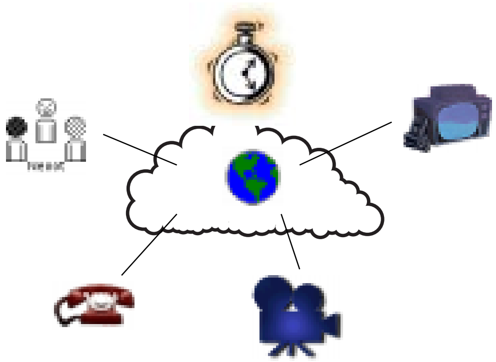
# Predicting Internet Telephony Call Setup Delay

Tony Eyers
TITR
University of Wollongong, Australia
tony@elec.uow.edu.au

Henning Schulzrinne
Department of Computer Science
Columbia University, New York
hgs@cs.columbia.edu



First IP Telephony Workshop, Berlin, April 12-13, 2000

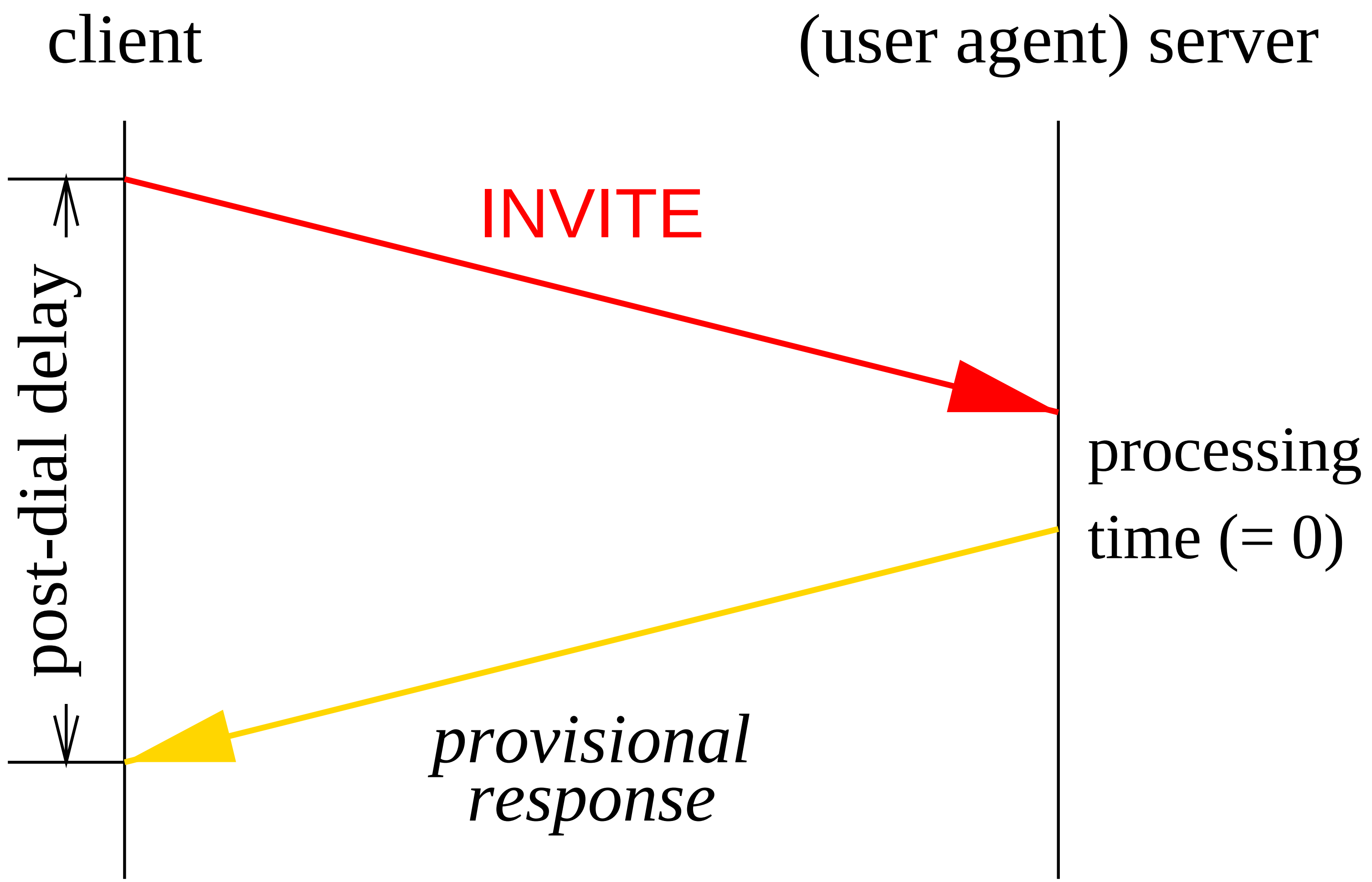
# Internet Telephony QOS

- Major focus on protocol development:
  - H.323, SIP/SDP, pint, ...
  - number mapping & gateway location (enum, TRIP)
  - as yet, no standard definitions
  - IETF work underway to adapt ITU performance targets
- focus so far on Internet *voice* QOS
- Internet Telephony call setup delay, a key QOS indicator, has not been widely studied:
  - IETF drafts on IP signaling transport (SIGTRANS WG)
  - Elwalid et al. consider delay performance of a Lucent H.323 GK

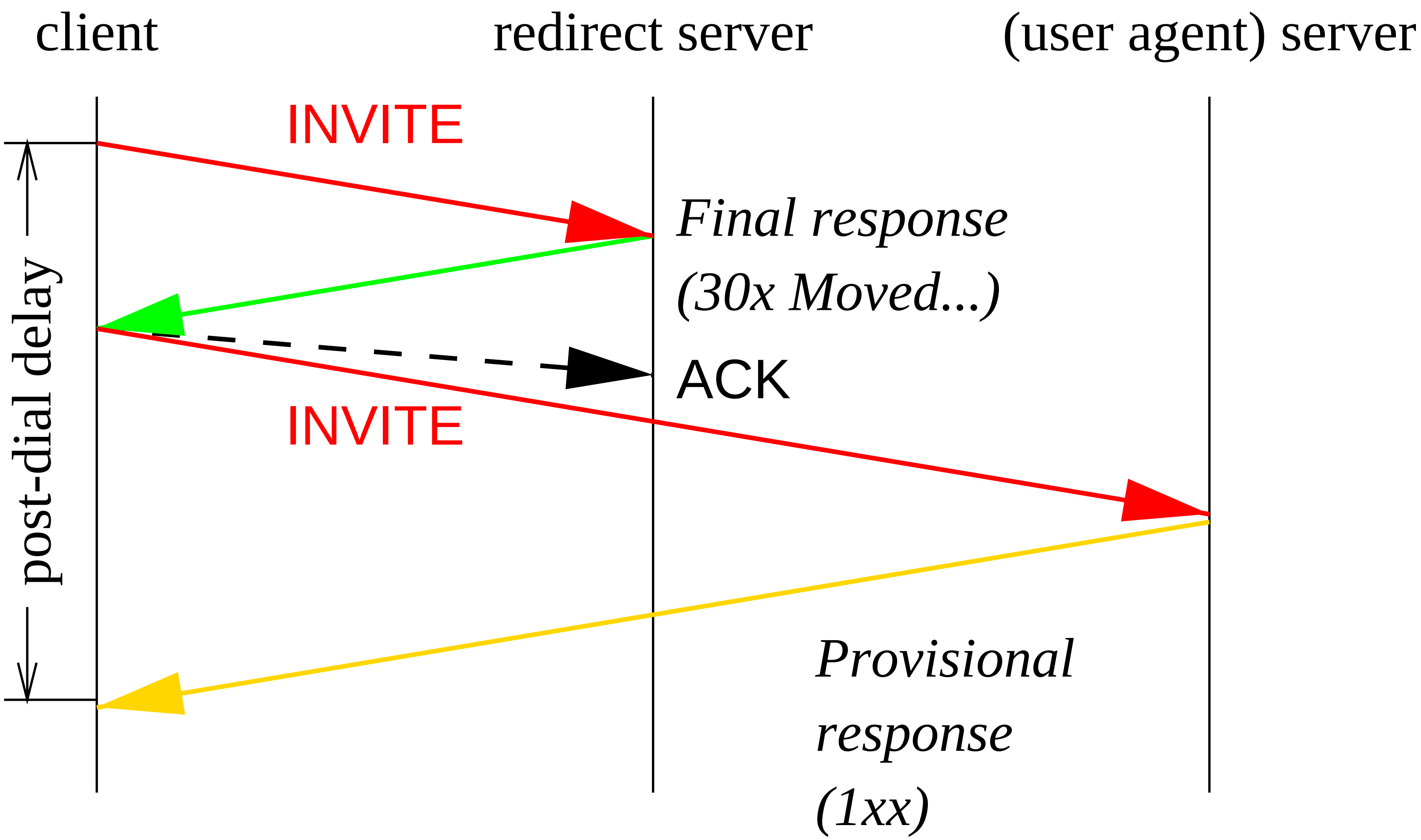
# Problem Statement

- Determine Internet telephony call setup delay distribution, over public Internet
- even with resource reservation or diff-serv for media, signaling likely to be sent best-effort!
- dial-to-ring delay
- key parameters: propagation delay, queueing delay, retransmission delay
- processing delay not considered
- also: compare SIP and H.323 setup delay performance
  - key difference: recovery scheme for lost messages
  - SIP uses UDP, with timeout/retransmission mechanism
  - H.323 uses TCP (mostly)

# SIP Call Setup



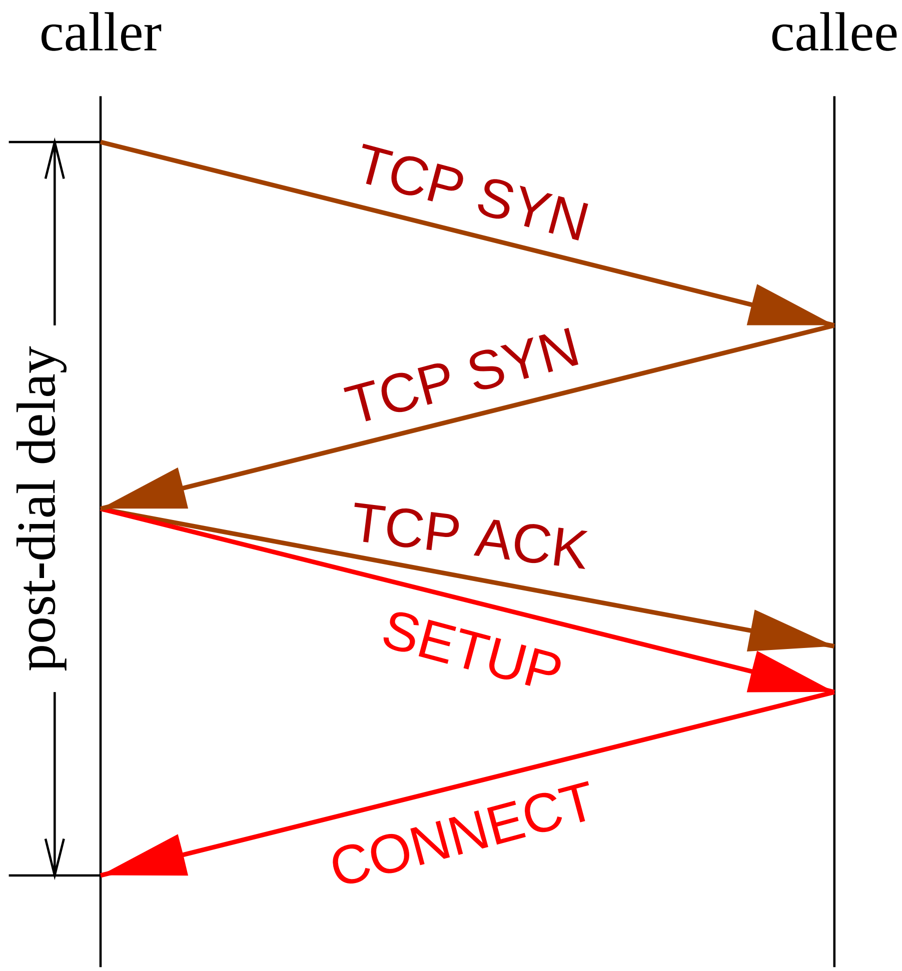
# SIP Call Setup: Extended



# SIP Characteristics

- INVITE over UDP retransmitted until first provisional response (1xx) arrives, at 0.5, 1, 2, 4, 8, 16 seconds intervals, then give up
- optional reliable provisional response mechanism (PRACK), not studied here
- final response (2xx or higher) retransmitted until ACK arrives, with interval capped at 4 s
- final responses also retransmitted when retransmitted INVITEs arrive
- retransmission limited to 7 packets
- other methods (BYE, OPTIONS, CANCEL) do not retransmit responses and cap time-out at 4 s

# H.323 Call Setup



# H.323 Characteristics

- TCP connection establishment adds at least one round trip
- TCP SYN, H.323 SETUP, CONNECT use ’standard’ TCP timeouts
  - recommended initial value 3 seconds (RFC 1122)
  - some implementations start at 6 seconds [Stevens94], Solaris allows initial timeout to be set *whole OS*
  - timeout value increases by a factor of two for each retransmission
- H.323 Annex E provides a UDP option, but not yet implemented

# Internet Performance Measurements: the Surveyor Project

- Run by Advanced Networks and Services
- measures one way UDP delay
  - currently around 40 sites (mostly USA, some in Europe, Korea, New Zealand)
  - each site exchanges 4 40-byte UDP packets each second (2 each way)
  - One way delay measured, resolution 50 $\mu$s
  - packet loss also recorded
- Surveyor site provides:
  - delay/loss histograms for each day, for all site pairs (measurements began in 1997)
  - located at http://www.advanced.org/csg-ippm/
- Surveyor traces provide packet delay/loss used in our simulations

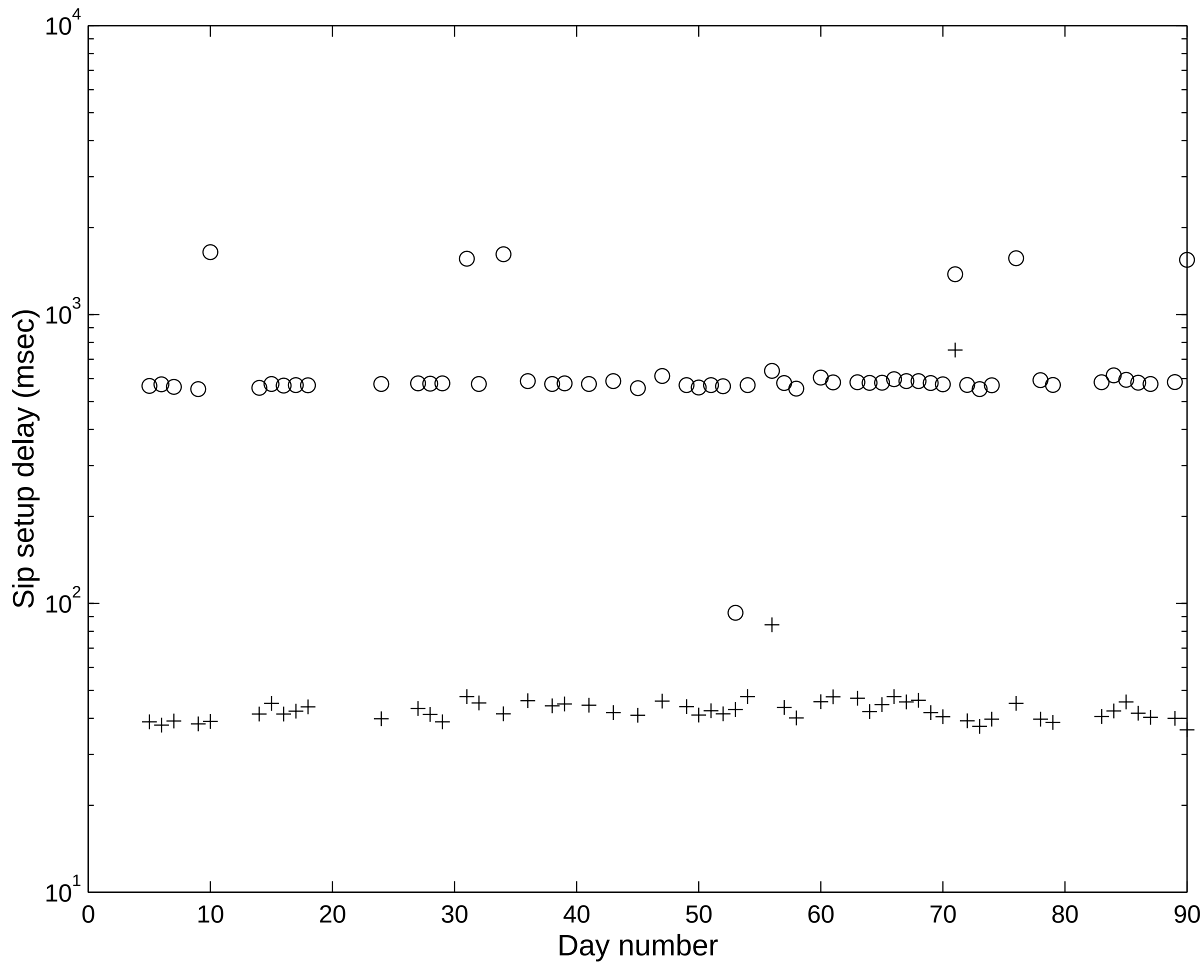
## Simulation Methodology

- Instantaneous delay/loss probabilities constructed from trace records
- two state error model used, to capture bursty errors
- Simulated calls arrive
- 50 calls/sec, Poisson call arrivals
- Call setup delay distribution/call loss rate recorded over simulated interval (e.g., one hour)

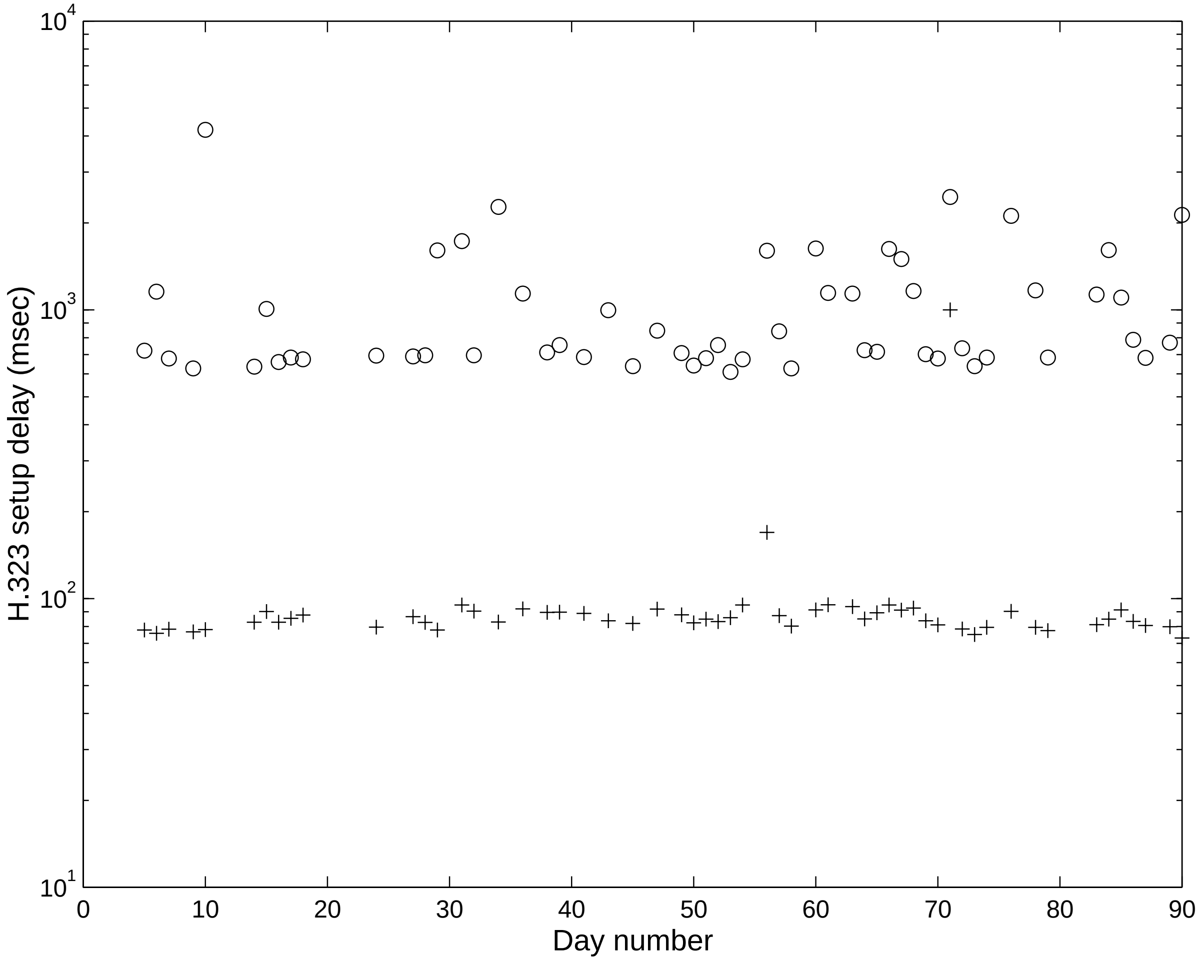
# Case Study

- simulation runs over the first 90 business days in 1999
- 60 minutes per day starting at 16:00
- source-destination pairs:
  - New York-Chicago
  - New York-West Coast
- variations:
  - all calls visit a redirect server in Washington, DC
  - all calls visit a redirect server in Washington, DC, then transit a proxy server in Indiana
  - H.323 (i.e., TCP)
- minimum and 95th percentile of call setup delay

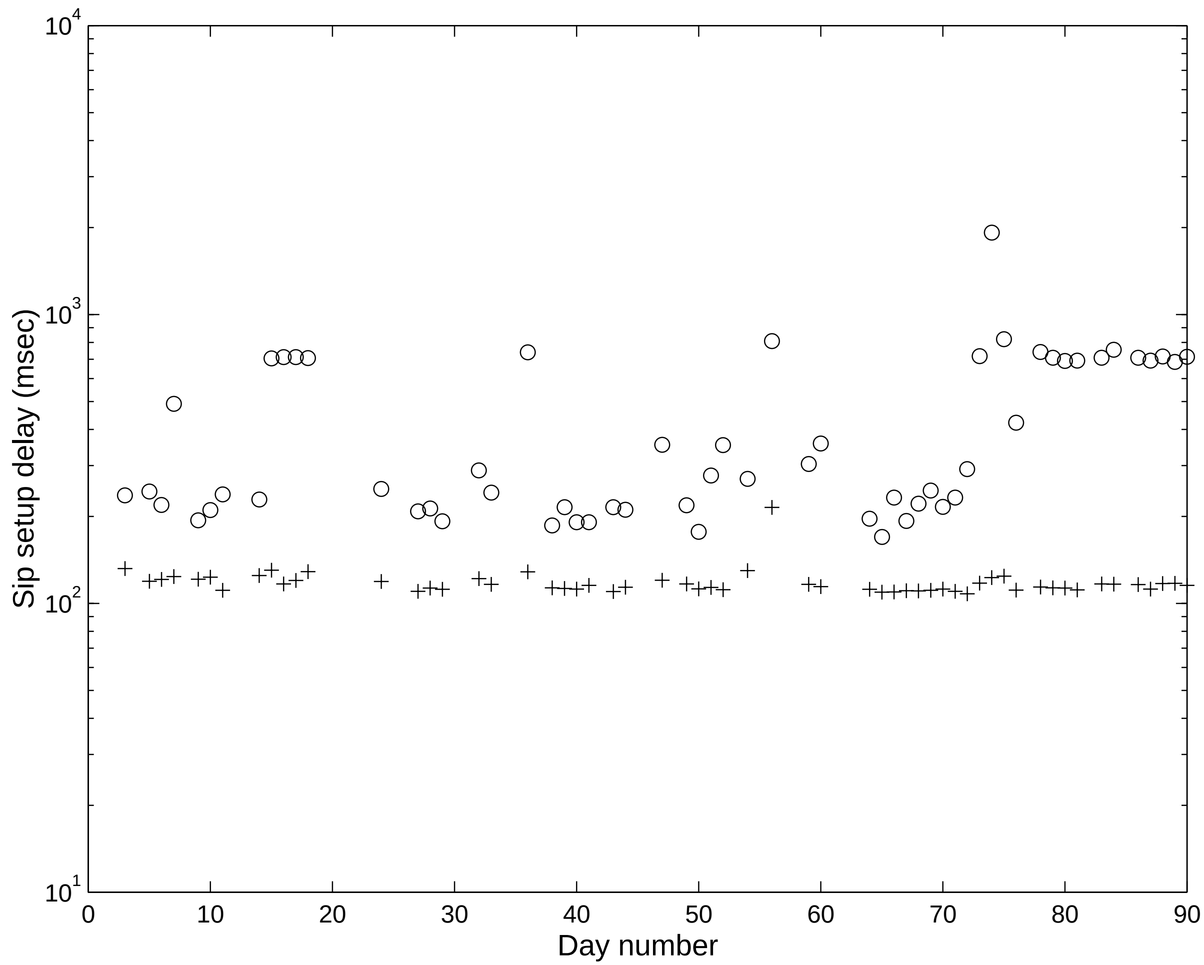
# SIP Call Setup Delay: New York-Chicago



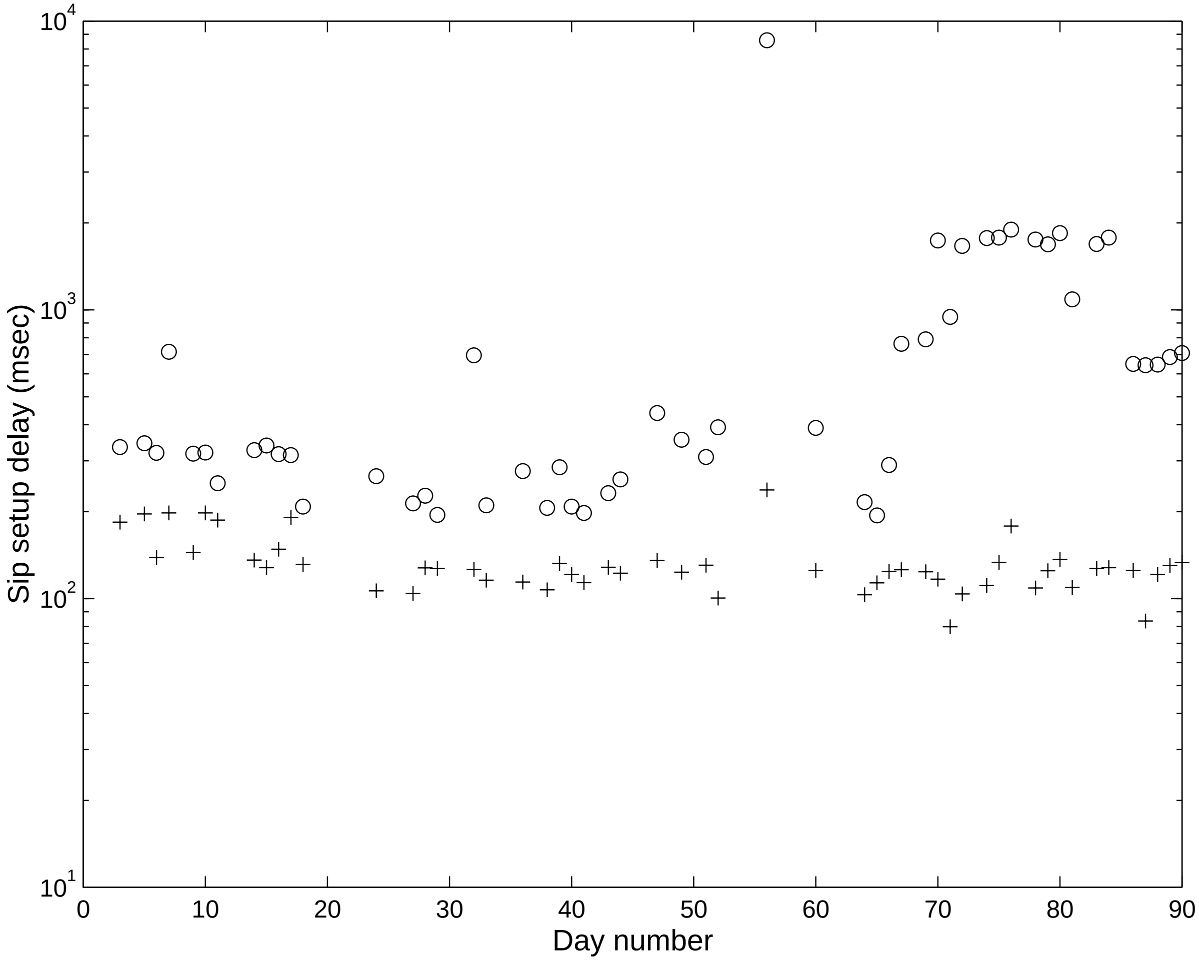
# H.323 Call Setup Delay: New York-Chicago



# SIP with Redirect Server: New York-West Coast



# SIP with Redirect Server: New York-West Coast, via Boston



# Blocking Probability

PSTN interworking: ISDN abandons calls after 2 s

| | Bos. | Chi. | West | Wash. | Colorado |
|---|---|---|---|---|---|
| New York | 20.3 | 77.2 | 32.3 | 9.1 | 15.4 |
| | 28.2 | 94.7 | 40.0 | 20.0 | 18.5 |
| Boston | | 1.6 | 31.5 | 0.0 | 5.4 |
| | | 1.6 | 31.5 | 0.0 | 10.8 |
| Chicago | | | 34.3 | 5.2 | 28.6 |
| | | | 34.3 | 6.9 | 61.4 |
| West Coast | | | | 33.3 | 45.3 |
| | | | | 36.7 | 57.3 |
| Washington State | | | | | 6.6 |
| | | | | | 6.6 |

% of days with PSTN/Internet telephony blocking $>$ 1%, SIP (top row) and H.323 (bottom row)

# Conclusions

- Acceptable SIP call setup delay, for simple calls
- H.323 delay worse at times, due to TCP
- TCP timers will need to be tuned for Internet telephony signalling
- complex call types exhibit variable delay performance
- high blocking probabilities likely if interfacing with PSTN
- related issue: post-pickup delay if call is gated