Improved channel allocation for voice and data traffic with resource reservation for voice traffic in EDGE system

Carlos M. Ramírez Casañas, Sònia P. Mansilla, Josep Paradells Aspas
Department of Telematics Engineering, University of Las Palmas de Gran Canaria, Department of Applied Mathematics IV, Department of Telematics Engineering, Technical University of Catalonia, Spain
Email: {cramirez, teljpa}@entel.upc.edu; sonia@ma4.upc.edu

Abstract— **This paper makes two main contributions. The first is a method that combines TDMA time slots for voice and data packet services to offer high peak rates and guarantee packet connectivity in the case of cell congestion. The method can also be used to analyse time slot reservation for voice or data packet services. The second contribution is a novel method for analysing the system. In addition to the traditional approach that consists in solving a large system of equations to compute state probabilities, a combinatorial method that simplifies the analysis and that can be applied to larger and more complex systems is introduced.**

Index Terms— **channel allocation, cellular networks, quality of service, voice and data traffic**

I. Introduction

Second generation mobile communication systems are widely used in the current market. These systems are based on standards such as the global system for mobile communications (GSM), the digital advanced mobile phone service (DAMPS) and cdmaOne have made possible to offer voice services across the entire country. In contrast, until now, data services have not enjoyed the same level of distribution. GSM [1] was enhanced by the incorporation of general packet radio service (GPRS) to meet the growing demand for data services [2]. GPRS uses idle time slots in GSM systems to define data channels. However, the initial rates reached by GPRS were comparable to those available using standard telephone lines with modems, and the service was therefore restricted. Two new systems were developed to reduce this restriction. The first is the universal mobile telecommunications system (UMTS) [3] which is suitable for multimedia services. Unfortunately, this system requires new frequency bands and additional radio infrastructure, which led to competition for the new spectrum (at an extremely high cost in some countries) and the need for a new base station subsystem. Such a high level of investment was required to implement a UMTS that some operators sought an alternative. The second solution is based on an improvement of the radio interface used in GSM and DAMPS, it is referred to as enhanced data rates for GSM evolution (EDGE) [4].

EDGE is an enhancement of current standards for delivering 3G services in existing spectrum bands. It can be considered a generic air interface that efficiently provides high bit rates and facilitates the adaptation of existing cellular systems to 3G capabilities. EDGE was developed simultaneously by the European Telecommunications Standards Institute (ETSI) and the Universal Wireless Communications Consortium (UWCC) to guarantee a high degree of synergy with the GSM and DAMPS standards. GSM and EDGE systems use the same time slot management mechanism for different communication types but EDGE provides up to a three-fold increase in bit rate. The spectrum is divided into 200 kHz carrier frequencies, each of which is divided into eight time slots using TDMA. A channel (used for voice or data communication) consists of one time slot from one carrier. Some key time slots are allocated to signalling but the rest can be assigned dynamically by the operator according the voice and data traffic. However, there are some common sense guidelines on the use of these channels. For example, since the beacon carrier (i.e. the one with the common control channel) is transmitted at constant power, it is advisable to allocate GPRS channels to this carrier because packet transmission does not usually apply power control on the downlink.

The comparative lack of radio channels justifies research aimed at determining how to share these resources efficiently. In the past, channels were allocated according to a fixed pattern to each different voice and data service [5]–[7]. However, permanent channel allocation has proved to be an inefficient solution for burst traffic. In fact, dynamic allocation schemes are often used when a channel can be shared by multiple users. There are many studies in the literature related to resource reservation management (RRM) [8]–[16], most of which use simulations to evaluate the blocking and packet delay probabilities. However, none of these studies provide an analytical framework for optimizing the allocation of time slots of a TDMA frame to maximize the peak rate offered

This paper is based on “Improving EDGE for rural environments through channel allocation for voice and data traffic,” by Carlos M. Ramírez Casañas, Josep Paradells Aspas, Sònia P. Mansilla, which appeared in the Proceedings of the 2nd International Symposium on Wireless Pervasive Computing (ISWPC’07), San Juan, Puerto Rico, February 2007 ©.

This research was supported by the Spanish “Comisión Interministerial de Ciencia y Tecnología” (CICYT) within the framework of project TEC2006-04504.

to the user and assuring a given voice call blocking probability.

The new framework developed in this study is able to reserve time slots for voice traffic and data packets. The rest of the available time slots can be shared between both services, but voice traffic pre-empts data slots in the event of slot starvation. Since GPRS and EGPRS require time slots in the frame to be sequential for simultaneous use [17] [18], the assignment strategy attempts to place voice calls close together. Commercial terminals can manage up to four time slots simultaneously but this constraint can easily be introduced into the model developed here. Our analytical model provides a method for calculating the probability of having consecutive time slots available for data transmission. These probabilities can not only be used to calculate the mean bit rate but also other relevant parameters, such as the probability of transmitting at a certain speed or the probability of transmitting a file in less than a given number of seconds. Several strategies may be compared for allocating time slots to voice and data communications and the results in terms of blocking probability or mean bit rate may likewise be compared. In addition, the proposed method makes it very easy to calculate state probabilities, which is a considerable improvement with respect to the methodology presented in [13] and [19]. Most analytical models for calculating the bit rate offered by GPRS or EGPRS only consider the mean bit rate, which depends exclusively on the number of time slots and users. If no QoS is applied (as is usually the case) each user will receive a proportional part of the total bit rate. The mean bit rate is an important parameter in terms of quality of service from a user perspective, but for certain applications such as web surfing, a higher peak rate can improve the quality of service perception because it increases the speed with which pages are downloaded and reduces the probability of a data transmission interruption due to handover or slot starvation.

The paper is structured as follows. In Section II we describe the novel method for maximizing the possible consecutive idle time slots in channel allocation. In Section III we introduce the analytical model used in the allocation of GSM/EDGE channels and present a method for deriving the probabilities of having consecutive idle time slots. In Section IV we describe how to find the sets of consecutive idle time slots, present the scenarios in which we analysed the model, and present the results. Finally, Section V contains the conclusions.

II. Description of the Allocation Mechanisms

In this section, the time slot allocation method for voice and data traffic is presented. Zukerman [19] and Ivanovich et al. [13] present several schemes for allocating channels in each time slot of the frame. Simple schemes such as the *Random* method assign a new call to any of the available idle time slots. In alternative methods, such as the *Best Fit* method and the *First Fit* method, each time slot in a frame is identified by a number and these procedures allocate each incoming call to an idle time slot using a criterion that depends on the slot identifier (for example, the time slot with smallest ID number is occupied first). In addition, the *Repacking* method uses the intracell-handover mechanism to reallocate calls between slots while the number of consecutive time slots that are idle or allocated to data packets is maximized. This latter method requires extra signalling, but since the two time slots (the old and the new) belong to the same base transceiver station (BTS), there is no degradation of voice quality.

A. Channel allocation strategy without resource reservation

There are eight time slots in basic scenarios and in time division multiple access (TDMA) the most suitable time slot allocation procedure must be determined. In the case of a carrier with a common control channel (CCCH), the same procedure is used but one time slot is allocated to the control channel and the number of available time slots is reduced to seven. The eight time slots in the frame are identified by ID numbers (from 0 to 7). A time slot is busy when it is used by a voice call; when the time slot is idle there is a hole. The frame state can be defined using a binary representation with 1 bit per time slot and as many bits as time slots per frame. In addition, for $i = 0, \ldots, 2^8 - 1$, we define State i by the same number but in base 10. That is:

$$\text{State } i \;:=\; \sum_{\text{ID number}=0}^{7} 2^{\text{ID number}} \cdot b(\text{ID number}),$$

where $b(\text{ID number}) = 1$ when the time slot with the corresponding ID number is busy and 0 otherwise. For example, Fig. 1 shows States 5 and 7.

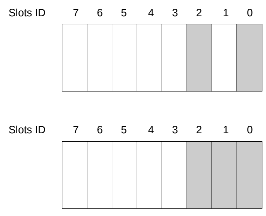

Figure 1. Example of time slot allocation

For this system we use the *First Fit* method [13], which consists in allocating the idle time slot with the smallest ID number to a new call. At the top of Fig. 1 we consider time slots 0 and 2 to be busy. As soon as a new time slot is allocated it will occupy the time slot with the smallest ID number, which in this case is the time slot with the ID number 1. At the bottom of Fig. 1, we show this latter state. When a time slot is released a hole is produced and no reordering is made, regardless of the place that the time slot occupies.

This method essentially avoids the excessive fragmentation of consecutive time slots. In theory, it seems that there is a high probability of finding consecutive idle time slots when using this method, so network operators have many potential locations to allocate to mobile stations with multislot capacity for GPRS. It should be noted that in this model channels are allocated to time slots for voice traffic, It is therefore also useful to estimate how many idle time slots can be used by data traffic.

The model is characterized by studying the transition of every possible system state and the time spent in each state.

B. Channel allocation strategy with resource reservation for voice traffic

In this case, the model distinguishes between voice and data traffic and is only able to reserve time slots for voice traffic. Consequently, there are two types of time slots: time slots reserved for voice calls (since handover operations are envisaged) and shared time slots for both voice and data calls. As mentioned above, radio resources are allocated sequentially in this model according to the type of traffic of the new call.

As shown in Fig. 2, data traffic will be allocated to the idle time slot with smallest ID number whereas voice calls will be allocated to the idle time slot with largest ID number.

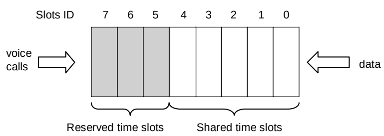

Figure 2. Example of time slot allocation with reservation

In addition, when all reserved time slots are occupied and a voice call occupies a shared time slot, if one of the calls in a reserved time slot terminates we can perform an intracell handover to free the shared slot. That is, we reassign the voice call occupying the shared time slot with the lowest ID number to the reserved time slot that has just been released. Consequently, it is not possible to have an idle reserved time slot and a voice call occupying a shared time slot at the same time. Note that voice calls in this configuration will always occupy the left-hand side of the frame whereas data calls will occupy the right-hand side. Consequently, we compact the remaining consecutive idle time slots in the centre of the frame. As above, this time slot allocation procedure can be modelled using a finite state diagram.

III. ANALYTIC MODELS

In this section we describe the analytical models used in the various scenarios considered. We make two assumptions regarding the distribution of the calls: firstly, that the call arrival times are independent and can be described by a Poisson process of mean λ; secondly, that the call holding time is random and can be described by a negative exponential distribution with mean $1/\mu$ [15].

In other parts of this section in which a distinction must be drawn between the different types of traffic, we will use λ_v and μ_v to denote the above parameters for voice calls and λ_d and μ_d to denote the same parameters for data calls. The system is Markovian due to the presence of memoryless interarrival and service time. If the system is in steady state, a linear equation system may be used to determine the state probabilities.

Given the offered voice traffic and the number of slots allocated to voice and data traffic, it is possible to compute the state probabilities. The blocking probability is determined by considering specific states.This is the systematic approach used in previous studies [9]. By analysing the system, a set of simple equations for calculating the state probabilities may be derived. This approach is preferable because the calculations become complex as the number of states increases. In Section IV we explain how to use the state probabilities to obtain the probabilities of finding bursts of any length in any scenario and in Section V we present the results of applying the equations to different scenarios.

A. Basic analytical model

In the basic model we do not make any distinction between voice and data calls and we do not reserve a number of time slots for voice traffic only. In addition, the method for allocating a new call to the idle time slot with the smallest ID number follows a Markovian model and is constructed as a system of state equations.

We can use the series of equations to describe the system analytically and we solve the system by using an iterative algorithm for an arbitrary number n of available time slots, which is not necessarily $n = 8$. We denote each frame as a base 2 number of length n, and consider that a position takes the value 1 if the time slot with the corresponding ID number is busy and 0 otherwise. For $r = 0, \ldots, 2^n - 1$, we will define State r using the same number, but in this case in base 10. For instance, if $n = 8$, State 5 is [00000101]. Hereafter, we will refer to State r indistinctly by r or by $[r_{n-1}, \ldots, r_1, r_0]$ with $r_i = \{0, 1\}$.

For $r = 0, \ldots, 2^n - 1$, the transient equation of the State r is:

$$[b(i_r < n) \cdot \lambda + i_r \cdot \mu]P(r) =$$

$$= b(i_r < n) \cdot \mu \cdot \sum_{\substack{j=k_r \\ j:r_j=0}}^{n-1} P(2^j + r) + b(k_r > 0) \cdot \lambda \cdot \sum_{j=0}^{k_r - 1} P(r - 2^j), \quad (1)$$

where i_r is the number of busy time slots of State r, k_r is the smallest ID number that is idle (or equivalently the number of consecutive busy slots at the beginning of the frame) and $b(Expression)$ is 1 if the $Expression$ is true or 0 otherwise.

To continue with the same example, if we consider $n = 8$ the transient equation of State 5 ([00000101]) reduces to:

$$\begin{aligned}[\lambda + 2 \cdot \mu] \cdot P(5) = \\ (P(7) + P(13) + P(21) + P(37) + \\ + P(69) + P(133)) \cdot \mu + P(4) \cdot \lambda.\end{aligned}$$

The first part of the equation determines that State 5 can either be converted into two different states if one of its calls is released or into a new state if a new call arrives. The second part of the transient equation denotes that State 5 arises whenever a certain call is released from States 7, 13, 21, 37, 69 and 133 or whenever a call reaches State 4.

B. Analytical model with reserved time slots for voice traffic

The second model distinguishes between voice and data traffic and can reserve time slots for voice traffic only. That is, there are two types of time slots: time slots reserved for voice calls and shared time slots for both voice and data calls.

As stated in the previous section, in this model radio resources are allocated sequentially according to the type of traffic. Data traffic will be allocated to the idle time slot with smallest ID number whereas voice calls will be allocated to the idle time slot with largest ID number. In addition, if there are voice calls occupying the so called "shared time slots" (Fig. 2), we perform an intracell handover when one of the calls in a reserved time slot is released.

We can derive the different state transition equations for this model, which is completely generic and not limited by the total number of available time slots or the number of reserved time slots in the frame for data traffic.

Let us consider that a frame is a sequence of length n, in which a position is 0 if the time slot with the corresponding ID number is idle, 1 if the time slot is busy with a data call and 2 if it is busy with voice traffic. Consequently, a state is defined by a sequence $[r_{n-1}, \ldots, r_1, r_0]$ with $r_i = \{0, 1, 2\}$. It should be noted that not all 3^n sequences of this type denote possible states because some time slots are reserved for voice traffic only. If m is the number of shared time slots for data and voice traffic and $n-m$ the number of reserved time slots for voice traffic only, in the state defined by $[r_{n-1}, \ldots, r_m, r_{m-1} \ldots, r_1, r_0]$, the time slots with ID numbers from 0 to $m-1$ are the shared slots and the time slots with ID numbers from m to $n-1$ are the reserved time slots.

The number of states is: $2^n + 3^m - 2^m$. One way of verifying this expression is to count all possible states in which no voice calls occupy the shared slots, which is 2^n. This is because a time slot with an ID number from 0 to $m-1$ is either 0 or 1 and a time slot with ID numbers from m to $n-1$ is either 0 or 2. We then count all possible states in which all reserved time slots are occupied and there may be voice traffic in the shared time slots, which is 3^m. Time slots with ID numbers from m to $n-1$ are all equal to 2 and time slots with ID numbers from 0 to $m-1$ are either 0 or 1 or 2. Finally, we double count the 2^m states in which all reserved time slots are occupied and all shared time slots take values of 0 or 1.

We now have to distinguish between two blocks of transient equations. The first block consists of the transient equations for states in which not all reserved time slots are occupied (and, by extension, those states in which there is at least one idle reserved time slot). The second block consists of the transient equations for states in which not all reserved time slots are occupied (and, possibly, with voice traffic in the shared time slots).

Let us enumerate the states as decimal numbers. First, we enumerate the states in which not all reserved time slots are occupied and represent them as numbers in base 2. For the conversion to base 10, we first consider that a position takes a value 1 if the time slot with the corresponding ID number is busy (either with a voice or data call) and 0 otherwise, as shown in the following table:

State number	Vector
0	$[\overbrace{00\cdots00}^{n-m}\overbrace{0\cdots00}^{m}]$
1	$[00\cdots000\cdots01]$
2	$[00\cdots000\cdots10]$
3	$[00\cdots000\cdots11]$
$\vdots$	$\vdots$
2^m-1	$[00\cdots001\cdots11]$
2^m	$[00\cdots020\cdots00]$
2^m+1	$[00\cdots020\cdots01]$
$\vdots$	$\vdots$
$2^{n-1}-1$	$[02\cdots221\cdots11]$
2^{n-1}	$[20\cdots000\cdots00]$
$2^{n-1}+1$	$[20\cdots000\cdots01]$
$\vdots$	$\vdots$
2^n-2^m-2	$[22\cdots201\cdots10]$
2^n-2^m-1	$[22\cdots201\cdots11]$

We then enumerate the states in which not all of the reserved time slots are busy, starting from State $2^n - 2^m$, which corresponds to $[2, \ldots, 2, 0, \ldots, 0]$. We consider the shared positions in base 3. A position takes a value of 2 if the time slot with the corresponding ID number is busy with a voice call, 1 if it is busy with a data call and 0 if it is idle, as shown in the following table:

State number	Vector
2^n-2^m	$[\overbrace{2\cdots2}^{n-m}\overbrace{0\cdots00}^{m}]$
2^n-2^m+1	$[2\cdots20\cdots01]$
2^n-2^m+2	$[2\cdots20\cdots02]$
$\vdots$	$\vdots$
$2^n-2^{m+1}+3^m$	$[2\cdots22\cdots21]$
$2^n-2^m+3^m-1$	$[2\cdots22\cdots22]$

Note that this classification of the states into two blocks is not strict. Some states from one block can be converted into states from the other block if a call is released or a new call arrives. For instance, if any call in State $[2, \ldots, 2, 0, \ldots, 0]$ from the second block is released - let us assume that this is the call in the leftmost position - the state becomes $[0, 2 \ldots, 2, 0, \ldots, 0]$, which is State $2^{n-1} - 2^m$ and one of the states from the first block. This type of base change is represented in the transient equations by the auxiliary term K, which will only be needed in boundary cases.

Since in this new case we distinguish between data and voice traffic, the system of equations will be much more complicated than Equation (1) and we must first define some general notations. Given a State r, let $i_{r,v}$ and $i_{r,d}$ denote the total number of time slots busy with voice calls and data calls respectively. Similarly, let $k_{r,v}$ and $k_{r,d}$ denote the number of consecutive busy time slots on the left and right of the frame respectively.

For all states r in which not all reserved voice slots are busy, the corresponding subsystem of transient equations is:

$$[\lambda_v + b(i_{r,d} < m) \cdot \lambda_d + i_{r,v} \cdot \mu_v + i_{r,d} \cdot \mu_d]P(r) =$$

$$= b(i_{r,v} < n - m - 1) \cdot \mu_v \cdot \sum_{\substack{j=k_{r,v} \\ j:r_{n-1-j}=0}}^{n-m-1} P(2^{n-1-j} + r) +$$

$$+ b(i_{r,v} = n - m - 1) \cdot \mu_v \cdot P(2^m(2^{n-m} - 1) + K) +$$

$$+ b(i_{r,d} < m) \cdot \mu_d \cdot \sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{m-1} P(2^j + r) +$$

$$+ b(k_{r,v} > 0) \cdot \lambda_v \cdot \sum_{j=0}^{k_{r,v}-1} P(r - 2^{n-1-j}) +$$

$$+ b(k_{r,d} > 0) \cdot \lambda_d \cdot \sum_{j=0}^{k_{r,d}-1} P(r - 2^j),$$

where

$$K = \sum_{j=0}^{m-1} r_j \cdot 3^j.$$

In addition, in the specific case in which no time slots are reserved for voice traffic ($n = m$), the transient equations are very similar to Equation (1); they are not identical because we distinguish between different types of traffic in this second model and consider 3^n different states (not 2^n, as in the basic model).

To illustrate the equations, take, for instance, $n = 4$ and $m = 2$. We therefore determine that State 9 is$[2001] \sim [V00D]$ and its transient equation is:

$$[\lambda_v + \lambda_d + \mu_v + \mu_d]P(9) =$$
$$\mu_v \cdot P(13) + \mu_d \cdot P(11) + \lambda_v \cdot P(1) + \lambda_d \cdot P(8),$$

The first part of the equation determines that State 9 can be converted into four different states: two are possible if one call of any type is released, and the other two are possible if a call of any type arrives. The second part of the transient equation denotes that State 9 arises whenever the voice call is released from State 13, whenever the data call is released from State 11, whenever a voice call reaches State 1, or if a data call reaches State 8 (Fig. 3).

We now define the transient equations of the states in which all reserved time slots are occupied. An additional

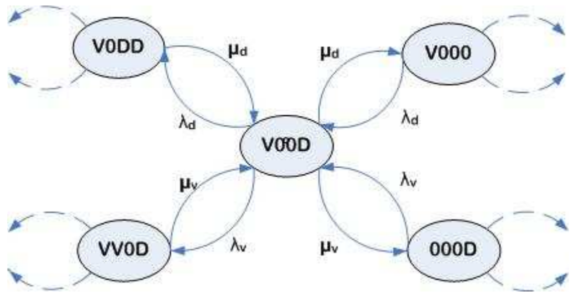

Figure 3. Diagram of transitions for State 9

notation is required for this block. First, since intracell handover will be performed each time that a reserved time slot is released (i.e. if there are voice calls occupying shared time slots) we need to know the smallest ID number of a time slot occupied by a voice call because the call in that time slot will be reassigned to the reserved time slot that was released. For a state r, let l_r. be the corresponding ID number. Secondly, let $\lambda_{v,p}$ be the loss rate of voice calls in the reserved time slots (which, given a *Blocking Probability* for the reserved time slots, BP_R, is simply $\lambda_{v,p} = \lambda_v \cdot BP_R$).

By following the notation above, for those states r in which all reserved time slots are busy, we obtain the following transient equations:

$$[b(i_{r,d} + i_{r,v} < n)(\lambda_v + \lambda_d) + i_{r,v} \cdot \mu_v + i_{r,d} \cdot \mu_d]P(r) =$$

$$= b(i_{r,v}+i_{r,d} < n) \left[\sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{m-1} (\mu_d P(3^j + r) + \mu_v P(2 \cdot 3^j + r)) + \right.$$

$$\left. + b(k_{r,d} < l_r) \cdot \sum_{\substack{j=k_{r,d} \\ j:r_j=0}}^{l_r-1} (n - m) \cdot \mu_v \cdot P(2 \cdot 3^j + r) \right] +$$

$$+ b(k_{r,d} > 0) \cdot \lambda_d \cdot \sum_{\substack{j=0 \\ j:r_j=1}}^{k_{r,d}-1} P(r - 3^j) +$$

$$+ b(k_{r,v} > n - m) \cdot \lambda_{v,p} \cdot \sum_{\substack{j=n-m-1 \\ j:r_{n-j}=2}}^{k_{r,v}} P(r - 2 \cdot 3^{n-j}) +$$

$$+ b(i_{r,v} = n - m) \cdot \lambda_v \cdot \sum_{j=m}^{n-1} P(K - 2^j),$$

where

$$K = \sum_{j=0}^{m-1} r_j \cdot 2^j + \sum_{j=m}^{n-1} 2^j.$$

To illustrate the equations, let us now assume that $n = 4$ and $m = 2$. The ID corresponding to State 20 is $[2222] \sim [VVVV]$ and the transient equation is:

$$4 \cdot \mu_v \cdot P(20) = \lambda_v \cdot (P(18) + P(14))$$

The equation determines that State 20 arises whenever a voice call reaches State 14 (=[2220]) or State 18 (=[2202]). The first part of the equation determines that State 20 can be converted into four states if one of the voice calls is released (not necessarily different states, because intracell handover is performed in this case). Note that [2022] is not a possible state because it has an idle reserved time slot and we perform intracell handover. The possible state corresponding to this situation would be [2220], because the voice call with the smallest ID number is moved to occupy the released call in the reserved time slots. Thus, if the calls occupying ID numbers 0, 2 or 3 are released State 20 is converted into State 14 (with a probability factor of $3\mu_v$) and when the call occupying ID number 1 is released State 20 is converted into State 18 (Fig. 4).

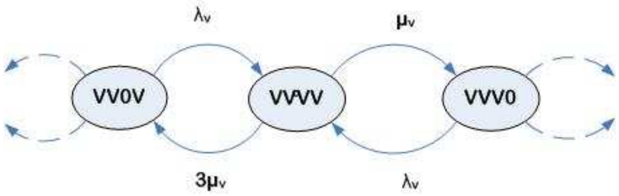

Figure 4. Diagram of transitions for State 20

IV. Bursts and results

In this section we present the concept of a burst and the different scenarios for determining the bursts of a frame (with and without resource reservation). Given the probabilities of the different states established in the previous section, we can define two ways of determining the probability of finding a given number of consecutive idle time slots, which we will call a *burst*. In other words, a *burst* is a group of contiguous time slots that are idle. Thus, when the first time slot of a burst is allocated to a given user, this user may use the time slots of the entire burst and will therefore benefit from greater throughput. This fact is particularly important since, if the network operator knows the disposition of consecutive idle time slots, it is possible to manage the burst lengths and improve the data rate for mobile users.

Each state consists of a certain number of bursts of different lengths. For instance, in the case of eight time slots, State 0 has a burst of eight consecutive idle time slots, State 1 has a burst of seven and State 2 has two bursts: one with six consecutive idle time slots and another with just one idle time slot (Fig. 5).

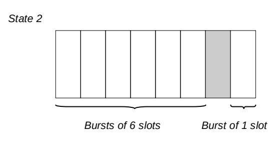

Figure 5. State 2 with bursts of 6 and 1 idle time slots

Therefore, if we want to determine the probability of finding a burst of a given length, we only need to know which states i have bursts of that length and the respective probabilities P_i of these states i.

In a previous study [18] we analysed the number of bursts and how this figure contributes to the overall probability of having bursts of a particular length. Using this information, the network operator is able to make a more accurate prediction of the QoS the system will be able to provide. The system has long bursts and can therefore offer high peak rates and is less likely to be interrupted by handovers or lack of resources.

We used the number of available time slots to program an algorithm that counts the different bursts that can be observed in all possible states (even with resource reservation for voice traffic) and their respective lengths. The information is saved as a matrix. The algorithm uses this matrix to calculate the probability of finding a burst of a given number of consecutive idle time slots.

The probabilities of the different states were determined by the method described in the previous section. Therefore, the next step is to add together the probabilities of all of the states in the matrix that contain bursts of the given length of consecutive idle time slots. There are two other ways of determining the probability of finding a burst of a given length:

1) *Probability of finding a burst of exactly length k.* To calculate this probability we add together the probabilities of all of the states in the matrix that contain bursts of exactly k, consecutive idle time slots. We have to take into account that, even if a state has two or more bursts of length k, the probability of the state will contribute only once to the global sum.
2) *Probability of finding a burst of at least length k.* In this case we count bursts with $\geq k$ consecutive idle time slots ("cumulative probabilities"). For example, if a state has a burst of length 4, it will contribute to the global probabilities of finding bursts of length 1 to 4. As above, even if a state has two bursts of at least m consecutive idle time slots, it will contribute only once to each global sum.

The only other requirement is to set another parameter that we call the *Blocking Probability*. When dimensioning this type of system, the network operator usually plans the resources (i.e. the number of time slots) according to a given maximum blocking probability. For example, given the typical blocking probability of 1% and an 8 time-slot frame, the offered voice traffic is 3.13 Erlangs. For every possible state i we obtain the corresponding probability P_i. Fig. 6 shows that P_{255}, which represents the probability that all time slots are busy, is 0.01, i.e. 1%. It is very important to note the trade off between blocking probability, data rate and burst length.

In each case we are able to calculate the maximum throughput when the mobile station has any number of available time slots to allocate to its data channels. Fig. 7 and Fig. 8 show some results for the case of finding bursts of a given length with eight available time slots.

Fig. 7 shows five scenarios plotted on a single graph.

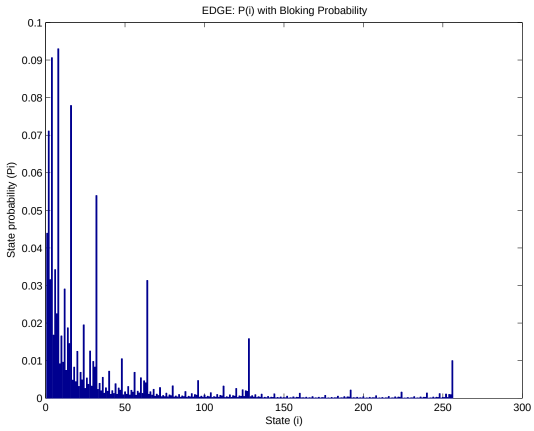

Figure 6. Probability of all states with a blocking probability of 1%

The first scenario consists of eight available time slots ($n = 8$) for any type of traffic and without resource reservation (using the notation given in the previous section: $n - m = 0$). In the following scenarios 1, 2, 3 or 4 time slots are reserved for voice traffic respectively. In this case, we determine the average burst length for a blocking probability of 1%.

If we define all bursts with a length equal to or less than 3 as short and all bursts a length greater than 4 as long, the major difference between the first scenarios in Fig. 7 and Fig. 8 is that the probability of finding long bursts is higher with a blocking probability of 1% than with 5%. Consequently, higher peak rates are obtained with the lower blocking probability (1%). In other words, when the system capability for voice and data traffic is high, there is also a high probability of data transmission with a high peak rate. In Fig. 7 and Fig. 8 we can see that the peak rates for long bursts are higher with a blocking probability of 1% than with a blocking probability of 5%. Note that this implies a considerable increase in user data rate with a blocking probability of 1%. Note that this fact implies a remarkable increase of user data rate in the case of blocking probability of 1%.

If we examine the scenarios with resource reservation for voice traffic in Figure 7 ($n-m > 0$), all cases share a common feature: the longest burst is concentrated at the burst of length 1 and the next longest burst corresponds to the highest possible burst value (m). However, the probabilities of finding consecutive idle time slots for the intermediate cases (burst length ($2 \leq x \leq m-1$)) are similar and lower than those for burst lengths of 1 and m.

A second observation of the cases with reserved time slots shown in Fig. 7 is that if the number of time slots reserved for voice traffic increases, the probability of finding bursts with a length equal to 1 also increases. This occurs because the number of shared time slots for data traffic and voice traffic decreases, thus reducing the number of possible consecutive idle time slots. In fact, these probabilities are concentrated on the burst of length 1, with the exception of the case $n-m = 3$ reserved time slots, in which the highest probability corresponds to the burst with a length of 5.

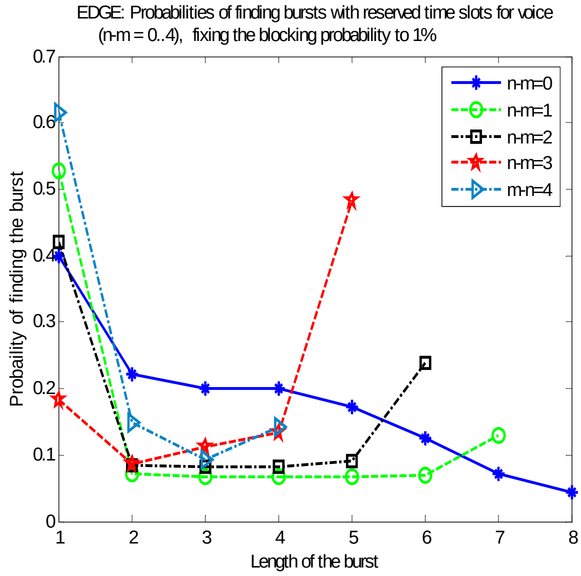

Figure 7. Comparison of scenarios with n-m= 0 to 4 reserved time slots and a blocking probability of 1%

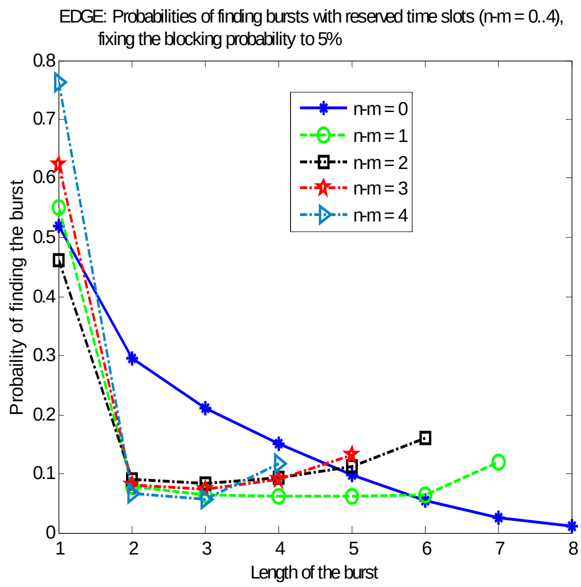

Figure 8. Comparison of scenarios with n-m= 0 to 4 reserved time slots and a blocking probability of 5%

Fig. 8 also shows the four scenarios with resource reservation but for a blocking probability of 5%. It can be seen that the probabilities of finding bursts with a length of 1 are greater than with a blocking probability of 1%, regardless of the number of reserved time slots. In contrast, the probabilities of finding bursts with the greatest possible length m are slightly lower than with a blocking probability of 1%. For the intermediate cases, in which $n-m = 2..m-1$, the probabilities of finding bursts of a given length are similar with blocking probabilities of 1% and 5%.

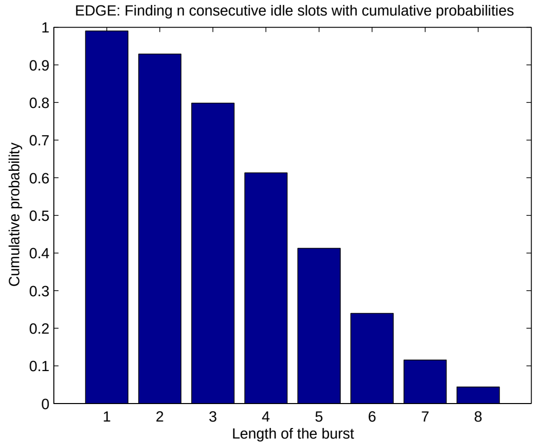

Figure 9. Cumulative probability of finding bursts with a blocking probability of 1%

In general, if the blocking probability increases and the number of shared time slots for data and voice traffics decreases, the total number of potential consecutive idle time slots also decreases. This demonstrates that an increase in the probability of all time slots being occupied considerably reduces the probability of finding consecutive idle time slots.

Based on the results, we believe that the proposed system configuration may provide a satisfactory peak data rate for a cellular mobile user who requires a regular minimum rate for transmitting or receiving data traffic, for example in an interactive application, with a burst of length 1. The user will also have a high peak rate (burst of length m) for downloading data traffic at the highest possible rate.

The network operator in the cumulative case without resource reservation shown in Fig. 9 will be able to estimate the total proportion of operating time during which a certain number of consecutive idle time slots are available in the system. For example, we can see that four or more consecutive idle time slots are available during approximately 60% of the operating time. Similarly, the figure shows that it is possible to find at least one idle time slot during almost the entire period. This is largely unsurprising, since there is at least one idle time slot in all possible states (except the one in which every time slot is busy, whose probability is the same as the blocking probability). In general, this case can be applied to the other scenarios with resource reservation.

V. Conclusions

In this paper we introduced a novel framework for radio resource management using schemes for allocating time slots in a cellular mobile network. We presented the models of these schemes and determined analytical solutions for channel allocation procedures in a TDMA environment.

Data transmission on a mobile terminal should use as much capacity as is available to improve the QoS perception. To meet this requirement, the network operator must allocate time slots consecutively. Therefore, radio resources are allocated using contiguous time slots to maximize the system utilization, which produces better results than other allocation strategies (for example, random or sequential allocation).

The method presented here can be used by planning teams to estimate whether they are able to fulfil QoS requirements other than the mean bit rate. We also presented a method for deriving equations from states diagrams, which can be used to determine the probabilities of the various states in a system.

We analysed the distribution of groups of consecutive idle time slots (bursts). By studying the distribution of the bursts, the network operator is able to evaluate the peak and average rates that can be assigned to network users. In addition, this analysis can be used to derive transmission times and relate them to transmission interruption due to handover and to determine the best strategy for using short transmissions.

We studied the results in detail and concluded that - without resource reservation for voice traffic - as the blocking probability is increased the probabilities of finding short bursts increase and the probabilities of finding long bursts decrease. In contrast to GSM (which is used predominantly for voice traffic), if we increase the blocking probability the efficiency of the system is reduced but if we decrease the blocking probability in our system (for voice and data traffic) there are more resources for data traffic and the bursts are longer. Consequently, these factors may improve the QoS requirements of the system.

We also found that the number of reserved time slots should always be as small as possible to create higher probabilities of finding long bursts of data traffic. This allows the network operator to find the optimal distribution of contiguous time slots to the different mobile stations.

Finally, we concluded that EDGE technology still requires a number of improvements in order to be considered a viable alternative to a UMTS. The findings of this study can be applied by network operators to optimize resource allocation for specific QoS profiles. Consequently, the burst length can be incorporated as a parameter in the scheduling algorithm in addition to the number of slots or priorities. For example, long bursts could be allocated to users or applications with low delay requirements to provide high peak rates.

References

[1] J. Eberspacher, H. Vogel, and C. Bettsbetter, “GSM-Switching, Services and Protocols,” in *New York. Ed. Wiley*, 2001.
[2] ETSI, “Technical Specification GSM 03.64. Overall description of the GPRS radio interface – stage 2,” in *ETSI V8.4.0*, 2000.
[3] E.T.S.I., “Selection procedures for the choice of radio transmission technologies of the universal mobile telecommunications system (umts),” in *ETSI UMTS 04-02 UTRA*, 1997.
[4] ETSI, “Technical Specification GSM/EDGE. Radio Access Network; overall description -stage 2,” in *3GPP TS 43.051 v.5.9.0*, (2003-2004).

[5] Y. Hwang, Y. Han, and Y. Kim, “Performance analysis of mixed voice/data service in a microcell-based PCS network,” in *IEICE Trans. Fundamentals, vol.E81-A, no. 6*, 1998, pp. 1136–1144.

[6] K. Mitchell and K. Sohraby, “An analysis of the effects of mobility on bandwidth allocation strategies in multi-class cellular wireless networks,” in *Proceedings of the IEEE INFOCOM’01*, 2001, pp. 1005–1011.

[7] J. Wang, Q. Zeng, and D. Agrawal, “Performance analysis of preemptive handoff scheme for integrated wireless mobile networks,” in *Proceedings of the IEEE GLOBECOM’01*, 2001, pp. 3227–3281.

[8] K. Kennedy and R. Litjens, “Performance evaluation of a hybrid radio resource allocation algorithm in a GSM/GPRS network,” in *Proceedings of the 10^{th} IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC ’99)*, 1999, pp. 3–26.

[9] H. Ferng and Y. Tsai, “Using priority, buffering, threshold control, and reservation techniques to improve channel-allocation schemes for the GPRS system,” in *IEEE Transactions on Vehicular Technology*, 2005, pp. 286–306.

[10] H. Araújo, J. Costa, and L. Correia, “Analysis of a Traffic Model for GSM/GPRS,” in *Proc. of the 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC ’01)*, 2001, pp. 124–128.

[11] P. Stuckmann and F. Müller, “GPRS radio network capacity and quality of service using fixed and on-demand channel allocation techniques,” in *Proc. of the IEEE VTC’01*, 2000, pp. 440–444.

[12] C. Lepschy, G. Minerva, D. Minervini, and F. Pascali, “GSM-GPRS radio access dimensioning,” in *IEEE 54th Vehicular Technology Conference*, 2001, pp. 558–561.

[13] M. Ivanovich, M. Zukerman, P. Fitzpatrick, and M. Gitlis, “Performance between circuit allocation schemes for half- and full-rate connections in GSM,” in *IEEE Transactions on Vehicular Technology*, 1998.

[14] P. Hiew and M. Zukerman, “Efficiency comparison of channel allocation schemes of digital mobile communication networks,” in *IEEE Transactions on Vehicular Technology, vol.49 NO.3*, May 2000, pp. 724–733.

[15] J. Itkonen, V. Salomaa, and J.Lempiäinen, “Air interface capacity for GPRS/EDGE over GSM traffic load,” in *IEEE 56th Vehicular Technology Conference*, 2002, pp. 396–400.

[16] P. Lin and Y.-B. Lin, “Channel allocation for GPRS,” in *IEEE Trans. on Vehicular Technology vol.50 no.2*, 2001, pp. 375–387.

[17] C. Ramírez, J. Paradells, and S. Mansilla, “Radio resource allocation strategies to guarantee data traffic in cellular networks,” in *3rd International Symposium on Wireless Communication Systems*, 2006.

[18] C. Ramirez and alt., “Analysis of consecutive idle time slots in mobile networks,” in *17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communication*, 2006.

[19] M. Zukerman, “Bandwidth allocation for bursty isochronous traffic in a hybrid switching system,” in *IEEE Transactions on Communications*, 1989, pp. 1367–1371.

Carlos M. Ramírez is currently a PhD candidate at the University of Las Palmas, Gran Canaria, Spain, where he also received his BSc and MSc degrees in Telecommunications Engineering. He has worked as a lecturer at the University of Las Palmas, Gran Canaria (1998-2002) and the Technical University of Catalonia (2002-2007), Spain. His research interests include mobile networks and radio resource management.

Sònia P. Mansilla received a BA in Mathematics from the University of Barcelona in 1996 and completed a PhD in Applied Mathematics in February 2001 at the Technical University of Catalonia. In 2001 she joined the Department of Computer Science at ETH Zürich as a post-doctoral research fellow and from 2002 to 2003 she was a Marie Curie post-doctoral fellow at the School of Mathematical Sciences, Queen Mary, University of London. She is currently a lecturer at the Technical University of Catalonia. Her research interests include combinatorics, in particular permutation groups, and the (finite or infinite) structures on which these can act (which may be graphs or digraphs, designs, geometries, etc.)

Josep Paradells Aspas was born in Barcelona, Spain. He received his PhD and his MSc degree in Telecommunications Engineering from the Technical University of Catalonia, Spain, where he is currently Professor of Telematics. His current research interests include mobile communications, heterogeneous networks, mesh and ad hoc networks and sensor networks.