Towards an Erlang-Like Formula for the Performance Evaluation of GPRS/EDGE Networks with Finite-Length Sessions

Bruno Baynat[1], Khaled Boussetta[1], Pierre Eisenmann[2], and
Nidham Ben Rached[2]

[1] Université de Paris VI, Laboratoire d'Informatique de Paris 6, Paris, France
{Bruno.Baynat, Khaled.Boussetta}@rp.lip6.fr
[2] Nortel Networks, Wireless Network Engineering, Chateaufort, France
{pierree, nbenrac}@nortelnetworks.com

Abstract. In this paper, we focus on a simple yet important objective: to provide Erlang-like models for GPRS/EDGE radio engineering. To this end, we develop a constructive model based on an original discrete-time Markov process that captures the detailed behavior of the radio allocator when a finite number of users are performing ON/OFF sessions in the cell. After a preliminary focus on infinite-length sessions, which provide a simpler analytical framework, we include finite-length sessions and study their specific effects.

1 Introduction

Several papers have been published on performance evaluation of GPRS or EDGE systems. A number of them were based on simulations (see e.g. [8,6]). Performance evaluation using analytical models were proposed in [7,5]. In these papers, GPRS session duration was supposed to be infinite. Finite GPRS sessions were investigated in [4,2,3]. All these studies rely on a number of relevant hypotheses and modeling assumptions, and quite usefully contribute to the understanding of the GPRS/ EDGE system. However none of them fully allowed us to reach our goal which was to develop an Erlang-like law for GPRS/EDGE network engineering relying on the most possible accurate description of existing GPRS infrastructure. So, we have developed a model based on an original constructive description of GPRS/EDGE with a discrete-time Markovian model at the granularity of the GPRS radio block (note that all previously mentioned works have developed continuous-time models). This first model is presented in [1]. The original contribution of our work in [1] is the detailed description of the GPRS/EDGE system in a discrete-time model which has allowed the derivation of a formula that has the simplicity of Erlang's B and C laws for voice but which applies to GPRS/EDGE.

In this paper, we extend our work presented in [1] by considering finite-length sessions. Our aim is to obtain, as in [1], closed form expressions of system performance parameters that could easily be used for GPRS/EDGE engineering, with a same focus on capturing the detailed behavior of the GPRS/EDGE system.

2   Model Description

Our study is focused on the analysis of the bottleneck i.e. the radio downlink, studied in a particular cell. Our GPRS/EDGE system is characterized by the following parameters:

- T: the fixed number of time-slots in the cell that are dedicated to GPRS.
- t_B: the radio block duration which is equal to 20 ms.
- x_B: the number of data bytes that are transferred during t_B. The value of x_B depends on the radio coding scheme (e.g. in GPRS CS2, $x_B = 30$ bytes).
- d: the maximum number of time-slots that can be simultaneously used by a mobile for the downlink traffic (all mobiles have same d value). Today, most mobiles have $d = 3$ or 4.
- tbf_{max}: the maximum number of mobiles that can simultaneously have an active downlink TBF, because of the GPRS system limitations.

We also assume that there is a fixed number M of GPRS mobiles on the cell. Their traffic is made of finite-length sessions: each mobile alternates between Session and Inter-session periods. During an Inter-session, a mobile doesn't generate any traffic while during a Session, a mobile generates an ON/OFF traffic:

- ON periods correspond to the download of an element (e.g. a www page, an email, a file, etc.). Its size in bytes is characterized by a discrete random variable X_{on}, with an average value of x_{on}. Note that the time-scale of the system requires the ON characterization to be discretized. Therefore, the X_{on} bytes of an ON period are divided into N_{on} blocks of x_B bytes each. In this study N_{on} is assumed to be geometric with parameter p, where p is the probability for a given block to be the last one of the considered packet. According to this modeling assumption[1], $p = 1 - e^{-\frac{x_B}{x_{on}}}$.
- OFF periods correspond to the reading time, which is modeled as a continuous random variable T_{off}, with an average value of t_{off}. The T_{off} seconds of an OFF period are divided into B_{off} units of t_B seconds. B_{off} is also assumed to be geometric with parameter q, where q is the probability for the current unit of time to be the last one of the OFF period. By equating the average OFF time, we obtain the value of $q = 1 - e^{-\frac{t_B}{t_{off}}}$.

Let us denote by N the number of mobiles in active session. N lies between 0 and M. Let us finally emphasize that there is a limitation $n_{max} = min(tbf_{max}, M)$ on the number of mobiles that can simultaneously be on active transfer in the cell. We chose to describe the succession of Session and Inter-session periods for each mobile independently from each others as follows:

- A Session period is characterized by its number of ON periods. For instance, for a WAP or a WEB traffic, each ON period corresponds to the download of a given page. This number is described by a discrete random variable

---

[1] The memory-less hypotheses on the traffic characteristics are strong assumptions that are discussed and validated in details in [1].

N_{pages}, with an average value of n_{pages}. In this study, N_{pages} is assumed to be geometric with parameter v, where $v = n_{pages}^{-1}$ is the probability for a given ON period to be the last one of the considered session.

– An Inter-session period is modeled by a continuous random variable T_{is}, with an average value of t_{is}. Again, the number of units of t_B seconds in an Inter-session period is assumed to be geometric with parameter r, where $r = 1 - e^{-\frac{t_B}{t_{is}}}$ is the probability for the current unit of time to be the last one of the Inter-session.

Finally, we also suppose that the allocator fairly allocates bandwidth between all active mobiles (no QoS is modeled so far). Thus, if there are n mobiles in an active downlink process, each one receives an average bandwidth equal to the total cell bandwidth divided by n as soon as its capacity (given by d) allows it.

3   The Infinite-Length Session Markovian Model

In this section, we first recall the simpler case of infinite-length sessions [1]. Every mobile in the cell is thus doing an ON/OFF traffic with an infinite number of pages (i.e. $t_{is} = 0$ or $n_{pages} = +\infty$). Therefore, $N = M$ anytime.

3.1   The Erlang-Like Model

The smallest time-scale of the system, namely t_B, will be accounted for in the modeling process, by associating with it a discrete-time Markovian model of equal time step. Thus, the state of the system will only be described at the end of each radio block period. As we assume no QoS and a fair allocation among all mobiles, we choose to describe the state of the model by the number n of mobiles in active transfer. Since there is a physical limitation n_{max} on n then the state space is given by the set $\{0, n_{max}\}$. We then define the following probabilities:

– $a_i^n =$ probability of i arrivals[2], provided that the state of the system is n;
– $d_j^n =$ probability of j departures[3], provided that the state of the system is n.

Considering the fact that there are at least two orders of magnitude between the time scales of the block duration (20 ms) and the ON/OFF duration (seconds), we make the assumption that between two steps of the process (i.e. 20 ms), there can be at most one arrival and one departure. The probabilities a_i^n and d_j^n can thus only take non zero values when i or j are equal to 0 and 1. In other words, $a_0^n + a_1^n = 1$ and $d_0^n + d_1^n = 1$.

In order to derive departure probabilities remember that each mobile can use a maximum of d time-slots in downlink. As a consequence if $nd < T$, all the mobiles in active transfer can get the maximum capacity with regard to their downlink capability. Otherwise, all the T time-slots of the cell are used, and each mobile receives a reduced downloading bandwidth. Let us also recall that, if n

---

[2] i.e. i new mobiles among $N - n$ ones entering an ON period
[3] i.e. j mobiles among the n ones in active transfer that ends up their download.

mobiles are currently in an ON period, $N - n$ are in an OFF period. Moreover, under realistic conditions $q \ll 1$ and $p \ll 1$. Therefore [1],

$$\forall n \in \{0, n_{max}\}, \; a_0^n \approx 1 - (N - n)\, q \;\text{ and }\; d_0^n \approx 1 - \min(nd, T)\, p \qquad (1)$$

According to these assumptions, the resulting Markov chain is linear and its transition probabilities can be calculated as:

$$
\begin{array}{ll}
p_{n\,n+1} = a_1^n d_0^n & \text{if } 0 \leq n < n_{max} \\
p_{n\,n-1} = a_0^n d_1^n & \text{if } 0 < n \leq n_{max} \\
p_{n\,n} = 1 - p_{n\,n+1} - p_{n\,n-1} &
\end{array}
\qquad (2)
$$

Because of the birth-death structure of the Markov chain the steady-state probabilities $p(n)$, $\forall n \in \{0, n_{max}\}$, are given [1] by the following closed form:

$$
p(n) = \begin{cases}
\dfrac{N!}{n!\, d^n\, (N-n)!} \left(\dfrac{q}{p}\right)^n p(0) & \text{if } 0 \leq n \leq n_0 \\[3mm]
\dfrac{N!}{n_0!\, d^{n_0}\, T^{n-n_0}\, (N-n)!} \left(\dfrac{q}{p}\right)^n p(0) & \text{if } n_0 < n \leq n_{max}
\end{cases}
\qquad (3)
$$

where $p(0)$ is obtained by normalization. The expression 3 has the simplicity of Erlang's B formula yet it applies to GPRS with two quality criteria, blocking and throughput. It has the same attributes: it provides generic reference curves and fruitful qualitative insight on the dimensioning of GPRS system.

3.2 Performance Evaluation

From the stationnary probabilities given by equation 3, we can compute the average utilization $\overline{U}$ of the cell as follow:

$$\overline{U} = \sum_{n=0}^{n_{max}} p(n) \min\left(\frac{nd}{T}, 1\right) \qquad (4)$$

Furthermore, we can deduce from the Markov chain the mean number of mobiles ending up their transfers per unit of time as well as the mean number of mobiles in active transfer. From Little's law, we can thus derive the average duration of an active transfer. Then, we can calculate the average throughput $\overline{X}$ obtained by each mobile in active transfer as

$$\overline{X} = \frac{x_{on} \sum_{n=1}^{n_{max}} p(n)\, d_1^n}{t_B \sum_{n=1}^{n_{max}} n p(n)} \qquad (5)$$

Finally, we can derive the last important performance parameter namely the so-called "blocking" (or "reject") probability P_r. P_r is the probability that an arrival is rejected, *i.e.* that a mobile that wants to switch to an ON state cannot do it because the limit of n_{max} mobiles in the cell is reached. As shown in [1],

$$P_r = \frac{\sum_{n=0}^{n_{max}} p(n)\, (a_1^n - d_1^n)}{\sum_{n=0}^{n_{max}} p(n)\, a_1^n} \qquad (6)$$

4   Erlang-Like Law for Finite-Length Sessions

We now come back to the finite-length sessions model as described in Section 2. Focusing on our goal to obtain a closed-form Erlang-like formula, we transform our finite-length sessions traffic into an "equivalent" infinite-length sessions traffic as follows: the inter-session period is "shared" among the inter-page OFF periods of an approximate infinite-length sessions model. So the equivalent OFF period duration is given by 7, all the other system parameters being equal.

$$t'_{off} = t_{off} + \frac{t_{is}}{n_{pages}} = t_{off} + v\, t_{is} \tag{7}$$

This is obviously an approximation as the distribution of the OFF period in the equivalent linear model are geometric and will not respect the variability of the actual succession of OFF and OFF+Inter-session. However, assuming the approximation holds, all the Erlang-like formulas of Sections 3.1 and 3.2 can be applied by performing the above transform on t_{off}. We will see in the next section that this approximation results in a very accurate estimation of all the average performance parameters.
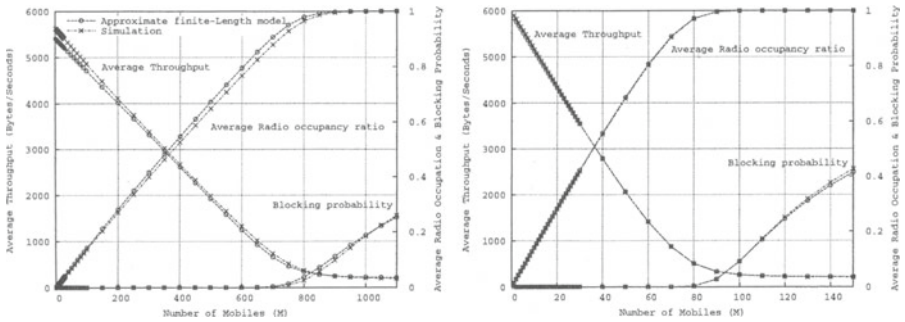
Fig. 1. Validation of Erlang-like law for finite-length sessions against simulations

4.1   Validation

We have compared the performance parameters obtained under the Erlang-like law for finite-length sessions against those obtained by simulations. To that end we have considered several test scenarios. In a first set of tests, we have considered the following parameters: $T = d = 4$ slots, $t_{off} = 7$ seconds, $t_{is} = 20$ minutes and $n_{pages} = 10$. Curves in the left side of figure 1 show average throughput, average radio occupation ratio and blocking probability obtained by both the Erlang-Like finite-length sessions model and simulations when $x_{on} = 1$ kB and M varies form 1 until 1100 mobiles. These curves show clearly that the values obtained by the approximate finite-length sessions model are extremely close to those obtained by simulations. In fact in these test scenarios the error remains less than 3%. In other comparison test scenarios we have observed that the error gets even smaller when x_{on} grows or when t_{is} decreases. We also have investigated the

influence of the average number of pages to be transmitted during a session on the equivalence between the finite-length sessions model and the equivalent infinite one. We ran the set of tests previously described while taking now n_{pages} = 3 and t_{is} = 2 minutes. The curves in the right side of figure 1 illustrate the average performance parameters obtained for x_{on} = 4 kB. These curves confirm the same observations we made before. They indicates that even for a small inter-session period of 2 minutes and a modest average number of transmitted pages set to 3 a good fit is obtained between simulation and the approximate finite-length sessions model. Other similar results[4], allows us to conclude that an Erlang-like law can be utilized for GPRS/EDGE network engineering with finite-length sessions traffic models. A framework quite similar to the standard voice dimensioning methods is thus available.

5   Conclusion

In this paper, we have developed an Erlang-like model for GPRS/EDGE network dimensioning with finite-length sessions. To that end, we have first derived a simple Erlang-like formula for the infinite-length sessions case. Then we have included finite-length sessions in our model. The simulation results demonstrated that finite-length sessions models could also be solved by a simple Erlang-like law with a very good accuracy: this is achieved by performing the appropriate transform in the input traffic, namely by increasing the OFF periods by a portion of the inter-session period.

References

1. B. Baynat and P. Eisenmann. Towards an Erlang-like formula for GPRS/EDGE network engineering. In *Proc. of IEEE ICC*, Paris, France, June 2004.
2. B. Wydrowski C. H. Foh, B. Meini and M. Zukerman. Modeling and Performance Evaluation of GPRS. In *Proc. of IEEE VTC*, pages 2108–2112, Rhodes, Greece, May 2001.
3. X. Fang and D. Ghosal. Performance Modeling and QoS Evaluation of MAC/RLC Layer in GSM/GPRS Networks. In *Proc. of IEEE ICC*, Alaska, USA, May 2003.
4. C. Lindemann and A. Thummler. Performance Analysis of the General Packet Radio Service. *Computer Networks*, 41(1):1–17, January 2003.
5. S. Ni and S. Haggman. GPRS performance estimation in GSM voice and GPRS shared resource system. In *Proc. of IEEE WCNC*, pages 1417–1421, New Orleans, USA, September 1999.
6. M. Oliver and C. Ferrer. Overview and Capacity of the GPRS (General Packet Radio Service). In *Proc. of IEEE PIMRC*, Boston, MA, USA, June 1998.
7. J. Romero S. Pedraza and J. Munoz. (E)GPRS Hardware Dimensioning Rules with Minimum Quality Criteria. In *Proc. of IEEE VTC Spring*, pages 391–395, Birmingham, Al, May 2002.
8. P. Stuckmann and O. Paul. Dimensioning Rules for GSM/GPRS Networks. In *Proc. of the Aachen Symposium on Signal Theory*, pages 169–174, Aachen, Germany, September 2001.

---

[4] Which we can not show here due to space limitation.