Hardware and Bandwidth Requirements Estimation for Mobile Networks Using Statistically Characterised Multi-service Traffic and Discrete Event Simulation

ZELADÚ Telecommunications Consulting

Juan E. Burgos, Consultant and Gregorio Delgado, Senior Consultant

1. Abstract

This paper presents a new method to dimension mobile access networks, here applied to the Node B hardware configuration and the Iub interface, using **statistically characterised multi-service traffic and discrete event simulation**. This method has been implemented in our dimensioning tool SUPRA (Simulador de UMTS para la Planificación de la Red de Acceso).

Although first applied to UMTS, the tool can be easily applied to GPRS, based on the event simulator, which is the core of the application.

2. Introduction

Typically, mobile operators calculate capacity in the radio access network using Erlang B as the basis of the engineering, even for data packet switched (PS) transactions. In this way, the bandwidth requirements for packet switched data are usually considered as best effort and estimated using a peak to average factor accommodated in the latent capacity available from the circuit switched (CS) connections and the blocking probability.

A similar approach is to translate the PS data into CS connection based on the time required to transfer the data; this approach is obviously not a natural simplification, and can only provide useful estimations if the amount of data to transfer is small, in systems with low utilisation of resources, requiring short transmission times.

With the introduction of GPRS and UMTS, where PS- and CS-connections are mixed on the same interfaces, the described engineering method is too rudimentary, and new methods based on traffic simulations are required to dimension hardware and interface bandwidth more accurately. The method presented in this paper is based on traffic simulation using discrete event simulation for statistically characterized applications. The method has been implemented in our tool SUPRA, and provides operators and network designers and planners with a practical tool which will be able to provide more accuracy in hardware and bandwidth requirements calculations, reducing costs in the network deployment and network growth. SUPRA can be applied to different areas as developing of engineering and planning guidelines, test and optimisation support, network development, or feasibility studies for the introduction of new applications, among others.

3. The air interface

The implementation of the air interface for both GSM and UMTS is supplier dependant in areas as Radio Resource Management. This means that the performance of the air interface and its influence on the capacity and quality of the network varies from supplier to supplier. For this reason, this influence has been usually not considered in the capacity calculations done to dimension the mobile networks. In this way, the capacity calculations are traditionally done to estimate the required hardware configuration, and the effect of the air interface performance, that could be named as "software influence on capacity" or "soft capacity" is considered as comparably small. This approach assumes of course that the software configuration of the air interface works properly and the capacity and quality is driven by the hardware configuration of the network elements.

This is specially true in the UMTS case, as the soft capacity of a cell depends mainly on the number of secondary scrambling codes assigned to it, and this configuration can be easily changed via remote configuration following the offered traffic growth. Giving that the air interface configuration is done properly, the capacity of a cell will be driven by the number of channel elements and the bandwidth of the Iub interface available. For more information see [1], [2], [3], [4] and [5].

We propose to continue using this assumption, as a reasonable and practical one. Additionally it is not viable to produce a general tool that considers in detail the implementation dependant effects of the air interface. As an alternative approach we propose to characterise the air interface performance effects on the offered traffic to the hardware resources and interfaces as an additional and parameterised distortion to this offered traffic. The tuning of the parameters of this model shall be achieved via simulations done by the suppliers or measurements.

4. Actual engineering practices

As outlined in the introduction, Erlang B continues to be the most widely traffic model used in mobile communications. Though, the introduction of PS traffic and several applications other than the CS voice on this networks make this model assumptions more difficult to be accepted. We will do a short review of the Erlang B assumptions and will analyse to what extend the simplifications made to continue using Erlang B as the basis for the engineering are reasonable. We will show that the old approach is not accurate enough.

It shall be noted that the accuracy of the traffic models is highly variable with the amount of offered traffic.

The Erlang B model has been applied to dimension de PSTN for voice traffic, and the assumptions underlying can be considered suitable for CS voice traffic over the PSTN. The Erlang B model is a M/M/c/c traffic model (following the Kendal nomenclature); this is:

- Exponential distribution of the time between arrival of service requests (given by the Poisson distribution of the call attempts) (M)
- Exponential distribution of call duration (holding time) (M)

- Limited number of service agents (servers) (c)
- Same limited number of calls or busy customers in the system (c)

Additional assumptions are:

- Incoming calls are served in random order
- Population is infinite (number of subscribers)
- Blocked calls are cleared (non persistent)
- Every source of traffic has the same priority
- Busy sources initiate no calls

One of the inherent limitation of the model is that no queuing is considered, and the blocking calls are cleared (non persistent). Many GSM networks provide queuing in the air interface and are anyhow dimensioned using Erlang B. The usual behaviour of the customers is to recall when blocked. Another assumption, that may be suitable for the PSTN, as the infinite population assumption, is also far to be acceptable for the air interface in cellular networks, where the number of total customers in a cell is often comparable with the number of servers or channels available.

After this evaluation, it gets clear that the reason for using Erlang B to dimension the air interface is not backed enough by the traffic theory. Its usage has been inherited by the engineering practices of the PSTN, considering also that it is an easy, well know and widely used method.

The limitations of the modelling have been usually palliated by careful network monitoring and time advanced network upgrades.

With the introduction of GPRS and UMTS the characteristics of the offered traffic are much more heterogeneous. PS and CS connections are served by the same resources, and the large number of different new applications ends up with a highly non heterogeneous traffic sources, which statistical characterisations differ strongly from each other. Additionally, new features as **priority** and **quality of service** are again new conditions against the Erlang B assumptions.

Different approaches to dimension the access network (as previously explained, the hardware and interface resources) have been applied to dimension GPRS networks and are also being used for UMTS having often in common the following characteristics:

- usage of latent capacity (from the Erlang B calculations) for PS traffic
- best effort assumption for PS traffic
- assimilation of PS traffic to CS traffic for Erlang B modelling based on transaction time
- constant average to peak correction factor for interface bandwidth estimation

A typical approach to dimension the GPRS Gb interface (Frame Relay) is to consider the average offered traffic in kbps and apply a peak to average ratio considering best effort. This peak to average ratio considers the peak characteristic of the aggregated traffic over time. But the peak characteristic of the aggregated traffic depends on the amount of traffic that is concentrated on an interface. The most suitable way to estimate the dimensioning factor is via traffic simulations that show the variability of this factor with parameters as delay and amount of traffic.

Looking at the considerations presented about the Erlang B model and the traffic characteristics of the new mobile networks it gets clear that the application of new engineering models and tools is required for an accurate dimensioning of the access network.

We propose a new model which applies traffic simulation using discrete event simulation for statistically characterised multi-service traffic. This approach has been implemented in our dimensioning tool SUPRA (Simulador de UMTS para la Planificación de la Red de Acceso). In the following we present some SUPRA features, the underling assumptions and some results obtained based on a practical case of network dimensioning, that show that SUPRA is a useful and versatile tool for the design of actual mobile networks.

5. SUPRA general description

As we have described above, SUPRA provides the network planners with an useful tool that takes into account the heterogeneous nature of the traffic generated by the new mobile communication services. It has been designed in order to allow an easy inclusion of different services and supplier hardware configurations.

Starting from a specified Node B hardware configuration, and given a set of services, SUPRA evaluates a set of QoS parameters that show if the chosen configuration is appropriate for the given traffic load. SUPRA implements a traffic simulator that processes the generated traffic using the rules defined by the Node B configuration.

SUPRA uses SIM as its simulation core. SIM is a general purpose Discrete Event Simulator created by A. Eliens y D. Bolier [6], that is being widely used in a lot of simulators.

The SUPRA planning tool contains three main modules:

- The **Network Editor Module**, in which the hardware configuration of a Node B can be explicitly configured: number of channel elements boards, number of channel elements per channel element board, the processing capacity of each channel element, possible reservation strategies in each board (number of channel elements reserved for a specified Access Service Class), the way in which the different AAL2 connections are opened over the Iub interface.... These multiple configuration options are necessary to match the Node B hardware configurations provided by the suppliers.
- The **Service Editor Module**, that allows the definition of different services: Access Service Class (CS, PS…), number of users, and traffic model. SUPRA supports traffic models whose parameters are defined in terms of 'time' (i.e. the idle period between calls, for voice traffic) as well as models whose parameters are defined in terms of 'size' (number of bytes of a web page). Each traffic parameter (for example, the idle time between voice calls) is defined by means of its statistical distribution (typically exponential for the voice example) and its associated parameters (in this example, an exponential distribution is absolutely defined throw the average). A set of statistic distributions is supplied, including infinite variance distributions (Pareto 1st and 2nd class) that are widely used for modelling burst traffic, as web traffic.

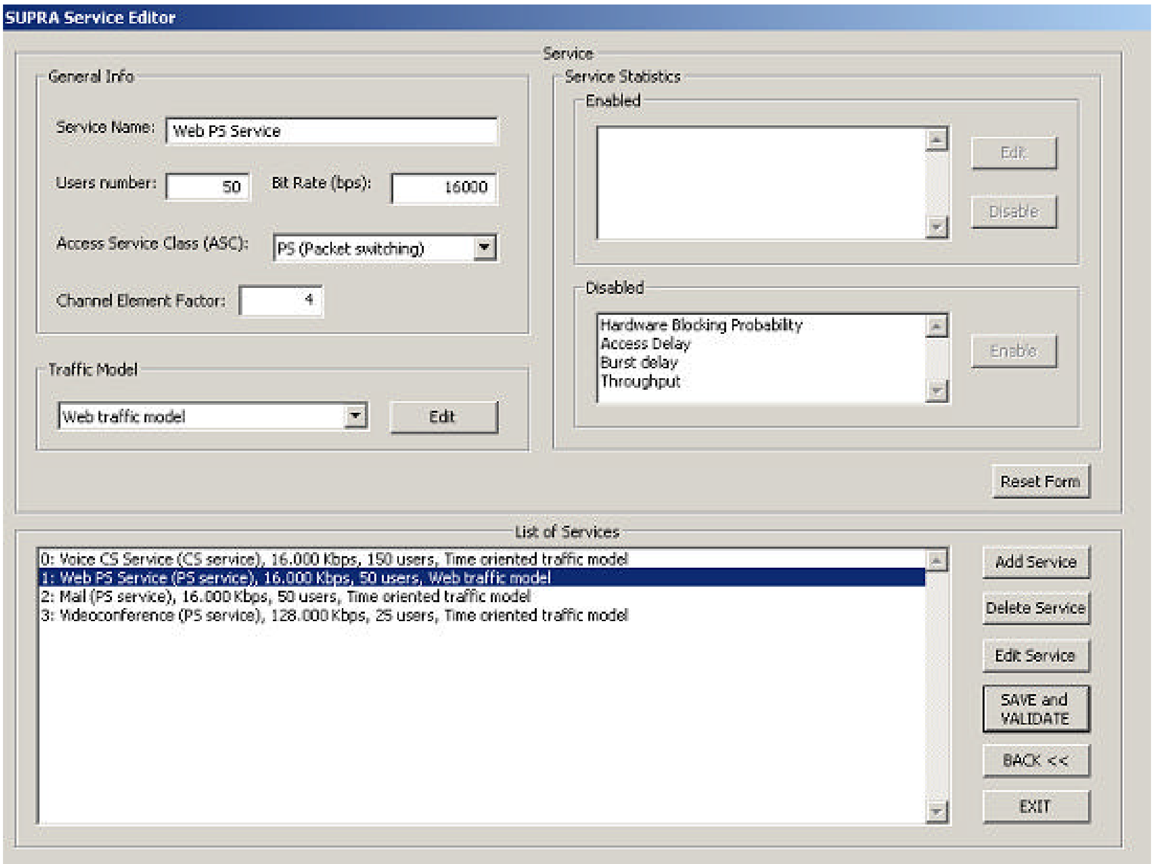

Figure 1. SUPRA Service Editor

- The **Simulation Editor Module**, whose main objective is the definition of the simulation settings (simulation time and random seeds for the inner random generator) and the monitoring of the current simulation when it is running.

At the end of the simulation, SUPRA provides a report in which the quality of the system (in terms of hardware blocking probability, throughput and access delay per service, and the bandwidth occupation of the Node B Iub interface) is measured for a given Node B hardware configuration.

The current release of SUPRA includes several assumptions, which are described below with the reasons of its implementation, and a brief analysis of its impact over the simulation results if necessary:

- SUPRA assumes that air interface effects can be neglected taking in account that strategies as Power Controls, the use of CRCs, an adequate random access dimensioning, etc. minimize strongly negative effects that are inherent to wireless communications (interferences, fading, shadowing, access collisions...). With this consideration, SUPRA provides valid results under good propagation conditions, so the realistic behaviour will be always a little worse.
- SUPRA is focussed on uplink. Uplink processing boards are more complex and expensive that downlink boards (there are several reasons for this: the inclusion of RAKE receivers, the

capability of synthesise the whole set of uplink scrambling codes....), so the cost of a specified Node B is closely related to the number of these boards. The high downlink radio capacity provided by 3rd generation systems and the possibility of adapting it via OMC (as we have previously commented) are another reason that justify this criteria. Downlink consideration requires some additional development and could be of relevance depending on the supplier of the Node B.

- The whole set of implemented traffic models are ON-OFF models, the most accurate for Discrete Event Simulation. Periods in which a specified user is active are called Session. Each session is compounded by several Bursts (or ON periods), when users are sending information, followed by OFF periods, when users are not sending any kind of information. Most of the most widely used traffic models are included in this scheme. (For more information and examples see Figure 2).
- Resources are occupied by CS users during the whole call/session, and they are released at the end. Instead of that, PS users performs a resource query and release with each data burst to transmit.
- Bandwidth occupied by high rate PS users can not be reallocated in order to accommodate new low rate CS users if no more resources are available (no perfect dynamic bandwidth allocation). This simplifies strongly allocation algorithms. The system will have a very similar behaviour most of the time (problems will appear under high load situations) and the results that SUPRA provides would always represent an upper limit to the PS users performances (the system will work as well or worse for PS services) and an lower limit for CS users (the system will work as well of better for CS services). This effect can be minimized via an adequate reservation strategy supported by SUPRA.

We continue to develop SUPRA adding new features that will increase the accuracy and flexibility of the simulations even more.

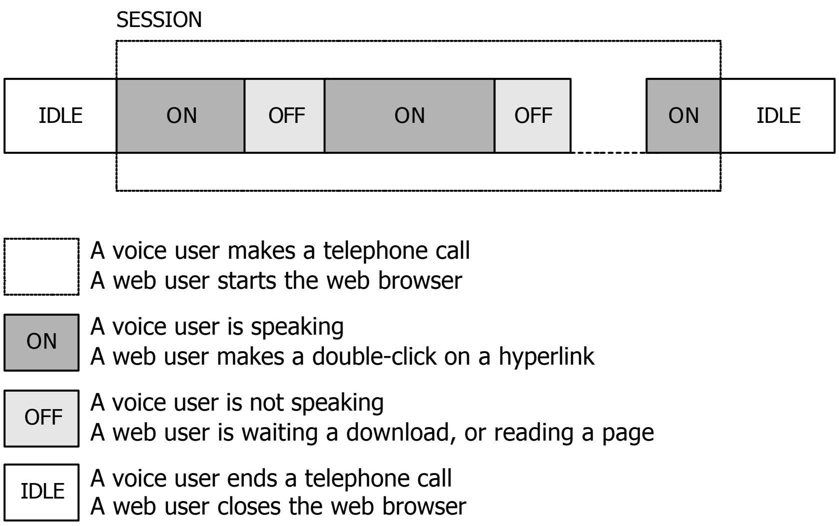

Figure 2. ON-OFF models and examples

6. SUPRA example: Assumptions

We have included a practical and realistic example in order to clarify how SUPRA works. Our initial data relates to a very highly populated Madrid district, Chamberi:

Network Growth	31/12/03	31/12/04	31/12/05	31/12/06	31/12/07	31/12/08
Population	144889	144850	144850	144850	144928	145044
Area (m^2)	4692200	4692200	4692200	4692200	4692200	4692200
Total number of subscribers	2942	7858	11458	19122	27561	37045
Cell range (m)	450	450	450	450	450	450
Number of cells (cloverleaf structure)	36	36	36	36	36	36
Total subscribers per cell	82	220	321	536	773	1038
Operator stake (%)	10	17	25	29	35	40
Operator's subscribers per cell	8	38	80	155	269	415

Table 1. Initial Data.

Some assumptions have been adopted:

- UMTS in this area will perform on average comparing to EUR15 (See [7] and [8]).
- We consider that network layout is planned for year 2008, and the cell range adopted is a practical estimation (this has not been checked with link budget calculations), valid for our simulation purposes.
- Network operator will increase its business stake with time.

Node B hardware features involved in our simulations are summarized below:

Hardware Feature	Value
Number of channel elements	32 channel elements per board
Processing capability (channel element/board)	16 Kbps / 512 Kbps
AAL2 multiplexer type	External (it is possible to combine channel elements from different boards to reach the necessary processing capability).
Reservation policy	None (all channels elements available for all services)

Table 2. Node B Hardware Features

We have considered two services: a voice service with a 16 Kbps CS carrier and data service, a web service that is supported by a 64 Kbps PS carrier.

For the **Voice service** a classic ON-OFF time-oriented traffic model has been adopted [9], whose parameters are:

Parameter	**Distribution**
Reattempt algorithm	Exponential (average: 0.1 secs). Maximum allowed delay: 5 secs
Idle period	Exponential (average: 900 secs)
Session period	Exponential (average: 180 secs)
ON period	Exponential (average: 1 secs)
OFF period	Exponential (average: 1.35 secs)

Table 3. Voice traffic model parameters

For the **Web service**, we have chosen an "always connected" model developed by the University of Malaga [9], [10]. This is an useful assumption in order to minimize the simulation time, because is equivalent to consider a certain number of web users during a busy hour.

It is an ON-OFF mixed model with three periods:

- Period 1: Request. It is an ON period. When a user clicks on a hyperlink to download a certain web page, the URLs corresponding with this page are sent (in this model all URLs are sent together). It must be included as Session period.
- Period 2: Acknowledgments. It is an ON-OFF period, that must be computed as Session period. As well as the user receives a TCP packet (which delay is modelled by means of the ACK interarrival time, that is an OFF period for uplink), it must send and ACK (ON period for uplink) to inform that it has been correctly received. Number of ACKs can be estimated as the page size divided by the packet size. It is the same to consider that the remaining page size decreases in an amount of bytes equal to the packet size with each ACK transmission.
- Period 3: Reading page period.

Parameter		**Distribution**
Reattempt algorithm		Exponential (average: 0.1 secs). Maximum allowed delay: 30 secs
Period 1	Request number	Pareto 2nd class (alfa: 2.22, beta: 6.42)
	Size of each request	Lognormal (average: 352 bytes, std: 109.85 bytes)
Period 2	Duration	Pareto 2nd class (alfa: 1.1, beta: 35.7 secs)
	Packet size	Constant (800 bytes)
	Time between ACKs	Exponential (average: 0.73 secs)
	Page size	Pareto 2nd class (alfa: 1.75, average 40600 bytes)

Parameter	Distribution
Period 3 (Reading)	Gamma (average: 35.286 secs, std: 147.39 secs)

Table 4. Web traffic model parameters

Note that web service uplink bit rate is not very high, because only URLs and ACKs must be sent.

The quality criteria are the following:

- For the voice service, the blocking probability must be lower than 2%. A voice call is considered blocked when its access delay is higher than 5 seconds.
- For the web service, the delay for 95% of bursts must be lower than the maximum allowed, fixed to 30 seconds. Due to data web characteristics, the number of access burst are not meaningful related to the rest of them, so the total burst delay can be appoximated by burst delay.

Finally, we have considered that each subscriber produces independently voice and web traffic.

The aim of the simulations is to determine when a processing capacity upgrade is required during the observation period (Dec. 2003 - Dec. 2008). SUPRA calculates the maximum number of subscribers that the configuration under test supports (situation in which QoS parameters are not reached), and this value will be translated into a date using operator's subscriber growth expectations via linear interpolation.

7. SUPRA example: Results

The simulation time was 50000 seconds, and we have considered the 99% confidence interval for all of them, using three simulations to compute each graph point (average value is shown).

We have computed the blocking probability for voice service and delay statistics (access delay and burst delay) for web service. Throughput statistics, as well as channel element usage, have been computed too in order to clarify how sharing of available resources is implemented by the system. The obtained results for a single channel element board have been summarized bellow:

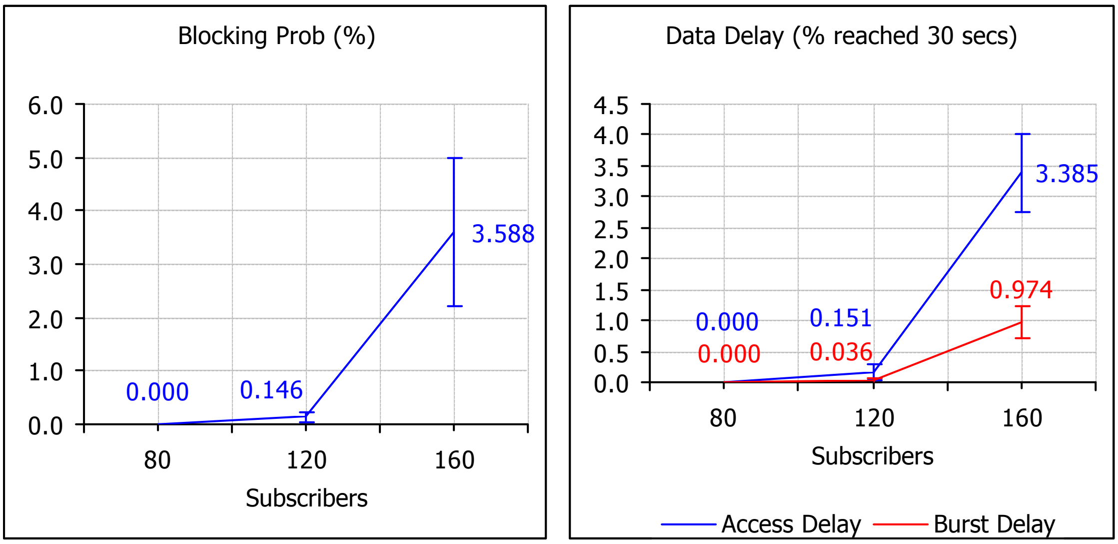

Figure 3. Voice Blocking Probability and Data Access and Burst Delay

Considering that total data delay is closed to burst delay value (as we have mentioned previously) we can see that the bottleneck is the voice service. Uplink web service is supported by a PS carrier and has not an important traffic load. Voice blocking keeps the quality criteria up to around 135 subscribers, so interpolating this value it is easy to follow that this configuration will be adequate until the third quarter of 2006 (see Section 8, SUPRA example: conclusions).

Throughput and average channel element usage have been also depicted:

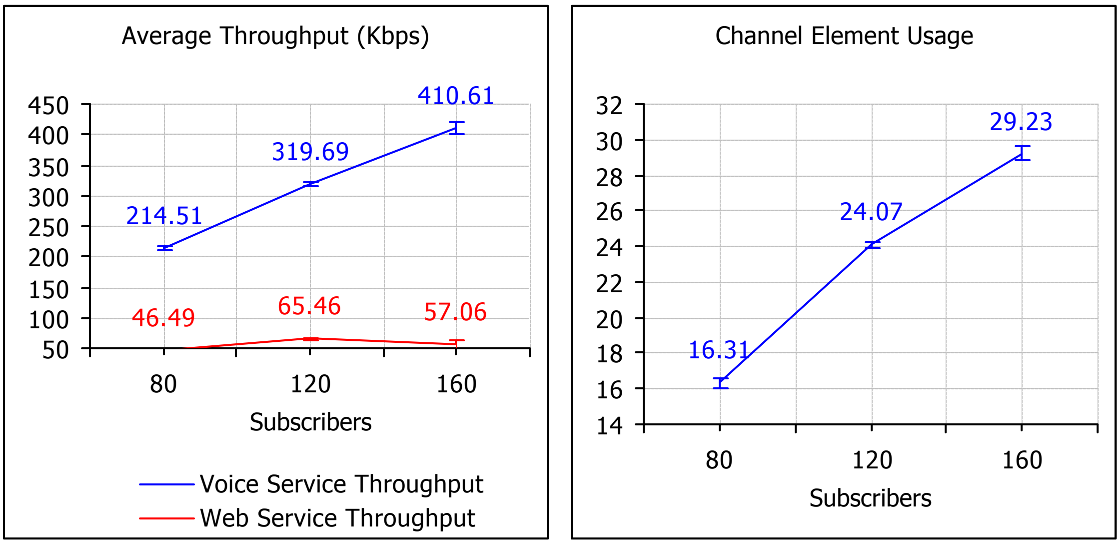

Figure 4. Average Throughput and Channel Element Usage

We can see how data throughput decreases with the number of users, according with the carrier performances because a higher number of CS users, when allocated, made use of resources until the end of the session, and there is less processing capacity available for PS subscribers. For 160 users, voice subscribers made use of 30 channel elements on average, so only 2 are available for data traffic. Note that a throughput increase of 16 Kbps means that a channel element is being used. The absence of blocking or delay for voice and web traffic with 80 subscribers is related to the low channel elements usage (17 on average). In fact, SUPRA channel element usage histograms shows how in this situation there are always available resources.

The next step is to calculate the maximum number of subscribers that could be successfully served using two channel elements boards. The results are the following:

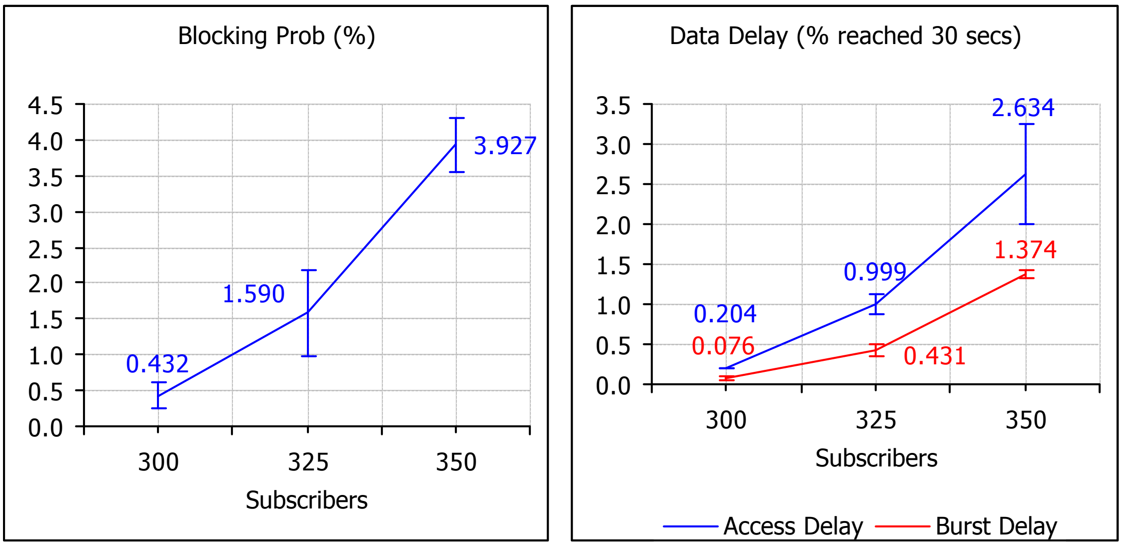

Figure 5. Voice Blocking Probability and Data Access and Burst Delay

Taking in account the upper limit of the confidence interval, the maximum number of subscribers that this configuration can serve is around 322. The observed behaviour for one channel element board (voice service as limiting and good data performances) is valid again. Note that the maximum number of allowed subscribers is more than twice than the obtained value for one board. This effect is due to the statistical multiplexing that this flexible configuration allows. Extrapolating this maximum number of users, we have found that the number of channel elements board must be increased (from 2 to 3) during the first third of 2008 (see Section 8, SUPRA example: conclusions).

Following figures show average throughput and channel element usage:

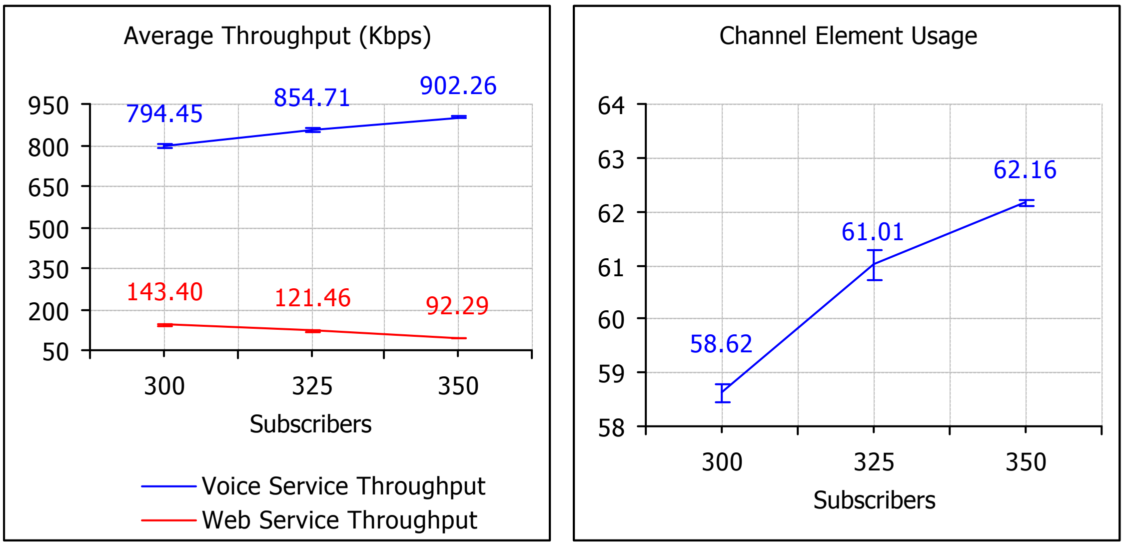

Figure 6. Average Throughput and Channel Element Usage

Note that channel element usage and throughput reach 95% of total available around the maximum permissible number of subscribers (for users, 61 occupied from 64 available and 975 Kbps from 1024 Kbps, respectively).

The results for three channel elements board are the following:

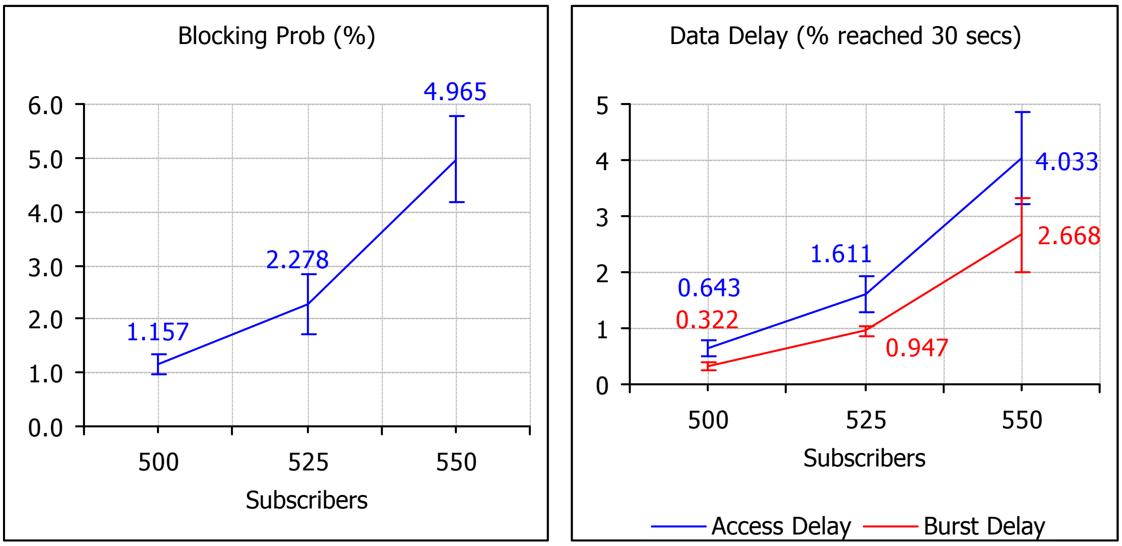

Figure 7. Voice Blocking Probability and Data Access and Burst Delay

These figures show that this configuration can accommodate more than 500 users. The voice service is the limiting factor again, while the data service maintains good performance also for 550 subscribers (total data delay, closed to 2.7%, is around 54% maximum allowed value). This processing capability is enough to serve subscriber expectations by the end of 2008 (415 users).

We have provided throughput and channel element usage statistics:

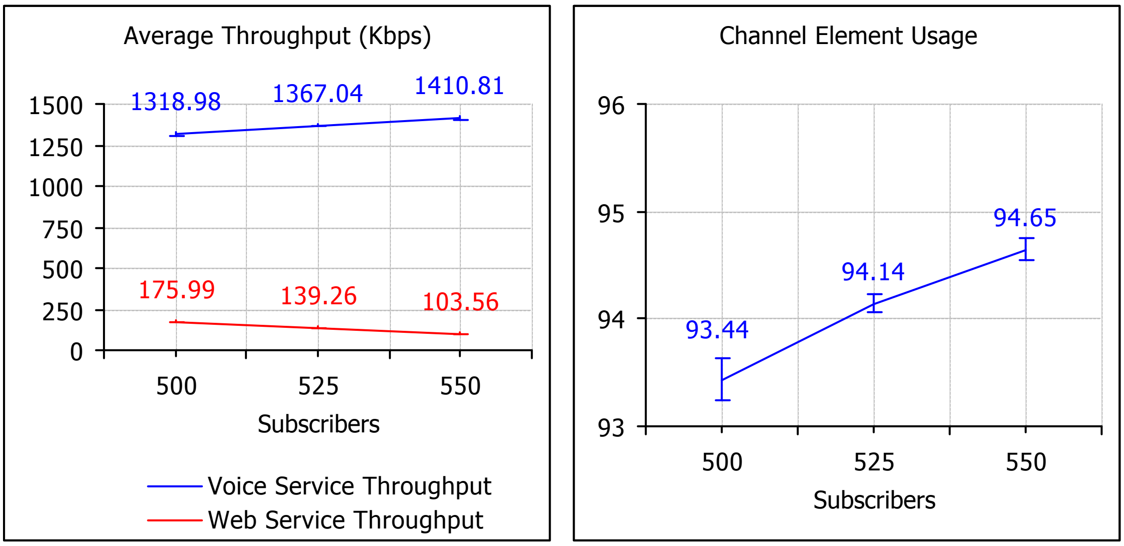

Figure 8. Average Throughput and Channel Element Usage

Note the high average channel element usage (94, when there are 96 available) for 500 subscribers while system maintain good performances (good voice and data QoS). This effect shows how this flexible configuration (external AAL2 MUX and no reservation strategies) made use of traffic statistical multiplexing to assure an appropriate system working.

8. Supra example: Conclusions

The practical example we have described shows how SUPRA can be applied as an useful tool to provide support for operators or suppliers during the network rollout. Next graphic describes the subscriber numbers expected by the operator from 2003 to the end of 2008, using data content from Table 1.

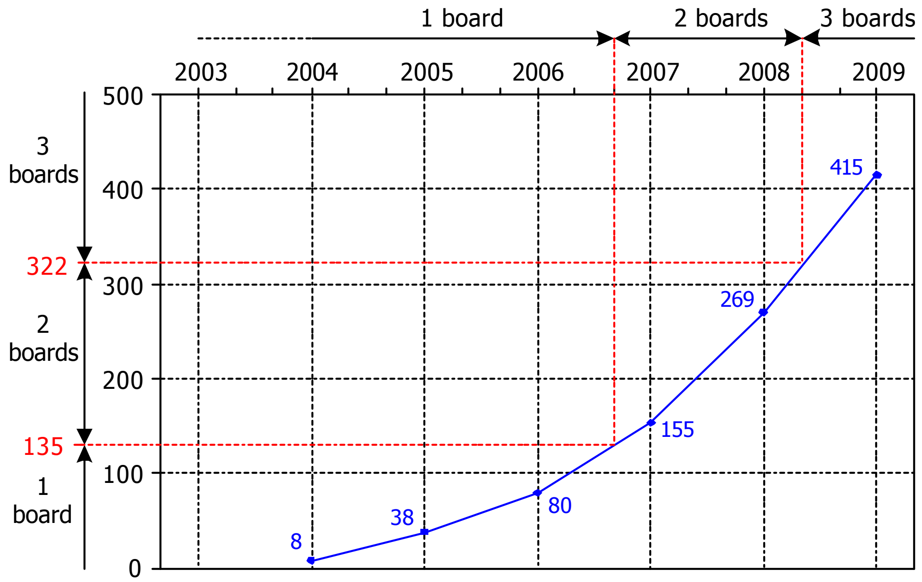

Figure 9. Subscribers expectations and needed processing capacity

We have depicted also over the y axis the calculated maximum number of subscribers that 1 and 2 channel elements boards can support, which have been obtained from simulations. Finally, we have interpolated this results over the x axis, that represents time. This method provides the dates when channel element board upgrades should be performed:

Network rollout hit	Number of Channel Elements Boards	Date
Network launch	1	1.1.2003.
First processing capacity upgrade	2	Before the end of 2006.
Second upgrade	3	During the first third of 2008.

Table 5. Time estimations for Network updating

These results are very useful for the operator to assure an adequate degree of service during the network operation, and could be used to perform an accurate equipment provisioning plan.

9. Summary

The new types of wireless telecommunication services available in 2.5 and 3 Mobile Generation Systems will made necessary new planning strategies in order to assure accurate predictions during the network design. We have shown that current strategies (based on Erlang B) are not appropriate to enclose the new traffic models that GPRS and UMTS networks must support. These new methods (that will reduce costs and time) must assume traffic simulation and shall allow a high degree of configuration flexibility to include different supplier implementations.

The SUPRA planning tool is an accurate response to this problem. It combines a high degree of flexibility (related to the Node B configuration) with the discrete event simulation for traffic models, that could be also parameterised. These attributes make SUPRA a powerful and versatile engineering tool for operators and suppliers that gives response to the actual necessities of the new generation network dimensioning and design.

In order to show the SUPRA capabilities, an hypothetic simulation environment has been described. Two meaningful services (voice and web services, both must be supported by the future networks) have been simulated, as well as a realistic network configuration. The obtained results show that this tool can be used to give answers to the new challenges that dimensioning and planning the new generation mobile networks means.

10. Contact Information

For consultancy services based on SUPRA contact with:

Gregorio Delgado, Senior Consultant, gdelgado@zeladu.com or

Juan E. Burgos, Consultant, jburgos@zeladu.com

To know more about the ZELADÚ service portfolio please visit www.zeladu.com

11. References:

[1] 3G TS 25.211 v3.7.0 (2001-06). Physical Channels and Mapping of Transport Channels onto Physical Channels. Release 1999. www.3gpp.org

[2] 3G TS 25.212 v3.6.0 (2001-06). Multiplexing and Channel Coding (FDD). Release 1999. www.3gpp.org

[3] 3G TS 25.213 v3.6.0 (2001-06). Spreading and Modulation. Release 1999. www.3gpp.org

[4] 3G TS 25.104 v3.3.0 (2000-06). UTRA (BS) TDD. Radio Transmission and Reception. Release 1999. www.3gpp.org

[5] "Comunicaciones Móviles de Tercera Generación". C. Lluch, J. M. Hernando. Telefónica Móviles España.

[6] "SIM: a C++ library for Discrete Event Simulation". A. Eliens, D. Bolier. Vrije Universiteit, Department of Mathematics and Computer Science. Amsterdam, Netherlands. http://www.cs.vu.nl/?eliens/sim.

[7] "The UMTS Third Generation Market Study Update". Report nº 17 from the UMTS Forum, August 2001. http://www.umts-forum.org/reports.html.

[8] "Long Term Potential Remain High for 3G Mobile Data Services". Report nº 18 from the UMTS Forum, February 2001. http://www.umts-forum.org/reports.html.

[9] "Multiaccess Strategies in an Integrated UTRA System". M. Ruiz, J. E. Burgos, J. M. Romero, A. Díaz. Departamento de Tecnología Electrónica, Universidad de Málaga. Congress ISCC-2000. July 2000. Paris, France.

[10] "Characterisation of Web Traffic". E. Calilari, A. Reyes-Lecuona, F. J. Gónzalez, A. Díaz, F. Sandoval. Departamento de Tecnología Electrónica, Universidad de Málaga. Congress Globecom 2001.