Teletraffic Engineering in UMTS FDD Mode Networks

Miguel A. Caramés[1,2], Fernando J. Velez[1,3], Luis M. Correia[1]

[1] Instituto de Telecomunicações
Instituto Superior Técnico
Av. Rovisco Pais, 1049-001 Lisboa, Portugal

[2] School of Telecommunications Engineering
Polytechnic University of Valencia
46071 Valencia, Spain

[2] MOTOROLA - ANDC
Martínez Villergas, 52 Block 3
28027 Madrid, Spain

[3] Department of Electromechanical Engineering
University of Beira Interior
6200 Covilhã, Portugal

M. Carames@motorola.com, fjv@ubi.pt, luis.correia@lx.it.pt

Abstract

This paper presents a tool for multi-service traffic engineering in UMTS FDD mode. For each of the WCDMA system scenarios, the power restrictions and a given mixture of symmetric applications are considered. The maximum number of supported codes was obtained by previous simulations, taking into account carrier-to-interference constraints, leading to a given available data rate per cell. Given the spread-factor, with the basic resource being considered of 15 kb/s, the various applications request a certain number of basic code channels. A standard algorithm was used to compute each application's blocking probability, P_b, as a function of the fraction of active users in cells. In the *urban*1 scenario (20 % data at 40 kb/s plus 80% voice), using one carrier per cell, UMTS can support 26 speech users plus 7 data users for $P_b = 2$ %. If two carriers were used, it would support 60 speech users and 14 data users. In the *urban*2 scenario, as data at 320 kb/s is considered, system performance degrades, which can only be overcome by using three carriers per cells, which seems impracticable in an actual system. The consideration of terminal mobility leads to a slight decrease on the number of supported users.

1 Introduction

UMTS is the realisation of a new generation of mobile communications technology for a world where services will be based on a combination of fixed and wireless applications to form seamless end-to-end services to the user. UMTS is conceived as a multi-function, multi-service, multi-application digital mobile system that will provide personal communications, at rates ranging from 8 kb/s up to 2 Mb/s, depending on the environment. From the two different modes in UMTS, this work focuses on the WCDMA/FDD one, since it is the most interesting for operators, for the time being: it will provide wide area coverage and allows considerable terminal mobility, while the TDD mode will only be devoted to pico-cells and hot spots.

One of the key points in UMTS engineering success will be the development of new cellular planning tools. In WCDMA systems the planning procedure is more complex than in TDMA ones, e.g., GSM, since a larger number of parameters are involved, they being dependent from one another: transmission power, spreading factor, bit rate, channelisation codes and interference level. Besides, depending on the load, the cell coverage area fluctuates, being smaller when the number of active users is higher; hence, teletraffic engineering is a key element in the implementation of connection admission techniques such that the cell range is kept within the desired limits.

In this paper one presents a simple tool that can be used as a first approach to traffic engineering in UMTS. The shared resource in a WCDMA system is power, more specifically the base station transmitted power. One focuses on the downlink, since the shared resources of interest to the planner are the code channels, or equivalently the channelisation codes, whose number is limited by the interference power level. In the uplink, there is no limitation of this kind since different scrambling codes are used for each user.

The power that is allocated at the downlink depends on the bit rate being transmitted or, equivalently, on the spreading factor. The spreading factor and the number of channelisation codes are directly related by means of the OVSF tree [1]. Hence, the traffic engineering allows us to obtain the upper bounds for the way these channelisation codes are distributed among the users of the various available applications [2].

One starts by presenting the performance measures and some assumptions in Section 2. In Section 3 one can find a brief description of the system simulations [3], [4], given the available data rate in the cell that will be used as an input parameter in the traffic model. Teletraffic engineering aspects are presented in Section 4. First, one defines the main parameters of the model, and then one presents the model itself. In Section 5, the scenarios under consideration are defined. In Section 6, results for the blocking probability and for the supported fraction of active users are presented for the *urban*1 and the *urban*2 scenarios, in the absence of mobility; finally, the impact of mobility is discussed for the *urban*1 scenario. Conclusions are drawn at the end.

2 Performance Measures

The considered model is being constructed over the main ideas of the model developed in [5], which provides a standard algorithm for multi-service traffic blocking probability computations.

ETSI considers that the performance measure for UMTS corresponds to 98% of the users being satisfied. A user is satisfied if the three following constraints are fulfilled simultaneously:

1. The user does not get blocked when arriving to the system.
2. The user has sufficiently good quality more than 95% of the session time.
3. The user does not get dropped due to BER requirements.

The WCDMA/FDD system simulation from [3] has been used as an input, since it copes with the last two requirements, while the traffic analysis deals with the first one (maximum blocking probability equal to 2%). The combination of the results from the three constraints can be used for a first capacity estimation in the cellular planning process.

The main assumptions and hypotheses under which the model can be considered are the following:

1. All service components are circuit switched, hence, packet switched services are assumed to behave somehow like circuit switched ones. When planning a cellular network, one must be conservative and consider the worst possible case. Considering circuit switched services implies that the activity factor is equal to 1 (for data), which means that once that a radio resource (channelisation code) is allocated to one user it keeps being used by this user until the communication is released. If one considered packet switched services the activity factor would decrease, hence, more users could be served.
2. The study focuses on the downlink. The performance measurement being considered is the blocking probability, P_b, it being the probability no resources (code channels) can be allocated to a user. At the uplink there is no limitation in the number of channelisation codes, because a different scrambling code is assigned to each active user, such that one has access to all the codes in the OVSF code tree. On the other hand, one BS transmits using one or two scrambling codes, hence, the number of channelisation codes is limited.
3. Variable rate services are evaluated using a fixed rate bearer.
4. When focusing on the downlink path, the fact of having services requesting different bit rates can be modelled by considering only one spreading factor (one basic resource) in each cell. Thus, only one kind of code channel is available, and only some data rates are available through code aggregation.
5. Basic code channels, with a bit rate of 15 kb/s, are being used in order not to loose efficiency in the voice service. Thus, one voice user requests one basic code channel.
6. Different applications can be multiplexed over a code channel if the aggregated data rate is lower than or equal to the considered basic channel data rate. When one user wants to transmit data of different services, he/she gets assigned a specific output power/rate threshold. The aggregate rate of all the services must be below this threshold.
7. Higher data rates are available through code aggregation, i.e., one user of the data service LCD 320 (320 kb/s of net data rate) requests 32 basic code channels.
8. Only dedicated channels are considered since common channels are only used to carry small amounts of traffic.
9. No mobility is considered at the beginning. When mobility is taken into account, a simple approach is used.

3 System Simulations Overview

Teletraffic engineering deals with the way available resources are shared among the users coping with the target performance measures. However, WCDMA systems are interference limited and there is not a simple analytical way to know the available resources in one cell. The overall data rate per cell depends on the location of the mobiles (distance mobile / base station), power levels, the power control algorithm parameters, etc.; this is the reason why a system simulation tool was used to obtain the available resources per cell before the traffic analysis.

The system copes with the bit error rate (BER) requirements, whilst guaranteeing that BER is higher than a $BER_{threshold}$ only in 5% of the time. This requirement is satisfied when the average signal to interference ratio (SIR) of the radio link (carrying the data from one user) is equal to a target value, SIR_{target}. The details on the structure of the simulator can be found in [3] and [4].

Its output is the available bit rate per MHz and per cell, in kb/s/MHz/Cell, for single and for mixed services. For an available bandwidth of 5 MHz, it is straightforward to obtain the available bit rate per carrier and per cell (or per sector, when sectorisation is used).

The uplink path is especially critical because the maximum transmission power of a mobile is limited (typically 21 dBm, against 40-46 dBm of the Node B). In [4] one can find the results obtained after simulating the uplink, while in [3] one also finds results for the downlink. As symmetric applications are considered in the FDD mode, the most critical path will limit the system performance in both links. In this way, the SIR requirements are guaranteed and also the maximum blocking probability.

4 Traffic Model

4.1. Main Parameters

One should start by defining the parameters being used in the model [5]:

N	is the number of available codes per cell. It is obtained from simulations [4], dividing the total amount of resources (in kb/s) by the basic code channel bit rate.
M	is the number of potential users in a cell.
C	is the capacity vector and gives the number of code channels that each application demands.
K	is the number of applications being available in a cell.
$U(t)$	is the vector that defines the number of active users of each application in a time instant t.
λ_k	is the arrival rate for the static case (Λ_k when mobility is taken into account); the arrival process follows a Bernoulli distribution.
μ_k	is the service rate for the static case (H_k when considering mobility), and is Poisson distributed.
A_k	is obtained by dividing λ_k by μ_k (or Λ_k by H_k when mobility is considered), and it is the traffic generated per user, for each application k.
$prop_k$	is the proportion of users of an application k among the K available ones.
b_k	is the application k data rate.

The arrival rate can be obtained through:

$$\lambda_k = (U_k - M_k)\cdot(-\beta_k)\,. \tag{1}$$

U_k is the number of active users of the application k, M_k is the number of potential users of the application k being in the system and β_k is the activation rate ($\beta_k < 0$ in the Bernoulli case of the Bernoulli-Poisson-Pascal model [5]).

Fig. 1 shows the way applications are activated. A user can be either in an idle state or using one of the K applications.

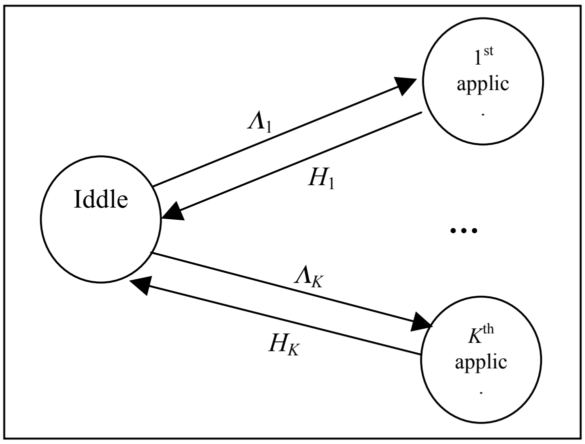

Fig. 1: Model for user applications activation.

The proportion of users of an application among all the available ones can be expressed by its usage:

$$prop_i = \frac{\Lambda_i / H_i}{\sum_{k=1}^{K} \Lambda_k / H_k} = \frac{A_i}{\sum_{k=1}^{K} A_k}\,, \tag{2}$$

hence,

$$A_i = A \cdot prop_i\,, \tag{3}$$

where

$$A = \sum_{i=1}^{K} A_i\,, \tag{4}$$

A being the total traffic. By writing the system of equations for the probability of transition between states (plus the normalisation equation), it is straightforward to obtain the probability of a user having an active application:

$$p_i = \frac{A_i}{1 + \sum_{k=1}^{K} A_k} = \frac{A}{1+A}\cdot prop_i = f \cdot prop_i \tag{5}$$

where f is the fraction of active users,

$$f = \frac{A}{1+A} \,. \qquad (6)$$

Multiplying f by the population of potential users, M, one obtains the number of users being simultaneously active in a cell. The system cell average load can be obtained by (in kb/s or Mb/s):

$$L = f \cdot c_{av} \cdot M \qquad (7)$$

where:

$$c_{av} = \sum_{i=1}^{K} prop_i \cdot b_i \,. \qquad (8)$$

c_{av} gives information about the average amount of resources (in kb/s or Mb/s) that is used by each user.

4.2. Theoretical model

The main objective of the model is to obtain an algorithm to compute P_b, having the parameters defined above (f and L) as inputs.

The number of channels used at an instant t is given by

$$Y(t) = U(t) \cdot C \,. \qquad (9)$$

The set of feasible states gives the number of active users of each application that can be served by the system and is defined by

$$U = \{ n \in N^K : n \cdot C \le N \} \qquad (10)$$

Blocking situations, i.e., the ones when a new user arriving to the system does not find enough resources available can be expressed by:

$$B_k = \{ n \in U : n \cdot C + C_k \le N \} \qquad (11)$$

where B_k is the set of blocking states for application k, and C_k is the number of channels requested by application k. In a blocking situation, the request is cleared, which means that the system remains in the same state.

Application k blocking probability is obtained by dividing the expectation of the number of blocked requests by the total number of class k requests,

$$P_b^k = \frac{\sum_{n \in B_k} \lambda_k(n_k) \cdot p(n)}{\sum_{n \in U} \lambda_k(n_k) \cdot p(n)} \qquad (12)$$

The state probability marginal function, $p(n)$, represents the probability of the system being in the state n or, equivalently, the probability of n users being in the system,

$$p(n) = \frac{\prod_{k=1}^{K} v_k(n_k)}{\sum_{n \in U} \prod_{k=1}^{K} v_k(n_k)} \,, \text{ for } n \in U \qquad (13)$$

where the non-normalised marginal probabilities, $v_k(n_k)$, are obtained for each applications and give the probability of having exactly n_k users of the application k in the system,

$$v_k(n_k) = \binom{M_k}{n_k} \cdot (-\beta_k)^{n_k} \qquad (14)$$

A standard algorithm for multi-service traffic [5] was used for P_b computations, which does it in a time-efficient way. A simplified flowchart of the algorithm can be seen in Fig. 2.

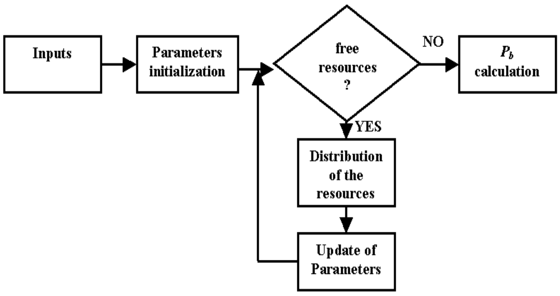

Fig. 2: Flowchart of the algorithm.

Results are going to be obtained for P_b as a function of the fraction of active users or, alternatively, as a function of the average load. Based on these results, a P_b threshold of 2% is then considered, and the maximum number of simultaneous active users supported by the system is obtained, as well as the cell resource occupancy (i.e., the spectral efficiency).

5 Scenarios Definition

In Table 1, one can find a description of the services being considered. LCD means Long Constraint Data, LCD40 being a circuit switched data service (i.e., Internet/Intranet access), while LCD320 may be a low/medium multimedia service (e.g., interactive games). The considered basic code channel bit rate is 15 kb/s. Each application requests a certain number of basic code channels.

Table 1: Description of the services.

Service	Average service duration [s]	Typical data rate [kb/s]	Physical channel rate [kb/s]	Channels per service
Speech	60	8	15	1
LCD40	156	40	60	4
LCD320	14	320	480	32

Two scenarios are considered for the two micro-cell structures one is assuming:

- In the *urban*1 scenario only Speech and LCD40 services are available. It corresponds to areas where high mobility can be accepted like, for example, the bounder limits of cities.

- In the *urban*2 one, Speech, LCD40 and LCD320 services can be requested. Only low mobility can be assumed, it being the scenario that characterises the central business district (higher data rates would be provided in an indoor environment by combining the TDD and FDD modes).

Both scenarios are defined in Table 2, the number of basic code channels being an input from previous system simulation [3], [4]. The usage refers to the proportion of expected users of each service, among all the services available in one cell. It is worthwhile noting that in the *urban*2 scenario more than one carrier should be allocated in order to provide the LCD320 service. One should also note that only the downlink is being examined, since at the uplink each mobile station transmits within a different scrambling code, thus, no limitation existing in the number of channelisation codes. Other scenarios have been considered, and results can be found in [6].

Table 2: Scenarios characteristics.

Scen.	**Carriers per cell**	**Ch. per cell**	c_{av} **[kb/s]**	**Services**	**Usage [%]**
Urban 1	1	90	24.0	S	80
				LCD40	20
	2	180		S	80
				LCD40	20
Urban 2	2	90	70.5	S	70
				LCD40	20
				LCD320	10
	3	135		S	70
				LCD40	20
				LCD320	10

6 Performance Evaluation

6.1. *Urban*1 Scenario

Two different situations have been analysed, considering one or two carriers per cell. When one carrier is considered, the curves for P_b as a function of the fraction of active users are the ones from Fig. 3, for both Speech and LCD40 users. P_b is obtained for both services for a set of values of f varying from 10 to 50 %. For a given P_b threshold, the most restrictive curve determines the supported fraction of active users; multiplying by M one obtains the number of active users in the system. Taking the usage of each application into account, it is straightforward obtaining the number of each application active users.

One concludes that, despite only 20 % of the users are LCD40 ones, data service limits the system, since its P_b is higher than the one associated with speech. From the curves it is also straightforward to obtain the fraction of active users supported by the system for a P_b threshold of 2 %. The LCD40 curve is the first one that reaches the 2 % value, and it happens for f = 33 %. Since one is considering a population of M = 100 users, one concludes that 33 users can be active simultaneously, corresponding to 26 speech users and 7 data ones.

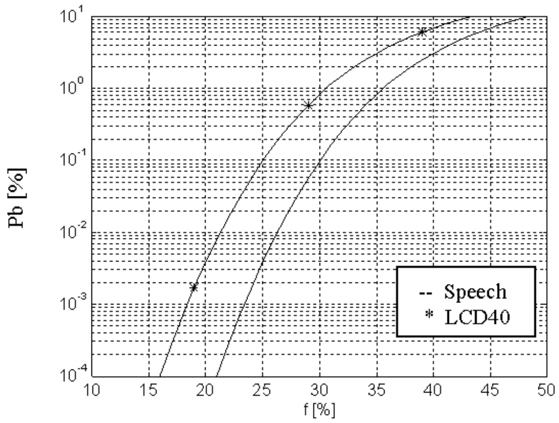

Fig. 3: P_b in terms of f, for 1 carrier per cell.

The system load, in kb/s, is obtained from (7), and gives information about the amount of resources that are used for a blocking probability threshold of 2 %. As it occurs for a system load of 800 kb/s, a spectral efficiency of 58 % is obtained taking into account that the total available bit rate in the cell is 1 360 kb/s.

An improvement in system capacity can be obtained by assigning one more frequency carrier to each cell. The results from Fig. 4 have been obtained by allocating a pair of carriers, and for a number of potential users of M = 200. Assuming one more time that the same P_b threshold (2 %) is being considered for both services, one obtains a maximum number of users being active at the same time in a cell of 74 users: 60 Speech users and 14 LCD40 ones.

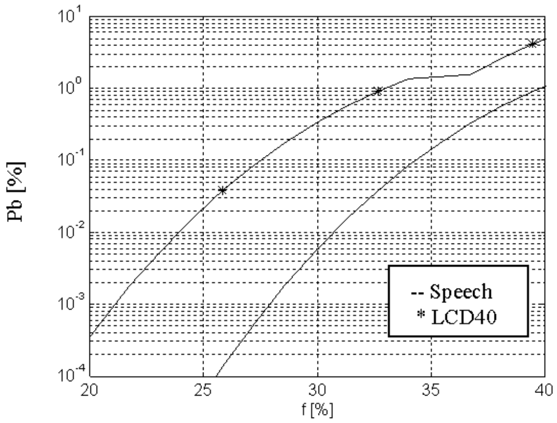

Fig. 4: P_b in terms of f, for 2 carriers per cell.

One should note that system performance improves to more than twice, as it could be expected. In Fig. 4 one observes an irregularity in the LCD40 curve for values of the fraction of active users in [33, 37] %. The reason for this is the resource allocation procedure being followed by the algorithm. In fact, the LCD40 curve modifies its slope near the points where one more data user "enters" in the system (f = 20, 25, 30, 35, 40 %). Following the same procedure described before, one obtained 66 % for the spectral efficiency, which represents an improvement relatively to one carrier per cell.

6.2. *Urban*2 Scenario

If only one carrier per cell would be considered, the system would be highly limited by the LCD320 service, not allowing for any active user of this service. Considering two carriers per cell, the curves for P_b as a function of f are the ones from Fig. 5.

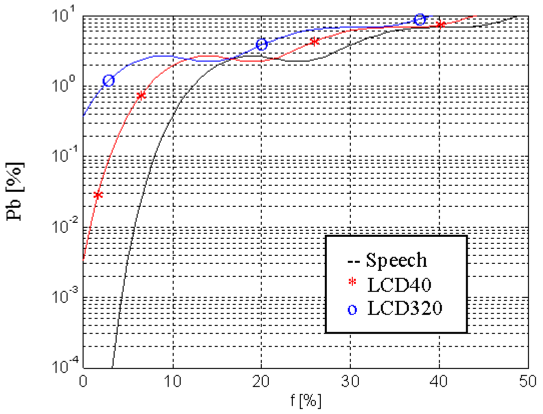

Fig. 5: P_b in terms of f, for 2 carriers per cell.

As all the channels are shared among all the users, when a new user enters in the system he demands some resources and the system will allocate them between both carriers, in a balanced way.

For P_b = 2% and M = 100, the maximum number of active users in the system is 5 (4 Speech, LCD40, and 0 LCD320 users), which is clearly unacceptable when two carriers are allocated to a cell: a possible solution for this limitation would be to consider a different threshold for data services. Therefore, the FDD mode is far from being the optimal solution to cope with data services requiring high bit rates, which aggravates if one considers that most of these services are highly asymmetric.

Due to the low values of the spectrum efficiency obtained from the algorithm, one has to consider the situation where three carriers are allocated to one cell, Fig. 6 (M = 100). In this case, for P_b = 2 %, the FDD mode supports a maximum of 19 users: 15 Speech, 3 LCD40, and 1 LCD320 ones.

6.3. The impact of mobility

The impact of mobility is considered only for the *urban*1 scenario. The change consists only in increasing the arrival and service rates, taking into account that two contributions exist, from new and handover calls [7].

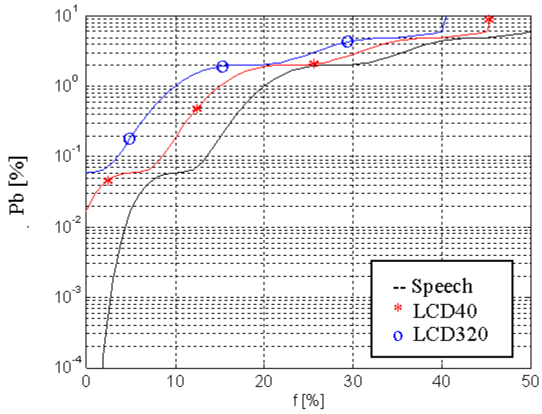

Fig. 6 P_b in terms of f, for 3 carriers per cell.

Mobility does not affect the computation of the blocking probability, for a given density of users and fraction of active users, determining however the proportion of new/handover calls [8]. Depending on the scenario, system capacity will be limited either by the handover failure probability, P_{hf}, or by P_b. The mobility requirement is usually given by the dropout probability threshold, $(P_d)_{max}$ = 0.5 %, being immediate to relate it with the handover failure probability one. As it was shown in the static case, the system is limited by the LCD40 service. Thus, the impact of mobility is being considered for this service. The mobility model is described in [7], [8] and [9], its main assumptions being the following: velocity follows a triangular distribution with an average value $\overline{V}$ and deviation Δ (both given in $m \cdot s^{-1}$). Assuming a cell coverage distance of 400 m, Table 3 shows the values for the cross-over rate (η) and for the remaining mobility model parameters.

Table 3: Mobility model parameters for various mobility environments.

Mobility environment	$\overline{V}$ $[m \cdot s^{-1}]$	Δ $[m \cdot s^{-1}]$	η $[s^{-1}]$	$(P_{hf})_{max}$ [%]
Pedestrian	1	1	0.0009	3.60
Vehicular	10	10	0.0090	0.36
Main Roads	15	15	0.0135	0.24

One verified that system performance was not limited by mobility when only low and pedestrian environments were considered; however, when high mobility is taken into account (vehicular and main roads environments), handover failure is the main limitation of the system. It is straightforward to obtain the fraction of active users for $M = 100$ using the curves for P_b from the *urban*1 scenario, Table 4. One assumed that for multi-service systems the equation for P_{hf} is equal to the one for P_b, as in single-service systems [7] (as no guard channels for handover are considered).

Table 4: Results for the fraction of active users and for the number of active users.

Mobility env.	f [%]	Speech active users	Data active users
Pedestrian	33	27	6
Vehicular	28	23	5
Main Roads	26	21	5

The fact of considering one isolated cell highly limits the study of the impact of mobility in UMTS. A more detailed work should include the soft handover feature, which is referred as "use before make" in contrast with the "make before use" definition of hard handovers. When one mobile is in soft handover, it is simultaneously using resources of 2 or more cells. For this to be possible, all the cells used in soft handover must use the same frequency. In order to decide to include/exclude cells of soft handover situation, the mobile measures and report the cell strength of a set of cells. The overall effect of being in soft handover should thus consider that:

a) Each mobile in soft handover is using resources from several cells.
b) Both the up- and downlink powers will decrease due to the combination mechanisms implemented in the mobile and in the UTRAN side, to get advantage from this situation.

Further work is necessary in this area.

7 Conclusions

A tool for teletraffic engineering was developed for the UMTS WCDMA/FDD mode. The performance measure being considered is the blocking probability, P_b, and one was interested in obtaining the number of supported active users for a given P_b threshold, the maximum accepted value being $P_b = 2$ %, derived from UMTS satisfaction requirements (98 % of the users must be satisfied). One mobile is blocked when the UTRAN can not allocate resources to this mobile. A previous system simulation determined the number of available resources in the cell according to signal quality requirements. Results are summarised in Table 5 for the *urban*1 and *urban*2 scenarios.

One concludes that the FDD mode presents high capacity in the *urban*1 scenario, spectral efficiencies of 58 and 66 % being obtained when one uses one or two carriers per cell, respectively.

Table 5: Summary of the results

Scenario	**Number of**				**Spectral efficiency [%]**
	Carriers per cell	**Speech users**	**LCD40 users**	**LCD320 users**	
*Urban*1	1	26	7	-	58
	2	60	14	-	66
*Urban*2	3	15	3	1	66

However, when high data rates services are considered (*urban*2 scenario), one would only obtain a high spectral efficiency if 3 carriers were used. Otherwise, with only two carriers, the spectral efficiency is 26 %.

System performance highly depends on the high data rate service users, i.e., the number of simultaneous users in the system is limited by the blocking probability curves for the data services, further investigation having to be done to allow higher blocking probabilities in data services, so that the number of active users may increase.

Taking terminal mobility into account, and assuming a maximum dropout probability of 0.5%, the results for the supported fraction of active users in the *urban*1 scenario were kept in the same order of magnitude of the values for the static case, being, however, slightly lower. Thus, one concludes that in this scenario considerable mobility is allowed, performance results being within acceptable margins.

8 References

[1] H. Holma and A. Toskala (eds.), *WCDMA for UMTS*, John Wiley, Chichester, UK, 2000.

[2] UMTS Forum, *Minimum spectrum demand per public terrestrial UMTS operator in the initial phase*, UMTS Forum Report No. 5, London, UK, Dec.1998.

[3] H.R. Pinto and J.G. da Silva, *Capacity Estimation of the FDD Mode of UMTS*, Graduation Thesis, IST - Technical University of Lisbon, Lisbon, Portugal, Mar. 2000.

[4] H.R. Pinto, J.G. da Silva and A. Rodrigues, "Uplink Capacity of the WCDMA FDD Mode in UMTS Networks for Mixed Services," in *Proc. of VTC'2000 Fall - IEEE Vehicular Technology Conference*, Boston, MA, USA, Sep. 2000.

[5] G.A.Awater and H.A. van de Vlag, "Exact Computation of Time and Call Blocking Probabilities in Large, Multi-traffic, Multi-resource Loss Systems," *Performance Evaluation,* Vol. 25, No. 1, Mar. 1996, pp.41-58.

[6] M. Caramés Garcia, *Analysis of Multi-service Traffic in UMTS FDD Mode Networks*, Graduation Thesis, IST - Technical University of Lisbon, Lisbon, Portugal, May 2000.

[7] B. Jabbari, "Teletraffic Aspects of Evolving and Next-Generation Wireless Communication Networks," *IEEE Personal Communications Magazine*, Vol. 3, No. 6, Dec. 1996, pp. 4-9.

[8] F.J. Velez, "Multi-service Traffic Analysis in Mobile Broadband Systems," in *Proc. of 4th ACTS Mobile Communications Summit,* Sorrento, Italy, June 1999, pp.239-244.

[9] F. Velez and L.M. Correia, "Capacity Trade-offs in Mobile Broadband Systems using Guard Channels for High Mobility Handover," in *Proc. of PIMRC'98 - 9th IEEE International Symposium on Personal Indoor, and Mobile Radio Communications*, Boston, MA, USA, Sep. 1998.