Asymptotic behavior of a GPRS/EDGE network with several cells controlled by a global capacity limit

Georges Nogueira[1], Bruno Baynat[1], and Pierre Eisenmann[2]

[1] Laboratoire d'informatique de Paris 6
LIP6 - CNRS
Paris, France
(e-mail: Georges.Nogueira@lip6.fr, Bruno.Baynat@lip6.fr)

[2] Nortel Networks
Wireless Network Engineering
Chateaufort, France
(e-mail: pierree@nortelnetworks.com)

Abstract. This paper is a contribution to the generic problem of having simple and accurate models to dimension radio cells with data traffic on a GPRS or EDGE network. It addresses the issue of capacity limitation in a given cell due to coupling with other cells because of a central equipment or transmission link of limited capacity. A mobile can't access the cell although it is alone, because the capacity limit is reached due to traffic on other cells. Our purpose is to extend our previously published Erlang-like law for data traffic to the constrained multiple-cell system and to derive asymptotic developments for all the average performance parameters that are necessary for the dimensioning of the multiple-cell system.
Keywords: GPRS, EDGE, modeling, performance evaluation, dimensionning, discrete-time Markov chain, Erlang, group of cells.

1 Introduction

GPRS is an overlay on GSM networks that allows end-to-end IP-based packet traffic from the terminal to e.g. the Internet. EDGE is an improvement over GPRS whereby the modulation scheme on radio is modified to allow higher throughputs thanks to advanced power amplifier and signal processing technologies. In a GPRS (or EDGE) cell, traffic is split between voice (on circuit) and data (on packet). Data uses a few dedicated circuits which are decomposed into 20 ms ÒblocksÓ carrying elementary packet traffic. The packet-based traffic is managed by the PCU (Packet Control Unit), a standardized network element in charge of the MAC layer (multiplexing of mobiles) and RLC layer (decomposition into elementary blocks and retransmission when radio errors occur). The PCU is connected to the SGSN (Serving GPRS Support Node) which manages the end-user mobility and hides it to the external world. It is linked to the edge router, called GGSN, by an IP tunnel in which traffic is encapsulated. The GGSN is the fixed anchor point to the Internet or service platforms, and a user may change SGSN while going from cell to cell. In this end-to-end chain, it is possible to have traffic limitation in an element in charge of managing several cells (typically a PCU or a SGSN module or a transmission link).

Several research works have investigated the analytical modeling of GPRS systems. Most of the studies develop complex Markovian models that require a numerical resolution in order to evaluate the system performance (see e.g. [Fang and Ghosal, 2003], [Lindemann and Thümmler, 2003], [Vornefeld, 2002], [Foh *et al.*, 2001]). Some of them use approximations to derive closed-form expressions (see e.g. [Ni and Häggman, 1999] and [Pedraza *et al.*, 2002]). However, these studies always focus on a single cell. In [Baynat and Eisenmann, 2004] we have developed a discrete-time Markov chain model for single-cell GPRS/EDGE network engineering. The model captures the main features of the GPRS/EDGE radio resource allocation and assumes an ON/OFF traffic (with infinite sessions) performed by a finite number of users over the cell. The Markov chain is simplified by Taylor series expansion and a simple and accurate Erlang-like law is obtained. Extensions to finite-length sessions traffic has been developed in [Baynat *et al.*, 2004]. In [Nogueira *et al.*, 2005] we study the impact of a capacity limitation imposed upon a group of cells. We assume it can be expressed by a maximum number of concurrent downlink transfers that are allowed in the group of cells. When this limit is reached, any transfer request on any of the cells will be rejected. However, even if the performance parameters can be obtained almost instantaneously with a very good accuracy, [Nogueira *et al.*, 2005] does not provide closed-form expressions. The goal of this paper is to further simplify the performance evaluation of a multiple-cell system, by deriving very simple asymptotic developments. Once again our objective is to obtain closed-form ÒErlang-likeÓ expressions to efficiently help network engineering.

Section 2 addresses single-cell systems, the basic hypotheses and the main results of the Erlang-like model developed in [Baynat and Eisenmann, 2004] are recalled. Then asymptotic developments are given for both low and high load cases. Section 3 deals with multiple-cell systems. In this section we quickly recall the principal steps developed in [Nogueira *et al.*, 2005] for multiple-cell systems. Asymptotic developments are then made to obtain closed-form expressions for the performance parameters. Section 4 presents numerical results.

2 Single cell system

2.1 System description

Our study is focused on the analysis of the bottleneck i.e. the radio link, studied in a particular cell. We are focused on the downlink, assumed to be the critical resource because of traffic asymmetry. We assume that the allocator fairly shares bandwidth between all active mobiles (no QoS is modeled so far).

We also make the following assumptions: there is a fixed number T of time-slots in the cell that are dedicated to GPRS; these time-slots are using a single TDMA. All mobiles have the same reception capability; they are "$(d+u)$" where d is the number of time-slots that can simultaneously be used for the downlink traffic and u is the number of time-slots that can simultaneously be used for the uplink traffic. Note that, as we are only interested in the modeling of the downlink traffic, only the parameter d is relevant for the model. This assumption is presently realistic as most of the time, less than four time-slots are reserved for GPRS ($T \leq 4$). Extensions to the case where $T > d$ are currently under investigation. Note that with this assumption, the parameter $n_0 = \lfloor \frac{T}{d} \rfloor$ used in [Baynat and Eisenmann, 2004] is such that $n_0 \leq 1$.

Our GPRS system is characterized by the following parameters:

- t_B: the system elementary time interval equal to the radio block duration, i.e. $t_B = 20$ ms;
- x_B: number of data bytes that are transferred during t_B over one time-slot. x_B/t_B is the throughput offered by the RLC/MAC layer to the LLC layer. The value of x_B depends on the radio coding scheme [Baynat and Eisenmann, 2004]. As an example, for GPRS CS2, $x_B = 30$ bytes;
- tbf_{max}: maximum number of mobiles that can simultaneously have an active downlink TBF (Temporary Block Flow). This limitation is due to the system hardware characteristics and ensures a minimum throughput per mobile (a TDMA can't be indefinitely shared).

2.2 Markovian analysis

Traffic is modeled as follows. There is a fixed number N of GPRS/EDGE mobiles in the cell, each of them doing an ON/OFF traffic with an infinite number of pages:

- ON periods correspond to the download of an element (a WAP, a WEB page, an email, a file, etc.). Its size is characterized by a discrete random variable X_{on}, with an average value of x_{on} bytes;
- OFF periods correspond to the reading time, which is modeled as a continuous random variable T_{off}, with an average value of t_{off} seconds.

Let us emphasize that there is a limitation $n_{max} = \min(tbf_{max}, N)$ on the number of mobiles that can simultaneously be on active transfer. It involves both the system constraint tbf_{max} and the total mobile population.

[Baynat and Eisenmann, 2004] develops an analytical model for the performance evaluation of a single-cell GPRS system. The smallest time-scale of the system, namely the radio block duration $t_B = 20$ ms, has been accounted for in the modeling process, by developing a discrete-time Markovian model of equal elementary time interval. The model assumes that both the size of ON periods and the duration of OFF periods have memoryless distributions.

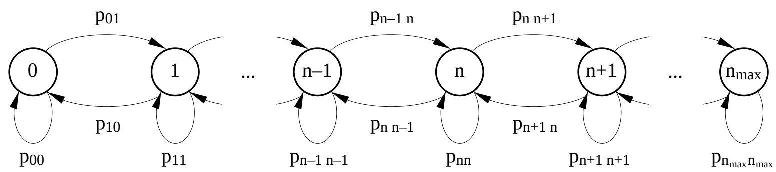

Fig. 1. Linear model

Several models with several levels of approximation have been developed. The simpler one makes the assumption that more than one mobile switching from one state (ON or OFF) to the other one during t_B is negligible, which transforms the model into the discrete-time birth-death process given in Fig. 1. As shown in [Baynat and Eisenmann, 2004], the stationary probabilities of having n mobiles in active transfer in the cell can easily be derived from this linear Markov chain. By further using Taylor series expansions, these probabilities can be expressed as a function of a single dimensionless parameter x as:

$$p(n) = \frac{N!}{T^n (N-n)!} x^n p(0) \qquad 0 \leq n \leq n_{max} \tag{1}$$

where x is given by

$$x = \frac{t_B \, x_{on}}{x_B \, t_{off}} \tag{2}$$

and $p(0)$ is obtained by normalization. Note that relation (1) is a simplification of the expression given in [Baynat and Eisenmann, 2004] that takes into account the fact that $n_0 \leq 1$.

The performance parameters of the cell can be derived from the stationary probabilities as follows [Baynat and Eisenmann, 2004]. The Ònormalized utilizationÓ $\tilde{U}$ of the cell, i.e. the mean number of time-slots used for GPRS, is directly obtained as:

$$\tilde{U} = T \sum_{n=1}^{n_{max}} p(n) = T\,(1 - p(0)) \tag{3}$$

The Ònormalized throughputÓ $\tilde{X}$, i.e. the average number of time-slots given to a mobile for its transfers is given by:

$$\tilde{X} = T \frac{\sum_{n=1}^{n_{max}} p(n)}{\sum_{n=1}^{n_{max}} n\, p(n)} \tag{4}$$

Finally, the ÒblockingÓ (or ÒrejectÓ) probability P_r, i.e. the probability that a mobile that wants to start the download of a new page cannot do it because the limit of n_{max} mobiles in the cell is reached, is obtained as:

$$P_r = 1 - \frac{T}{x} \frac{\sum_{n=1}^{n_{max}} p(n)}{\sum_{n=0}^{n_{max}} p(n)(N-n)} \tag{5}$$

2.3 Asymptotic analysis

We first define the quantity $v = x/T$. We then rewrite the expression of the steady-state probabilities and the performance parameters (relations (1) to (5)) as a function of v:

$$p(n) = \frac{v^n N(N-1)(N-n+1)}{1 + vN + v^2 N(N-1) + ... + v^{n_{max}} N(N-1)...(N-n_{max}+1)} \tag{6}$$

$$\tilde{U} = T \frac{vN + v^2 N(N-1) + ... + v^{n_{max}} N(N-1)...(N-n_{max}+1)}{1 + vN + v^2 N(N-1) + ... + v^{n_{max}} N(N-1)...(N-n_{max}+1)} \tag{7}$$

$$\tilde{X} = T \frac{1 + v(N-1) + ... + v^{n_{max}-1}(N-1)...(N-n_{max}+1)}{1 + 2v(N-1) + ... + n_{max} v^{n_{max}-1}(N-1)...(N-n_{max}+1)} \tag{8}$$

$$P_r = \frac{v^{n_{max}}(N-1)...(N-n_{max})}{1 + v(N-1) + v^2(N-1)(N-2) + ... + v^{n_{max}}(N-1)...(N-n_{max})} \tag{9}$$

From these expressions, we can easily obtain the asymptotes for low and high load. Note that the quantity vN characterizes in an aggregate way the load of the system, as it increases when the size of the downloaded pages or the number of mobiles increase, and decreases when the number of time-slots dedicated to GPRS or the reading time of a page increase. It is in fact the equivalent to the Erlang load factor for a finite number of users doing ON/OFF sessions in data traffic.

When $vN \ll 1$ (low load), we directly obtain from the previous performance expressions the following asymptotic developments:

$$\tilde{U} \approx TvN = xN; \qquad \tilde{X} \approx T(1 - v(N-1)); \qquad P_r \approx 0 \tag{10}$$

When $vN \gg 1$ (high load), we get the following asymptotic developments:

$$\tilde{U} \approx T; \qquad \tilde{X} \approx \frac{T}{n_{max}}; \qquad P_r \approx 1 - \frac{1}{v(N - n_{max})} \tag{11}$$

3 Multiple cell system

3.1 System description

We now assume, as described in Section 1 and [Nogueira *et al.*, 2005], that traffic may be limited because of a capacity constraint in a network element that controls traffic over a group of P cells. Let M_{max} be the total number of mobiles that can currently be in active transfer in the P cells. In order to simplify the derivation of the asymptotes, we assume that the P cells are identical. In each cell, there is a fixed number N of GPRS mobiles generating an ON/OFF traffic as described in Section 3.2 and having the same characteristics (the average page size is x_{on} and the average reading time is t_{off} for all mobiles in all cells). Note however that non identical cells can be handled by the model [Nogueira *et al.*, 2005]. Of course, if $\sum_{i=1}^{P} n_{max}^{i} \leq M_{max}$, the limit does not generate any additional constraint over the system, and each cell can thus be analyzed using the single-cell model described in Section 2. As a consequence, we only consider here the case where $\sum_{i=1}^{P} n_{max}^{i} > M_{max}$. In such a system, a mobile in a given cell i will not be able to start a new transfer, not only because the cell capacity (n_{max}^{i}) is reached, but also because the global system capacity (M_{max}) is reached.

3.2 Model description

We consider a particular cell of the system. As all cells are identical, without loss of generality we can focus on the last one, cell P. When needed, a superscript i will be added to the notations to refer to the parameters of a cell i. The first step of the analysis consists in developing the so-called Òaggregate Markov chainÓ [Nogueira *et al.*, 2005] associated with the considered cell (cell P). As shown in Fig. 2 This aggregate model has the same structure as the single-cell linear Markov chain model (Fig. 1), but the transition between any state n and state $n+1$ is now multiplied by a factor $(1 - r(n))$. $r(n)$ is the probability that an inactive mobile in cell P that wants to start a new transfer cannot do it because the system limit M_{max} is reached, assuming that there are n mobiles currently in active transfer in the cell.

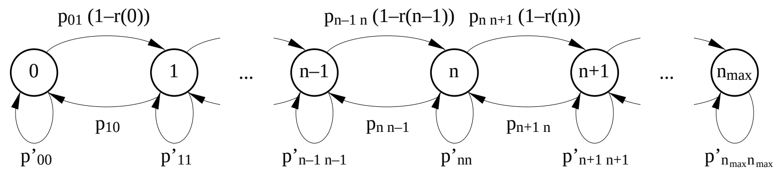

Fig. 2. Aggregate model

As a consequence, $r(n)$ is the probability that the system is full when there are n mobiles in the considered cell, and can thus be estimated by the probability that the $P-1$ other cells (cells 1 to $P-1$) contain $M_{max} - n$ mobiles. It is shown in [Nogueira *et al.*, 2005] that the probabilities $r(n)$ can be estimated by the following expression:

$$r(n) = \frac{p_{uc}^{\{1,\ldots,P-1\}}(M_{max} - n)}{\displaystyle\sum_{k=0}^{M_{max}-n} p_{uc}^{\{1,\ldots,P-1\}}(k)} \tag{12}$$

where the probabilities $p_{uc}^{\{1,...,P-1\}}(k)$ are obtained by convolution over the steady-state probabilities of the $P-1$ other "*unconstrained*" cells, where unconstrained cell i is a virtual cell having the same characteristics as cell i but that is not subjected to the overall constraint M_{max} (see [Nogueira *et al.*, 2005] for derivations):

$$p_{uc}^{\{1,...,P-1\}}(k) = \sum_{\substack{n^1+...+n^{P-1}=k \\ n^j \le n_{max} \\ \forall j=1,...,P-1}} \left(\prod_{j \in \{1,...,P-1\}} p_{uc}^j(n^j) \right) \tag{13}$$

We can then inject the $r(n)$ parameters in the aggregate model and analyze it. The resulting steady-state probabilities of the aggregate model are thus:

$$p_{agg}(n) = \frac{N!}{T^n(N-n)!} x^n \left(\prod_{k=0}^{n-1} (1-r(k)) \right) p_{agg}(0) \qquad 0 \le n \le n_{max} \tag{14}$$

where x is given by (2) and $p_{agg}(0)$ is obtained by normalization.

Finally, we can derive the normalized utilization $\tilde{U}$ of any cell as well as the normalized throughput $\tilde{X}$ offered to a mobile for its transfers and the blocking probability P_r, from relations (3), (4) and (5), by replacing the probabilities $p(n)$ by the aggregate probabilities $p_{agg}(n)$ obtained from relation (14).

3.3 Asymptotic analysis

In this section we are only interested in the high load case ($vN \gg 1$ and $P \gg 1$). Indeed, in the low load case ($vN \ll 1$), the system constraint does not affect the behavior of the system, and the asymptotes are those of the single-cell case developed in Section 2.3.
We first develop the expression of the probability $p_{uc}^{\{1,...,P-1\}}(k)$:

$$p_{uc}^{\{1,...,P-1\}}(k) = \Psi(k) \left(p_{uc}^i(0) \right)^{P-1} \tag{15}$$

where $\Psi(k)$ comes from relation (13):

$$\begin{aligned} \Psi(k) &= \sum_{\substack{n^1+...+n^{P-1}=k \\ n^j \le n_{max} \\ \forall j=1,...,P-1}} \left(v^{n^j} N...(N-n^j+1) \right) ... \left(v^{n^{P-1}} N...(N-n^{P-1}+1) \right) \\ &= \sum_{\substack{n^1+...+n^{P-1}=k \\ n^j \le n_{max} \\ \forall j=1,...,P-1}} v^k \left(N...(N-n^j+1) \right) ... \left(N...(N-n^{P-1}+1) \right) \end{aligned} \tag{16}$$

When $vN \gg 1$, we can give the following first order approximation for $p_{uc}^{\{1,\ldots,P-1\}}(k)$:

$$p_{uc}^{\{1,\ldots,P-1\}}(k) \approx \mathrm{C}_{k+P-2}^{k}(vN)^k \left(p_{uc}^{i}(0)\right)^{P-1} \tag{17}$$

Even if it is not exact, we have empirically checked that when this expression is replaced into the $r(n)$ probabilities, it results in a very good approximation:

$$r(n) = \frac{p_{uc}^{\{1,\ldots,P-1\}}(M_{max}-n)}{\sum_{j=0}^{M_{max}-n} p_{uc}^{\{1,\ldots,P-1\}}(j)} \approx \frac{\mathrm{C}_{M_{max}-n+P-2}^{M_{max}-n}(vN)^{M_{max}-n}}{\sum_{j=0}^{M_{max}-n} \mathrm{C}_{j+P-2}^{j}(vN)^j} \tag{18}$$

By replacing the development of the $r(n)$ probabilities into the aggregate model, we get the aggregate steady-state probabilities. Here we only give the expression of $p_{agg}(1)$ and $p_{agg}(2)$:

$$p_{agg}(1) \approx \frac{M_{max}}{M_{max}+P-2} p_{agg}(0); \qquad p_{agg}(2) \approx \frac{N-1}{N} \frac{(M_{max}-1)}{(M_{max}+P-3)} p_{agg}(1) \tag{19}$$

When $P \gg 1$, it appears that the probabilities $p_{agg}(n)$ decrease very fast with n: $p_{agg}(0) \gg p_{agg}(1) \gg p_{agg}(2) \gg \ldots$. As a consequence we can obtain the asymptotic developments for the performance parameters by only taking into account the preponderant values of the aggregate probabilities in the summation of expressions (3), (4) and (5). By doing that, we obtain the following expressions for the performance parameters of any cell i:

$$\tilde{U}^i \approx T p_{agg}(1) \approx T \frac{M_{max}}{M_{max}+P-2} \tag{20}$$

$$\tilde{X}^i \approx T \frac{1+\frac{p_{agg}(2)}{p_{agg}(1)}}{1+2\frac{p_{agg}(2)}{p_{agg}(1)}} \approx T\left(1 - \frac{N-1}{N}\frac{(M_{max}-1)}{(M_{max}+P-3)}\right) \tag{21}$$

$$P_r^i \approx 1 - \frac{p_{agg}(1)}{vN p_{agg}(0)} \approx 1 - \frac{M_{max}}{vN(M_{max}+P-2)} \tag{22}$$

4 Numerical Results

In this section, we compare the asymptotic developments to the results obtained with the analytical models developed in [Baynat and Eisenmann, 2004] and [Nogueira *et al.*, 2005]. The constraint M_{max} is set to 40. The number T of time-slots reserved to GPRS traffic in each cell is chosen equal to 2. We have tested different values for T (1 to 4) and M_{max} (20 to 100), and the results obtained were very similar to those shown here. All mobiles are assumed to be able to use the T time-slots of the TDMA ($d \geq T$) and generate the same traffic load.

4.1 Influence of the mobile population per cell

First, we investigate the influence of the GPRS mobile population. We set the number of cells to $P = 30$, the dimensionless parameter to $x = 0.268$ (corresponding e.g. to $x_{on} = 4000$ bytes, $t_{off} = 10$ s and $x_B = 30$ bytes), and vary the number N of GPRS mobiles in each cell. We compare the normalized utilization $\tilde{U}$ of a cell, the normalized throughput $\tilde{X}$ offered to a mobile and the blocking probability P_r obtained by the analytical model to those derived from the asymptotic developments. As shown in Fig. 3, 4 and 5, the asymptotic curves are made of two parts. First, when N is low, the system is almost empty, i.e. the bottleneck is not reached. Every cell has the same behavior as if it were alone in the system. Thus, we use the asymptotic expressions (10) developed for low traffic load in single cells. Second, when N is high, the system is saturated because of the important traffic load. Thus, we use the asymptotic expressions (20), (21) and (22) developed for high traffic load in a multiple-cell system. Fig. 3, 4 and 5 show the very good fit between the asymptotic and the analytical curves, for both low and high traffic load. In both cases, the cells performance reach the asymptotic limit quickly. These expressions given by two simple functions are very useful to quickly analyze the qualitative behavior and quantitative bounds on the system performance.

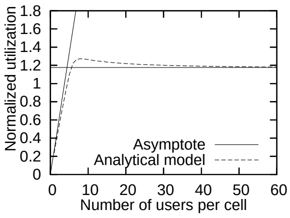

Fig. 3. $\tilde{U}$ function of N

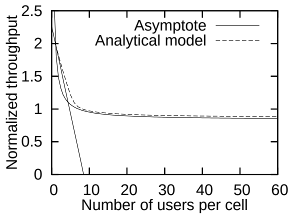

Fig. 4. $\tilde{X}$ function of N

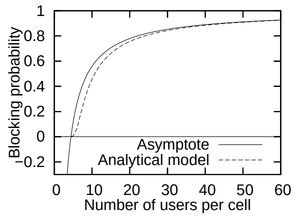

Fig. 5. P_r function of N

4.2 Influence of the number of cells

We now focus on the influence of the number P of cells in the system. We study the system performance evolution for different traffic load profiles ($vN = 0.13, 0.27, 0.54, 1.07$). The asymptotic curves are obtained from relations (20), (21) and (22). We notice a systematic behavior on analytical utilization and throughput curves (Fig. 6 and 7). They nearly follow a horizontal line until they reach the asymptotic curve. The horizontal line corresponds to the performance parameter of the cell that is not subject to the capacity constraint and that can thus be analyzed with the single-cell model of Section 2. Performance remains unchanged when P increases until the capacity constraint starts influencing the cell. We can thus cut out the construction of the performance curves as follows:

i) compute the reference performance parameter for a single-cell system;

ii) draw the asymptote for high traffic load in multiple-cell system;
iii) bind the reference point to the asymptote with a horizontal line.

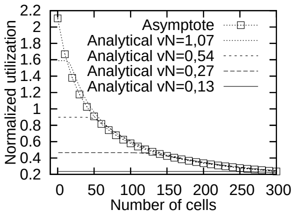

Fig. 6. $\tilde{U}$ function of P

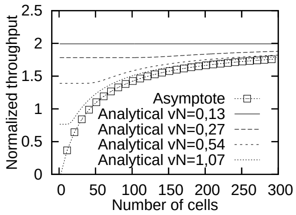

Fig. 7. $\tilde{X}$ function of P

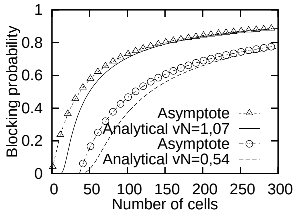

Fig. 8. P_r function of P

5 Conclusion

We have first been able to provide a computationally simple model of the a priori complex system made of a group of cells in a cellular network coupled by capacity limitation in a centralized equipment handlink packet traffic. This model has been further simplified by developing asymptotic expansions for low and high load traffic. The resulting close-form Erlang-like expressions that have been derived allow the construction of even simpler dimensioning models. Indeed, in a dimensioning situation, the problem consists in finding the optimal input system parameter that fulfill a given performance criterion. Our proposal offers simple functions that can easily be inverted in order to obtain directly the required solution without any iteration process. The complexity of such problems is thus drastically reduced.

References

[Baynat and Eisenmann, 2004]B. Baynat and P. Eisenmann. "Towards an Erlang-Like formula for GPRS/EDGE network engineering". In *IEEE International Conference on Communications (ICC)*, 2004.

[Baynat *et al.*, 2004]B. Baynat, K Boussetta, P. Eisenmann, and N. Ben Rached. "Towards an Erlang-Like formula for the performance evaluation of GPRS/EDGE networks with finite-length sessions". In *3rd IFIP-TC6 Networking Conference*, 2004.

[Fang and Ghosal, 2003]Fang and D. Ghosal. "Performance Modeling and QoS Evaluation of MAC/RLC Layer in GSM/GPRS Networks". In *Proc. of IEEE International Conference on Communications*, 2003.

[Foh *et al.*, 2001]C. H. Foh, B. Meini, B. Wydrowski, and M. Zukerman. "Modeling and Performance Evaluation of GPRS". In *Proc. of IEEE VTC*, 2001.

[Lindemann and Thümmler, 2003]Ch. Lindemann and A. Thümmler. "Performance Analysis of the General Packet Radio Service". In *Computer Networks*, pages 1–17, 2003.

[Ni and Häggman, 1999]S. Ni and S. Häggman. "GPRS performance estimation in GSM voice and GPRS shared resource system". In *Proc. of IEEE Wireless Communication and Networking Conference(WCNC)*, pages 1417–1421, 1999.

[Nogueira *et al.*, 2005]G. Nogueira, B. Baynat, and P. Eisenmann. "An analytical model for the dimensioning of a GPRS/EDGE network with a capacity constraint on group of cells". *Unpublished, Technical report in progress*, 2005.

[Pedraza *et al.*, 2002]S. Pedraza, J. Romero, and J. Muoz. "(E)GPRS Hardware Dimensioning Rules with Minimum Quality Criteria". In *Vehicular Technology Conference*, pages 391–395, 2002.

[Vornefeld, 2002]U. Vornefeld. "Analytical Performance Evaluation of Mobile Internet Access via GPRS Networks". In *European Wireless*, 2002.