Quality aspects of Internet telephony

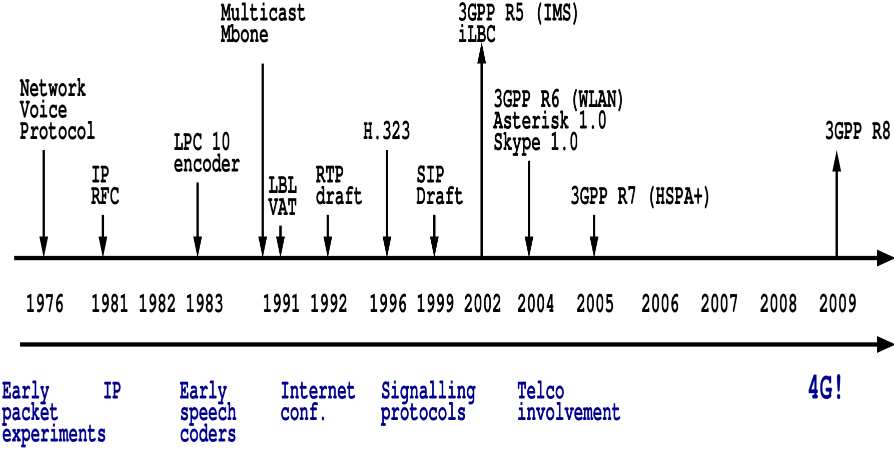
Ian Marsh

Ph.D defence
June 5th 2009

The structure of the presentation

- A very brief history of Internet telephony development
- Some words about deployment
- Problems facing Internet telephony quality
- A very simplified walk-through of a Voice over IP system
- What quality *really* means
- The dissertation's contributions
- The dissertations' conclusions

A historical time-line of packetised voice



Network
Voice
Protocol

IP
RFC

LPC 10
encoder

Multicast
Mbone

LBL
VAT

RTP
draft

H.323

SIP
Draft

3GPP R5 (IMS)
iLBC

3GPP R6 (WLAN)
Asterisk 1.0
Skype 1.0

3GPP R7 (HSPA+)

3GPP R8

1976 1981 1982 1983 1991 1992 1996 1999 2002 2004 2005 2006 2007 2008 2009

Early packet experiments

IP

Early speech coders

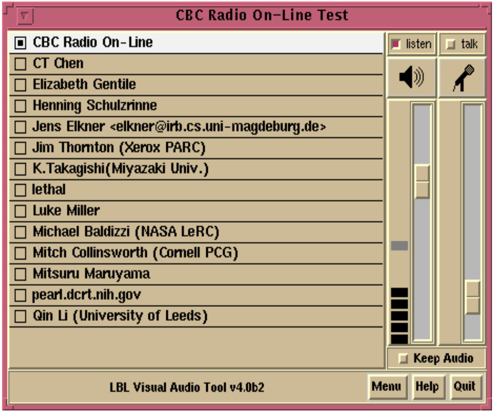
Internet conf.

Signalling protocols

Telco involvement

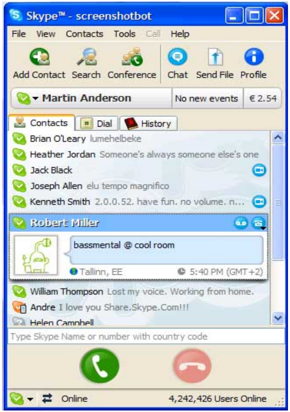
4G!

From a graphical perspective (1992 and 2005)



- ▣ CBC Radio On-Line
- ☐ CT Chen
- ☐ Elizabeth Gentile
- ☐ Henning Schulzrinne
- ☐ Jens Elkner <elkner@irb.cs.uni-magdeburg.de>
- ☐ Jim Thornton (Xerox PARC)
- ☐ K.Takagishi(Miyazaki Univ.)
- ☐ lethal
- ☐ Luke Miller
- ☐ Michael Baldizzi (NASA LeRC)
- ☐ Mitch Collinsworth (Cornell PCG)
- ☐ Mitsuru Maruyama
- ☐ pearl.dcrt.nih.gov
- ☐ Qin Li (University of Leeds)

LBL Visual Audio Tool v4.0b2



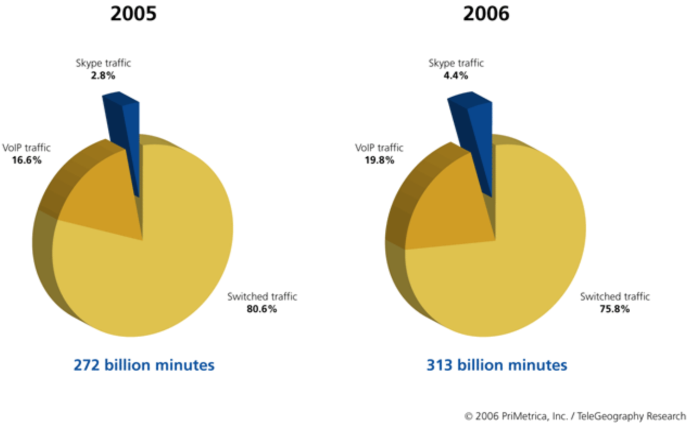
Martin Anderson | No new events | € 2.54

- Brian O'Leary lumehelbeke
- Heather Jordan Someone's always someone else's one
- Jack Black
- Joseph Allen elu tempo magnifico
- Kenneth Smith 2.0.0.52. have fun. no volume. n...
- Robert Miller
  - bassmental @ cool room
  - Tallinn, EE 5:40 PM (GMT+2)
- William Thompson Lost my voice. Working from home.
- Andre I love you Share.Skype.Com!!!
- Helen Campbell

Type Skype Name or number with country code

Online 4,242,426 Users Online

IP telephony deployment



© 2006 PriMetrica, Inc. / TeleGeography Research

- Telephony grew 15% (in total) from 2005 to 2006
- VoIP grew from 19.4% in 2005 to 24.2% during the same period
- This still leaves 75% for migration
- And motivation for this work!

Problems facing VoIP quality

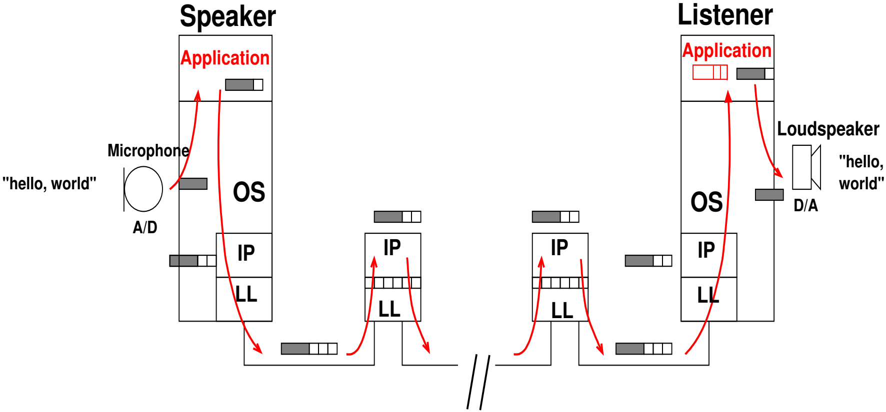
First a *non-problem*. There is nothing inherently problematic with sending real-time voice across a packet switched network, however, there is

1. Disruption caused by competing traffic
    - At the single flow level, TCP induces loss by increasing its sending rate until a loss event
    - Increasing volumes of traffic, specifically, P2P, video streaming and IPTV, *without* appropriate capacity changes
2. 802.11-based communication
    - Interference, environment & mobility can induce problems using *unlicensed* spectrum
    - Poor quality radios in some wireless terminals
3. Additional problems due to
    - Poor quality infrastructure in some countries
    - Poorly designed end-systems
    - Individual human tolerances

Simple VoIP walk-through

- On the next slide I am going to explain the path that real-time conversations take across the Internet from mouth to ear
- I will describe:
  1. The important processing steps
  2. How loss, delay and jitter are introduced into a VoIP stream
- The illustration will be used to describe, in turn, each of the dissertations' contributions later

A simple VoIP walk-through



What quality *really* means...

According to an ITU standard, **one-way delay** can be quantified as:

- "good" if the delay is less than 180 ms
- "acceptable" if the delay lies between 180 ms and 400 ms
- "unacceptable" if the delay exceeds 400 ms

Loss

- For telephony voice, acceptable **losses** can range from 1-10%
- Losses can be "hidden" using concealment techniques

What you are about to hear are 3 speech samples with:

1. Original sample
2. The original corrupted by 2% loss
3. The original corrupted by 5% loss

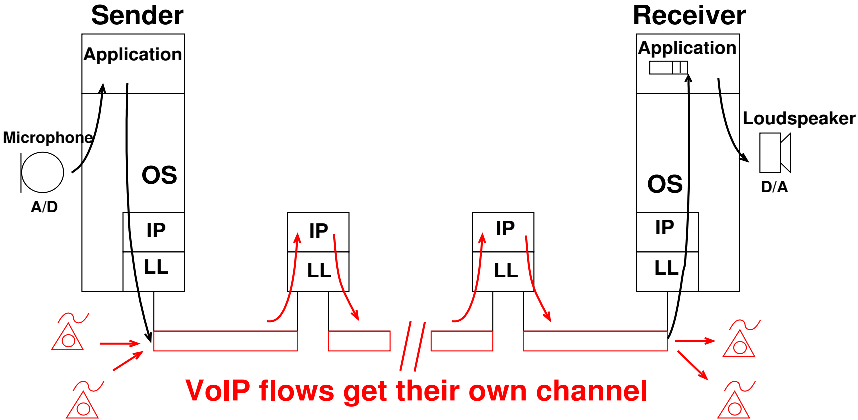
Jitter

- Buffers are required to handle jitter in VoIP systems
- Therefore they result in **additional** delay and/or loss

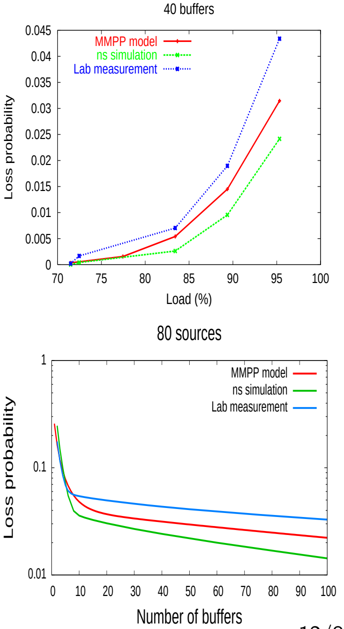
Now, presentation of the contributions

- Each contribution will now be shown in red using the previous figure
- Presentations will be brief, please consult the dissertation for full details

Protecting VoIP traffic (paper A)



Protecting VoIP traffic

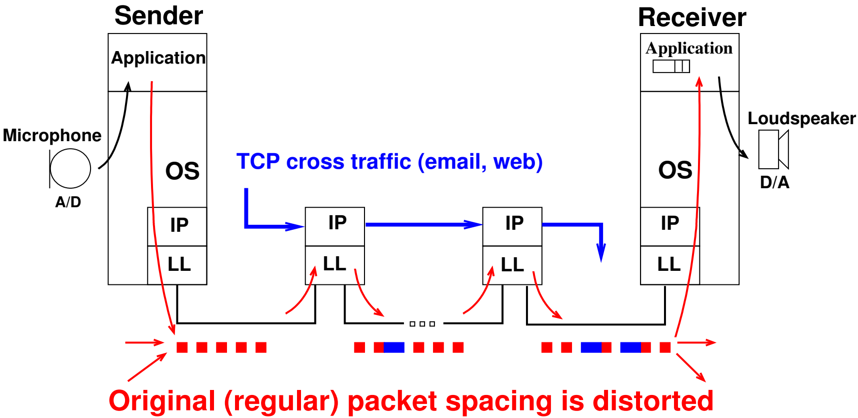


- Contribution: Three methods to dimensioning a link for statistically multiplexed VoIP flows
- The three colours show:
  1. A FreeBSD implementation (blue)
  2. A Markov-Modulated Poisson arrival model (red)
  3. A ns-2 simulation (green)
- Top figure is the loss probability against differing loads, calculated as follows:

$$Load = \frac{\#Sources \cdot ProbOn \cdot PeakRate}{Link\ \ capacity}$$

- Peak rate allocation would yield 25 sources for 85% load, whereas statistical multiplexing allows 72 sources
- Lower plot show loss against buffer sizes

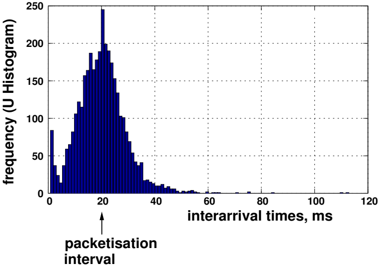
Mixing VoIP and data traffic (paper B)

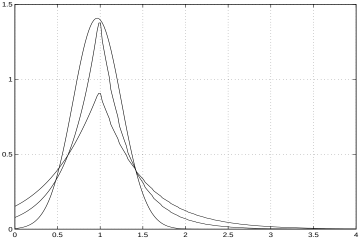


Original (regular) packet spacing is distorted

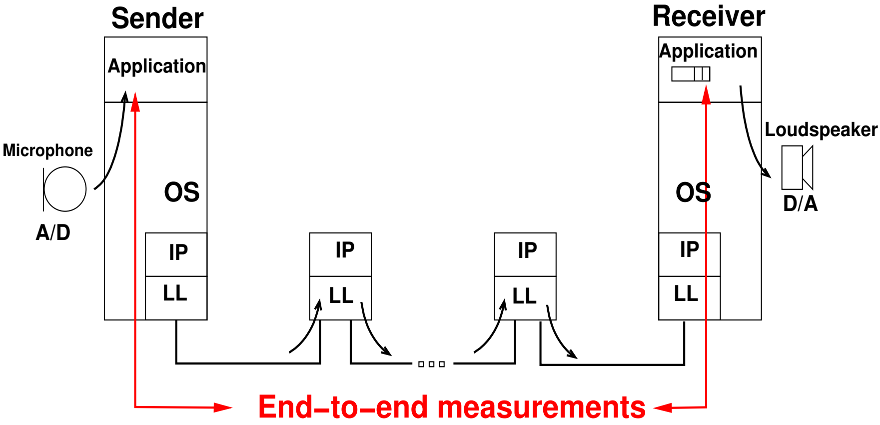
Mixing VoIP and data traffic (paper B)

- Contribution: Extensible model describing the arrival process of VoIP packets
- Most of the traffic on the Internet is TCP, therefore VoIP must coexist with bulk data
- Top plot shows the interarrival times for a single VoIP session
- The interarrival times of a packet delay *model* have been generated in the bottom plot
- Features:
  - $x = 0$ (origin) back to back packets
  - $x = 1$ (maximum) packet. interval
  - $x > 1$ (tail) of the distribution
- The model requires an *estimate* of the networks delay distribution (3 shown)



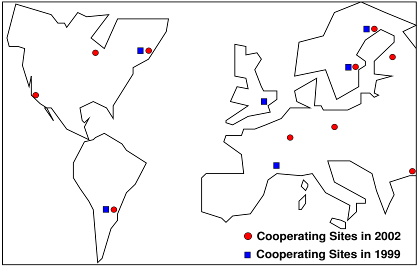


Measuring wide-area VoIP quality (papers D and E)



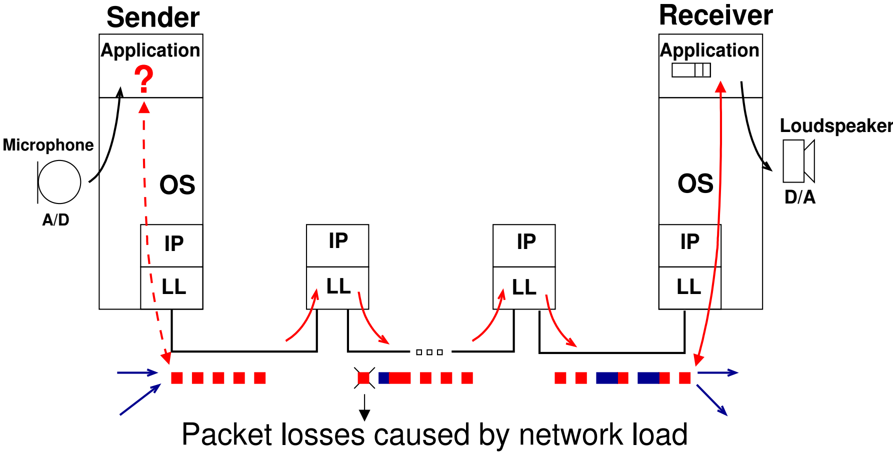
Measuring wide-area VoIP quality (papers D and E)

- Contribution: comprehensive measurement VoIP effort from 1998 and 2002
- Measured:
  - Loss
  - Delay
  - Jitter
  - Hops
- Hosts located at universities
- Most *sessions* showed good quality
- Two *sites* did not provide adequate quality
- Quality has slightly improved for the same sites since 1998
- Data used in papers B and F



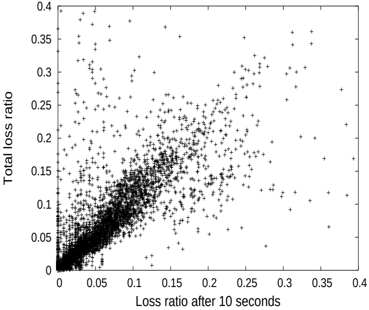
Transmitted session	
Call duration	70 seconds
Payload size	160 bytes
Packetisation time (ms)	20ms
Data rate	64kbits/sec
Without silence suppression	3653 packets
With silence suppression	2043 packets
Coding	8 bit PCM
Recorded call size	584480 bytes
Obtained data	
Number of hosts used	9
Number of traces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

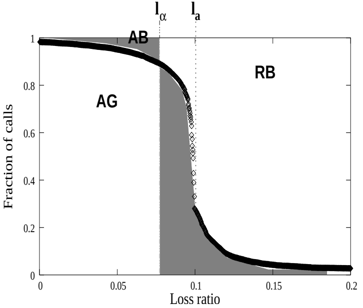
Self-admission control (paper F)



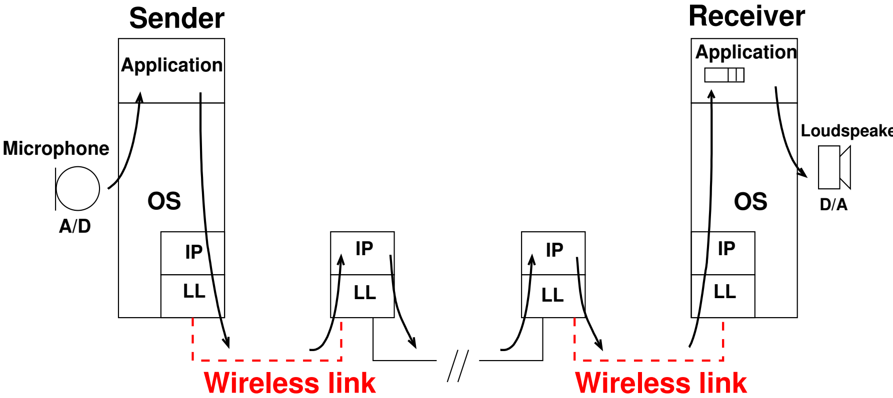
Self-admission control (paper F)

- Contribution: Free-standing endpoint admission control mechanism
- Goal is to estimate the quality of a call from its initial 'few' seconds
- Top plot shows the loss process for a number of calls
- Given the decision to admit or reject a flow, there are four possible outcomes:
  - Accepted and good quality (AG)
  - Accepted but bad quality (AB)
  - Rejected and bad quality (RB)
  - Rejected but good quality (RG)
- Right plot shows AB and RB for a *target* loss rate of 2%
- For a higher admission threshold, the number of incorrectly admitted calls (AB) will increase, whilst the number of rejected good (RG) will decrease





802.11 wireless access (paper G)



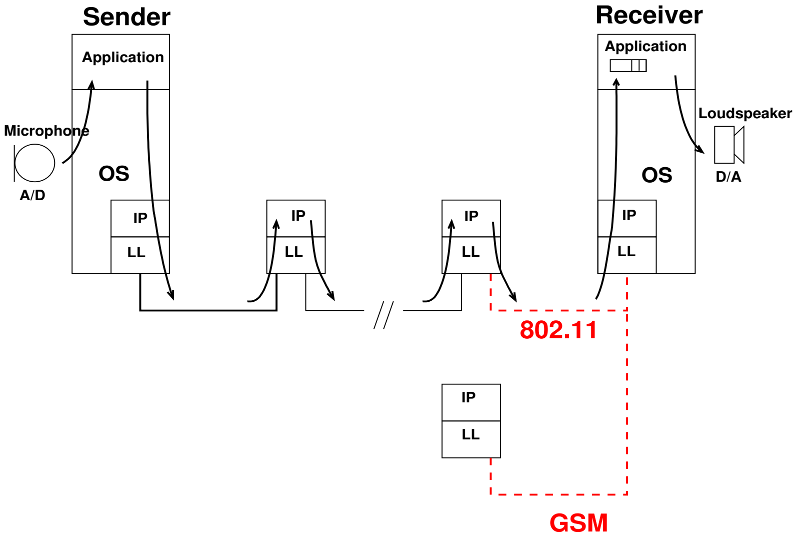
Investigate VoIP quality with local wireless access

802.11 access for VoIP

- Contribution: A comprehensive study of 802.11b networks for voice
- Studied different scenarios in a methodical manner:
    1. Line of sight (outside)
    2. Non-line of sight (inside an office)
    3. With & without background load
    4. Infrastructure mode (with an access point)
    5. Using request to send/clear to send
- Cross-layer techniques to infer VoIP quality at application proved useful
- Need to do perform radio pre-planning and measurements before deployment
- Immediate rate changes beneficial for VoIP
- Work proved an important precursor for the next contribution

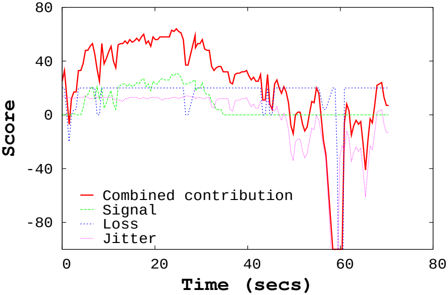
Pos.	Loss % [min, mean, max]	Round-trip delay (
O_2	[0, 0.0, 0.1]	[1.9, 2.2, 2.4]
C	[0, 0, 0]	[1.9, 2.1, 4.0]
D	[0, 0, 0]	[2.1, 2.6, 3.1]
E_1	[0, 0.2, 2.6]	[2.8, 3.2, 5.4]
E_2	[9.4, 36.3, 89.1]	[5.3, 12.2, 24.3
E_3	[0, 0.0, 0.2]	[2.8, 2.8, 4.0]
F_1	[20.1, 54.9, 88.7]	[5.4, 13.4, 24.6
F_2	No signal	No signal
F_3	[1.8, 22.8, 84.9]	[4.6, 11.7, 13.7
G	[0, 0.2, 1.9]	[1.9, 2.2, 3.8]
H_1	[0.4, 5.4, 30.2]	[3.5, 3.9, 8.1]
H_2	[3.4, 11.0, 28.3]	[5.9, 6.1, 11.7
H_3	[0, 0.2, 2.7]	[3.4, 3.4, 5.4]

802.11-based voice with alternative access (paper H)



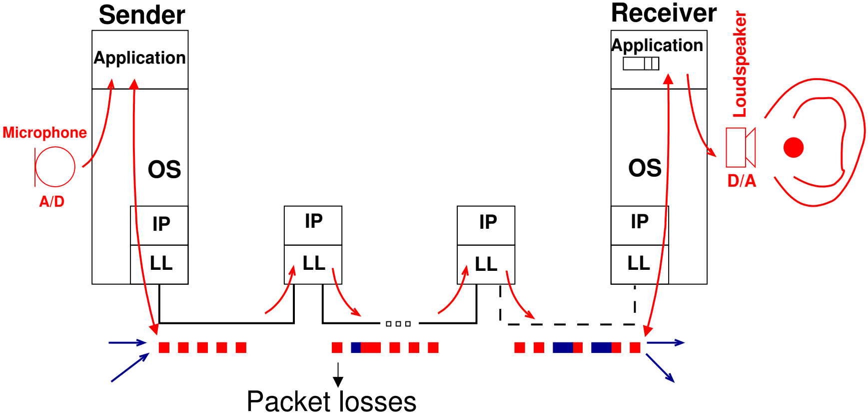
A handover solution

- Contribution: Fully automated handover mechanism from 802.11 to GSM
- Implemented a function to estimate the network parameters in a receiver
- When the quality drops below a given threshold, we switch from 802.11 to GSM
- The difficulty is in the prediction needed (i.e. the PSTN call setup)
- Tested the handover in a real implementation and conducted 100 user tests in an office environment (bottom table)



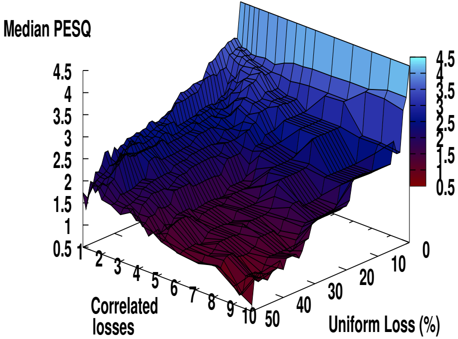
Quality started good and became bad	Timely HO **68**	Late HO 10
Quality started good and remained good	Unnecessary HO 7	No HO **15**

Single-sided measure based on PESQ (paper I)

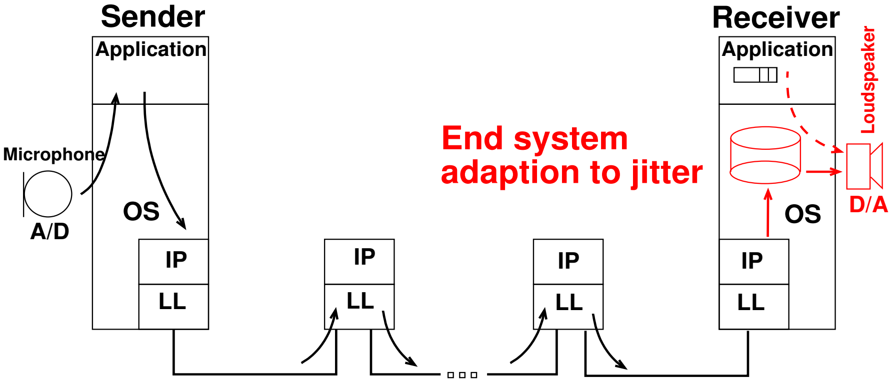


Single-sided measure based on PESQ

- Contribution: Single-sided real-time estimation for loss-based quality prediction
- Idea is to derive an off-line estimation of voice quality
- Compute off-line response to typical network losses using PESQ
    1. PESQ = Perceptual Evaluation of Speech Quality
    2. Compares a reference and degraded signal
    3. Outputs a value between 0.5 and 4.5
- 3D plot shows response to correlated and uncorrelated losses
- PESQ tends to underestimate the impact of bursty losses



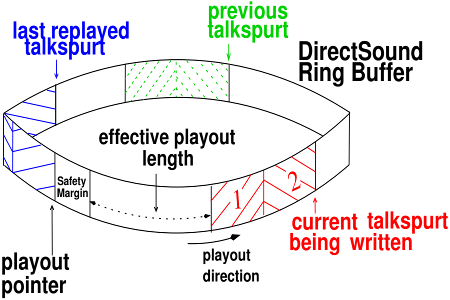
End-system adaption to network jitter (paper C)



End system adaption to jitter

End-system adaption to network jitter

- Contribution: General guidelines in delay reduction techniques for end-systems
- End systems can add considerable delays to network stream
- Carefully engineered solutions make end-system optimisations possible
  1. Remove unnecessary data copying
  2. Setup memory transfers in advance (DMA)
  3. Optimise buffer playout for appropriate coding
  4. Require as few hardware interrupts as possible
- 100's of milliseconds can potentially be saved
- Paper describes a 3-tier solution to inhibit fluctuations



Dissertation conclusions

- A near complete system study for Voice over IP
- Carefully engineered systems can provide good quality voice
  1. Controlled load for VoIP traffic, using dimensioning
  2. End system self-admission
  3. Monitor and measure the network performance
     - For good engineering designs
     - To gain insight into the problem
     - Collate data for further investigation
  4. Rate control within 802.11 networks
  5. Where alternatives are available, use them, e.g. handover
  6. *Estimate* the quality as near to the user as possible
  7. Delay aware end-systems
- Used many different techniques:
  1. Analysis and modelling
  2. Simulation
  3. Measurements
  4. Implementation
  5. Subjective tests
- A byproduct of our research is a suite of software tools, some in commercial systems (handover module) and a large measurement data repository

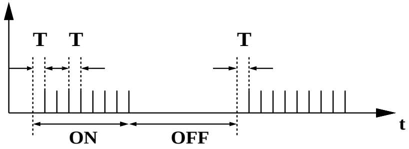
Extra slides

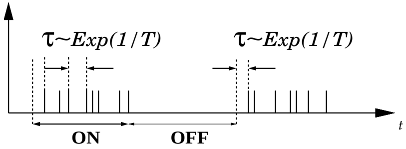
Protecting VoIP traffic

- Problem becomes capacity allocation for a required quality
    - Telephony and ATM research fields have suggested solutions
    - Largely ignored by the IP community
- We investigated an existing proposal and applied it to IP networks
- Implemented a computationally efficient model
- We modelled the superposition of independent sources
    - Markov Modulated Poisson Process (MMPP)
    - See next slide
- Measured the loss probability through a finite buffer
    - For different loads and buffer sizes
- Compared the model, simulation and a laboratory setup
- Generally quite unusual to try all three approaches

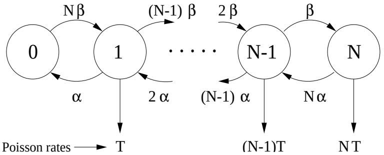
A more tractable model

- Model the inter-packet distance as exponentially distributed (as in the figure below)
- Deterministic arrival times during talkspurts are in theory manageable in terms of losses, however sources can become de-synchronised, then the determinism breaks
- Need to verify that the burstiness of exponential is indeed captured
- Each call (i.e. a sequence of ON and OFF states) as a state in a Markov chain.
- Within each state
- More precisely, birth-death process (only changes to neighbouring states are allowed)





Modelling the arrival flows

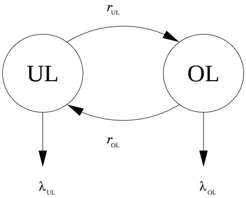


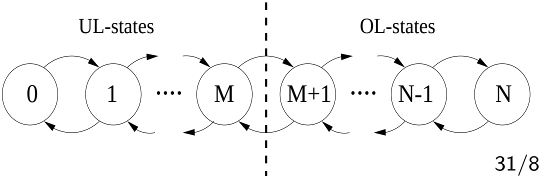
Birth-death process

- Chain of transition probabilities
- However, can be computationally expensive to compute

Asymptotic matching

- Method to reduce the complexity of calculating





Burstiness, index of dispersion for intervals (IDI)

IDI is defined as:

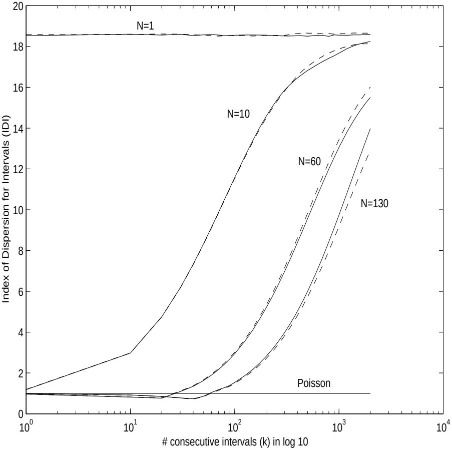
$$D = \frac{\sigma^2}{\mu}$$

- For Po(x) $= 1$ (eq. σ^2 and μ)

It is observation time (k) dependent:

$$c_k^2 = \frac{k Var(S_k)}{[E(S_k)]^2}$$

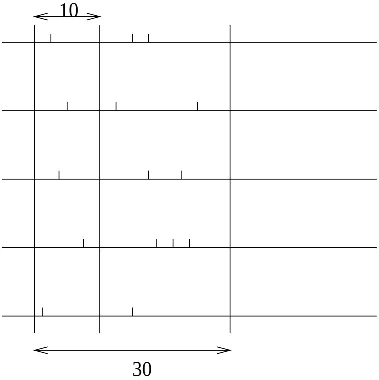
- Poisson "burstiness" is still 1 $(\forall k)$
- In the figure: N is the number of sources
- Short observation time, the superposition process $\approx$ Po(x)
- Solid lines are deterministic and the dashed lines are exponential interarrival times during a talkspurt



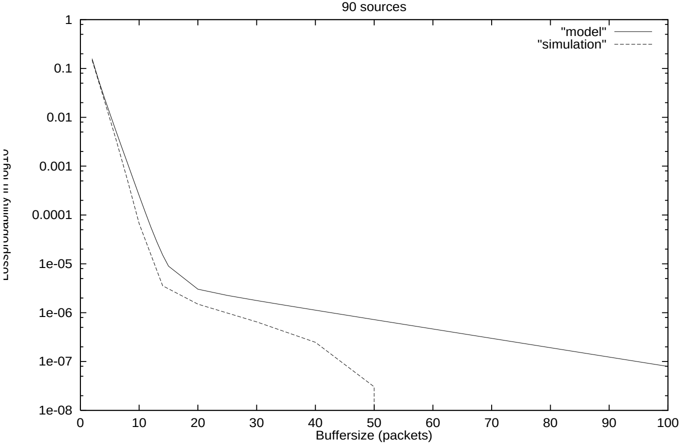
Quantifying burstiness, intuitively

Looking at 10 consecutive inter-arrivals is more likely to show 10 packets from different sources. If one looks at more than 10 arrivals the probability that a given source is on is higher. This results in a correlation between arrivals which gives rise to a skewed behaviour, illustrated below.

Where in the larger time frame it is possible for packets from the same source to be included and thus affect, as stated, the correlation, hence the skew.

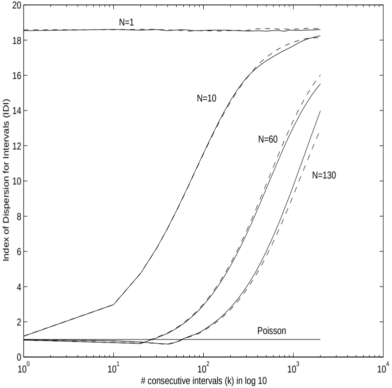


Differences between model and simulation

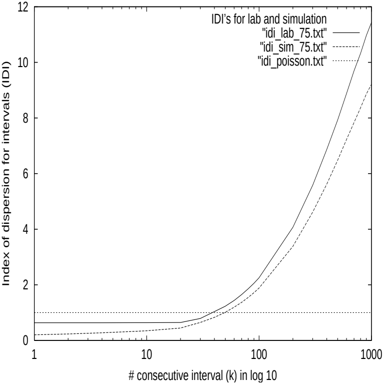


- A 100 packet buffer with 90 sources has a loss probability of 10^{-8}
- We needed over 1 day to get the losses shown above and still not close

Differences between Model & Lab & Sim

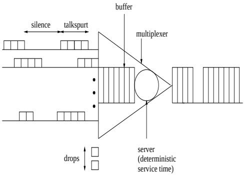


Model only



Simulation and laboratory

Loss calculations for a MMPP/D/1/K queue



The number N of multiplexed voice sources is fixed. We denote the *output link capacity* as C. The *peak rate* is defined as

$$h = \frac{B}{T},$$

where B denotes packet size. Assuming peak rate assignment, is the value

$$M = \lfloor C/h \rfloor$$

the maximum number of sources that can be accommodated in the MUX. Furthermore, we assumed that the following stability condition is satisfied

$$\rho = \frac{Nhp_{on}}{C} < 1,$$

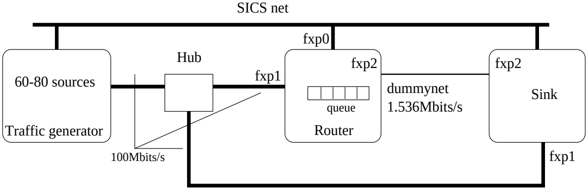
where ρ is the traffic intensity(or load factor). And to insure that we are in the statistical multiplexing region, we assume that $Nh > C$.

Loss calculation

$\Pi(K)$	The loss probability of the MMPP/D/1/K queue.
ρ	The mean offered load.
$H(t)$	The distribution function of the service time.
$\tilde{H}(a)$	The *Laplace Stieltjes Transform*(LST) of $H(t)$.
θ	The mean service time.
L	The packet emission rate of single source while in talkspurt.
$\mathbf{R}$	The intensity matrix of the phase process.
$\boldsymbol{\Lambda}$	The diagonal matrix, whose element Λ_{ij} is equal to the mean arrival rate while in phase i.
$\mathbf{U}$	The matrix given by $(\boldsymbol{\Lambda} - \mathbf{R})^{-1}\boldsymbol{\Lambda}$ which accounts for the evolution of the phase process during server's idle periods.
$\mathbf{e}$	The unit column vector.
$\pi_{\mathbf{K}}(\mathbf{i})$	The row vector whose j-th element is the limiting probability at the embedded time instants of having i users in the system and being in the phase j of the MMPP, $i = 0, 1, ..., K$.
$\mathbf{q}$	The row vector containing the limiting state probabilities of the phase process. It can be obtained as the unique solution of the system $\mathbf{qR} = 0$ and $\mathbf{qe} = 1$.

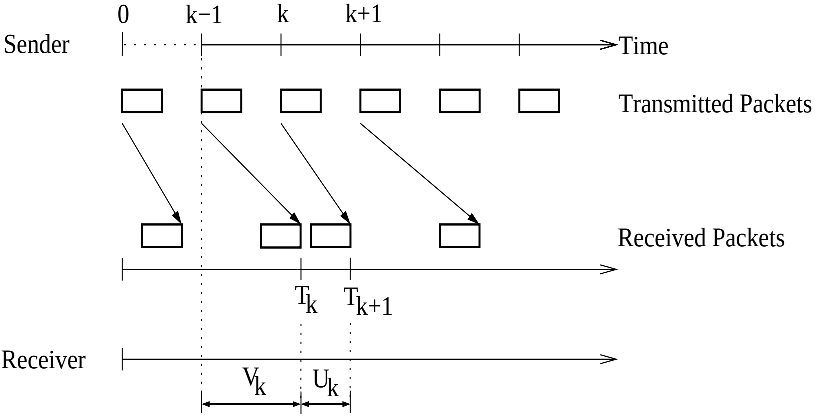
$$\Pi(K) = 1 - \frac{1}{\rho[1 + \pi_{\mathbf{K}}(0)\mathbf{U}\boldsymbol{\Lambda}^{-1}\theta^{-1}\mathbf{e}]}.$$

Laboratory setup



- Traffic generator for VoIP
- Router with constrained output link (Dummynet)
- Receiver sees 'lost' and original stream
    - mostly for delay measurements
- Superposed sources pre-generated to save processing

Packet delay model



- k packet number (time, but not random)
- T_k arrival times
- V_k observed delays
- U_k observed interarrival times
- Y_k network delays, not directly 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)

Proof of interarrival times $E(U_k) \approx 1$

$$T_k = \max(Y_1, 1 + Y_2, \ldots, k - 1 + Y_k) \quad k \geq 1, \tag{1}$$

$$T_k = \max(Y_1, 1 + T'_{k-1}), \quad k \geq 2 \tag{2}$$

$$T'_{k-1} = \max(Y_2, 1 + Y_3, \ldots, k - 2 + Y_k) \tag{3}$$

$$\{T_k > t\} = \{1 + T'_{k-1} > t\} \cup \{Y_1 > t, 1 + T'_{k-1} \leq t\}. \tag{4}$$

$$\begin{aligned} P(T_k > t) &= P(1 + T'_{k-1} > t) + P(Y_1 > t, 1 + T'_{k-1} \leq t) && (5) \\ &= P(1 + T_{k-1} > t) + P(Y_1 > t) P(1 + T_{k-1} \leq t) && (6) \end{aligned}$$

$$\begin{aligned} E(T_k) &= \int_0^\infty P(T_k > t)\, dt \\ &= E(1 + T_{k-1}) + \int_1^\infty P(Y_1 > t) P(T_{k-1} \leq t - 1)\, dt. && (7) \end{aligned}$$

$$E(U_k) = 1 + \int_1^\infty P(Y_1 > t) P(T_{k-1} \leq t - 1)\, dt \to 1, \quad k \to \infty \tag{8}$$

(since $\nu = \int_0^\infty P(Y_1 > t)\, dt < \infty$ and $T_k \to \infty$, the dominated convergence forces the integral to vanish in the limit).

Dominated convergence theorem

- One of the main results from Lebesgue integration (a branch of analysis). Lebesgue integration allows a wider range of functions to be integrated than Riemann integrals, particularly the sums of functions.
- Dominated convergence theorem says that:
- One can evaluate the limit of the integrals of a sequence of functions as the integral of the point wise limit of the functions as soon as the sequence of functions is dominated in absolute value by an integrable function.
- For example, if f_n is a sequence of integrable functions, convergent almost everywhere to f, and if there is an integrable g such that:

$|fn| < g$ for all n, then f is integrable then:

$$\lim_{n\to\infty} \int_S f_n \, d\mu = \int_S \lim_{n\to\infty} f_n \, d\mu.$$

In our case $f_n = T_k$ and the dominating integrable function is $P(Y_1 < t)$

Steady state distributions

By

$$\begin{aligned} T_1 &= Y_1 \\ T_k &= \max(T_{k-1},\, k-1+Y_k), \quad k \geq 2. \end{aligned} \tag{9}$$

$$P(T_k \leq x) = \prod_{i=1}^{k} P(i+Y_i \leq x+1) = \prod_{i=0}^{k-1} F(x-i),$$

and therefore the sequence (V_k), which we defined by $V_k = T_k - k + 1$, $k \geq 1$, satisfies

$$P(V_k \leq x) = \prod_{i=0}^{k-1} F(x+k-1-i) = \prod_{i=0}^{k-1} F(x+i) \quad x \geq 0.$$

This shows that (V_k) is a Markov chain with state space the positive real line and asymptotic distribution given by

$$P(V_\infty \leq x) = \prod_{i=0}^{\infty} F(x+i) \quad x \geq 0. \tag{10}$$

Furthermore, for $x \geq 0$

$$\begin{aligned} P(U_k \geq x) &= P(k-1+Y_k - T_{k-1} \geq x) = P(V_{k-1} \leq Y_k + 1 - x) \qquad (11) \\ &= \int_0^\infty P(V_{k-1} \leq y+1-x)\, dF(y), \qquad (12) \end{aligned}$$

where in the step of conditioning over Y_k we use the independence of Y_k and V_{k-1}. Therefore the sequence (U_k) has the asymptotic distribution

$$P(U_\infty \leq x) = 1 - \int_0^\infty \prod_{i=1}^{\infty} F(y-x+i)\, dF(y) \quad x \geq 0, \tag{13}$$

Modelling 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)

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

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

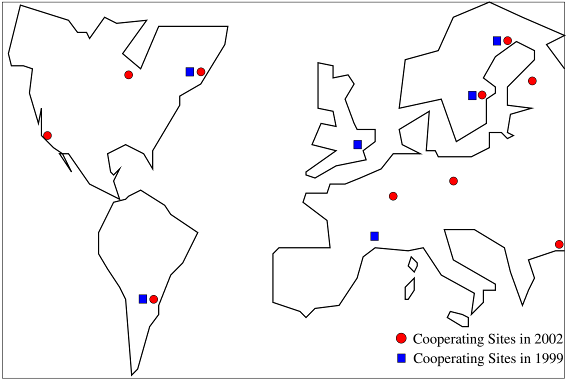
$$T_1 = S_1 + Y_1, \quad T_k = \max(T_{k-1},\, k-1+S_k+Y_k), \quad k \geq 2,$$

Modelling loss

Probability p that a packet is subject to loss is independent $Y(x)$
K_k is the no. of attempts between successful packets $k-1$ & k, $\quad k \geq 1$
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 sequence of variables with a negative binomial distribution
The arrival times are now adapted:

$$T_1 = K_1 - 1 + Y_{K_1}, \quad T_k = \max(T_{k-1}, L_k - 1 + Y_{L_k}), \quad k \geq 2. \quad (15)$$

Co-operating test sites used in 1999 and 2002



An active measurement method was used, a 70 second pre-recorded conversation between the marked sites was sent once/hour

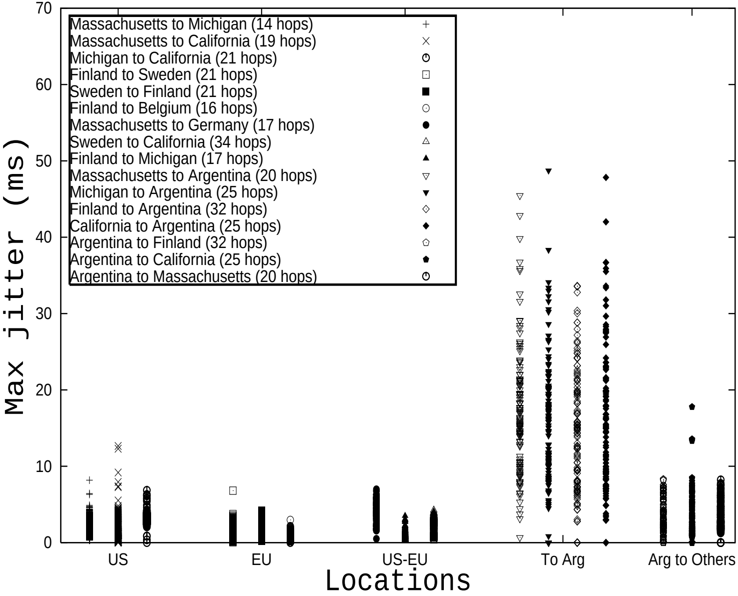
Correlated and non-correlated losses

Rows are:

- Loss in %
- Non-correlated loss %
- Correlated loss %

Receiver Sender	Massachusetts	Michigan	California	Belgium	Finland	Sweden	Germany	Turkey
Massachusetts		0.578471 0.001293 0.777848	0.265175 0.002555 0.038911	0.372212 0.002923 0.217494	0.120434 0.000933 0.226214	0.076086 0.000627 0.176692	0.000000 0.000000 0.000000	7.402258 0.052835 0.339062
Michigan	0.037677 0.000363 0.036697		0.752838 0.006357 0.161900	0.385696 0.003012 0.222011	0.752838 0.006357 0.161900	0.561914 0.000781 0.861876	0.144367 0.001355 0.062500	6.271999 0.038896 0.418742
California	0.279106 0.001451 0.481572	0.369602 0.000597 0.838990		0.788673 0.007022 0.116706	1.406078 0.012887 0.096354	0.344077 0.003411 0.012092	4.349055 0.034402 0.243383	6.768777 0.055885 0.230252
Belgium	0.005043 0.000050 0.000000	0.002781 0.000028 0.000000	0.471160 0.004586 0.031142		0.007750 0.000078 0.000000	0.000000 0.000000 0.000000	0.149653 0.001275 0.149123	8.544666 0.078624 0.158476
Finland	0.000226 0.000002 0.000000	0.035876 0.000146 0.593137	1.285342 0.012256 0.058756	0.269181 0.002188 0.189320		0.000927 0.000009 0.000000	0.001262 0.000006 0.500000	4.910323 0.038820 0.248246
Sweden	0.000789 0.000008 0.000000	0.148330 0.000223 0.849727	0.132657 0.001287 0.031469	0.197712 0.001851 0.065421	0.000540 0.000005 0.000000		0.000000 0.000000 0.000000	4.281929 0.037661 0.158125
Germany	0.000464 0.000005 0.000000	0.323992 0.000045 0.986301	2.844235 0.027520 0.059940	0.006313 0.000063 0.000000	0.000000 0.000000 0.000000	0.000000 0.000000 0.000000		6.333008 0.045419 0.328241
Turkey	13.167093 0.090965 0.400113	13.061485 0.094192 0.373051	8.829068 0.081313 0.160348	13.251261 0.089106 0.416672	12.119530 0.088442 0.358694	13.073778 0.093331 0.379449	12.185872 0.093076 0.329274	

Mean jitter values - grouped by region



Measurement goals and brief results

Comprehensive loss, delay and jitter assessment of VoIP in 2002

- Some indication of the trend by comparing the results with those taken in 1999
  - Improvements from 1999's measurements: full-mesh topology, totally automatic invocation and more hosts (next slide)
  - Also investigated some new items: asymmetry issues, time of day effects, different packet sizes and silence suppression
- Collected a repository of VoIP sample sessions (approx. 22,500 calls)
- Most calls $< 2\%$ loss & 10 ms jitter (delay is location dependent)
- Quality has slightly improved over the past three years
- VoIP still not usable on a global scale, two sites showed poor quality
- Infra-structure not distance (or the number of hops) is important
- Hundreds of downloads, at least six publications from our data

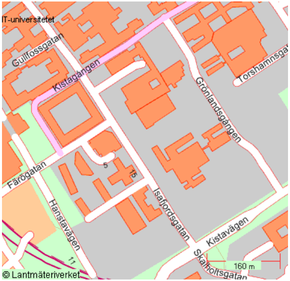
Wireless measurements

Now we look at IEEE 802.11b access, and again use active measurements of the channel. In particular we look at the MAC layer behaviour, environment and the role of competing traffic.

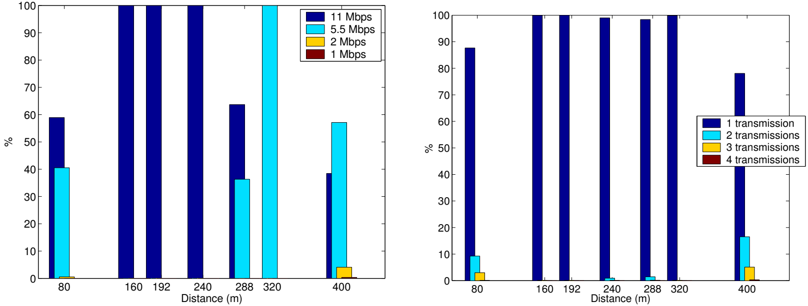
- Pure distance effects using line-of-sight between a single sender and receiver (outside)
- Distance effects with line-of-sight (same premises)
- Distance effects without line-of-sight (same premises)
- Competing traffic effect in ad-hoc mode (same room)
- Competing traffic effect in infra-structure mode (same room)
- IEEE specific bitrate selection, RTS/CTS (same premises)

Line-of-sight setup

We took eight measurements at distances from 80 to 400 meters We recorded the loss, delay, jitter, bitrate & no. of MAC transmissions The sender was stationary and the receiver moved

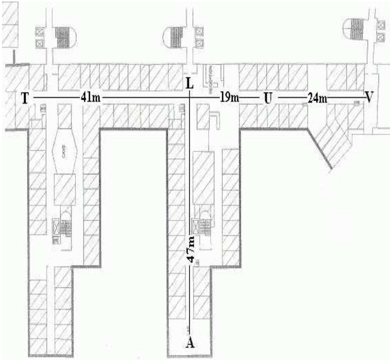


Line-of-sight results



Left histogram shows the rates at which each frame was sent, the right one is the number of retransmissions at each distance. VoIP quality is good.

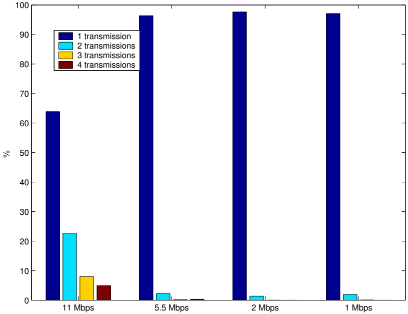
Line-of-sight in the office I



Locations	fraction of losses (%)	round-trip time (ms)	jitter (ms)
$A- > L$ (47m)	[0.0, 0.0, 0.0]	[1.9, 2.0, 2.2]	[0.1, 0.1, 0.2]
$T- > U$ (60m)	[0.0, 0.2, 0.9]	[1.9, 2.0, 2.9]	[0.2, 0.4, 0.9]
$T- > V$ (84m)	[0.0, 0.4, 1.4]	[1.8, 2.2, 3.5]	[0.1, 0.2, 0.5]

Non line-of-sight office results

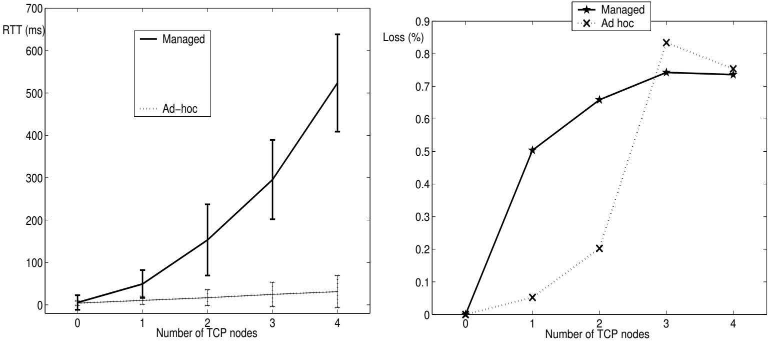
In this setup, the receiver is around a corner. This time we fixed the maximum bitrate and observed the rates selected.



5.5 Mbits in this case is a reasonable choice

Voice quality and competing TCP traffic

If we now look at competing TCP traffic and its implications in both ad-hoc and managed modes



Left plot is the delay & the right plot the loss, the delay is very high

RTS/CTS

- RTS/CTS (Request to Send / Clear to Send) is used by 802.11 wireless to reduce frame collisions introduced by the hidden terminal problem.
- The hidden node (or terminal) problem occurs when one node is visible from an access point, but not from the other nodes communicating with said AP. Generally leads to difficulties in media access control.
- A node wishing to send data initiates the process by sending a Request to Send frame (RTS). The destination node replies with a Clear To Send frame (CTS).
- The amount of time the node should wait before trying to get access to the medium is included in both the RTS and the CTS frame.
- This protocol was designed under the assumption that all nodes have the same transmission range.
- 802.11 relies on physical carrier sensing only which is known to suffer from the hidden node problem.
- RTS/CTS packet size threshold is 0-2347 octets. Typically, sending RTS frames is turned off by default (threshold > 2347).
- If the packet size the node wants to transmit is larger than the threshold, the RTS/CTS handshake gets triggered.

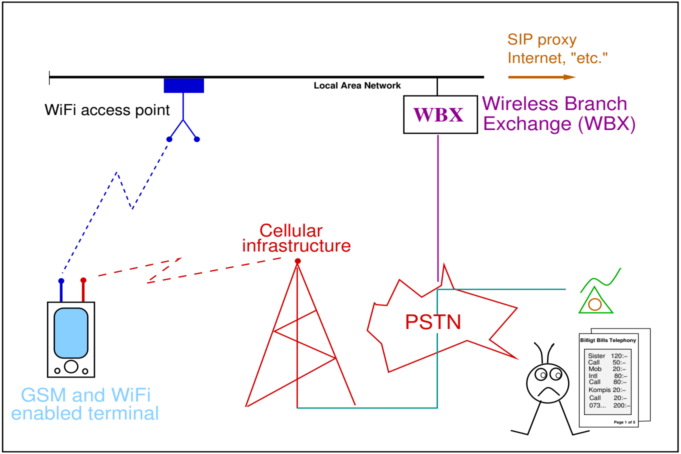
Conclusions for real-time voice and 802.11b access

- Generally reasonable quality
- 802.11b performs better than we expected (or reported)
- However even line-of-sight vulnerable to interference
- Packet loss caused by intervening obstacles
- Delay and jitter arise from competing traffic
- The access point can “add” high delays (scheduling & queueing)
- Deeper layer (radio layer) investigations needed
- If not for 802.11 (WiFi) then for 802.16 (WiMAX)
- Use information from as many layers wherever possible

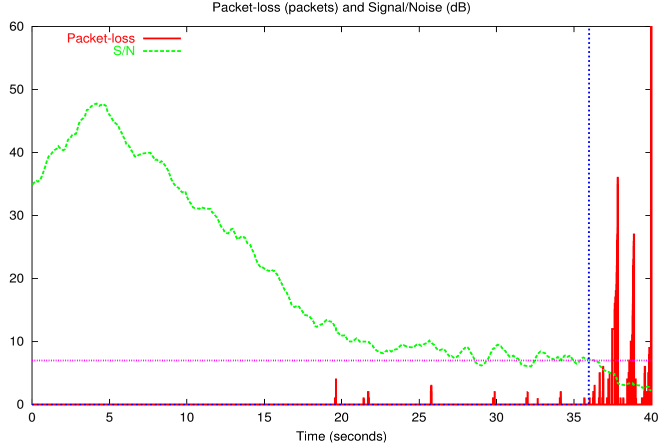
Quality prediction problems

- The problem is the 5s setup to the PSTN cellular network
- Therefore some *prediction* is needed in the WiFi network
- One needs to initiate a handover to the network before poor quality is experienced
- But also not to confuse this with short-term audible (but tolerable) glitches
- Need to handle leaving and entering WiFi network
- Don’t want to flip between the two networks
- Seems desirable to have more or less aggressive handover strategies
- Therefore financial incentives become important

Handover infra-structure



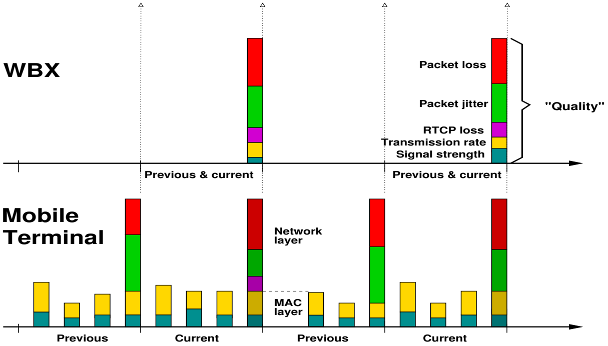
Loss and signal to noise ratio



Explanation

- The magenta horizontal line shows the SNR, it is acceptable but could mislead to waiting too long before a handover
- The blue vertical line shows the start of loss that would be unacceptable, but it is too late to initiate a successful handover
- So one has to combine the parameters and weigh them in a spam filter-type approach
- Also it is clear degrading quality over consecutive time intervals is more serious than ones in a single interval

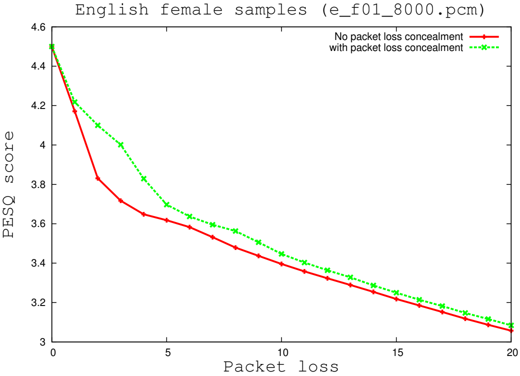
Combining parameters



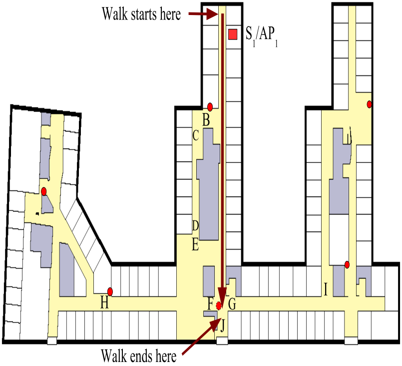
Component	interval 1	interval 2	both intervals
Loss ($> 4\%$)	-1	-1	-1
Jitter ($> 85ms$)	-1	-1	-1
SNR ($> -70db$)	-	+1	-
SNR ($< -90db$) & rate ($< 2Mbits$)	-	-1	-
Rate ($> 3Mbits$)	-	+2	-
2 lost messages	-	-	-1
Total	$\sum X$	$\sum Y$	≥ 0 do handover

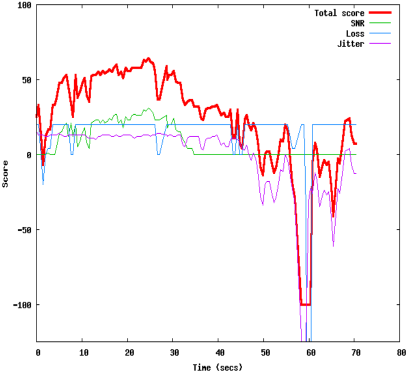
Small diversion - human tolerance to loss

- How much loss do we know is acceptable?
- Well, the ITU have developed a psychoacoustic model for human speech quality evaluation called PESQ
- No need for human listeners, the model estimates the quality close to a humans rating 4.5 (excellent) to 1 (very poor)
- 10 losses seem to be approximately a lowering of one rating



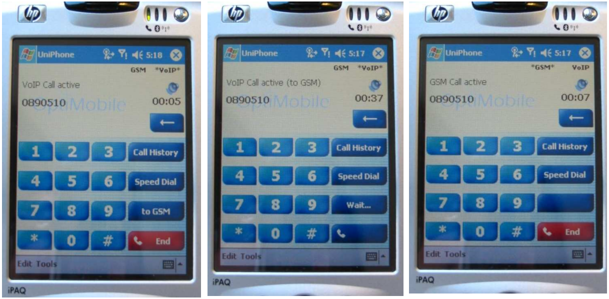
Handover evaluation





Handover implementation

- We have tested with a single cell with one access point (AP)
- By walking away from the AP we decrease the reception signal & and hence the communication quality
- An estimation of the quality is done in the handset and also sent to the WBX
- Using the score system described earlier it will initiate a handover when the quality falls below acceptable levels



VoIP Call active

VoIP Call active (to GSM)

GSM Call active

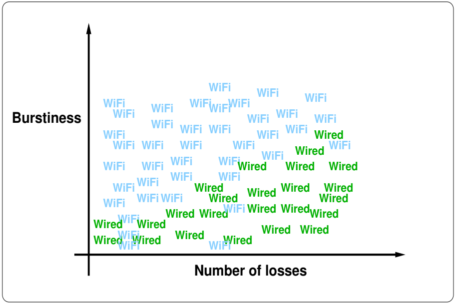
Voice quality estimation

Estimating the quality of telephony streams over IP networks in real-time is useful for:

1. Reporting quality back to provider (application writer too)
2. Calls do not need to be continued and use valuable resources
3. Give the user some feedback what is happening (is it the network, computer or me?)
4. Not easy to deduce the quality when using different codecs
5. Path characteristics during and after the call
6. Skype also allows some human feedback
7. Help determining a handover point (more later)

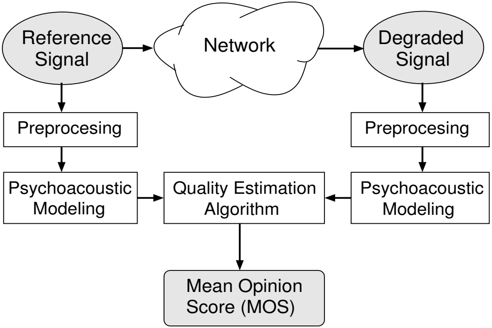
Basic technique: Off line pre-calculation

- The idea is pre-calculate the most common loss patterns
- Using this loss information we can "calibrate" the receiver
- And interpolate values that we didn't calculate
- Possible to "parametrise" the receiver for different codecs
- The secret is to choose the right combination



PESQ in a nutshell

- Perceptual Estimation of Speech Quality (PESQ)
- PESQ compares a reference signal to a degraded one
- Gives a human-understandable rating (MOS)



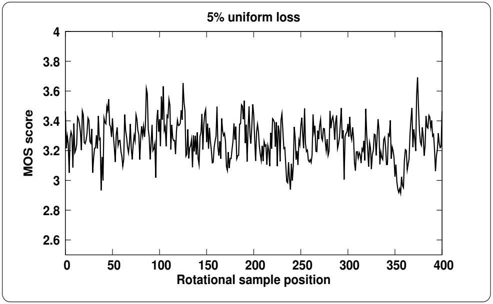
ITU's PESQ standard in words and numbers

PESQ MOS	Linguistic equivalent	Quality degradation
4.5	Excellent	None
4	Good	
3.5	Good/Fair	"Moderate"
3	Fair	
2.5	Fair/Poor	"Severe"
2	Poor	
1	Bad	

- A MOS score of ≥ 0.5 is usually audible
- 0.5 scores are not in the ITU standard (we added them)
- We also added an extra column, quality degradation

Uniform losses

- Losses are from a uniform distribution from 1% to 50%
- Losses are applied to 8 second ITU standard samples
- We rotated the loss pattern one packet at a time
- With 5% loss, the PESQ scores can vary up to 0.7 MOS points



MLBS and LR

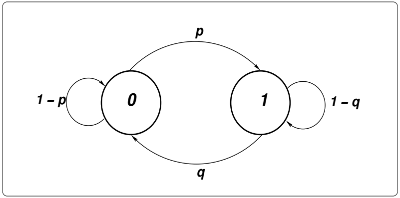
- Use WiFi network for voice whilst in "good" conditions and switch to the cellular network when the quality is not
- Switch with minimum disruption to the user
- LR Loss rate - uniform losses
- MLBS - Mean Loss Burst Size bursty losses

$$p = \frac{1}{\mathrm{MLBS}} \frac{\mathsf{LR}}{1 - \mathsf{LR}} \tag{16}$$

$$q = \frac{1}{\mathrm{MLBS}} \tag{17}$$

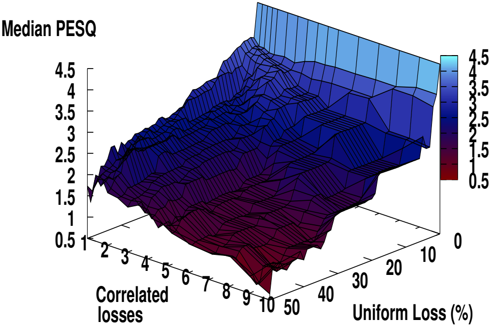
Bursty losses

- We consider losses up to 50% and bursts up to 10 packets
- Basically it covers the wireless network scenario
- Losses are generated according to a Gilbert model
- p is the *uncorrelated* loss and 1-q the *correlated* losses



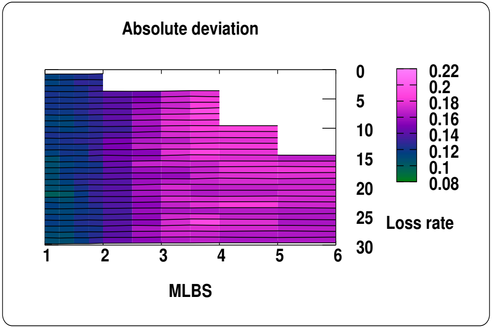
Bursty losses I

- The loss rate (LR) is the major factor in degrading the quality
- Burstiness is also significant (shown as MLBS)
- There is an initial fall in quality then flattens somewhat
- Shown below are the results without packet loss concealment



Absolute errors

If one uses the median, then some errors exist:

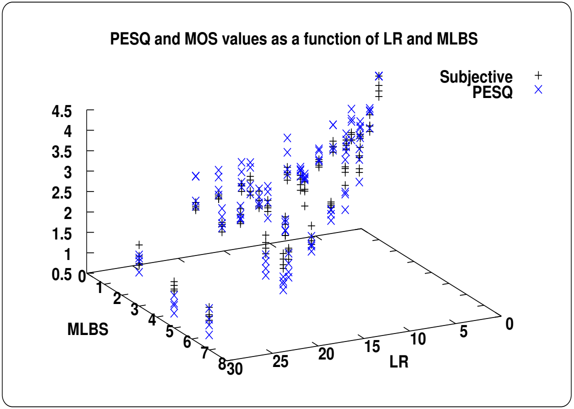


Subjective assessment

A P.800 based test (ITU designed) was carried out with

- 11 subjects
- 168 samples, corresponding to 42 configurations
- We chose a 9–point scale.
- Anchors were used during the warm–up session.
- We compared PESQ to our subjective scores
- Correlation was quite high, given the losses considered (0.86)
- PESQ's estimation did not degrade at the same rate as the network conditions degraded
- It also tends to over–estimate when the burstiness is low, and under–estimate when high
- Our estimation is robust & close to human perception

PESQ and subjective scores



We also compared the scores with P.563 (single sided metric)

Wireless VoIP challenges

- WiFi voice is not yet a true competitor to the cellular network
- Lack of true handsets (currently only PDAs), also people *like* their phones as fashion accessories, cameras etc.
- Security issues exist WiFi, WEP is known to be broken
- Quality issues related too:
  - PDAs are not voice optimised
  - WiFi primarily a data communication technology
  - Quality related to the environment, walls, windows etc.
  - Use of unlicensed spectrum (2.4 Ghz)
  - Other interfering devices: access points, users on the same channel, Bluetooth devices, microwaves ovens and so on
  - User mobility is unpredictable

More conclusions

- We have looked at the complete system in separate studies
- We do not assume any network QoS support
- Simply “how is the system performing today?”
- Good quality voice is possible on today’s Internet (as you know)
- Operators want isolate VoIP traffic or give it higher priority
- We found well-provisioned links work fine
- Certain infrastructures need upgrading for reasonable VoIP
- End-systems should not be ignored
- Wireless IP VoIP has not made any impact
- Due to poor quality handsets and a data-centric approach (WiFi)
- Radios are well not designed, nor is 802.11 speech coding

Short summary of publications

Research & prototypes of VoIP systems over 5 years

- 2001: Started with call quality and link dimensioning
- 2003: Moved onto the end-systems, performance issues
- 2003: Investigated WAN measurements, backbone (1999 too)
- 2003: Some modelling of network delays and VoIP
- 2003: Prediction aspects: can we predict poor quality calls
- 2003: Licentiate thesis
- 2004: Moved to wireless, 802.11 measurements
- 2005: Handovers from 802.11 to cellular networks (ongoing)
- 2006: Estimating subjective voice quality using ITU methods
- 2008: Wrote PhD thesis
- 2009: Defence :-)

Current(ish) research issues

- Measurements on latest technologies (WiMAX)
- PESQ-like assessment on incoming calls
- Improve handover performance
    - Prediction from noisy signals, robustness of parameters
    - Time series analysis, e.g. Kalman filter
- Wide-scale deployment of IP telephony
    - Wired (mass migration to IP)
    - Wireless (fon like, DECT wide-range)
- 3D telephony
    - Better spatial separation
    - Higher quality interactions
    - Tools for separating background noise

DMA

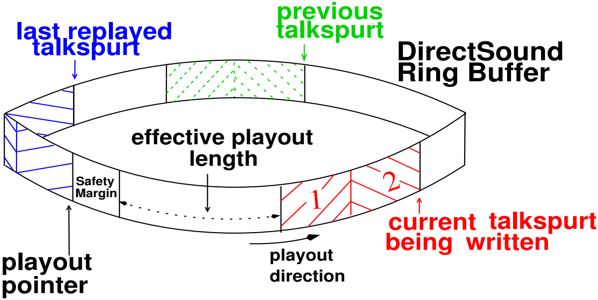
- Direct memory access (DMA) allows hardware subsystems to access system memory for reading and/or writing independently of the central processing unit.
- Many hardware systems use DMA including disk drive controllers, graphics cards, network cards and sound cards.
- Without DMA, using programmed input/output (PIO) mode for communication with peripheral devices. The CPU is typically fully occupied for the entire duration of the read or write operation,
- With DMA, the CPU initiates the transfer, does other operations while the transfer is in progress, and receives an interrupt from the DMA controller once the operation has been done.
- Especially useful in real-time or near real-time computing. For stream processing one needs data processing and transfer in parallel, in order to achieve sufficient throughput, e.g. VoIP applications.

Some obvious things...

Research issues exist from both the operator & end-user perspectives:

- Operator viewpoint
  - Multiservice networks will happen, separate networks are costly for operators
  - Voice still is the biggest revenue earner for Telco's, so VoIP needs accommodating in the multi-service the network
  - Some issues exist here
- End-user viewpoint
  - Internet Telephony needs wireless access (802.11x or WiMAX) to seriously compete with the cellular systems
  - More subjective assessment is needed in the VoIP domain
  - Loss, delay and jitter are not really sufficient in order to get a *real* human understanding

Directsound playout support



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

Mouth-to-ear delay measurements

Audio Tool	Latency (ms)
Sicsophone 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 rather than a speech sample for easier triggering and delay calculations
- Point here is to show the end-systems should not be neglected, and one can change them relatively easily (unlike the network!)