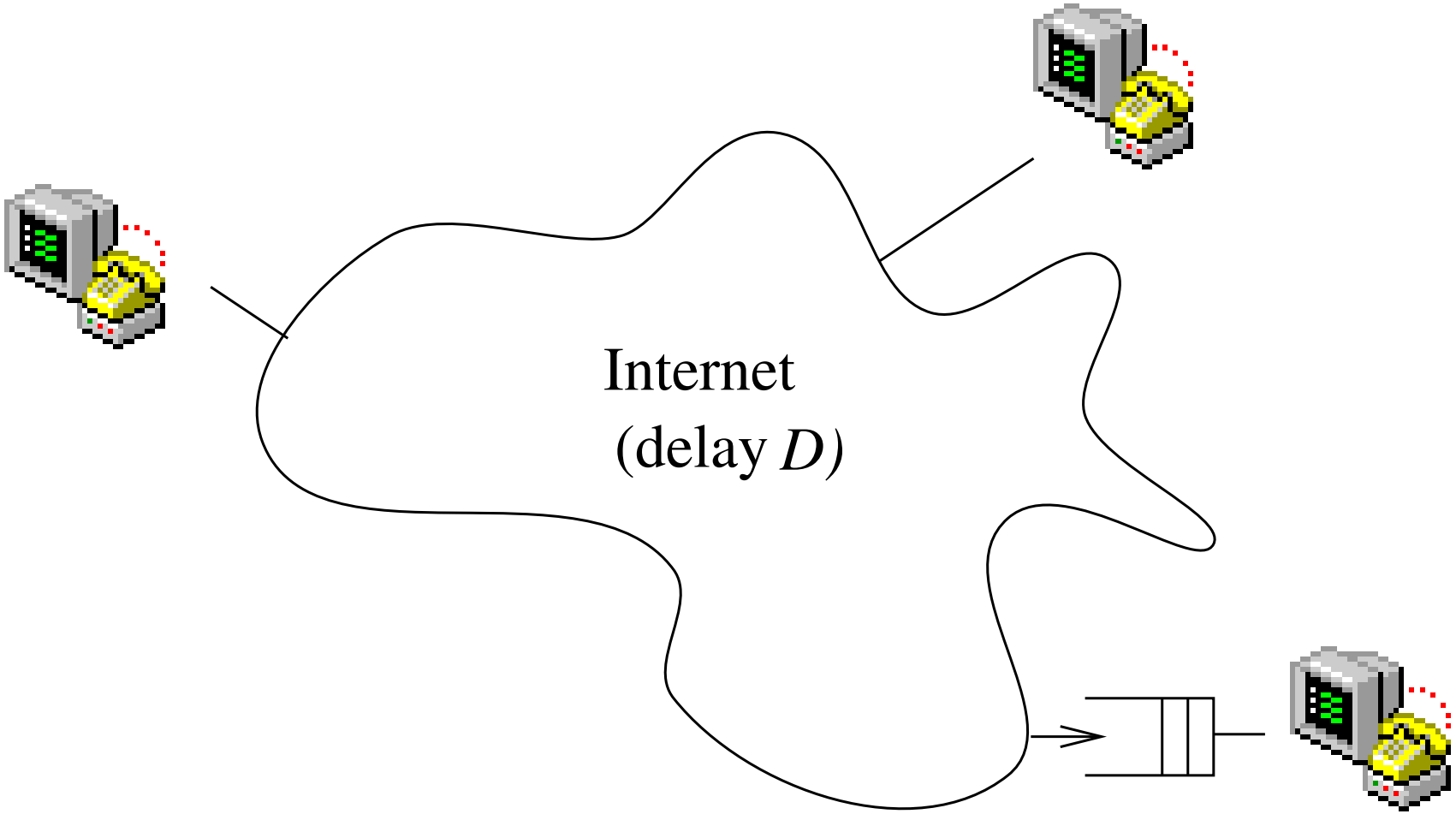
- send in RTCP packet: observed number of users
- compute interval $T = f(\max(\text{observed}, \text{allannounced}))$
- must track announcement for each member

Characteristics:

Learning curve: $L(t)$ = estimate at each member

Packets sent: ideally (no loss): = $L(t)$

Reconsideration: Network Model

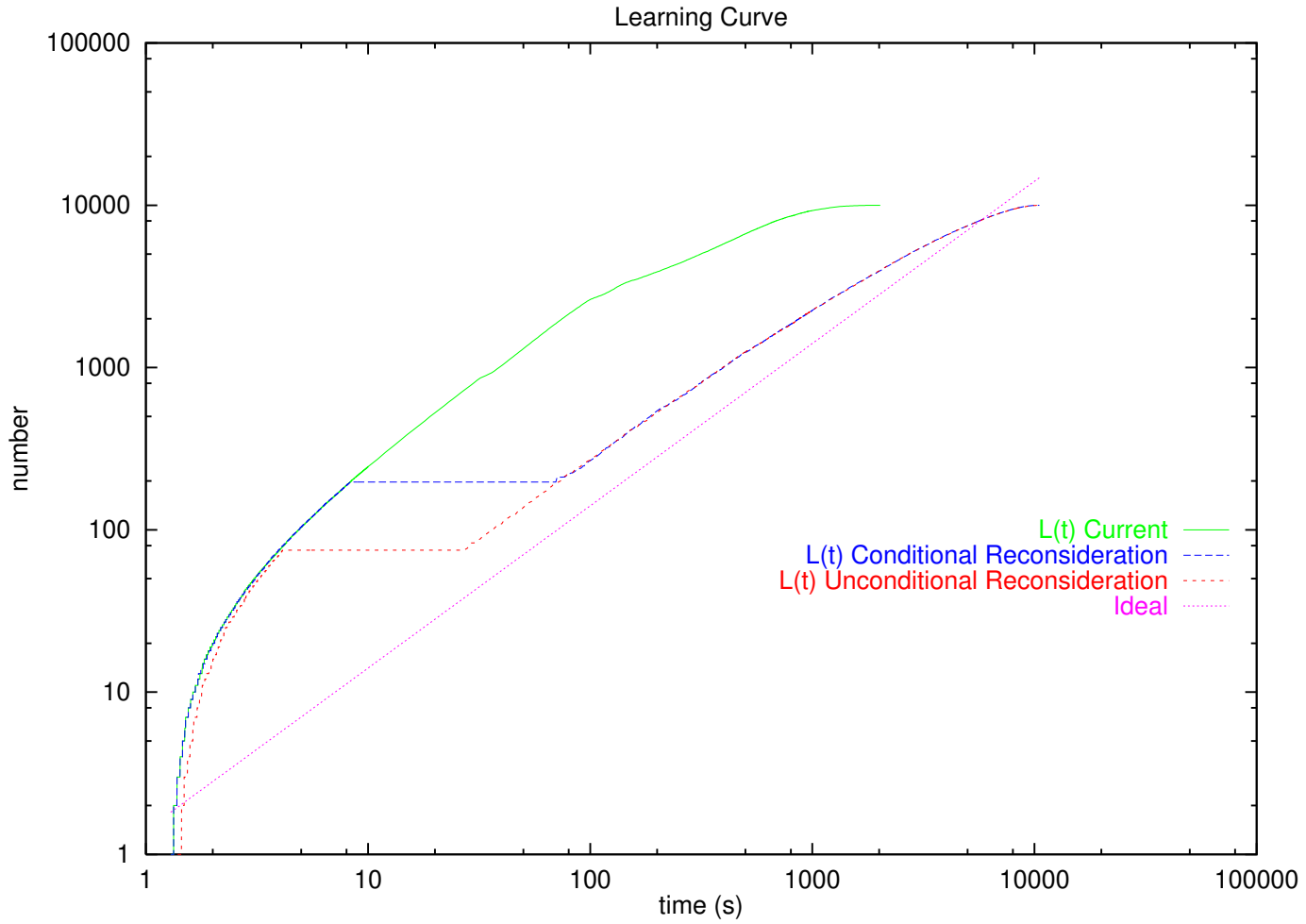


Reconsideration: Experiments

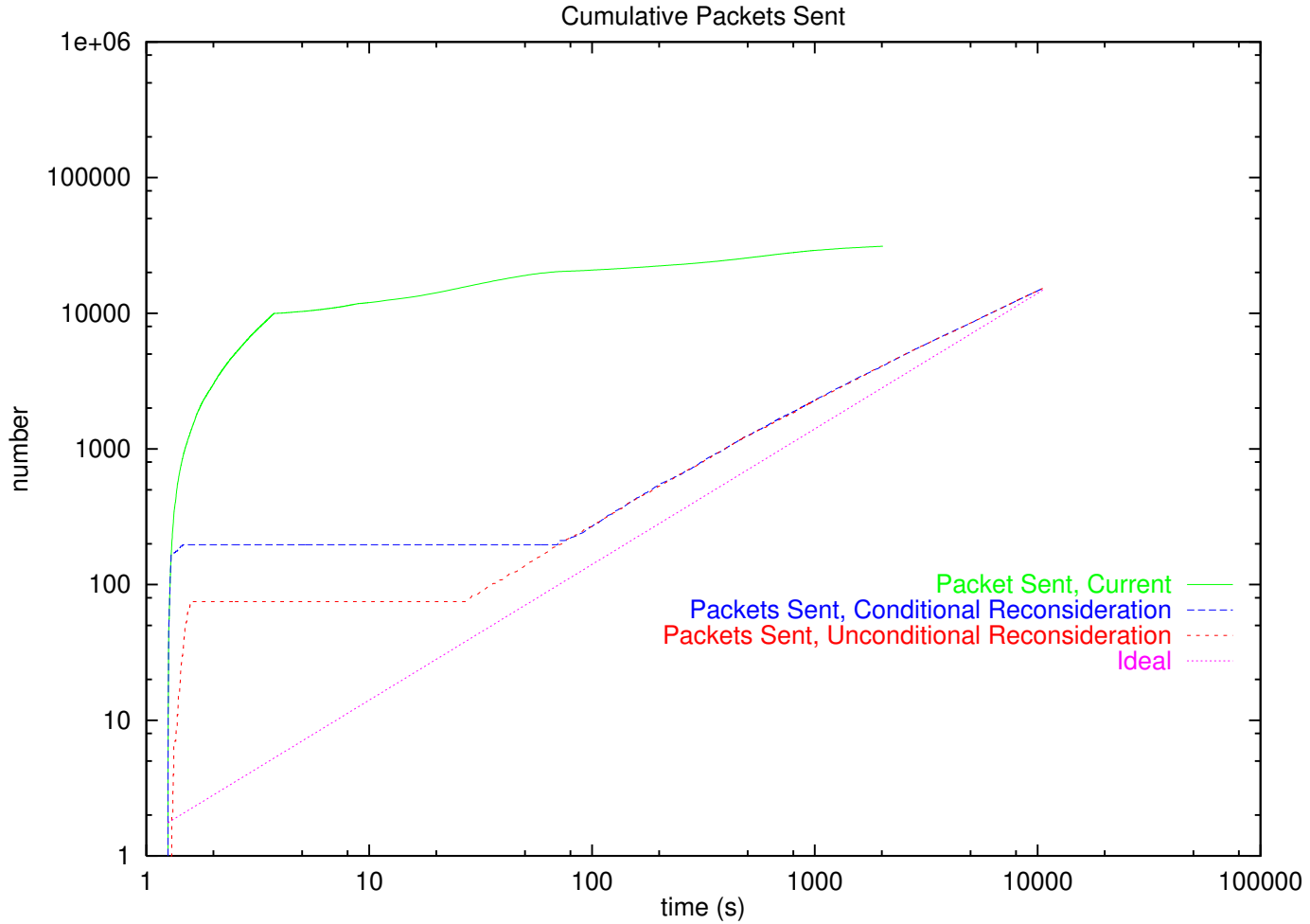
- 10,000 modem (28.8 kb/s) users in group
- join at $t = 0$ or 200 users every 200 ms
- 300 ms average delay, uniformly on [0 ms, 600 ms]
- 100 kB network buffer
- 128 byte RTCP packets

Note: backwards-compatible with existing spec

Reconsideration: Step Join



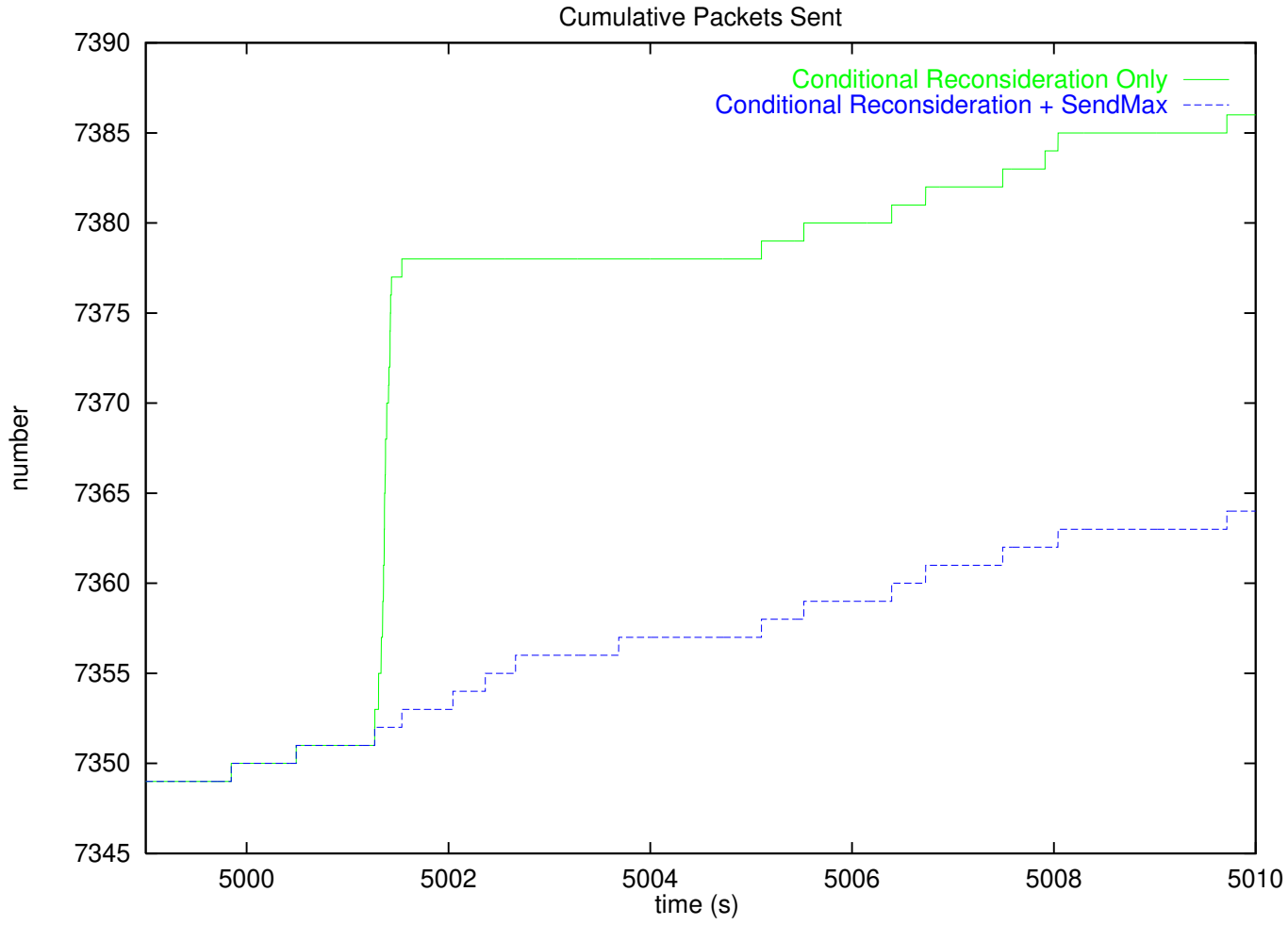
Reconsideration: Step Join



delay ➠ worse; slight dependence on distribution

Reconsideration: Linear Join

$t = 0$: +5000, $t = 5000$: +5000



Internet Telephony Signaling

Session Initiation Protocol (SIP)

- don't need to know IP address of user's host
- many dial-up users do not have a permanent IP address
- invite to new and on-going conferences
- allow for manual and automatic forwarding
- *personal mobility* (complements data link/IP mobility) ➠ change of terminal (PC, digital cordless, palmtop), location
- call center: reach first (load distribution) or reach all (department conference)
- possibly change permissible encodings during conference
- integrate media-on-demand servers and recorders
- separate issue: QoS ➠ RSVP (but: provide parameters)

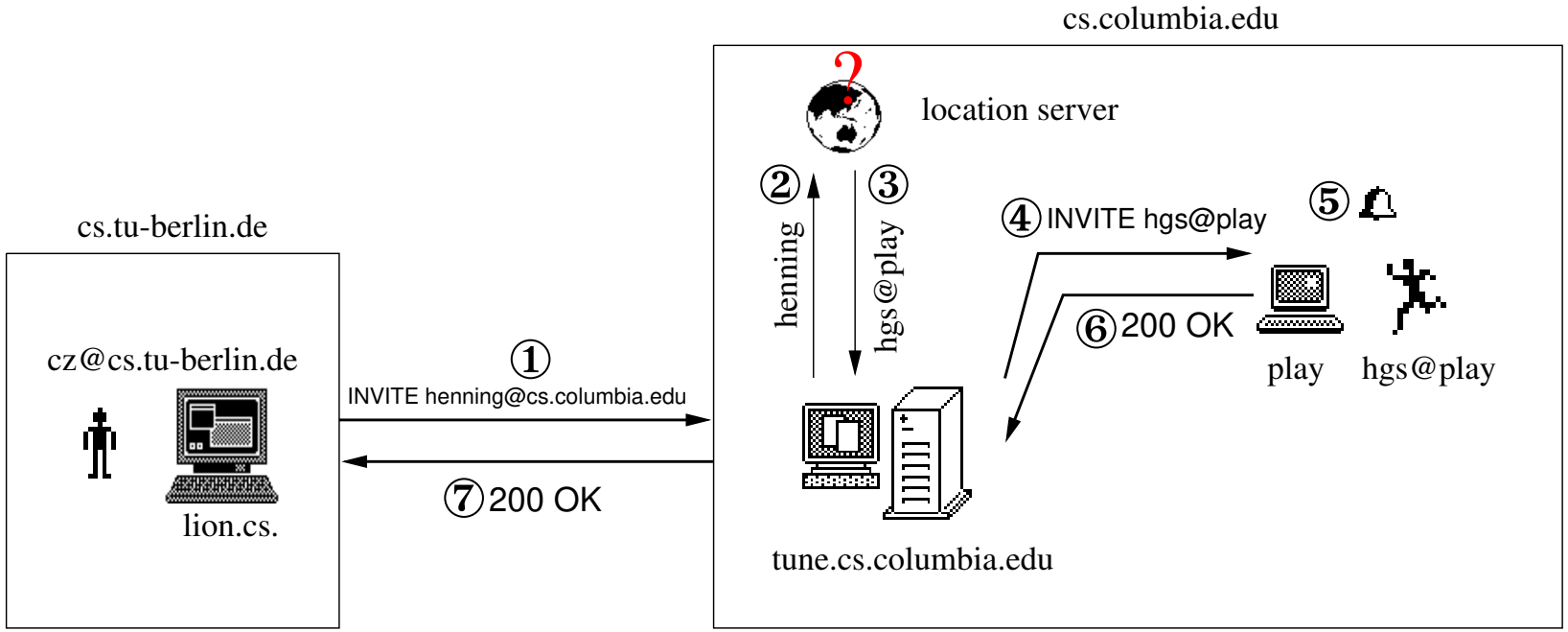
Protocol

similar to HTTP (and RTSP) ➠

- HTTP = “the Internet RPC protocol”
- simple servers are easy, Apache for industrial-strength
- re-use HTTP extensions:
  - authentication
  - PICS = content labeling
  - JEPI = electronic payments
  - PEP = protocol extensions

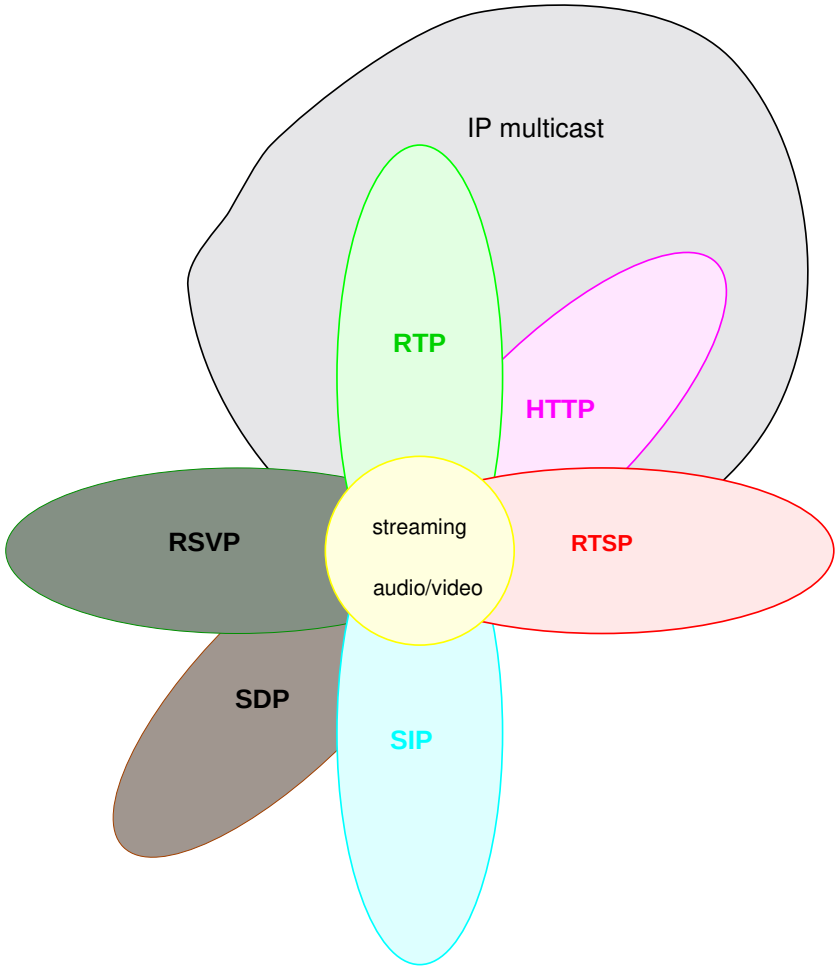
but: also UDP (multicast!), bidirectional

Operation in Proxy Mode



RTSP – Real-Time Stream Protocol

Streaming Multimedia

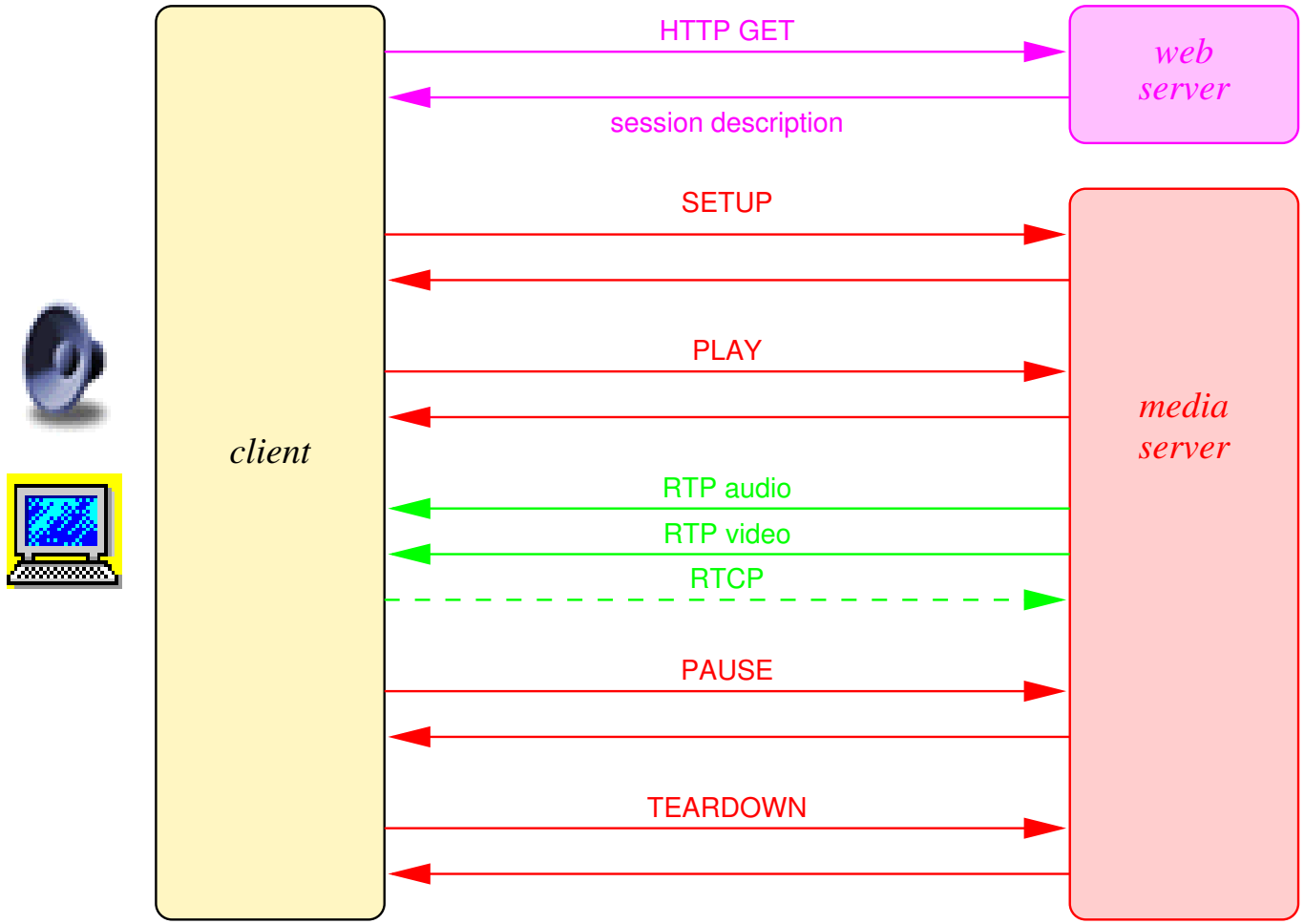


(similar functionality, much simpler than DSM-CC)

RTSP

- similar design as SIP (TCP + UDP, HTTP, . . .)
- TCP connection $\neq$ RTSP association
- “passing around the remote”
- connection/stream, connection/command, UDP
- session by identifier
- supports any session description
- control “tracks” (audio, video) and “presentation” (movie)
- remote digital editing

RTSP Operation



Charging

- currently:
  - access bandwidth for individual, organization, ISP
  - peak and long-term average (MCI T3's)
- multicast reservation: receiver pays (like CATV)
- multicast pipe: ISP loses if it replicates
- telephony: reservations

but: several providers (instead of local/long distance)

- NJ Turnpike: pay at entrance, settlements
- GSP: pay every provider along the way
- 800 services? ➠ need bidirectional reservation

Reliability

- phone switch: 120 seconds/year
- AOL: 88 hours/year for 1996
- ANS: 44 hours/year promised
- Ethernet switch: MTBF $\approx$ 20 years
- router configuration, route flap
- on-line software upgrades
- end-system auto-configuration (already easier than ISDN. . .)

Summary

- price, features ➠ Internet as ISN
- second (last?) chance to design telecom networks
- scaling: routing, reservations, charging, . . .
- reliability