Quality aspects of audio communication

Ian Marsh

SICS and KTH

Friday June 13th, 2003

Outline of the talk

- Vision, quality and a statement of the problem
- Voice over IP (VoIP) traffic in isolation
- Mixing VoIP traffic with data traffic
- Measuring VoIP quality
- End system adaption to jitter
- Contributions and conclusions

My vision

The Internet will carry a *signifi cant proportion* of the world's telephony traffic

and

the quality should be *no worse* than that offered by the traditional telephony system.

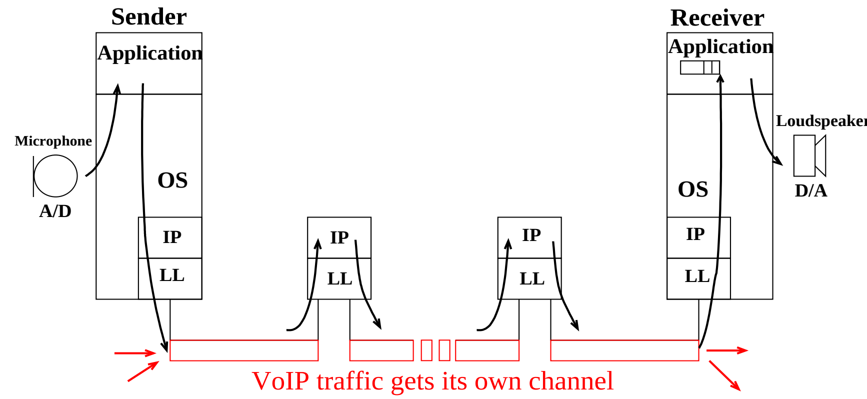
Quality terms used in this presentation

- Delay
  - Delay under 180 ms rated as good
  - Delay between 180 ms and 400 ms rated as acceptable
  - Delays over 400 ms rated as unacceptable
- Loss
  - Losses for simple audio coding e.g G.711 PCM
    - A loss rate of 1% with no packet loss concealment
    - A loss rate of 10% with packet loss concealment
- Jitter
  - Smoothing buffers used to deliver uninterrupted speech, therefore add delay

Problems facing VoIP quality

- Disruption caused by competing TCP traffic
  - Probing available capacity by inducing loss
  - Traffic varies on different time scales
- Poor quality infrastructure
- The end systems
- Human tolerances, the *real* end systems

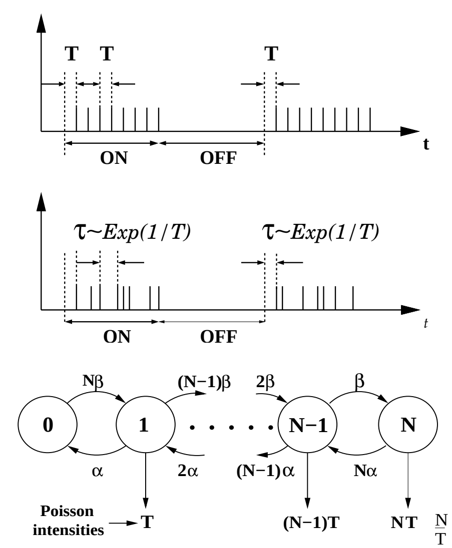
Protecting VoIP traffic



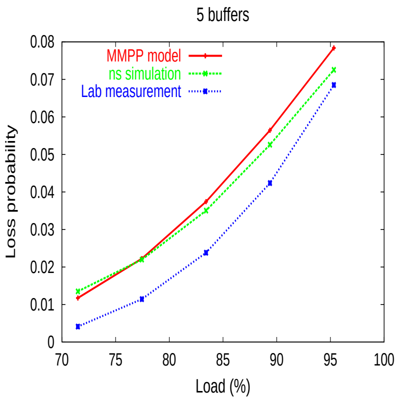
Protecting VoIP traffic

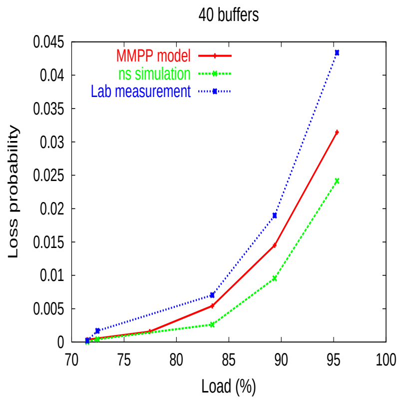
- One solution is to give VoIP its own channel
- Problem becomes capacity allocation for a required quality
  - Telephony and ATM research fields have suggested solutions
  - Largely ignored by the IP community
- Investigate an existing proposal and applied it to IP networks
  - Implement a computationally efficient model
- Use a Markov Modulated Poisson Process (MMPP)
  - To model superposition of independent sources
- Measure the loss probability through a finite buffer
  - Compare the model, simulation and a laboratory setup

A tractable model



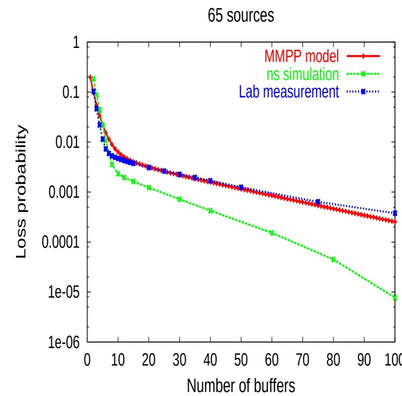
Loss probability against load

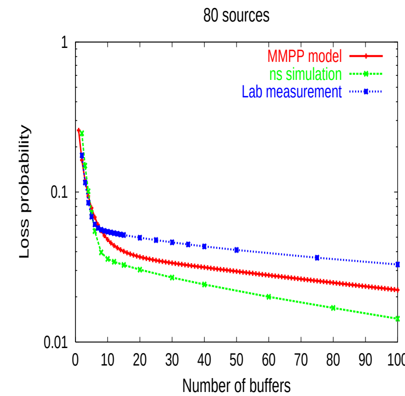




$$Load\ (\rho) = \frac{No.\ sources \times Sourceon_{prob} \times Rate_{peak}}{Link\ capacity}$$

Loss probability against buffer size



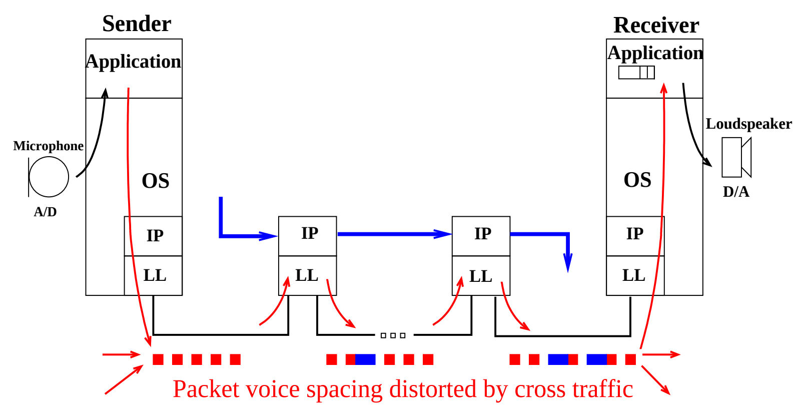


Choose buffer ≈ 8 Admission control ($>$1% loss)

Conclusions for isolating VoIP traffic

- Good agreement between model, simulation and laboratory for small buffer sizes
- Some difference in the larger buffer cases
- Model is flexible, calculate number of sources, buffer size, capacity, coding rate etc.
- Statistical multiplexing gains
  - Peak rate allocation gives 25 sources for 85% load
  - 72 sources can be admitted for 1% loss for 85% load
- Joint work with B. Ahlgren, A. Andersson and O. Hagsand
- My contribution: Initial idea, laboratory environment, writing

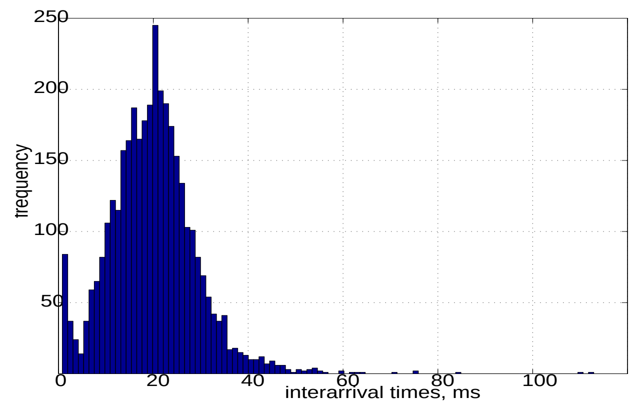
Mixing VoIP and data traffic



Mixing VoIP and data traffic

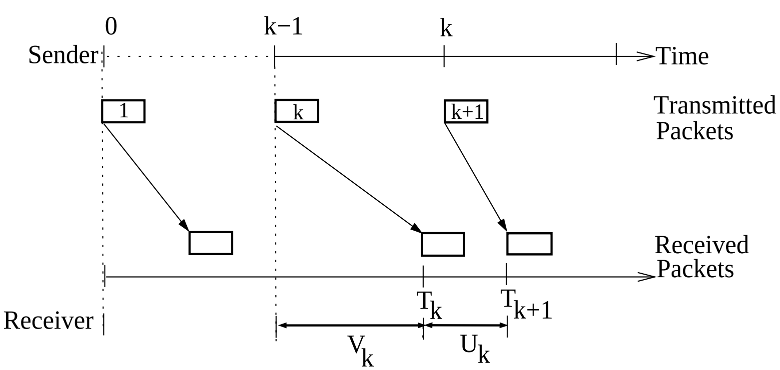
- Vast majority of the data on the Internet is TCP
  - Approximately 90% packets, 96% bytes (Sprint labs)
- So VoIP must coexist with bulk data in non isolated networks
- Our goals are:
  - Gain insight into how VoIP is affected
  - Understand what causes the packet delays in voice sessions?
  - Use this knowledge to write more realistic traffic generators
- Use the artificial generators to test new playout algorithms

VoIP interarrival histogram



A VoIP session from Argentina to Stockholm, packetisation time 20 ms

Packet delay model



- Y_k is a random network delay sequence, with a general unknown distribution $F(x)$
- It has a finite mean e.g. 20 to 40 time slots (400 ms to 800 ms)
- Y_k is *non-observable*, whereas T_k, V_k and U_k are observable

Packet delay model

Packet k needs Y_k to propagate through network
If it is not delayed it arrives at $T_k = k - 1 + Y_k$
If it is delayed by packets in front of it the arrival times satisfy:

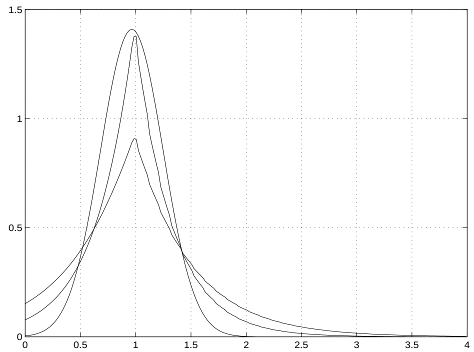
$$T_1 = Y_1, \quad T_k = \max(T_{k-1}, k - 1 + Y_k) \quad k \geq 2$$

(Note: packetisation time k used in the model but not the length)

Since T_{k-1} and Y_k are independent, the arrival times (T_k) form a transient Markov chain, note two conditions needed for Markovity:

- Probability of the future state only depends on the present state (memoryless property)
- Probabilities don't change at each step (time homogeneous)

Density function of U

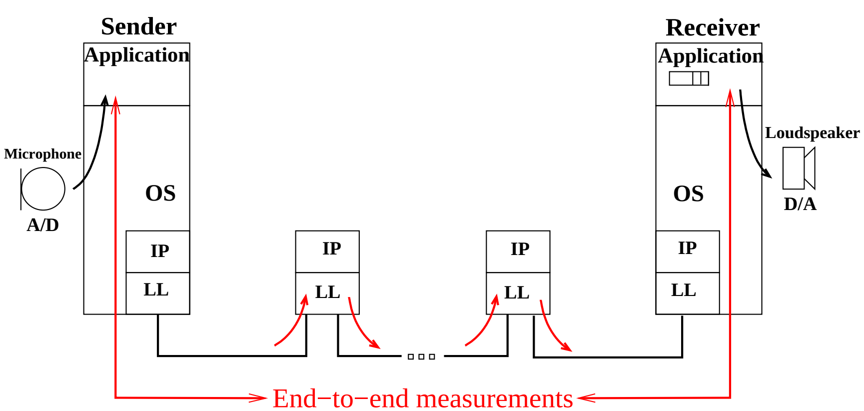


- Density function $\frac{d}{dx}P(U_\infty \leq x)$ for three choices of $F(x)$
- Choices of $F(x)$ are 2 exponentials and 1 Gaussian
- $x = 0$ back-to-back packets, $x = 1$ the packetisation interval and $x > 1$ spread out

Conclusions when mixing VoIP and data traffic

- By observing the arrivals we can make some assumptions and to model the process
- Model includes silence suppression and loss
- Some insight into the possible cause of delay
  - Cross traffi c causing the tail
  - Back to back packets at point mass $x = 0$
- Can produce a traffi c generator from the distribution of U_k
  - For testing new jitter playout algorithms
- Joint work with I. Kaj, Uppsala University
- My contribution: Data collection, writing

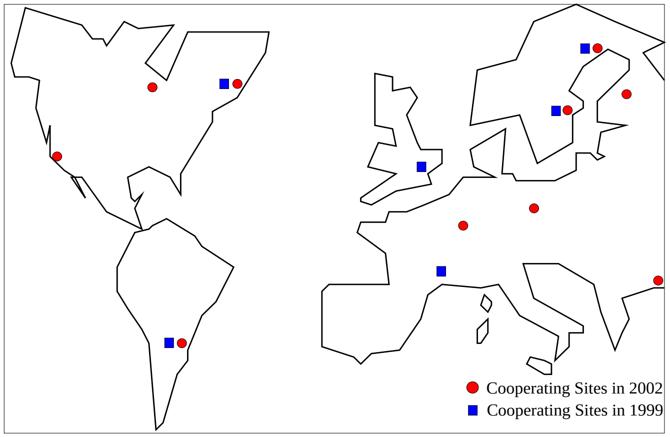
Measuring VoIP quality



Measurement goals

- Loss, delay and jitter measurements for VoIP
  - Give some indication of the trend in VoIP quality
  - Compared to 1999: Use a bidirectional full-mesh
    - ∗ Investigate asymmetry issues
    - ∗ Time of day effects
  - Compared to 1999: Vary VoIP parameters
    - ∗ Different packet sizes
    - ∗ Silence suppression
- Construct a robust measurement infra-structure
- Collect a repository of VoIP sample sessions

Co-operating Test Sites in 1999 and 2002



A 70 second pre-recorded speech pattern was sent between the sites once an hour

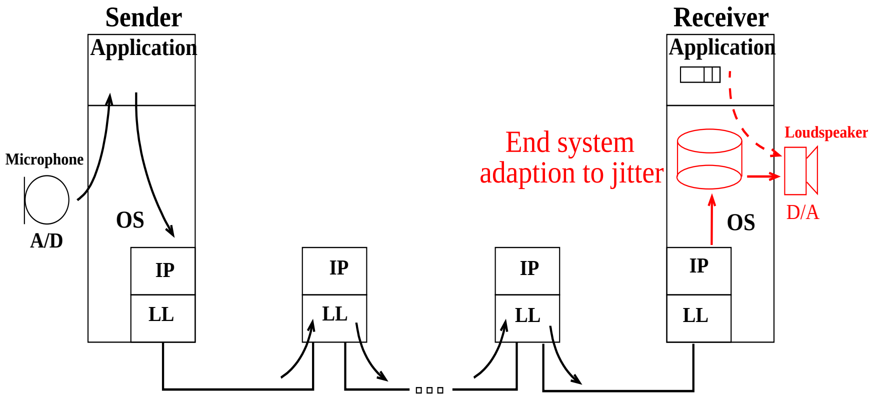
Used and obtained data in 2002

Test "signal"	
Call duration	70 seconds
Payload size	160 bytes
Packetisation time (ms)	20ms
Data rate	64kbits/sec
Without silence suppression	3653 packets
With silence suppression	2043 packets
Coding	8 bit PCM
Recorded call size	584480 bytes
Obtained data	
Number of hosts used	9
Number of traces obtained	22436
Number of data packets	32,771,021
Total data size (compressed)	411 Megabytes
Measurement duration	12 weeks

Measurement Results

- Most sessions show less than 5% loss and under 10 ms jitter
  - VoIP quality is good using academic networks
  - Quality has improved slightly over the past 3 years
  - VoIP not usable on a global scale, 2 sites show poor quality
  - Infra-structure not distance (or number of hops) important
    * Most losses are single losses
    * Suppression, packet sizes and asymmetry not big factors
- Not much time for the end systems to deliver the packets
- Joint work with F. Li
- My contribution: initial idea, data processing, writing

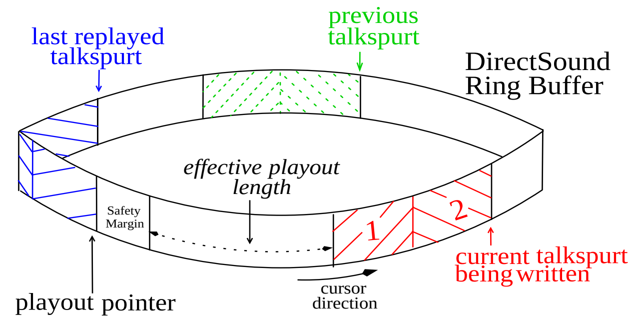
End system adaption to jitter



Three tier solution

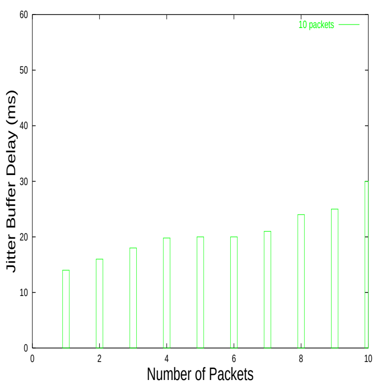
1. Use low-level functionality of the sound card
   - Memory in the audio card *becomes* the playout buffer
   - Sound card timers determine whether packets are early/late
2. Statistical per packet playout algorithm
   - Standard algorithm as proposed by Van Jacobson (1994)
   - Keep a (sorted) running history of the arrival times
   - If loss can be tolerated, the delay can be reduced
3. Limit the frequency of buffer size changes
   - Jitter calculations are bounded to upper and lower limits
   - Alleviates unnecessary changes in the system

Directsound playout support

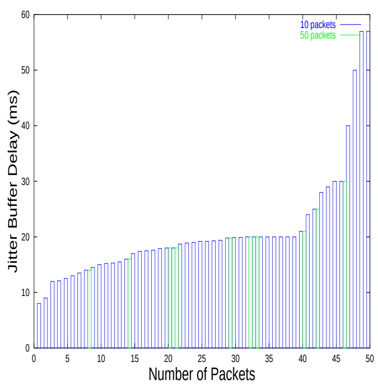


- Implemented using DirectX interface by Microsoft
- Circular buffer, pointers rotate anti-clockwise
- Talkspurts written contiguously, adapt during silence periods

Benefit of history

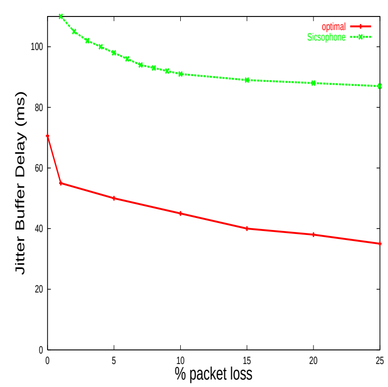


10 packets

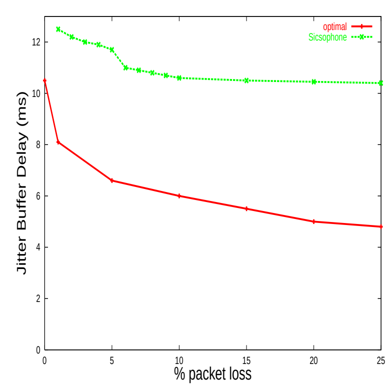


Original 10 plus 40 new arrivals

Sics*o*phone compared to optimal playout



California to France



Amherst to Berlin

Mouth-to-ear delay measurements

Audio Tool	Latency (ms)
Sics*o*phone prototype	25-100
Vocal Internet Phone 4.5	450-550
NetMeeting 2.1	620
VAT 3.4 (Solaris)	1200
RAT 3 (Solaris)	1500

Windows (98, NT) operating systems with SoundBlaster audio cards
Simple square wave used, easier triggering and delay calculation

End system adaption to jitter conclusions

- Other tools were not tuned, just used out of box
- Precious time can be saved by careful buffer management
- Inter-layer copying slows down delivery of audio
- Standard algorithm used, further delay reduction is possible
- Joint work with O. Hagsand and K. Hansson
- My contribution: RTCP protocol, optimal playout, writing

Contributions of this thesis

- Addressed topics are real-world problems, and implementable
- Implemented approaches to reduce the mouth-to-ear delay
- Used a variety of methods to address the problems:
  - Modelling, simulation, measurement and implementation
- Suggested approaches not dependent on ongoing QoS research
- Largest publicly available VoIP repository
- Sics*o*phone and measurement work:
  - Together yield an estimate of the mouth-ear wide-area delay
  - Used in other efforts, probing measurements, loss process, early estimation of quality

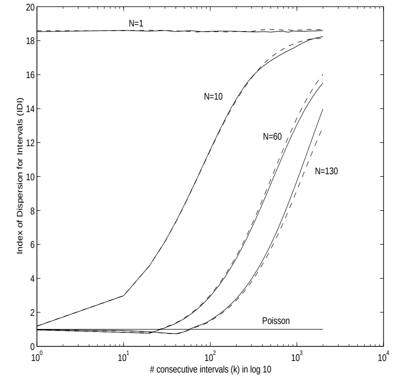
Conclusions

- Good quality Voice over IP is possible on todays Internet
- Best solution is to isolate VoIP traffic
- Well provisioned links are almost as satisfactory
- Certain infrastructures need to be upgraded for VoIP
- End systems should not be ignored
- Problems exist outside of this thesis
  - Reliability of routers and end systems
  - Understanding the nature of bulk traffic
- But, we are infinitely adaptable :-)

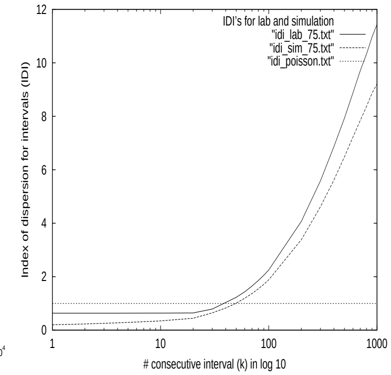
Future Work

- Will use more techniques to solve the remaining problems
  - Control theory (feedback from VoIP receiver to sender)
  - Network calculus (delay bounds for VoIP system)
  - Optimisation technqiues (adaption jitter buffer)
  - Queuing theory (variable portion of the mouth-to-ear delay)
- More measurements (access networks)
- Deeper analysis of measurement data:
  - Loss pattern detection (Gilbert model)
  - Correlation in jitter patterns
- Implementation work within Sics*o*phone: SIP, iLBC

Differences in Model/Lab/Sim - Burst Evaluation

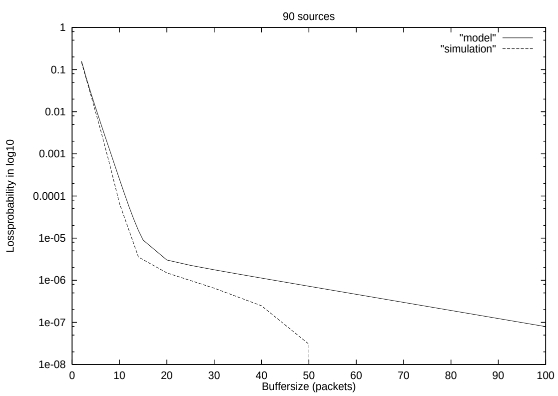


Model only



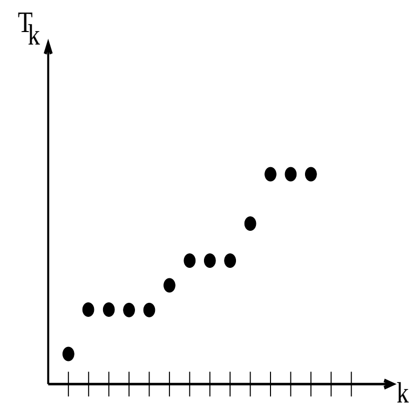
Simulation and laboratory

Differences between model and simulation



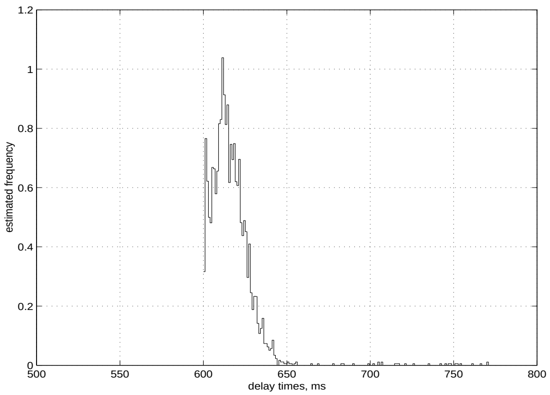
A 100 packet buffer with 90 sources loss probability 10^{-8} rare
We needed over 1 day to get the losses shown above and still not close

T_k process



Shows the progression of the interarrival time T_k for each packet k

Estimation of Y(x)



Estimated from observed delays $\bar{F}_V$ of V by taking: $\bar{F}(x) = \frac{\bar{F}_V(x)}{\bar{F}_V(x+1)} \quad x \geq 0$
x is measured in 20 ms intervals

Silence Suppression

Introduced to reduce load on network when talker is silent

Silence suppression is an additional source of random delay

X_k = duration of silence between packets $k-1$ and k.

Assume silence suppression intervals are independent of (Y_k)

$$G(x) = P(X_k \leq x),$$

$S_k = \sum_{i=1}^{k} X_i$ = total time of silence suppression affecting packet k

$$T_1 = S_1 + Y_1, \quad T_k = \max(T_{k-1},\, k-1+S_k+Y_k), \quad k \geq 2,$$

Loss

Probability p that a packet is subject to loss, independent $Y(x)$

K_k no. of attempts between successful packets $k-1$ & k, $k \geq 1$

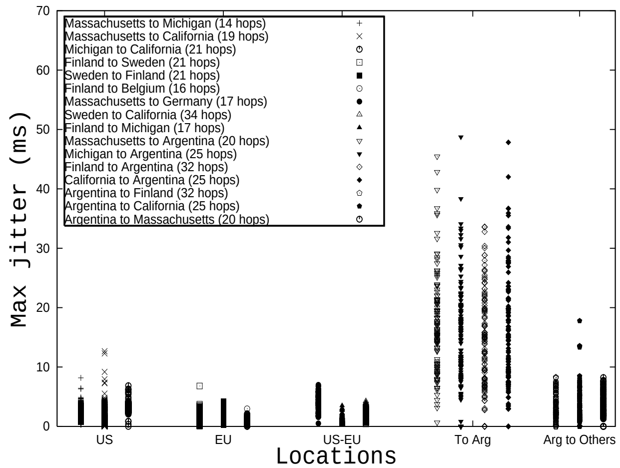
Which gives a sequence $(K_k)_{k\geq 1}$ of iid variables with the geometric distribution $P(K_k = j) = (1-p)p^j, \quad j \geq 0.$

L_k is a seq of variables with a negative binomial distribution

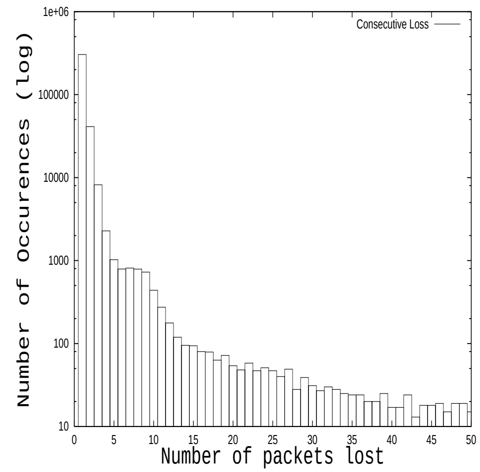
Arrival times are now:
$T_1 = K_1 - 1 + Y_{K_1}, \quad T_k = \max(T_{k-1},\, L_k - 1 + Y_{L_k}), \quad k \geq 2.$

receiver sender	Massachusetts	Michigan	California	Belgium	Finland	Sweden	Germany	Turkey	Argentina	Mean
Massachusetts	*	D:38.0 (17.1) J:2.4 (1.7) L:0.1 (0.6) H:14 (+1) T:0	D:54.2 (15.8) J:2.4 (1.8) L:0.1 (0.9) H:19 T:-3	D:67.1 (15.5) J:3.6 (1.5) L:0.1 (0.8) H:11 T:+6	D:97.1 (2.6) J:2.5 (1.5) L:0.1 (0.8) H:15 T:+7	D:99.5 (8.5) J:3.2 (1.7) L:0.04 (0.2) H:21 T:+6	D:58.4 (5.0) J:4.5 (1.4) L:0.0 (0.0) H:17 (+3) T:+6	D:388.2 (43.2) J:10.4 (4.9) L:4.9 (4.7) H:20 T:+7	D:99.7 (4.9) J:19.9 (8.4) L:8.9 (7.2) H:25 T:+1	D:112.8 J:6.1 L:1.2 H:17
Michigan	D:36.4 (15.4) J:4.7 (0.8) L:0.0(0.2) H:14 (+1) T:0	*	D:40.4 (4.5) J:4.4 (0.8) L:0.2 (1.1) H:20 (+1) T:-3	D:63.5 (4.2) J:4.3 (0.7) L:0.0 (0.1) H:11 T:+6	D:88.2 (8.0) J:4.1 (0.7) L:0.1 (1.1) H:17 T:+7	D:86.7 (4.7) J:5.2 (0.6) L:0.1 (2.2) H:23 T:+6	D:63.6 (8.2) J:7.3 (1.9) L:0.2 (0.9) H:16 (+1) T:6	D:358.9 (44.9) J:5.6 (1.7) L:3.0 (1.9) H:20 T:7	D:112.1 (10.6) J:18.7 (7.9) L:6.5 (7.0) H:25 T:+1	D:106.2 J:6.8 L:1.3 H:18
California	D:54.5 (16.7) J:2.0 (1.0) L:0.1 (0.36) H:18 (+1) T:+3	D:40.6 (5.1) J:1.2 (0.6) L:0.1 (1.9) H:21 T:+3	*	D:81.0 (2.2) J:1.6 (0.8) L:0.2 (0.8) H:20 T:+9	D:106.0 (3.0) J:1.4 (0.8) L:0.6 (1.4) H:25 (+1) T:+10	D:108.0 (2.4) J:2.1 (0.9) L:0.2 (0.3) H:30 (+2) T:+9	D:81.5 (1.8) J:4.9 (1.5) L:2.8 (3.0) H:23 T:+9	D:386.9 (60.5) J:5.3 (1.7) L:4.4 (2.4) H:23 T:+10	D:123.9 (12.4) J:18.1 (9.9) L:8.9 (8.2) H:25 T:+4	D:122.2 J:4.6 L:2.2 H:23
Belgium	D:65.2 (10.1) J:1.6 (0.6) L:0.0 (0.0) H:16 T:-6	D:63.4 (3.3) J:0.6 (0.1) L:0.0 (0.0) H:17 T:-6	D:84.0 (1.3) J:0.9 (0.8) L:1.2 (1.0) H:23 T:-9	*	D:31.3 (0.6) J:0.9 (0.5) L:0.0 (0.0) H:17 T:+1	D:33.4 (0.2) J:1.6 (0.9) L:0.0(0.0) H:22 T:0	D:16.6 (10.4) J:3.4 (1.5) L:0.21 (0.7) H:13 T:0	D:341.1 (24.7) J:6.9 (2.0) L:3.8 (2.7) H:16 (+2) T:+1	D:136.5 (7.1) J:NA L:NA H:19 T:-5	D:96.4 J:2.0 L:0.6 H:17
Finland	D:97.8 (4.2) J:1.7 (0.8) L:0.0 (0.1) H:15 (+1) T:-7	D:86.8 (1.9) J:1.1 (0.6) L:0.0 (0.3) H:17 (+1) T:-7	D:109.9 (4.7) J:1.4 (0.8) L:0.7 (1.4) H:24 (+2) T:-10	D:30.7 (0.3) J:1.4 (0.6) L:0.1 (0.3) H:16 T:-1	*	D:13.6 (1.0) J:1.9 (0.9) L:0.0 (0.0) H:20 T:-1	D:26.8 (7.3) J:3.9 (1.1) L:0.0(0.0) H:20 (+1) T:-1	D:321.2 (39.3) J:3.4 (1.7) L:3.2 (1.7) H:17 (+2) T:0	D:161.5 (12.2) J:17.4 (8.2) L:7.5 (6.5) H:19 T:-6	D:106.3 J:4.1 L:1.4 H:18
Sweden	D:99.3 (8.8) J:3.0 (1.9) L:0.0 (0.0) H:22 (+1) T:-6	D:84.9(1.9) J:2.5 (2.0) L:0.03 (0.4) H:25 T:-6	D:105.6 (2.1) J:3.2 (1.96) L:0.1 (0.1) H:30 T:-9	D:33.3 (0.4) J:2.8 (1.6) L:0.1 (0.3) H:24 T:0	D:13.5 (0.5) J:2.4 (1.8) L:0.0 (0.01) H:21 T:+1	*	D:29.8 (12.8) J:4.8 (2.5) L:0.0 (0.0) H:25 T:0	D:322.2 (30.3) J:3.2 (1.49) L:2.9 (1.0) H:26 T:+1	D:165.6 (17.9) J:NA L:NA H:41 T:-5	D:107.8 J:2.8 L:0.4 H:26
Germany	D:63.5 (9.6) J:1.72 (0.7) L:0.0(0.0) H:15 T:-6	D:60.4 (0.5) J:0.7 (0.3) L:0.0 (0.0) H:16 T:-6	D:84.4 (1.0) J:1.8 (0.7) L:2.5 (1.9) H:22 T:-9	D:11.1 (0.2) J:0.8 (0.3) L:0.0 (0.0) H:12 T:0	D:27.8 (7.3) J:1.0 (0.5) L:0.0 (0.0) H:17 T:+1	D:29.2 (7.6) J:1.5 (0.6) L:0.0 (0.0) H:22 T:0	*	D:300.7 (39.7) J:4.8 (2.1) L:3.7 (2.5) H:16 T:+1	D:149.8 (15.6) J:NA L:NA H:18 T:-5	D:90.9 J:1.6 L:0.8 H:17
Turkey	D:379.1 (47.1) J:8.6 (0.7) L:8.1 (2.8) H:18 (+1) T:-7	D:387.9 (35.5) J:8.9 (1.2) L:8.0 (2.9) H:20 T:-7	D:410.9 (43.9) J:8.8 (2.5) L:7.6 (6.8) H:19 T:-10	D:330.2 (28.6) J:9.2 (2.0) L:7.10 (4.0) H:17 T:-1	D:318.9 (42.4) J:8.8 (0.6) L:7.8 (2.7) H:19 T:0	D:311.1 (8.3) J:9.1 (0.7) L:8.4 (3.1) H:25 T:-1	D:378.2 (49.3) J:10.7 (1.2) L:8.0 (3.1) H:16 T:-1	*	D:490.8 (26.0) J:NA L:NA H:18 T:-6	D:375.9 J:8.0 L:6.9 H:19
Argentina	D:117.0 (30.8) J:4.2 (2.0) L:0.5 (1.4) H:NA T:-1	D:146.7 (44.2) J:4.3 (2.3) L:0.5 (1.5) H:NA T:-1	D:152.0 (47.8) J:3.1 (2.4) L:0.6 (1.8) H:NA T:-4	D:NA J:4.2 (2.0) L:0.5 (1.4) H:NA T:+5	D:164.1 (27.2) J:3.9(2.2) L:0.5 (1.4) H:NA T:+6	D:160.9 (47.7) J:2.9 (0.8) L:0.0 (0.1) H:NA T:+5	D:180.5 (50.5) J:4.7 (1.5) L:0.1 (0.1) H:NA T:+5	D:NA J:6.0(1.2) L:5.8 (3.0) H:NA T:+6	*	D:115.2 J:4.2 L:1.1 H:NA
Mean	D:114.1 J:3.4 L:1.1 H:14	D:113.6 J:3.4 L:1.1 H:16	D:115.7 J:3.2 L:1.6 H:19	D:77.1 J:3.5 L:1.0 H:13	D:105.8 J:3.1 L:1.1 H:16	D:105.2 J:3.4 L:1.1 H:20	D:104.4 J:5.5 L:1.4 H:16	D:345.6 J:5.7 L:4.0 H:17	D:180.0 J:9.3 L:4.00 H:23	D:136.2 J:4.1 L:1.8 H:18

Mean jitter values - grouped by region



Consecutive Packet Loss



Measurement conclusions

- Diffi cult to envisage the future use of the data, hence gather right level of associated or detailed information:
  - Number of hops as the experiment was conducted
  - State of the routing tables
- Contact with remote/network operators is important, did they have problems when anomalies were observed?
- Current tools are insuffi cient