End-to-end IP Service Quality and Mobility

- Lecture #12 -

Special Course in Networking Technology

S-38.215

vilho.raisanen@nokia.com

Vilho Räisänen

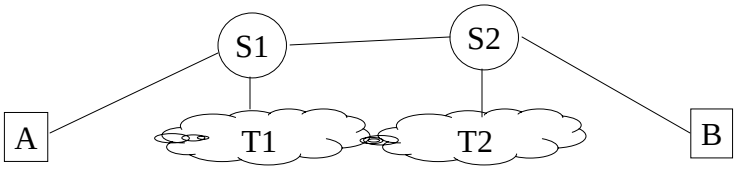
Agenda

- Service quality negotiation.
- TSpec multiplexing.
- Mobility modelling.
- E2e QoS provisioning.
- Access network provisioning.
- Fairness.

Vilho Räisänen

Exercise A: service quality negotiation

- Negotiation:
  - (A-B: capability negotiation).
  - A-S1: QoS request
  - S1-T1: available QoS levels
  - S1-S2: QoS request.
  - S2-T2: Available QoS levels.
- S2-S1: Available QoS levels.
- S1-S2: Inform QoS.
- S1-T1: QoS commit.
- S2-T2: QoS commit.
- S1-A: QoS acknowledgment.
- S2-B: QoS acknowledgment.



Vilho Räisänen

Exercise A, cont'd

- Assume: QoS request between MS and agent or between two agents: A ms.
- Query QoS levels from transport: B ms.
- Inform available QoS levels: C ms.
- Commit transport QoS: D ms.
- Acknowledge QoS: E ms.

2A+
2B+
2C+
2D+
2E
=
2x(A+B+C+D+E).

- Imaginary example:
  - A=10 ms.
  - B=30 ms.
  - C=10 ms.
  - D=200 ms.
  - E=10 ms.

⇒ 520 ms.

Number of messages optimised.

Set up classifiers & conditioners.

Capability negotiation time needs to be added.

Vilho Räisänen

Exercise B: TSpec multiplexing

- Assume: TSpec parameters = *r*, *B*, *M*.
- N streams with same TSpec parameters.
- Transmission time of a MTU sized L2 frame = M/l (l = link layer speed).
  - Ethernet MTU, 256 kbit/s => 47.5 ms.
- Worst case: whole burst comes at once => $t_{\max} = \frac{NM}{l}$
- Example:
  - r = 30 kbit/s.
  - M = 12160 bit.
  - l = 256 kbit/s.
  - N = 8.

  $\Rightarrow t_{max}$ = 380 ms.
- ON/OFF traffic => ~ 405 ms separation for bursts.

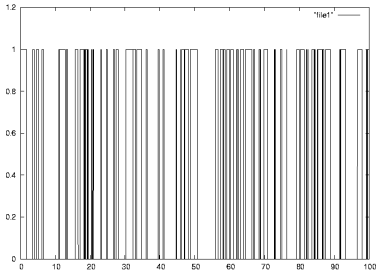
Vilho Räisänen

Exercise B, cont'd.

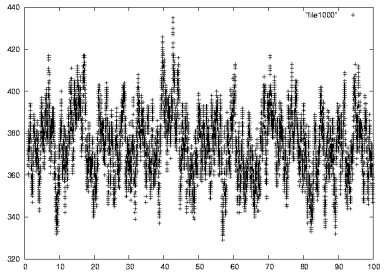
- Maximum delay variation between two adjacent packets = 380 ms in the previous example.
  - Experienced by lower priority class.
  - Even high priority class may experience delay variation up to 335 ms.
- VoIP multiplexing: codec with Voice Activity Detection (VAD) produces ON/OFF patterned streams.
  - Assume maximum bit rate = CBR for payload; VAD active 60% of the time.
  - Earlier lecture:
    - r = 0.375 x CBR.
    - b = CBR x 1 sec.

Vilho Räisänen

CLT (half duplex)



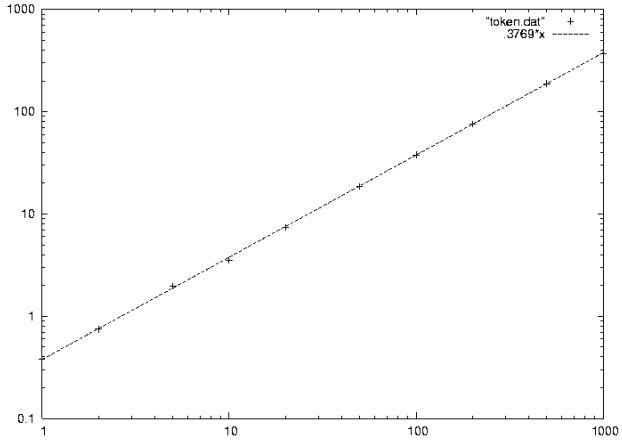
N=1



N=1000

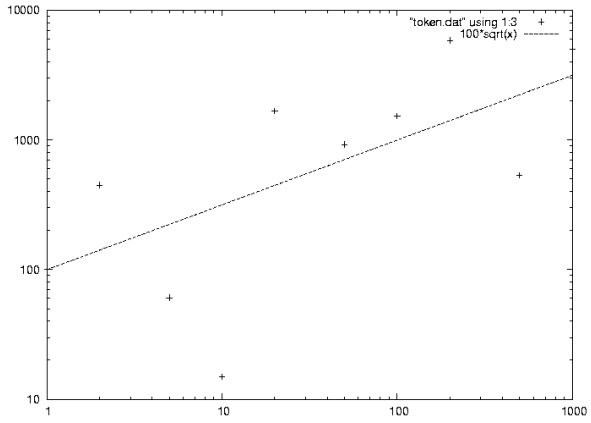
Vilho Räisänen

Token rate



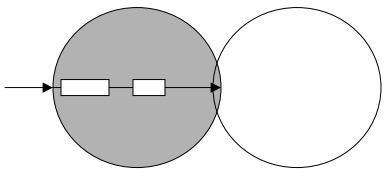
Vilho Räisänen

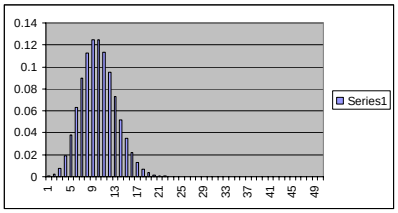
Maximum bucket depth (l=0%)



Exercise C: Mobility modelling

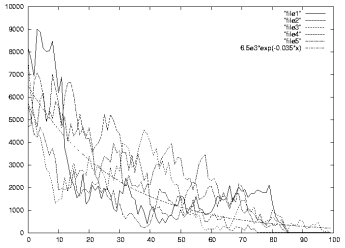
- Assumptions:
  - Stable loading: arrival rate = r, exponential distributions.
  - Exponentially distributed staying time s within the cell.
  - Call arrival rate = c, length exponentially distributed.
  - Call duration = d, length exponentially distributed.



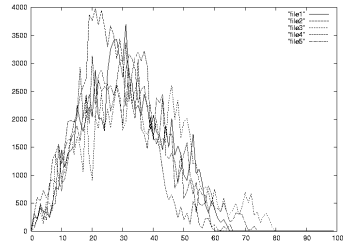


Exercise C, cont'd

- Aggregate rate in/out = 1/60, simulate over 200000 time steps =>

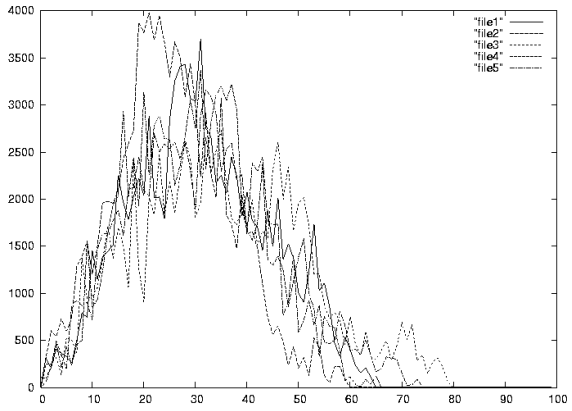


- Aggregate rate in = 1/60, rate out for **single** mobile = 1/1800 =>



Vilho Räisänen

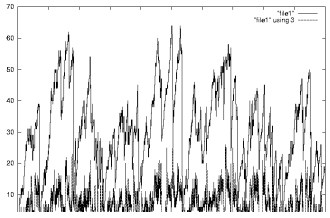
Frequency distribution: #(MN)

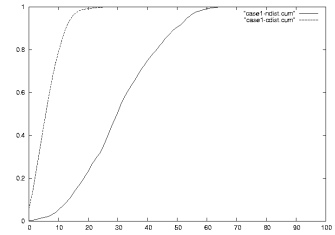


Vilho Räisänen

Exercise C, cont'd

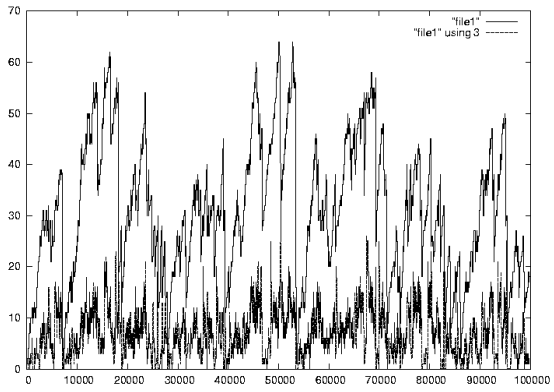
- Parameters:
  - R = 1/60
  - S = 1/1800
  - C = 1/600
  - D = 1/180
- N(MN):
  - 50% = 30
  - 99% = 60
- N(call):
  - 50% = 6
  - 99% = 19





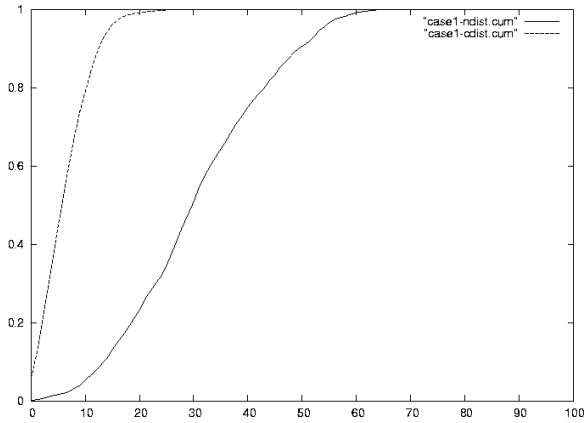
Vilho Räisänen

Time series (#MN, #call)



Vilho Räisänen

Cumulative distribution: #MN, #call



Exercise C, cont'd

- Next level of details:
  - Handover probabilities to adjacent cells defined.
  - Can maintain a call during handover
  - ⇒ Amount of calls handed over between cells.
  - ⇒ Variability of inter-cell traffic.
- Further elaborations:
  - Different types of applications.
  - Define routes for mobiles.
  - Use velocity distributions.
  - …

Exercise D: e2e QoS provisioning

- **Independent & random packet loss:**
  - Probability of single packet loss across M domains: $1 - P_{total} = \prod_{i=1}^{M} (1 - p_i)$
  - Probability of exactly N consecutive packets being lost in a domain. $P(N) = p^N (1-p)^2$
  - Exactly N packet losses happen in M domains: $P_{tot}(N) = \left[1 - \prod_{i=1}^{M} (1-p_i)\right]^N \prod_{i=1}^{M} (1-p_i)^2$
- Access networks: packet loss correlated => independent loss assumptions don't hold.

  => full description: probability of different combinations of loss patterns. *{P(a,l)}*

E2e QoS provisioning, cont'd

- **Delay:**
  - Average delay additive.
  - End-to-end delay using percentiles: $P\left(d_{tot} > \sum_i d_i^p\right) \geq p^M$
  - Whole distribution is convolution of individual distributions. $S(s) = \int_0^\infty d_1(d) d_2(s-d) dd$
  - Delay variations typically depend on time of day.
- **Delay variation:**
  - Simplest estimator based on estimators for delay distribution: $\sigma_{e2e}^2 = \sum_{i=1}^{N} \sigma_i^2 + 2\sum_{i=1}^{N-1} \sigma_{i,i-1}^2$
  - Does not take into account temporal correlations.

E2e provisioning, cont'd

- **Throughput:**
  - E2e throughput = min(t_i).
  - Meaning of per-domain throughput needs to be defined precisely.
  - TCP throughput has temporal dependencies due to CW control mechanism.
    - Effect of RTT.
    - Effect of packet loss.

Exercise E: AN provisioning

- Mobile IP access network using DiffServ transport.
  - VoIP calls use EF.
  - Browsing uses AF.
  - Other traffic uses BE.
- Assume AP with mobility pattern similar to Exercise C, use 99% dimensioning:
  - #(MN) <= 60.
- Assume VoIP calls follow the usage pattern of Exercise C => #(calls) <= 19 (99% dimensioning).
  - Assume AMR codec, 20 ms frame size, single frame / pkt.
    - => maximum bit rate 12.2 kbit/s
    - Add IPv6 headers => 36.2 kbit/s on IP layer.

Exercise E, cont'd

- Maximum VoIP traffic from single cell = 678.8 kbit/s without VAD.
  - Conservative provisioning for this speech pattern (full duplex).
- Assume 2Mbit/s link => 1321.2 kbit/s for AF and BE traffic.
  - If browsing usage follows the same pattern than VoIP, get >60 kbit/s average throughput per user.
    - Probably OK for browsing => multiplexing of bursty request/reply traffic.
    - Probably not enough for large downloads.
    - Possible strategy for browsing: token rate = 60 kbit/s for streaming, relatively large bucket size.

Exercise E, cont'd

- Add support for streaming, keep old services.
  - Alternatives: map to EF or AF PSC.
- Provide support for 128 kbit/s token rate for streaming, maximum 1Mbit/s of cell capacity available for streaming.
  - Assume maximum single stream bandwidth = 256 kbit/s.

  ⇒Tight delay dimensioning in AN: ~ 4 simultaneous streaming users.

  ⇒Loose delay dimensioning in AN: ~ 8 simultaneous streaming users.
- Conclusion: Loose delay dimensioning in AN, use lower-delay SLA towards transport provider.
  - Map streaming to AF PSC with token bucket policing.

Exercise E, cont'd

- Summary:
  - VoIP to EF, map to low-delay SLA in external transport.
  - Streaming to AF PSC, map to low-delay SLA in external transport.
  - Browsing to AF PSC, map to medium delay SLA in external transport.
  - E-mail to BE PSC, map to BE SLA in external transport.

Exercise F: fairness

- Fairness can be used for comparing service quality support for different users.
- *Fairness index* compares QoS indices computed for individual streams or users.
- To be noted:
  - Effect of other domains.
  - Sampling and measurement methodology.
  - "Distribution tail" behaviour vs. average behaviour.

$$f = \frac{\sum x_i^2}{n(\sum x_i)^2}$$