A Virtual Time Simulator for Studying QoS Management Functions in UTRAN

by David Soldani
Contents

- Simulator structure
- Traffic models
- QoS management functions
- Simulation assumptions
- Simulation results
- Conclusions
All calls/sessions are generated at the beginning and subsequently processed (played back) taking into account the corresponding arrival times and service activities, whence the name virtual time simulator.
Traffic models

- **Real Time (RT) traffic mix [7]**
  - Conversational Speech (CS), Conversational Data (CD), and Streaming (S) Traffic Classes (QoS Classes)
  - Calls are generated according to a Poisson process
  - For each service a separate process
  - Each call is held for an exponentially distributed service time
  - The call inter-arrival period follows the same distribution

- **Non-real time (NRT) traffic mix [7]-[8]**
  - Interactive and Background QoS Classes
  - Arrival of session set-ups is modelled as a Poisson process
  - For each service there is a separate process
  - Interactive class
    - Each session is modelled as an ON/OFF process
    - ON periods: log-normal distributed with a cut off
    - OFF time: Pareto distribution with a cut off
  - Each session of Background traffic class is assumed to be in service for an exponentially distributed time and the arrival rate has been approximated with the same type of distribution
Admission Control (1/2)

- Each Resource Request (RR) is assigned a resource request priority (RRP) based on radio access bearer attributes provided by the core network, and a cell based **Priority Class** parameter, which can be set differently for the different traffic classes.

- In case of overload, the radio resource requests are arranged into a queue, and served following the strict priority principle, and, at given priority, taking into account the corresponding arrival times (FIFO).

- Resource requests cannot stay in the queue longer than the maximum allowed queuing time, i.e. **Max Queuing Time** seconds (TC based management parameter), and are immediately rejected if the maximum allowed queue length, **Max Queue Length** (cell based management parameter, in number of RABs), is exceeded.

- Except for the overload situation NRT traffic is always admitted.
Admission Control (2/2)

- Overload state (1)
  \[ P_{TxTotal} = P_{NRT} + P_{RT} > P_{TxTarget} + Offset \]

- RT traffic is not admitted, if either (1) or the following inequation is satisfied
  \[ P_{RT} + \Delta P_{RT} > P_{TxTarget} \]

- The load increase and wide band power estimates are based on the downlink fractional load equation presented in [5]
  \[ \eta_{DL} = \frac{1+SHO}{W} \cdot \rho_k \cdot R_k \cdot v_k \cdot ((1-\alpha_k) + i_{k,DL}) \]
Load (Congestion) Control Function

- The only congestion control actions supported by the simulator is the reduction of the bit rates of NRT bearer services, whenever an overload situation occurs, i.e. when inequation (1) is satisfied.

- The bit rates are downgraded starting from the bearer services with lowest priority, and at given priority, based on their arrival times (FIFO), but none of the sessions are released.

- Though connections, RT included, may be dropped following the same principle.
Packet Scheduler

- Bit rates of admitted NRT bearer services are fast scheduled in a *Round Robin fashion (fair throughput)*, based on resource request priorities and arrival times (FIFO)
- Packet scheduler follows the best effort model and relies on the capacity left by the RT traffic, i.e.
  \[ P_{\text{NRT Allowed}}^{\text{NRT}} = P_{\text{TxTarget}} - (P_{\text{NRT}} + P_{\text{RT}}) \]
  which will be fairly shared among the different capacity requests whenever is available, whence *fast scheduling*
- For a *fair resources utilisation*, the maximum transmission rate (RMax), at given position of the terminal, needs to be computed as a function of the geometry factor (G) and the downlink required transmission power (PRL) for that particular radio link, which yields:
  \[ R_{\text{Max}} = \frac{W}{E_b/N_0} \cdot \frac{P_{\text{RL}}}{P_{\text{TxTotal}}} \left[ \frac{1}{1 - \alpha + 1/G} \right] \]
Simulation assumptions (1/2)

- **Simulation resolution**: 500ms
- **Simulation time**: 6h
- **Simulation parameters**:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity factor ($\nu$)</td>
<td>0.67 (CS), 1 (CD &amp; Str.) and 0.5 (Background)</td>
</tr>
<tr>
<td>CCH transmission power</td>
<td>20%</td>
</tr>
<tr>
<td>Chip rate ($\nu$)</td>
<td>3.84 Mcps</td>
</tr>
<tr>
<td>DCH DL allowed bit rates for Interactive and Background</td>
<td>0, 16, 32, 64, 128, 256 and 384 kbps</td>
</tr>
<tr>
<td>DCH DL guarantee bit rates for Conversational Data</td>
<td>64 kbps</td>
</tr>
<tr>
<td>DCH DL guarantee bit rates for Conversational Speech</td>
<td>12.2 kbps</td>
</tr>
<tr>
<td>DCH DL guarantee bit rates for Streaming</td>
<td>32 kbps (Audio)</td>
</tr>
<tr>
<td></td>
<td>64 kbps (Video)</td>
</tr>
<tr>
<td>Downlink orthogonality ($\alpha$)</td>
<td>0.5, for ITU Vehicular A</td>
</tr>
<tr>
<td>Downlink required EbN0</td>
<td>5-7 dB @ 20% FER, depending on bearer service</td>
</tr>
<tr>
<td>Max Queue length</td>
<td>10 RABs</td>
</tr>
<tr>
<td>Max Queuing Time</td>
<td>20 s (Em. Call), 10 s (CS &amp; CD), 15s (Str.), and 5s (Int. &amp; Backgr.)</td>
</tr>
<tr>
<td>Offset</td>
<td>10%</td>
</tr>
<tr>
<td>Other-to-own cell interference ($i$)</td>
<td>0.55</td>
</tr>
<tr>
<td>Priority Class</td>
<td>1 (Em.Call), 2 (CS), 3 (CD), 4 (Str.), 5 (Int.), and 6 (Backgr.)</td>
</tr>
<tr>
<td>ARP</td>
<td>1 (Gold), 2 (Silver), and 3 (Bronze)</td>
</tr>
<tr>
<td>PtxTarget</td>
<td>70%</td>
</tr>
<tr>
<td>Simulation time</td>
<td>6 h</td>
</tr>
<tr>
<td>Soft handover overhead</td>
<td>20%</td>
</tr>
</tbody>
</table>
Simulation assumptions (2/2)

- Radio Resource Priority values:

<table>
<thead>
<tr>
<th>Code</th>
<th>Call/Session type</th>
<th>Gold (ARP=1)</th>
<th>Silver (ARP=2)</th>
<th>Bronze (ARP=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Signalling</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>Emergency call</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Conversational Speech</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>Conversational Data</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>Streaming</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Interactive</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>Background</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

- Main parameters of simulated traffic:

<table>
<thead>
<tr>
<th>Call/Session type</th>
<th>Mean service time [s]</th>
<th>Mean arrival rate [s]</th>
<th>User traffic [mErl]</th>
<th>N. of users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signalling</td>
<td>120</td>
<td>4800</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Emergency call</td>
<td>60</td>
<td>60000</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Conversational Speech</td>
<td>120</td>
<td>4800</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>Conversational Data</td>
<td>120</td>
<td>24000</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>Streaming</td>
<td>300</td>
<td>10000</td>
<td>30</td>
<td>60</td>
</tr>
<tr>
<td>Interactive</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>70</td>
</tr>
<tr>
<td>Background</td>
<td>1200</td>
<td>7200</td>
<td>166</td>
<td>70</td>
</tr>
</tbody>
</table>

* see [8]

- ARP and TC values are randomly allocated with equal probability
Calls/Session arrivals distribution

- 395 users lead to 1125 connection attempts
Load status

- All supported QoS management functions work as intended
Load distribution

![Graph showing normalised distribution function for Lrt, Lnrt, and Ltot in percentage. The x-axis represents Lrt(r), Lnrt(g), and Ltot(k) in percentage, while the y-axis represents the normalised distribution function, ranging from 0 to 1.](image-url)
Cell throughput

Cumulative Distribution Function

Total traffic

NRT traffic

NRT traffic

RT (r)  NRT (g)  Total (k)  Throughput [kbps]

0  200  400  600  800  1000  1200
AC queue length distribution

Queue length

DF (*) and CDF (o)
Call Block Ratio

- This demonstrates the impact of prioritising the CS service over the CD service, which in turn has higher priority than the Streaming service, on the selected traffic mix.
Served RT traffic in Erlangs

- This reflects the corresponding offered loads and call block ratios
Interactive and background user throughput

- This demonstrates the impact of prioritising the Interactive service over the Background service
Interactive user throughput

- This demonstrates the impact of prioritising Gold, Silver and Bronze users within the Interactive traffic class.
Background user throughput

- This demonstrates the impact of prioritising Gold, Silver and Bronze users within the Background traffic class
NRT mean user throughput in kbps

- The differentiation among Gold, Silver, and Bronze users is due to the priority based bit rate scheduling, which works as good as expected.
Conclusions

- A virtual time simulator for assessing QoS management functions in UTRAN has been presented.
- As a part of this framework, Admission Control, Load Control, and Packet Scheduler functions, based on priorities and fast scheduling, have been analysed in terms of offered and served traffic mix, throughput, queuing time and call block ratio by means of simulations.
- From the simulation results, the proposed solution turns out to perform as intended, and appears to be a good trade of between the complexity and the simplicity of an advanced dynamic and static simulator, respectively; and thus has the potential for investigating any QoS management function in UTRAN, before its deployment throughout a radio access network.
References


[7] ETSI, TR 101 112 v.3.2.0, “Selection procedures for the choice of radio transmission technologies of the UMTS”, UMTS 30.03 v.3.2.0.
