Modified Stochastic Knapsack for UMTS Capacity Analysis

●Mona Jaber ●Syed Ammar Hussain ●Adel Rouz

(Manuscript received June 6, 2002)

This paper is a brief introduction to the complex task of UMTS (Universal Mobile Telecommunications System) network dimensioning, with emphasis on capacity analysis and the innovative solution developed by FLE for this purpose.
Capacity analysis (an integral part of dimensioning) is a study that allows the estimation of required resources that can guarantee a defined level of satisfaction for the users. Compared to fixed and 2G mobile networks, this task is much more challenging in 3G systems due to two main reasons. First is the fact that the cell edge in UMTS is continuously moving according to the traffic load, making dimensioning ambiguous and leaving room for under or over dimensioning. The second challenge is an outcome of the multiple data rates and switching modes supported by the 3rd generation networks. Traffic and capacity analysis for such mixed services is a new concept which was addressed by Fujitsu by the use of a modified Stochastic Knapsack traffic model. The Stochastic Knapsack is a multi service traffic model by definition and is generally used in ATM multiplexer dimensioning. This paper introduces the steps involved in UMTS air interface dimensioning and describes in detail how the basic stochastic knapsack was modified to provide a traffic model that captures the peculiarities of UMTS traffic.

1. Introduction

Dimensioning of the radio access network for W-CDMA (Wideband Code Division Multiple Access) is a very important but complex task. The purpose of dimensioning is to estimate the amount of hardware or resources required to provide a satisfactory grade of service (GoS) to users. It is important because it is the initial step considered in network implementation; consequently, radio dimensioning will impact the network significantly. Its complexity arises from the fact that the cell edge in a CDMA environment is continuously moving. A CDMA cell radius varies with changes in the amount of interference in that area. These changes occur due to traffic variations; for example, the cell radius for 10 users using a video service will not be the same as the cell radius for 10 voice users. Similarly, for a 2G CDMA system, the cell radius for 10 voice users will not be the same as the cell radius for 30 voice users. This characteristic is known as Cell Shrinking.

W-CDMA faces a new challenge compared to IS95[note 1)] due to its support for multiple data rates and switching modes. This feature further complicates the estimation and prediction of interference levels and therefore the position of the cell edge.

Having said that, the similarity between W-CDMA and CDMA is still significant and Fujitsu is very well positioned to utilize its extensive experience in CDMA-one system design and optimization to improve the approach to W-CDMA

note 1) A 2G CDMA standard

network design.

This paper briefly discusses the entire dimensioning process with an emphasis on capacity analysis with the new traffic model.

2. Radio dimensioning

The objective behind the analytical dimensioning process is to simplify the complex task of dimensioning by making some assumptions and approximations. The amount and accuracy of these hypotheses impact on the reliability one can expect from this process. The major approximation is obtained by considering a flat and homogeneous landscape. That is to say, the target area is divided into different morphology classes, or clutter types, which are considered separately on the assumption that each clutter forms an exclusive homogeneous area. This assumption results in identical sites per clutter, which may be considered "prototype" sites in the planning stage. The flow chart in **Figure 1** describes the analytical radio dimensioning procedure.

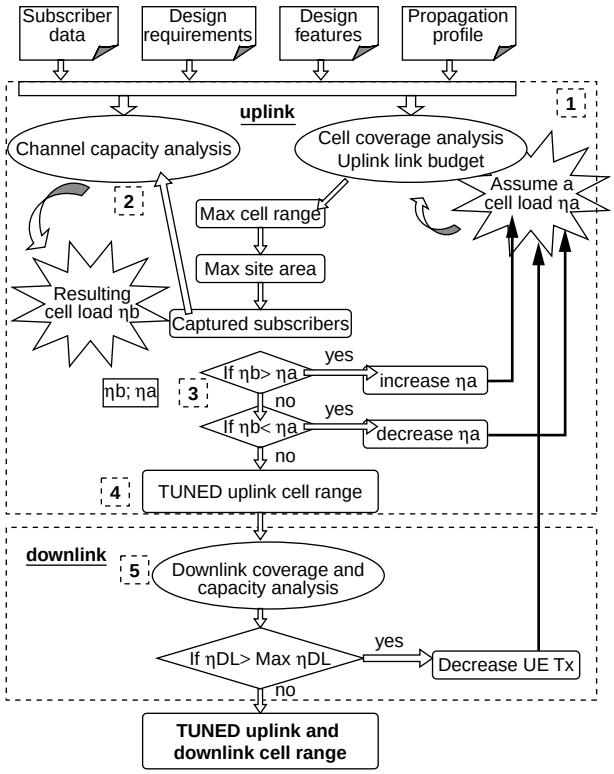

Figure 1
Analytical dimensioning.

2.1 Traffic modelling with binomial approach

Traffic models like Erlang-B, Erlang-C, etc., are established models which can model single-service, circuit-switched traffic quite accurately. However, there are no established ways for modelling multi-service traffic in UMTS. In the early stages of developing a dimensioning procedure for W-CDMA, it was considered that the traffic generation probability conforms to a binomial distribution for both packet-switched and circuit-switched services. This lead to the use of the Binomial model for air interface dimensioning.

However, the Binomial model could not differentiate between packet switched and circuit switched services. Furthermore, it only considered an aggregate GoS; that is, it did not guarantee a certain GoS per service type. For instance, securing a 2% aggregate GoS would probably result in a higher GoS for data services and an equal or lower GoS for voice services. Another issue with the binomial approach is that the probability of "success" or "failure" is always the same, regardless of the current load of the system.

The binomial approach was considered pessimistic and therefore a safe approximation for traffic modelling. However, through a detailed comparison of the Binomial and Knapsack models, we found that this assumption was not always true.

The binomial distribution is a discrete distribution (**Figure 2**). It gives the likelihood of finding x failures (or dropped or blocked calls) as opposed to success (or admitted calls). The findings are the results of n experiments (or call attempts). The binomial formula that represents the probability of k successes (admitted calls) in N experiments (call attempts) is:

$$\Pr(k, N) = \frac{N!}{(N-k)! \cdot k!} \cdot (1-p)^{N-k} \cdot p^k \qquad (1)$$

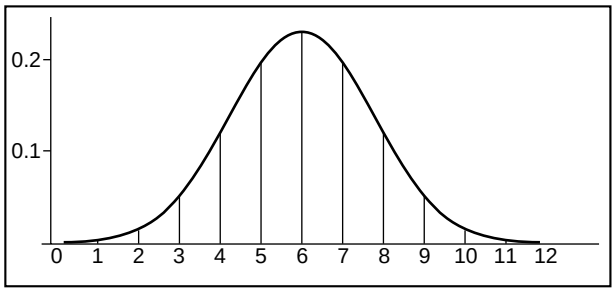

Figure 2
Binomial distribution curve.

Consider an experiment in which a coin is flipped n times. Each trial has two possible outcomes, heads or tails, and these outcomes are coded as 1 and 0, respectively. The trials are independent. The probability of heads p is constant from one trial to another. For example, when a coin is flipped 12 times, n is 12 and p is 0.5.

The black vertical lines in Figure 2 represent the discrete binomial output. The y-axis represents the probability of each outcome. As expected, getting 6 heads is the most likely outcome of 12 flips. The trend (bell-shaped) line represents the normal distribution (continuous distribution), which is often used to approximate the discrete binomial.

Example: Simplified UMTS traffic

1) Two services are equally used: voice at 12.2 kb/s and circuit-switched data at 64 kb/s.
2) The traffic load for voice is 30 mErl per user.
3) The traffic load for data is 3 mErl per user.
4) The pole capacity[note 2)] is 40 voice users, which represents the peak number of voice users.
5) The required GoS-blocking rate is 2%.

In this application, the binomial distribution is approximated to a normal distribution. First, the inverse of the standard normal cumulative distribution (mean = 0, standard deviation = 1) is derived for a probability of 98%. This gives a value of 2.054, and the following formula applies:

$$\text{Peak traffic} = \text{mean traffic} + \sigma \cdot 2.054 \qquad (2)$$

note 2) The pole capacity is the maximum capacity of a UMTS cell (capacity is limited by the amount of interference).

Where:
Peak traffic = pole capacity = 40
Mean traffic = $N \times 0.03 + N \times (64/12.2) \times 0.003 = 0.0457 \cdot N$

In the above equation, it is assumed that a data user consumes 64/12.2 resources, whereas a voice user consumes 1 resource. N is the number of attempts, and s is the standard deviation for blocked/admitted calls on the assumption that there are equal numbers of voice and data attempts.

$\sigma^2 = N \times 0.03 \times (1 - 0.03) + (64/12.2)^2 \times N \times 0.003 \times (1 - 0.003) = 0.1114 \cdot N$

Peak traffic can be rewritten as:

$$\text{Peak traffic} = 0.0457N + 2.054 \times \sqrt{0.1114} \times \sqrt{N} \qquad (3)$$

Solving for N gives the following results:
N = 530 attempts
Mean traffic = 24

This implies that one W-CDMA carrier with a pole capacity of 40 is dimensioned for 744 subscribers at an aggregate blocking probability of 2%.

This method may be acceptable in some cases where all services are circuit switched and a global grade of service is sufficient to estimate the performance of the system. Unfortunately, this is not the case for UMTS air interface traffic, and consequently an improvement to this approach is needed.

2.2 Stochastic knapsack

Stochastic models are popular in ATM (Asynchronous Transfer Mode) and other packet switch traffic analysis. The model considers a fixed resource and simulates the respective GoS of the various supported services. Due to its fixed capacity feature, the knapsack is not valid for UTRAN capacity calculation in its original form. We therefore modified the basic stochastic knapsack model so that it can capture the peculiarities of the UTRAN air interface by allowing flexible resource availability and variable data rates.

Our study not only proved that the Knap-

sack model allows for a more realistic representation of the actual traffic behaviour, but it also showed some hidden flaws in the previous binomial approach. For instance, the binomial approach has an uncontrollable Max GoS and, despite its reputation as a safe approach, it can result in under-dimensioning.

The original Knapsack model is similar to Erlang-B,[note 3)] except that the objects handled by the network can be grouped into K different classes. In class k, each object has a size of b_k, an arrival rate of λ_k, and a mean call duration of μ_k. The size of each object is the amount of network resource it occupies, (i.e., the data rate of the connection: 12.2 kb/s, 384 kb/s, etc.).[1)]

The stochastic knapsack has a fixed capacity C. A call is blocked if the knapsack does not have enough free capacity to take the call, as shown in **Figure 3**.

The model gives us the following equation.[2)]

$$P_{B,k} = 1 - \frac{\sum_{set\ Sk}\left[\prod_{j=1}^{K} \frac{n_j}{\rho_j} / n_j !\right]}{\sum_{set\ S}\left[\prod_{j=1}^{K} \frac{n_j}{\rho_j} / n_j !\right]} \qquad (4)$$

Where:

- $P_{B,k}$ is the probability of blocking an object in class k.
- S_k is the set of states that can accommodate an object from class k.
- S is the set of all the knapsack states.
- ρ_j (λ_j/μ_j) is the network load from class j.
- n_j is the number of objects in the knapsack from class j.

2.3 Fujitsu knapsack model and uplink (UL)[note 4)] capacity analysis

The basic stochastic knapsack cannot be applied directly for UMTS air traffic modelling. As explained in the previous section, the Knapsack model is based on a fixed resource (e.g., a buffer or a fixed-capacity link). This is not the case with the UMTS air interface, because there is no clear concept of resource unit. If we try to apply the Knapsack concept to 2G systems, then it is easy to assume that the maximum allowed number of users is the peak capacity; hence, the resource unit is an active user. In this scenario, the stochastic knapsack capacity C would be equal to say 7 (1 GSM carrier supports 8 timeslots: 1 time slot is reserved for control and the other 7 are used for user data) and the object size b_k would be equal to 1. Alternatively, the knapsack size could be equal to the maximum throughput per carrier (spectrum efficiency) and b_k could be equal to the data rate.

Both applications are possible to implement because only one type of traffic is considered. However, for UMTS it is not a straightforward application for two reasons. Firstly, the maximum capacity of one carrier (pole capacity) is not a fixed

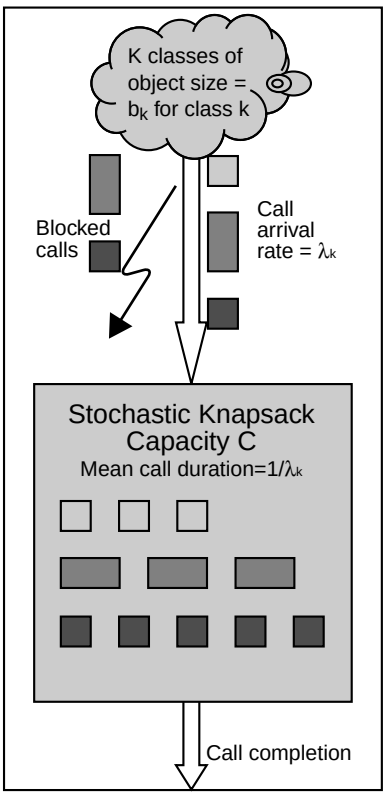

Figure 3
Stochastic knapsack.

note 3) This is a famous traffic model for circuit switched traffic.

note 4) Radio link when the mobile is transmitting and the base station is receiving.

value; it depends on the traffic mix, the quality of service requirements within the cell, and the traffic activity in neighbouring cells. Secondly, it is not obvious how to define the resource unit in this case. If the resource unit is equal to one user, is the user a voice, CS data, or PS data user? If the resource unit is kb/s, then the object size could be the data rate, in which case, the knapsack size would be the maximum throughput per carrier (spectrum efficiency). However, this application overlooks one important factor—the QoS (Quality of Service).

Let us consider a scenario with two services at 64 kb/s. Service 1 is for e-mail (QoS is low ⇒ Eb/No[note 5)] is low [Eb/No is energy per user bit over spectral noise density]).[3)] Service 2 is for video conferencing (QoS high ⇒ Eb/No high). A higher QoS means a lower block error rate, which would require a higher level of energy per user bit to secure a lower rate of errors at the receptor. A higher Eb/No requirement would necessitate a higher transmit power from the terminal, which would generate a higher level of interference. This aspect of multi-services cannot be represented by assuming a resource unit related to the throughput (kb/s).

The only unit that is independent of the service type and captures the different QoS requirements is the power level. The modified stochastic knapsack represents the power or noise level of a given UMTS carrier. The model is tightly related to the pole capacity concept of spread spectrum systems, which represents the infinite noise level at which the system diverges.

2.3.1 UL pole capacity

Determining the pole capacity of a UMTS carrier (cell) is an important step in capacity analysis. Since the frequency reuse of the W-CDMA system is 1, the system is typically interference-limited by the air interface and the amounts of interference and delivered cell capacity need to be estimated. The theoretical spectrum efficiency of a W-CDMA cell can be calculated from the pole capacity equation, which refers to the 100% cell load or the soft capacity limit.[3)] The pole capacity is derived from the definition of Eb/No. The final expression of pole capacity in terms of the number of active users is as follows:

$$N_{pole}^{UL} = \frac{W}{(1+f_{UL}) \cdot \sum_{s=1}^{S} p_s^{UL} v_s^{UL} R_s^{UL} \dfrac{\left(E_b/N_o\right)}{\left(1+\left(E_b/N_o\right)_s \cdot R_s^{UL}/W\right)}} \quad (5)$$

Where:

- N_{pole} is the maximum number of active users allowed.
- S is the number of offered services.
- W is the chip rate.
- f_{UL} is the average other-cell interference on the uplink.
- V_s is the activity factor of service s.
- P_s is the percentage of active users using service s.
- $(E_b/N_o)_s$ is the required Eb/No for service s (assuming ideal power control, the Eb/No can be regarded as constant for all mobiles using service k).
- R_s is the data rate of service s.

A cell load concept is defined which is related to the pole capacity as follows:

$$\eta_{UL} = \frac{NumberofActiveUsers}{PoleCapacity} \quad (6)$$

The cell load represents the amount of traffic intensity within a carrier, which directly results in what is called the shrinking phenomena in UMTS. In essence, what is happening is that because all users are using the same frequency at the same time, the other users' activity is seen by Node B as interference. This means that a terminal needs to transmit at a higher level to overcome the interference level from other users and allow the signal to reach Node B with the required QoS.

note 5) Energy per bit/noise power density

If we consider a user who is at the cell edge and already transmitting at maximum power, the user's signal will not be stronger than the interference from other users and will not reach Node B with the required QoS. In this case, the user is considered to be out of coverage due to cell shrinkage. This tight cause-and-effect relationship between the capacity and coverage can be represented, respectively, by the relationship between the cell load and noise rise[note 6)] (**Figure 4**).

2.3.2 Modified stochastic knapsack

As explained in the previous section, it is necessary to find a cell edge that incorporates both the capacity limitation (cell load) and coverage limitation (noise rise) at the required GoS. The GoS for circuit-switch services is very well represented by the blocking probability. For packet-switch services, blocking is not a performance measure: indicators such as throughput and delay are more representative of the users' satisfaction. For the capacity analysis, which is the step that estimates the required cell load to guarantee the required GoS, we have adopted the modified knapsack as explained below:

- The knapsack capacity C represents the cell load. C = 0 to 100%.

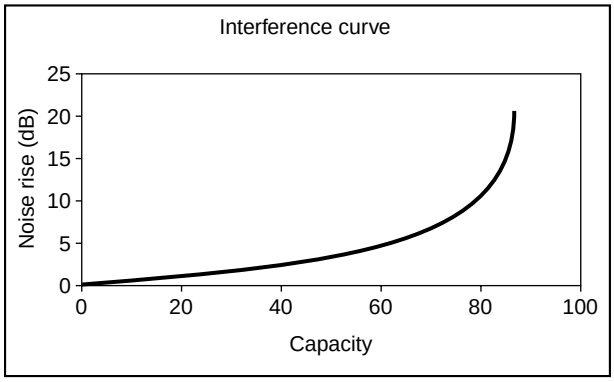

Figure 4
Noise rise versus cell load.

- The object size b_k is the interference contribution by a user of class k:

$$b_k = 100 \times \frac{W}{(1+f_{UL}) \cdot v_s^{UL} R_s^{UL} \dfrac{\left(E_b/N_o\right)_k}{\left(1+\left(E_b/N_o\right)_k \cdot R_k^{UL}/W\right)}} \quad (7)$$

Where,

- W is the chip rate.
- f_{UL} is the other-cell interference on the uplink.
- V_k is the activity factor of service k (circuit switch services would have a very different value than packet switched services).
- $(E_b/N_o)_k$ is the required Eb/No for service k (assuming ideal power control, the Eb/No can be regarded as constant for all mobiles using service k).
- R_k is the data rate of service k.

The GoS for circuit switch services will be measured as a blocking probability using the following formula:

$$PB_k = 1 - \frac{\sum_{n \in S_k} \prod_{j=1}^{K} \rho_j^{nj} / n_j!}{\sum_{n \in S} \prod_{j=1}^{K} \rho_j^{nj} / n_j!} \quad (8)$$

And the GoS for packet switch services will be expressed in terms of throughput, which represents the long-run throughput per service. The throughput is derived by considering the long-run rate at which users of this service enter the air interface (or Knapsack) and the data rate per user by employing the following formula[1)]:

$$TH_k = R_K \lambda_k (1 - PB_k) \quad (9)$$

Where,

- λ_k is the arrival rate of service k assuming a Poisson process.
- PB_k is the probability of blocking for service k.
- R_k is the data rate of service Blocking probability of service.
- n_j is the number of objects in the knapsack from class j.

note 6) Noise rise over thermal noise due to interference = $-10 \log_{10} (1 - \eta UL)$.

The UL analysis is an iterative process that continues as we tune between the capacity and coverage analyses.[3] UL coverage analysis refers to calculating the UMTS link budget for the uplink. Like all radio link budgets, the purpose of the UMTS link budget is to find the maximum allowed path loss. However, it has one main difference from other radio link budgets in that it also includes the noise rise due to captured subscribers. The value of noise rise is not always equal to the noise rise calculated from our loading equation, which is why we need to tune these two analyses. **Figure 5** shows the tuning procedure for the UL.

2.4 Downlink analysis

Once the uplink is tuned (i.e., step 4 in Figure 1 is reached), the downlink analysis is performed. The downlink analysis differs slightly from the uplink analysis in that the coverage and capacity are considered simultaneously, not iteratively. The downlink analysis checks whether the Node B power allocation is enough to cover the captured subscribers (from the uplink analysis). If the power is not enough, the uplink cell range needs to be reduced (by reducing the TX power of UEs) and the whole process is repeated from step 1).

The downlink load factor can be defined based on a similar principle as the one for the uplink; however, there are some additional parameters involved. The DL pole capacity is also derived from the definition of Eb/No – energy per user bit over spectral noise density, which can be expressed as follows for a user j:

$$(E_b / N_o)_j = \frac{W}{R_j \cdot v_j} \cdot \frac{P'_j}{\alpha \cdot I_{own} + \sum_{c=2}^{Ctotal} I_{external,\,c} + N_{thermal}} \qquad (10)$$

Where,

$P'j$ = Received signal power by user j intended for that user (watts)

$Iown$ = Total received wide-band power from own-cell base station

α = Orthogonality factor (signals from the cells themselves are partially received as interference due to imperfect orthogonality)

We can say that the received power is proportional to the own-cell transmitted power: $P'j = Iown \times \zeta j$, where ζj, which is the ratio of power allocated to user j, and the total received power can be calculated as follows:

$$\zeta_j = \frac{(E_b / N_o)_j \cdot v_j}{(W / R_j)} \left(\alpha + \frac{I_{other}}{I_{own}} + \frac{N_{thermal}}{I_{own}} \right) \qquad (11)$$

This ratio between other-cell and own-cell interference can be written as follows:

$$f_{DLj} = \left(\frac{I_{external} + N_{thermal}}{I_{own}} \right)_j \qquad (12)$$

Here, ζ j becomes:

$$\zeta_j \le \zeta_{\max} \qquad (13)$$

The maximum value of ζ_j would mean the maximum power that is allowed to be allocated to user j for normal operation.[4] There will be an outage when ζj exceeds ζmax. The probability of an outage for user j can be stated as follows:

$$P_{outagej} = P_r\,(\zeta_j > \zeta_{\max}) \qquad (14)$$

The downlink loading is the ratio between the total power that the BS transmits to the maximum power that it is capable of transmitting:

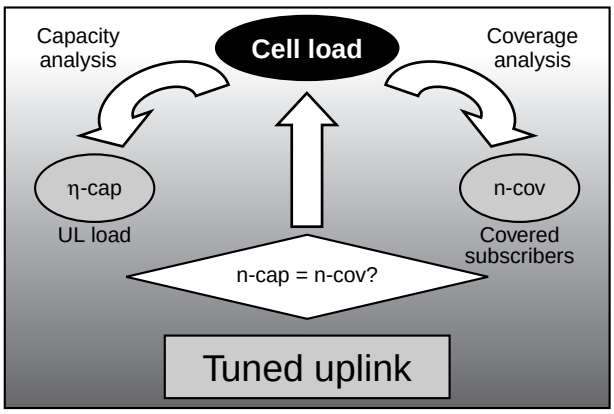

Figure 5
UL tuning.

$$\eta_{DL} = \frac{P_{total}}{P_{\max}} \tag{15}$$

Since ζj represents the allocation ratio for one user, then:

$$\sum_{j=1}^{N'} \zeta_j \le \beta \tag{16}$$

Where β is the percentage of traffic channels and 1 – β would be the ratio of the common channels. The equation can be rewritten as follows:

$$\sum_{j=1}^{N'} \zeta_j / \beta \le 1 \tag{17}$$

Where N' represents the total number of connections, including the connections due to soft handover[note 7)]:

$$N' = N \times (1 + SHO) \tag{18}$$

Where N is the number of active users and SHO is the overhead due to soft handover. The loading formula can now be written as follows:

$$\eta_{DL} = \sum_{j=1}^{N \times (1+SHO)} \frac{(E_b / N_o)_j \cdot v_j}{(W / R_j)} (\alpha + f_{DL\,j}) / \beta \tag{19}$$

3. Conclusions and future work

In this paper, we described a UMTS air interface dimensioning procedure. This newly developed model solves many problems which could not be answered by the previous approach. However, there are still improvements which can be made to our dimensioning procedure.

Currently, our model is a loss model; that is, we do not incorporate queuing for packets. Therefore, the reader should keep in mind that this queuing is different from the one which may be present in the core network or the Radio Network Controller. We are referring here to queuing solely in the air interface. This is an important issue, but the problem is not as big as it may seem. On the UL, Node B has no means to make queues for packets, but it is possible to have a queuing mechanism for the DL packet traffic.

We are currently studying this possibility and hope to improve our model based on the results.

note 7) A state where one mobile terminal is connected to multiple base stations.

References

1) Keith W. Ross: Multiservice Loss Models for Broadband Telecommunication Networks. Springer, 1995.
2) J. M. Pitts and J. A. Schormans: Introduction to IP and ATM Design and Performance, John Wiley & Sons Ltd., 2000.
3) H. Holma and A. Toskala: WCDMA for UMTS Radio Access for Third Generation Mobile Communications. John Wiley & Sons, 2000.
4) Andrew J. Viterbi: CDMA: Principles of Spread Spectrum Communication (Addison-Wesley Wireless Communications). 1st ed, Prentice Hall PTR, June 1995.

Mona Jaber received the B.E. in Computer and Communication Engineering from the American University of Beirut, Beirut, Lebanon in 1996. She joined MSI (Metapath Software International) as a radio consultant in London in 1998 and later joined the Fujitsu Europe R&D Center, UK in 2000 as a Senior Engineer. In October 2001, she joined the Mobile System division of Fujitsu Laboratories of Europe Ltd., UK as a Principal Research Engineer. She is in charge of 3G network planning research and business support at that division.

E-mail: M.Jaber@fle.fujitsu.com

Syed Ammar Hussain received the B.E. in Telecomunication Engineering from the National University of Sciences and Technology, Pakistan in 1999. He started his carrier as a researcher with the Army Signals Research & Development Establishment (Pakistan). Later he joined Motorola Pakistan. In 2001, he joined Logica Mobile Networks UK as an RF Consultant Engineer. Since 2002, he has been working with FLE as a Research Engineer for Mobile Systems Development.

E-mail: S.Hussain@fle.fujitsu.com

Adel Rouz received the Msc & Ph.D. in Computer Science from the University of Electro-Communications, Tokyo, Japan in 1991 and 1992, respectively. He joined Fujitsu Ltd., in 1991 and then joined the Fujitsu Europe R&D Center, UK in 1996 as a Principal Engineer. Since October 2001, he has been the Director of the Mobile Communication Research Division in Fujitsu Laboratories of Europe Ltd., UK, where he is in charge of the 3G & Beyond research activities in Europe; Open Mobile Architecture; IP based mobile networks; Radio Network Planning and Terminal Development.

E-mail: A.Rouz@fle.fujitsu.com