Traffic Modeling and Performance Evaluation in GSM/GPRS Networks

Cornel Balint, Georgeta Budura, Marza Eugen
“Politehnica” University of Timisoara
Bd. V. Parvan. No. 2
ROMANIA
cornel.balint@etc.upt.ro, georgeta.budura@etc.upt.ro

Abstract: - The sharing of resources between different users and different services is a key concept of radio resources dimensioning in GSM/GPRS networks. In this paper we present dimensioning rules based on traffic evaluation and quality of service level for GSM/GPRS users, focused on two radio resources allocation strategies: CP (Complete Partitioning) and PP (Partial Partitioning). The quality of service is given by the blocking probability according to Erlang-B formula for voice users meanwhile for GPRS users the individual throughput and the blocking probability represent the performance parameters. The preemption probability of voice over data users is also considered.

Key-Words: - GSM, GPRS, modeling, blocking probability, Erlang-B law, Engset law, throughput

1 Introduction
The Internet popularity creates a large number of services available online. These include, World-Wide-Web browsing, E-Mail,... The integration of General Packet Radio Service and Enhanced GPRS into GSM system world wide raises many problems. For networks operators, equipment vendors and system integrators dimensioning rules have to be developed to plan and estimate the radio capacity that is needed for the predicted amount of data users when the radio resources are shared between circuit and packet switched services.

GSM operators have been dimensioning their networks for voice service in terms of offered voice traffic and blocking probability. The reference model for this system is the Erlang-B formula [1]. This formula gives the proportion of calls that are blocked as a simple function of system capacity and traffic intensity. The GPRS network is designed for supporting several types of data traffic such as Wap, Web, E-Mail, etc. Therefore GPRS traffic process characterization is very demanding. Usually the data traffic is very bursty and relies to the application. A communication session may last for an extended period of time with intermittent packet transmissions.

On the other hand the GPRS service allows dynamic allocation of bandwidth resources. Wireless channels are allocated to a mobile terminal based on its traffic demands which results in a better resource utilization. This traffic behavior coupled with flexible bandwidth allocation in a GPRS network represent the starting point for constructing an appropriate model in order to evaluate the performance and to establish dimensioning rules for the system. Another major problem of GSM/GPRS networks dimensioning is the choice of strategy to partition the available cell capacity between traditional GSM and new GPRS services. The Radio Resources Manager (RRM) is in charge of optimizing the usage of radio resources, based on a specific resource sharing algorithm. In this paper we will consider two static resources sharing schemes:
- In the first one, called Complete Partitioning (CP), time-slots (TS) are divided into two sets and each type of traffic is allowed to use only its dedicated set.
- The second scheme known as Partial Partitioning (PP), contains the following channel sets: one set shared between voice and data traffic and two sets each one being reserved for strict usage of its dedicated traffic: voice or data. This scheme, offers many advantages: first, reserving a set of time-slots for each type of traffic allows guaranteeing, as in CP a minimum QoS for each type of traffic. Second, PP scheme provide a better efficiency than CP which is not suitable for maximizing radio utilization, especially in highly varying demand. Due to these advantages PP is widely implemented in a number of actually operating GSM/GPRS networks.

Several papers have been published on traffic modeling and performance evaluation in GSM/GPRS networks. The major works in this field
are based on analytical models using queuing theory and continuous-time Markov chains, and assuming an infinite number of users in the cell [1], [2] - [5]. In [7]-[9] analytical models based on discrete-time Markov chains have been proposed and a single type of traffic (data traffic) is considered. It is assumed to be generated by a finite number of users and modeled by an Erlang-like law. Other work is based on the modified Engset model [6]. In our study we first introduce some performance parameters for GPRS networks and then we propose a dimensioning method for estimating the number of TS allocated for each traffic type.

2 System Description

Our paper considers a single cell submitted to two different types of traffic: GSM voice calls and GPRS data flows. In traditional circuit-switched GSM networks, on each frequency carrier a 200 kHz bandwidth is shared between 8 voice calls. Each voice call is given a circuit, also called time-slot (TS) because it is a Time-Division multiplexing scheme (TDMA). Each voice call needs the assignment of a single time-slot for its entire duration. GPRS data traffic uses the same radio interface as GSM voice calls hence radio resources available in the cell have to be shared among GSM and GPRS traffics. As a reminder GPRS is a packet switching technology over circuit-switching based GSM system. In GPRS technology a mobile station can use several time-slots simultaneously for one application data flow to perform its transmission with a higher throughput. Each time-slot can be shared among several users by assigning different Temporary Flow Identities (TFI) to the mobiles. Each TFI identifies a GPRS physical connection called Temporary Block Flow (TBF). Up to 32 TFI’s can be allocated per TDMA frame due to the 5 bits allocated for TFI encoding at TRX level. Data flows are multiplexed by a PCU-based scheduling algorithm. In addition to time-slot partitioning, GPRS system allows time-slot aggregation: for a single mobile user the system can allocate up to d-time-slots simultaneously for downlink and up to u-time-slots simultaneously for uplink, depending on mobile station capability class (d+u). The choice of the number of TBF’s that a PDCH can have in uplink and downlink depends on the operator’s choice. For example, in Alcatel-Lucent technology up to 6 uplink and 10 downlink TBF’s are allocated per PDCH.

Our study is focused on the radio allocator which distributes the downlink radio channels among voice calls and GPRS data flows.

We make the following assumptions regarding the system to be modeled:

- TS: the number of time-slots of the TDMA partitioned into a contiguous set of \( T_{SV} \) time-slots dedicated to voice calls, \( T_{SD} \) time-slots shared between voice and data and \( T_{ST} \) time-slots dedicated to GPRS; time-slots used by data \( T_{SD} + T_{ST} \) are on a single TDMA which has a total of 8 time-slots.

- \( d \) (resp. \( u \)): is the number of time-slots that can be used simultaneously for downlink (resp. uplink) traffic. All GPRS mobiles have the same radio capability, denoted \( d + u \).

- The RLC radio block size and the data rate according to the GPRS coding are indicated in TABLE 1.

<table>
<thead>
<tr>
<th>GPRS Schemes</th>
<th>CS-1</th>
<th>CS-2</th>
<th>CS-3</th>
<th>CS-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLC block radio (bytes)</td>
<td>23</td>
<td>33</td>
<td>39</td>
<td>53</td>
</tr>
<tr>
<td>Data rate: ( \mu_{GPRS} ) (kbits/s)</td>
<td>9.05</td>
<td>13.4</td>
<td>15.6</td>
<td>21.4</td>
</tr>
</tbody>
</table>

- Voice calls have a preemptive priority over data flows on the shared part of the TDMA due to the fact that they generate the largest amount of revenue in most actual operating systems. As a consequence, if all \( T_{SV} \) time-slots dedicated to voice are occupied and all \( T_{SD} \) time-slots are in use with at least one of them allocated to data, then one time-slot assigned to GPRS traffic in the shared part of the TDMA will be reallocated to voice on the arrival of a GSM request.

3 System Model

3.1 Cells with Complete Partitioning Strategy

Complete Partitioning strategy allows two separate sets of time-slots dedicated to voice respectively to data. As a consequence the GSM and GPRS systems can be analyzed separately.

3.1.1 Voice traffic model

The classical Markov chain model applies for voice. The steady-state voice probabilities given by relation (1) are generated by the birth-death structure of this model shown in Fig.1.
of active users in progress and represents a finite state space.

\[
\lambda_j = (N-j)\lambda_D = (N-j) \frac{1}{E[\tau]},
\]

for \( j = 0,1,\ldots,n_{\text{max}} - 1 \)

We can express the transition rate of death process as:

\[
\mu_j = \min(\sigma,TS_D)\mu_D = \min(\sigma,TS_D) \frac{\mu_{\text{GPRS}}}{E[\tau]},
\]

for \( j = 1,\ldots,n_{\text{max}} \)

As indicated in Fig.2, the state \( j \) of the Markov chain corresponds to the number of the data mobiles that are simultaneously in active transfer (in ON state). The maximum bandwidth capacity they can use is \( TS_D \). Because of the maximum downloading capacity \( d \) of each GPRS mobile, two situations can be distinguished:

1. If \( jd < TS_D \), the available bandwidth is not fully utilized by data mobiles. As a consequence the transition rate from state \( j \) to state \( j - 1 \), given by the accomplished transfer of one mobile, is

\[
\frac{\mu_{\text{GPRS}}}{E[\tau]};
\]

2. If \( jd \geq TS_D \) the allocator has to share the \( TS_D \) time-slots among the \( j \) data mobiles and the transition rate from state \( j \) to state \( j - 1 \) is

\[
\frac{\mu_{\text{GPRS}}}{E[\tau]};
\]

Let \( p_D(j) \) be the steady-state probability that \( j \) users are in active transfer. According to the Engset model it is given by the closed form here below:

\[
p_D(j) = p_D(0) \prod_{i=1}^{j-1} \min(d,TS_D) \frac{\mu_{\text{GPRS}}}{E[\tau]} \frac{1}{E[\tau]} \]

3.1.2 Data Traffic Model

Data traffic is modeled assuming that there is a fixed number \( N \) of data mobiles in the cell. Each mobile is doing an ON/OFF traffic with an infinite number of pages:

- ON periods correspond to the download of an element like a WAP, a WEB page, an email, a file, etc. Its size is characterized by a discrete random variable \( X_{\text{on}} \), with an average value \( E[X_{\text{on}}] \).
- OFF periods correspond to the reading time of the last downloaded element, which is modeled as a random variable \( T_{\text{off}} \) with an average value of \( E[T_{\text{off}}] \) seconds.
- The maximum number of GPRS users in active transfer is given by:

\[
N_{\text{max}}(TS_D) = \min(N,32,mTS_D)
\]

\( m \) is the maximum number of users that can use a single time-slot.

The system model is based on the Engset model [6] and includes particular specifications as indicated in Fig.2. This stochastic process describes the number

\[
p_r(t) = \frac{\rho_r^t}{t!} \sum_{i=0}^{\rho_r} \frac{\rho_r^i}{i!}, \quad t \in [0,TS_r]
\]

\[
B_{r,CP} = \frac{\int_{0}^{TS_r} \rho_r(t) dt}{TS_r} \sum_{i=0}^{\rho_r} \frac{\rho_r^i}{i!}
\]

Fig.1 The birth-death model applied to voice traffic process

\( \rho_r = \frac{\lambda_v}{\mu_v} \) represents the voice traffic.

Obviously the Erlang-B [1] formula gives the call blocking probability necessary to dimension the cell in order to guarantee a minimum QoS for voice traffic.

Fig.2 The Engset model applied to data traffic process

As can be seen from Fig. 2 the transition rate from state \( j \) to \( j + 1 \), \( \lambda_j \), is given by:

\[
\lambda_j = (N-j)\lambda_D = (N-j) \frac{1}{E[\tau]},
\]

for \( j = 0,1,\ldots,n_{\text{max}} - 1 \)

We can express the transition rate of death process as:

\[
\mu_j = \min(\sigma,TS_D)\mu_D = \min(\sigma,TS_D) \frac{\mu_{\text{GPRS}}}{E[\tau]},
\]

for \( j = 1,\ldots,n_{\text{max}} \)

Let \( p_D(j) \) be the steady-state probability that \( j \) users are in active transfer. According to the Engset model it is given by the closed form here below:

\[
p_D(j) = p_D(0) \prod_{i=1}^{j-1} \min(d,TS_D) \frac{\mu_{\text{GPRS}}}{E[\tau]} \frac{1}{E[\tau]}
\]
We can express the steady-state probability in terms of data traffic $\rho_D$, defined by relation (7):

$$\rho_D = \frac{E[\sigma]}{E[\tau]} \mu_{\text{GPRS}}$$  (7)

$$p_D(j) = p_D(0) \prod_{i=1}^{\min(d,TS_D)} \rho_D^j$$  (8)

As can be seen the steady-state distribution depends only through the ratio $\frac{E[\sigma]}{E[\tau]}$ on data traffic parameters $E[\sigma]$ and $E[\tau]$.

Based on this distribution we have calculated the average performances of the system. As performance measures the average downlink throughput per user ($X_u$), the average downlink total throughput ($X_T$) and the blocking probabilities ($B$) for data are considered. All these parameters are function of the load $\rho_D$, the available cell capacity $TS_D$, the user capability $d$ and the total number $N$ of users.

The average total throughput was determined using the expression below:

$$X_{CP} = \sum_{j=1}^{n_{\text{max}}} p_D(j)/r(j)$$  (9)

where $r(j)$ represents the effective bandwidth received by each user:

$$r(j) = \min(d,TS_D) \mu_{\text{GPRS}}, \text{ for } j = 1, \ldots, n_{\text{max}}$$  (10)

From formula (9) we derive the average throughput per user:

$$X_{uCP} = \frac{X_{CP}}{E[j]} = \frac{\sum_{j=1}^{n_{\text{max}}} p(j) \min(jd,TS_D)}{\sum_{j=1}^{n_{\text{max}}} \mu_{\text{GPRS}}}$$  (11)

The data blocking probability can be expressed based on Engset model as follows:

$$B_{CP} = p(0) \frac{C_{n_{\text{max}}}^{n_{\text{max}}}}{\prod_{i=1}^{\min(d,TS_D)} \rho_D^i}$$  (12)

and represents the probability that $TS_D$ time-slots are being used by $n_{\text{max}}$ users among the other $(N-1)$ users.

### 3.2 Cells with Partial Partitioning Strategy

In this case, the available time slots (TS) of the TDMA are partitioned into $TS_V$ time-slots dedicated to voice, $TS_D$ time-slots dedicated to data and $TS_{VD}$ time slots shared between voice and data with a total preemptive priority of voice over data on the shared part as mentioned before in Section 2.

$$TS = TS_V + TS_{VD} + TS_D$$  (13)

#### 3.2.1 Voice traffic model

The voice traffic process is modeled as mentioned before by relation (1). As voice has a preemptive priority over data, the blocking probability for voice can be obtained by Erlang-B formula (because data traffic is transparent to voice), with a number of resources equal to $TS_V + TS_{VD}$:

$$B_{V,PP} = \frac{\rho_V(TS_V + TS_{VD})}{(TS_V + TS_{VD})!} \sum_{i=0}^{\rho_V} \frac{\rho_V^i}{i!}$$  (14)

#### 3.2.2 Voice traffic model

The model applied to this system has to deal with two traffic processes: voice and data, sharing the same air interface and using the same physical channels.

In our previous work [10] we have implemented the bi-dimensional model proposed in [8] and have deduced the blocking probabilities formulas according to the PP strategy.

In this paper we applied the modified Engset model proposed in [6] to analyze the GPRS performances and to establish dimensioning rules. The basic idea in constructing the model relies on 2 assumptions: the voice calls are independent of GPRS connections and the voice and data traffic evolves at different time scales. The time required to transfer data is about several seconds and should be shorter than the mean call duration which is about several minutes. As a consequence between two variations of the number of voice calls, the number of data transfers reaches its stationary regime.

For voice calls the steady-state probabilities are given by relation (1) with the maximum number of available resources equal to: $TS - TS_D$. Among the $TS_{VD}$ time slots, those not used by the voice calls may be used for data traffic with a probability...
equal to the probability that $TS - TS_D - s$ are used by GSM users: $p_V ((TS - TS_D - s))$.

The data traffic model is constructed with respect to the Engset model and the particularities imposed by the resources sharing capability as presented above. The total number of available resources is given by:

$$n_{\text{max}} = \min(N, 32, mTS_D^*)$$  \hspace{1cm} (15)

We denoted by $TS_D^*$ the available cell capacity for data traffic.

The average performance parameters are determined as follows:

- the average throughput per user can be expressed as [6]:
  $$X_{u,PP} = \sum_{s=0}^{TS-2TS_D} p_V (s) X_{u,CP} (\min(TS-2TS_D, TS-s))$$  \hspace{1cm} (16)

- for the total average throughput we propose the formula:
  $$X_{PP} = \sum_{s=0}^{TS-2TS_D} p_V (s) X_{CP} (\min(TS-2TS_D, TS-s))$$  \hspace{1cm} (17)

- the data blocking probability is deduced similarly as in [6]:
  $$B_{PP} = \sum_{s=0}^{TS-2TS_D} p_V (s) B_{CP} (\min(TS-2TS_D, TS-s))$$  \hspace{1cm} (18)

- finally we propose a formula for estimating the preemption probability due to voice calls priority over data users:
  $$B_{P,PP} = \sum_{s=TS_D}^{TS-2TS_D} p_V (s) B_{CP} (\min(TS-2TS_D, TS-s))$$  \hspace{1cm} (19)

4 Experiments and results

The models have been implemented by simple programs written in Matlab and various scenarios were experimented. For the data traffic we have adopted the following parameters: $E[\sigma]=5KB$, $E[\tau]=12x$, GPRS mobile class: 4+1 and CS2 coding scheme ($\mu_{\text{GPRS}} = 13.4 \text{kbits/s}$). In the first scenario we have considered a cell with CP strategy, equipped with a single TRX that provides: $TS = 8$, $TS_V = 7$, and $TS_D = 1$.

In the second scenario we have considered a cell with PP strategy, equipped with a single TRX that provides: $TS = 8$, $TS_V = 3$, and $TS_D = 1$. We have considered three loading situations, according to different offered voice traffic values: $\rho_V = 0.602$, $\rho_V = 2.94$ and $\rho_V = 20$. The first two values correspond to the different occupancies of the time-slots that can be used by voice: 3 and 7, considering in that case a 2% for the lost voice traffic according to Erlang-B formula. The third traffic value corresponds to 7 voice TS occupancy but with a blocking probability closed to 90%.

In Fig. 3 we have plotted the total average throughputs $X_{CP}$ and $X_{PP}$, the average throughputs per user $X_{u,CP}$ and $X_{u,PP}$ as well as the blocking probabilities $B_{CP}$ and $B_{PP}$.

Fig.3 Performance parameters

The PP scheme takes advantages regarding the throughput and the blocking probability comparing with CP scheme.

The typical network dimensioning problem is to compute how many time-slots to allocate for data traffic.

Current used dimensioning rules need to estimate the number of active users per cell [11]. Our dimensioning method avoids this estimation using dimensioning criteria based on the throughput per active user combined with blocking probabilities in an interactive algorithm. Blocking probabilities in commercially GPRS networks must be small enough in order to avoid interruption caused by TBF rejection. As we can see in Fig 3 for a reasonable number of users in the cell the minimum accepted throughput is reached, that means that the limiting factor for dimensioning is the throughput rather then the blocking probability. To illustrate the proposed rule, in Fig.4 we have
plotted as a detail of Fig.3 the throughput per user and the blocking probability as function of number of GPRS active users in the cell, for two different total time-slots values. The same previous PP scenario was used for dimensioning algorithm illustration, except that the initial value of TS number was set to 7 (3 dedicated for voice, 1 dedicated for data and 3 shared between voice and data).

![Fig.4 Dimensioning algorithm illustration](image)

The A point gives the number of active users that can obtain a desired throughput value. According to this number the C point indicates the reached blocking probability. If that probability is greater than an imposed value, an increment of the number of time slots gives the new users number (point B) and the corresponding smaller blocking probability (point D). The algorithm will be repeated until the desired value for blocking probability is reached.

5 Conclusion

In our paper we have introduced a dimensioning algorithm for GSM/GPRS networks with PP strategy. The proposed algorithm is based on performance parameters as the average downlink throughput per user ($X_v$), the average downlink total throughput ($X$) and the blocking probabilities ($B$). All these parameters are function of the cell load (voice and data), the available cell capacity, the user capability $d$ and the total number of users. As quality criteria we have considered the blocking probability, according to Erlang-B law, for voice and the average downlink throughput per user for data.

References: