Analytical Comparison of Different GPRS Introduction Strategies

M. Ermel, K. Begain, T. Müller, J. Schüler, M. Schweigel
Chair for Telecommunications, Dresden University of Technology
01062 Dresden, Germany
{ermel,begain,muellert,schuelj,schweige}@ifn.et.tu-dresden.de

Abstract

The ongoing introduction of GPRS services in existing GSM networks by mobile network providers raises the question of the best strategy to partition the available cell capacity. The paper describes three different strategies – complete partitioning, partial sharing and complete sharing. An analytical call/burst level model of one cell of a homogeneous multiservice GSM/GPRS network is used to investigate these strategies with respect to important performance measures like new and handover blocking probabilities, system utilization and mean data rates of the services. Numerical results show, that the complete sharing strategy in connection with the proposed admission policy achieves a high system utilization while even for low priority data connections some minimum quality of service is guaranteed.

1 Introduction

The General Packet Radio Service (GPRS) is a new service, that has been added to the traditional Global System for Mobile Communications (GSM), which is the world’s most popular second generation digital cellular system. The evolution of GSM technology has been led by the European Telecommunications Standard Institute (ETSI) and can be divided into three phases. In Phase 1 (1992), the basic commercial services, like telephony and short message service, were introduced. Phase 2 (1996) has completed the original work and established a framework for future enhancements. In order to satisfy the increasing user requirements and to preserve competitiveness, one major concern of the GSM Phase 2+ has been the specification of GPRS, to accommodate data connections with high bandwidth efficiency.

With the introduction of GPRS in existing GSM networks, service providers are able to offer new services with a bursty traffic nature. The GPRS standard [12] provides packet based access, in which resources are efficiently shared between active users. Scarce radio resources are occupied only when necessary. Using GPRS, many users can share a radio link by adapting their individual data rates. On the other hand, higher bit rates can be achieved, when resources are available. Consequently, GPRS is supposed to be the main development trend of GSM networks towards third generation mobile systems like UMTS/IMT-2000 [13, 10, 11], which are based on Wideband CDMA.

Although the initial work on GPRS standardization has begun in 1994 there are many open questions to be studied. One of the most recent questions is how to divide the cell capacity between traditional GSM and GPRS services. There are very few experiences so that network providers can follow different introduction strategies. Based on an analytical call/burst level model and an efficient admission policy originally described in [4], this paper describes and compares different introduction strategies with respect to call and burst level performance measures like new and handover blocking probabilities, achievable average data rates and utilization of cell capacity.

From performance point of view, there are some works in this area like [5, 6]. Both papers present performance analysis using simulation. Analytical studies can be found in many publications on GSM networks (e.g. [3, 1]), but only few deal with multiple service networks and different partitioning strategies. In [1], an analytical model of one cell of a GSM network, with voice, data, and video services, all operating on the traditional circuit switching, is introduced.

2 System Description

The ETSI GSM standard [12] introduces GPRS as a GSM Phase 2+ feature for efficient packet switched data transfer. With the separate addressing of up- and downlink, bitrates up to 171.2 kbit/s per user can be achieved under good radio conditions. Specific point-to-point and point-to-multipoint services are defined and the interworking with standard packet data protocols like X.25 or IP is supported. A quality of service (QoS) profile provides tracking of service parameters within the entire Public Land Mobile Network including the radio link. According to [12] the QoS profile defines the expected quality of service in terms of the...
following attributes: Precedence Class, Delay Class, Reliability Class, Peak Throughput Class and Mean Throughput Class.

A QoS profile is associated with each Packet Data Protocol (PDP) context. A PDP context is the set of parameters related to a ‘session’ between the mobile and the Packet Data Network, containing the QoS parameters, PDP type, mobile station address, etc. A session will be established by initiating a so called PDP context activation procedure. During a session, user packet data are transferred transparently end-to-end between the mobile station and the external data network using encapsulation and tunneling methods.

The GPRS air interface, which can be assumed to be the bottleneck of the system, provides a maximum data rate of 21.4 kbit/s per occupied packet data channel (PDCH). This data rate includes overhead for the GSM physical layer (synchronization, guard band) as well as protocol overhead for the GPRS air interface, but excludes additional channel coding for forward error correction (FEC) and overhead of higher protocol layers. FEC channel coding is done adaptively to the actual measured radio condition and changes dynamically between four proposed schemes. The reachable bit rate ranges from 9.05 kbit/s (coding scheme 1) up to 21.4 kbit/s (coding scheme 4).

The adaptive FEC mechanism together with the Automatic Repeat Request (ARQ) protocol of GPRS makes it impossible to foresee the net data rates seen by the applications. This strongly depends on the current radio conditions, the geographic environment and the movement behavior of the users. Only a long term average data throughput could be applied, making more or less realistic assumptions.

For this paper, we decided to base all network related calculations on the maximum achievable data rate of one occupied PDCH under good radio conditions (21.4 kbit/s). The additional overhead for FEC, ARQ and higher layer protocols are taken into account at the service description. This allows to calculate a service dependent overhead. An appropriate QoS profile can be specified for every single service, including a reliability class, which indicates the used ARQ strategy. Also, for each application a different user behavior can be supposed.

The maximum number of PDCH a single user can occupy is 8, according to the number of timeslots per radio frequency. This sets the maximum data rate a single user can achieve to 171.2 kbit/s for unprotected data transmission, if the network resources are available.

3 Classification of Services

The performance in a GSM/GPRS system depends on the mix of services and the user behavior. However, typical services in GPRS differ from those in networks like traditional GSM, the telephone network or the Internet. The feasibility of a service depends on constraints like available bit rate and packet delay.

Possible GPRS services are listed in [5]. Among them, there are typical Internet applications like WWW, FTP, email and telnet as well as real time services like video conferencing or Road Traffic and Transport Informatics (RTTI) applications. In the future, there may be many services, which have not been defined yet. For example in the Internet, we expect new innovative services to be implemented very fast, if there is a need for them. Therefore, it is not feasible to include every possible service in our investigation. We have considered 3 service classes:

Voice service
The voice service is treated as in traditional GSM systems. Calls arrive according to a state dependent Poisson process. The call holding time is negative exponentially distributed.

Road Traffic and Transport Informatics (RTTI)
Examples for such services are route guidance, parts of transport planning applications or positioning systems. We expect such applications to become very popular in GPRS. In a circuit switched network, this would mean high overhead and costs for establishing connections for these small messages. In the packet oriented GPRS, such services could be transported with reduced costs.

We use a modified version of the ETSI railway model [5]. Messages arrive according to a Poisson process. The message size is geometrically distributed with mean 170 byte. Compared to the original model, our model does not include the maximum message size, which would have complicated the analysis.

Internet
Because the Internet is very popular, we expect the usage of GPRS also for typical Internet applications. We model Internet traffic on burst level. Bursts of geometrically distributed size are generated which arrive according to a Poisson process. The mean burst size has been chosen according to measurements of typical flow lengths [15, 7] to 10 kByte. In reality, heavy tailed burst sizes can be expected, because of the distribution of the file size [9]. However, whether users retrieve large files or not, depends on the pricing. If pricing is based on volume or time, users will rarely download large files. For the near future, we assume Internet over GPRS to be more expensive than in a fixed network. Therefore, we believe that the assumption of non-heavy tailed distribution of burst size is acceptable. Similar effects have been reported for the holding time of Internet dialup traffic [14]. In general the holding time is shown to be heavy-tailed, but during expensive periods of the day, the measurements do not show heavy tails.

All models in this study capture the call or burst level.
They do not incorporate the data transport itself. However, this has a big influence on the performance, which has been shown in [6]. The coding mechanism as well as mechanisms like ARQ influence the overhead for error correction and therefore the available net bit rate. For all results of this investigation, this has to be taken into account. In the following all types of arrivals (calls and bursts) are commonly denoted as calls.

4 Partitioning Strategies

Figure 1 shows three strategies that can be used to introduce GPRS into an existing GSM network. Partitions are not created for individual services but for GSM and GPRS traffic. Inside a GPRS or common partition complete sharing is used for the different services. Admission control and resource allocation are done as described in subsection 5.2. In common partitions voice service is always given the highest priority.

![Partitioning Strategies Table]

**Figure 1. Partitioning Strategies**

*Complete Partitioning (CP)*

Complete partitioning divides the total cell capacity into two parts, one for GSM and one for GPRS traffic. From the analytical point of view the system behaves like two separate systems that can be analyzed separately. In each partition resources have to be reserved for incoming handover calls. Because there is no sharing of resources across partition borders, the utilization is expected to be low.

*Complete Sharing (CS)*

With complete sharing no capacity is exclusively assigned to GSM or GPRS traffic. Only for incoming handover calls capacity is reserved. Because voice traffic is assigned the highest priority, it can occupy the whole system capacity. Therefore, this strategy promises the lowest influence of the introduction of GPRS on traditional GSM service.

*Partial Sharing (PS)*

Partial sharing divides the total cell capacity into three parts, one for GSM, one for GPRS traffic and one common partition. Capacity for incoming handover can be reserved in the individual partitions separately or for all types of handover in the common partition. The GPRS partition has to be big enough to guarantee a certain QoS and small enough that no bandwidth is wasted in the case of very low GPRS traffic.

Like with complete partitioning there is a minimum capacity that is exclusively used by GPRS traffic. Therefore, a certain QoS can be guaranteed. On the other hand, the utilization is expected to be much higher than with complete partitioning because of the common resources.

5 The Model

5.1 Model Description

A GSM/GPRS cell can be modeled as a loss system with total capacity $C$, offered a finite number $k$ of traffic streams, each describing a different service. Depending on the partitioning strategy, $C$ is divided into at most three partitions, $C_{GSM}$, the capacity exclusively used for traditional GSM voice calls, $C_{GPRS}$, the capacity exclusively used for GPRS, and $C_{Com}$, the capacity shared by all services. The total cell capacity of the system $C$ depends on the actual number of frequencies, used in the cell.

Let $X_s(t)$ denote the number of calls of type $s$ in progress in the system at time $t$. Then $X(t) = (X_1(t), \ldots, X_k(t))$ defines a continuous-time stochastic process with finite discrete state space. The set of allowed states $\Omega$ depends on the admission policy to be defined later. Assuming, the process $X(t)$ reaches a steady state for $t \to \infty$, let $p(\mathbf{x})$, $\mathbf{x} = (x_1, \ldots, x_k), \mathbf{x} \in \Omega$ be the steady state probabilities.

**Bit Rate Thresholds**

The vector $\mathbf{b}(\mathbf{x}) = (b_1(\mathbf{x}), \ldots, b_k(\mathbf{x}))$, $\mathbf{x} \in \Omega$ defines the individual state dependent bit rates, where $b_i(\mathbf{x})$ is the bit rate of one type $s$ call in state $\mathbf{x}$. It is assumed, that calls of the same type transmit with the same bit rate somewhere between 0 and $b_i^{(\text{max})}$. This behavior can be achieved by the base station, because it has complete control over the capacity of the system in up- and down link. With the limitation, that a mobile station can receive only one frequency at a time, $b_i^{(\text{max})}$ will not be greater than $C_f$, with $C_f$ as the maximum capacity of one frequency.

To introduce as much flexibility as possible into the bit rate allocation, a *bit rate threshold vector* $b_i^{(\text{min})}, \text{min} = (\text{min}_1, \ldots, \text{min}_k)$ is assigned to each service type $s$. The threshold $b_i^{(\text{min})}$ in a call of type $s$ is the minimum bit rate, that calls of type $s$ have to reduce to in favor of arrivals of type $i$ calls. A value of $b_i^{(\text{min})} = 0$ means, an arrival of a type $i$ call can cause a disruption of an ongoing type $s$ call. A *softly occupied bit rate* of a call of type $s$ is defined as the difference $b_s(x) - b_i^{(\text{min})}$. This difference is the...
arrival rate for calls of one idle service type \( s \), respectively. The state dependence of the arrival rate for new calls will be due to the finite number of potential service type \( s \) users \( M = (M_1, \ldots, M_k) \) in the cell, which are assumed to be constant numbers (as many users entering the cell as many leave it on average). Then, the state dependent arrival rate for new calls will be \( \lambda_s^{(i)}(x) = (M_s - x_s) \lambda_s^{(i)} \) with \( \lambda_s^{(i)} \) as the mean arrival rate for calls of one idle service type \( s \) user. As can be seen, \( \lambda_s^{(i)} \) is proportional to the number of idle users \( M_s - x_s \).

Contrarily, the handover arrival rate \( \lambda_s^{(h)}(x) \) depends on the number of active users \( x_s \) (if there is no active user, there will be no handover). With the assumption of a homogeneous network structure (adjacent cells have the same \( M \), the same load and handovers are uniformly distributed among their neighboring cells), the handover arrival rate can be expressed as \( \lambda_s^{(h)}(x) = x_s \lambda_s^{(h)} \), with \( \lambda_s^{(h)} \) as the mean handover rate for one active service type \( s \) user.

Regarding the service time of voice calls it was already mentioned in section 3, that it is reasonable to assume a negative exponential distribution with parameter \( \mu_s \). Therefore, \( \mu_s(x) = x_s \mu_s \).

Data services are not described by their holding time distribution, but a geometrically distributed amount of data to be transmitted. Nevertheless, this can be mapped into a negative exponentially distributed service time with parameter \( \mu_s \) that depends on the state \( x \) through the bit rate \( b(x) \) of this service. For a high bit rate the service time will be lower than for a low one. With mean message size of \( m_s \) for type \( s \) calls, the state dependent mean service rate \( \mu_s(x) = x_s \frac{b_s(x)}{m_s \mu_s} \).

### Capacity Reservation Thresholds

The values \( C_s^{(n)} \) and \( C_s^{(h)} \) for new calls and handover calls, respectively, are introduced. They are called capacity reservation thresholds. A partition, where calls of type \( s \) can be admitted is named \( s \)-partition. Then \( C_s^{(n)} \) (\( C_s^{(h)} \)) is the maximum value for the sum of the minimum occupied bit rates of all services in all \( s \)-partitions, after the acceptance of a new (handover) type \( s \) call. For reserving some capacity for handover arrivals we assume the relation \( C_s^{(n)} < C_s^{(h)} \) to be true for all services \( s \).

Let the voice service be the only one in the GSM-partition, denote it by \( s = 1 \) and denote the GPRS services by \( s = 2, \ldots, k \). To use the whole system capacity the following conditions must hold:

- Complete Partitioning
  \[ C_{GSM} = C_1^{(h)} \]  and  \[ C_{GPRS} = C_2^{(h)} \]
  \[ \ldots, \]
  \[ C_k^{(h)} \]

- Complete Sharing
  \[ C_{Com} = C_1^{(h)} \]
  \[ \ldots, \]
  \[ C_k^{(h)} \]

- Partial Sharing
  \[ C_{GSM} + C_{Com} = C_1^{(h)} \]
  \[ \ldots, \]
  \[ C_k^{(h)} \]

### 5.2 Admission Policy

In the following, an admission policy is introduced, that can reserve capacity for high priority calls (e.g. handover voice calls), while maintaining a high utilization of the partition resources. The main idea is to utilize the reserved capacity by softly occupied bit rate that can be released upon request immediately.

In state \( x \) a new arrival of type \( s \) is accepted, if after its admission the sum of the minimum occupied bit rates of all services in all \( s \)-partitions is not greater than \( C_s^{(n)} \). Depending on the admission policy the following conditions must hold:

- Complete Partitioning (CP)
  \[ x_i^{(min)} \leq C_1^{(n)} - \frac{b_s}{\mu_s} \]
  \[ \sum x_i^{(min)} \leq C_s^{(n)} - b_s \]
  \[ s = 2, \ldots, k \]  (GSM)

- Complete Sharing (CS)
  \[ \sum x_i^{(min)} \leq C_s^{(n)} - b_s \]
  \[ s = 1, \ldots, k \]  (GSM, GPRS)

- Partial Sharing (PS)
  \[ \sum x_i^{(min)} - C_{GPRS} \leq C_s^{(n)} - b_s \]
  \[ x_i^{(min)} \leq C_s^{(n)} - b_s \]
  \[ s = 2, \ldots, k \]  (GSM)

For handover arrivals the same admission conditions apply, except that the capacity reservation threshold for handover \( C_s^{(h)} \) is used, instead of \( C_s^{(n)} \).

There are different approaches to distribute the capacity of the system among ongoing calls, whereof the following is chosen for our investigations. As already mentioned above, calls of the same type transmit always with the same bit rate. Assume the services are ordered according to their priorities in descending order with voice calls having the highest priority and a state independent constant bit rate. Then the bit rate of one type \( s \) GPRS call in state \( x \) is set to

\[ b_s(x) = \min \left( \max \left( A, b_s^{(min)} \right), b_s^{(max)} \right) \]
\[ A = C \cdot \prod_{i=1}^{x} \frac{1}{b_i^{(m)}} \prod_{j=1}^{x} \frac{1}{b_j^{(m)}} \]

where \( b_i^{(m)} = \min(b_i^{(m_1)}, \ldots, b_i^{(m_k)}) \) and

\[
C^* = \begin{cases} 
C_{GPRS} & \text{for CP} \\
C - x_i b_i^{(m)} & \text{for CS} \\
C_{GPRS} + C_{CS,om} x_i b_i^{(m)} & \leq C_{GSM} \\
C - x_i b_i^{(m)} x_j b_j^{(m)} & > C_{GSM} & \text{for PS}
\end{cases}
\]

This means, at any time there is at maximum one type \( j \) of calls, transmitting with a bit rate somewhere between \( b_j^{(m_{\min})} \) and \( b_j^{(m_{\max})} \). All type \( i \) calls with a higher priority than type \( j \) calls transmit with maximum bit rate \( b_i^{(m_{\max})} \), \( i < j \) and calls with a lower priority than type \( j \) calls transmit with their minimum bit rate \( b_i^{(m_{\min})} \), \( i > j \). The total amount of resources occupied in the cell at state \( x \) is \( C_{occ}(x) = x \cdot b(x) \).

### 5.3 Performance Measures

For the evaluation of the different introduction strategies following performance measures have been observed with respect to the offered load. A more detailed description can be found in [4].

**New Call Blocking Probability** \( B^{(n)}_s \), defined as the time congestion probability regarding new calls of type \( s \).

**Handover Failure Probability** \( B^{(h)}_s \), defined as the time congestion probability regarding incoming handover requests of type \( s \).

**Cell Utilization** \( U \), defined as the ratio of the average occupied capacity \( C_{occ} \) to the total capacity \( C \).

**Average Data Rate** \( b^{(aw)}_s \), defined as the average data rate of a randomly chosen connection of type \( s \).

### 6 Numerical Example

In this section, we give an example on the results, that can be obtained using the presented analytical model. The model has been written in MOSEL language [2] and the resulting Markov chain has been solved using the SPNP package [8]. The examined system has the characteristics according to the ETSI standards as described in section 2. Table 1 summarizes values for the main characteristics of the GSM/GPRS system.

**Traffic**

During the early phase of GPRS introduction into existing GSM systems, voice calls are still supposed to be the major source of traffic. RTTI calls will not contribute a lot to the overall load because they represent short messages and hence have relatively small holding times. Nevertheless, they will be much more frequent than voice calls and, therefore, contribute a lot to the signalling load. There is very few experience about mobile Internet traffic. For the early phase we assume Internet traffic to contribute about one tenth to the overall load. Following this argumentation the offered traffic for the numerical investigations is composed of 90% voice, 1% RTTI and 9% Internet traffic.

**Partitioning**

The sizes of the individual partitions for the three strategies reflect the offered traffic ratio (90% GSM, 10% GPRS) and the assumptions given in section 4. The values are given in table 2. There has to be no extra partition for voice service with partial sharing. Because of the priority scheme used, voice calls can rule out GPRS calls in the common partition and, therefore, they have not to be protected.

### Table 1. Network parameters

<table>
<thead>
<tr>
<th></th>
<th>GSM</th>
<th>GPRS</th>
<th>Common</th>
</tr>
</thead>
<tbody>
<tr>
<td># frequencies</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td># channels</td>
<td>8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td># signaling</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td># data channels</td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>Overall gross bandwidth</td>
<td>470.8 kbit/s</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. PDCH per user</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max. gross data rate per single user</td>
<td>171.2 kbit/s</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Partition sizes in channels/PDCH

<table>
<thead>
<tr>
<th></th>
<th>Partition</th>
<th>incl. HO</th>
<th>Partition</th>
<th>incl. HO</th>
<th>Partition</th>
<th>incl. HO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CP</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>1/4</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>CS</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>PS</td>
<td>–</td>
<td>1</td>
<td>–</td>
<td>21</td>
<td>2</td>
</tr>
</tbody>
</table>

### Thresholds

Because the partitioning strategies are very different, comparable results are not easy to obtain. When setting the bit rate thresholds and the capacity reservation thresholds, the following design goals were in mind. At first, voice calls should not suffer any disadvantage because of the introduction of GPRS traffic. Furthermore, since data calls (RTTI calls and Internet bursts) consist of short messages, it is better to allow these messages to be transmitted as fast as possible. Therefore, data calls are allowed to occupy a part of the channels reserved for handover. Finally, handover blocking probabilities should be a factor 5 to 10 lower than new call blocking probabilities. Data call handover are expected to be very infrequent because of the short holding time. The resulting values for the systems investigated are given in table 3.
Table 3. Data rate and capacity reservation threshold values in kbit/s

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Partition</th>
<th>Service</th>
<th>s</th>
<th>$b_s^{max}$</th>
<th>$b_s^{min}$</th>
<th>$b_s^{min}$</th>
<th>$C_s^{(n)}$</th>
<th>$C_s^{(h.o)}$</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compl. Partitioning</td>
<td>GSM, Voice</td>
<td></td>
<td>1</td>
<td>21.4</td>
<td>21.4</td>
<td></td>
<td>385.2</td>
<td>428.0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GPRS, RTTI</td>
<td></td>
<td>2</td>
<td>42.8</td>
<td>5.35</td>
<td>5.35</td>
<td>37.45</td>
<td>42.8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GPRS, Internet</td>
<td>3</td>
<td>42.8</td>
<td>5.35</td>
<td>5.35</td>
<td>37.45</td>
<td>42.8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Complete Sharing</td>
<td>Common, Voice</td>
<td></td>
<td>1</td>
<td>21.4</td>
<td>21.4</td>
<td>21.4</td>
<td>428.0</td>
<td>470.8</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Common, RTTI</td>
<td></td>
<td>2</td>
<td>171.2</td>
<td>10.7</td>
<td>10.7</td>
<td>449.4</td>
<td>470.8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Common, Internet</td>
<td>3</td>
<td>171.2</td>
<td>10.7</td>
<td>10.7</td>
<td>449.4</td>
<td>470.8</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Partial Sharing</td>
<td>GSM/Com., Voice</td>
<td></td>
<td>1</td>
<td>21.4</td>
<td>21.4</td>
<td>21.4</td>
<td>406.6</td>
<td>449.4</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>GPRS/Com., RTTI</td>
<td>2</td>
<td>171.2</td>
<td>10.7</td>
<td>10.7</td>
<td>449.4</td>
<td>470.8</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GPRS/Com., Internet</td>
<td>3</td>
<td>171.2</td>
<td>10.7</td>
<td>10.7</td>
<td>449.4</td>
<td>470.8</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Service parameters

<table>
<thead>
<tr>
<th>s</th>
<th>Service</th>
<th>$1/\mu_s$</th>
<th>$1/\lambda_s^{(e.o)}$</th>
<th>$m_{s0}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Voice</td>
<td>100 s</td>
<td>80 s</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>RTTI</td>
<td>–</td>
<td>40 s</td>
<td>170 Byte</td>
</tr>
<tr>
<td>3</td>
<td>Internet</td>
<td>–</td>
<td>600 s</td>
<td>10 kByte</td>
</tr>
</tbody>
</table>

Services

Numerical results are given for three classes of services as discussed in section 3. The first traffic class represents voice calls, which are treated in a similar manner as in traditional GSM circuit switched networks. As can be seen in table 3 the bit rate thresholds of this class are always set in such a way, that the bitrate of voice calls is not influenced by the current load of the cell. Therefore, the values are set to the gross data rate of one PDCH. This class is associated to the highest priority. Furthermore, the values for the capacity reservation thresholds are set to give incoming voice handovers priority over new calls. For the voice service, we assume a mean call duration of 100 s, a dwell time of 80 s and that there are 500 users per cell (Table 4).

Figure 2. Utilization

Results

All results in the following figures are given with respect to the normalized offered load. A normalized offered load of 1.0 means, that the offered bitrate traffic equals the capacity of the system. This point is called critical load. Also we define a normal load as the normalized offered load with a voice call blocking probability of 1%.

In this numerical example, the blocking probabilities of the RTTI calls and Internet bursts (Figures 3, 4, 5) are equal for all three strategies, because of the equal thresholds of the two services (Table 3). As can be seen in the figures, the results for complete sharing (CS) and partial sharing (PS) are very similar, PS having slightly better blocking probabilities than CS for data calls because of the individual GPRS-partition.

Complete partitioning (CP) can be evaluated as two independent systems. The small GPRS-partition promises low blocking probabilities for low load, whereas it becomes not acceptable for high load (Figure 3). Because of the small GPRS-partition (2 PDCH), RTTI calls and Internet bursts get a low average data rate (Figure 6), resulting in longer assumed not to be frequent (Table 4).
holding times. This is reflected in the higher mean number of RTTI calls and Internet bursts, as calculations had shown. Whereas the mean number of voice calls lies in the same region for all strategies, for data calls (bursts) this number is a factor 5 to 7 higher with CP.

As expected, CP has the lowest utilization. Comparing the strategies regarding the utilization at the normal load (CP: 0.369; PS: 0.404; CS: 0.427) the utilization for CP is about 6.4% lower, for PS around 2.4% and for CS around 9.1% higher. Whereas the carried traffic of the system was about 90% voice, 1% RTTI and 9% Internet, in terms of the number of calls the ratio was about 5% voice, 83% RTTI and 12% Internet bursts.

Summarizing it can be said, that the complete sharing strategy in combination with the proposed admission policy gives the best results in terms of utilization of the system resources.

Additionally, the CS strategy was investigated with allowing voice calls to push out data calls completely. This was achieved by setting the bandwidth thresholds \( b_2^{(\text{min})} = 0 \) and \( b_3^{(\text{min})} = 0 \). At normal load, Internet bursts (RTTI calls) did not get any bandwidth for 0.07% (0.0002%) of the time on average. These values change to 2.8% (0.008%) for the critical load. The distribution of these time periods is of great interest for investigations of the interactions with lower layers, e.g. outrunning timer of the TCP layer during such periods could cause retransmissions and hence increase the load to the system. This can be considered as a major drawback of such an approach.

7 Conclusions

The contribution of the paper is threefold. First, an efficient call admission policy has been proposed for multiservice GSM/GPRS networks, comprising voice and different data services. Secondly, an analytical model, based on a multi-dimensional Markov chain, has been introduced. The model is quite general and capable of capturing interesting call and burst level dynamics of the GSM/GPRS system. Thirdly, three different strategies of introducing GPRS traffic into existing GSM networks have been described and investigated.

The results of the numerical examples allow a comparison of the three strategies. They have verified the efficiency of the proposed call admission policy and have shown, that with an appropriate setting of the parameters individual blocking probabilities and QoS requirements of the different service types can be achieved. It has also been shown, that with the use of the partial sharing or complete sharing strategy the flexible GPRS service contributes to a better utilization of the GSM system.
Figure 6. Average data rates for Complete Partitioning

Figure 7. Average data rates for Partial Sharing

Figure 8. Average data rates for Complete Sharing

References


[12] ETSI. GSM Specification 02.60, 03.60, 03.64: Digital cellular telecommunications system (Phase2+); General Packet Radio Service; Service description, Stage 1, 1999 (02.60); Service description, Stage 2, 1999 (03.60); Overall description of the GPRS radio interface, Stage 2, 1999 (03.64).

