Admission and GoS control in multiservice WCDMA system

Jean Marc Kelif[1] and Eitan Altman*[2]

[1] France Telecom R&D, 38-40 Rue du General Leclerc, 92794 Issy les Moulineaux Cedex 9, France
[2] INRIA, BP 93, 06902 Sophia Antipolis Cedex, France

Abstract. We consider a WCDMA system with both real time (RT) calls that have dedicated resources, and data non-real time (NRT) calls that use a time-shared channel. We assume that NRT traffic is assigned the resources left over from the RT traffic. The grade of service (GoS) of RT traffic is also controlled in order to allow for handling more RT call but at the cost of degraded transmission rates during congestion periods. We consider both the downlink (with and without macrodiversity) as well as the uplink and study the blocking probabilities of RT traffic as well as expected sojourn time of NRT traffic. We further study the conditional expected sojourn time of a data call given its size and the state of the system. We extend our framework to handover calls.

Keywords: WCDMA, call admission control, HSDPA, HDR, handover.

1 Introduction

An important performance measure of call admission policies is the probability of rejection of calls of different classes. In order to be able to compute these probabilities and to design the call admission control (CAC) policies, a dynamic stochastic approach should be used. A classical approach for CAC is based on adaptively deciding how many channels (resources) to allocate to calls of a given service class [2,7,8]. Then one evaluates the performance as a function of some parameters (thresholds) that characterize the admission policy. Where as this approach is natural to adapt to TDMA or FDMA systems in which the notion of "channel" and of "capacity" is clear, this is not the case any more with CDMA systems in which the capacity is much more complex to define. For the uplink case in CDMA, the capacity required by a call has been defined in the context of CAC see [12,5,3].

We focus on two type of calls, real-time (RT) and non-real time (NRT) data transfers. Where as all calls use CDMA, we assume that NRT calls are further time-multiplexed[3]. We propose a simple model that allows us to define in the

* Supported by a CRE research contract with France Telecom R&D.

[3] Time multiplexing over CDMA is typical for down link data channels, as the High Data Rate (HDR) [1] and the High Speed Downlink Packet Access (HSDPA) [11]

downlink case (also in presence of macrodiversity) the capacity required by a call when it uses a given GoS (transmission rate)[4]. We then propose a control policy that combines admission control together with a control of the GoS of real-time traffic. Key performance measures are then computed by modeling the CDMA system as a quasi birth and death (QBD) process. We obtain the call blocking probabilities and expected transfer times (see [3] for the uplink case). We further obtain the expected transfer time of a file conditioned on its size. We study the influence of the control parameters on the performances.

2 The downlink

We use the model similar to [4]. Let there be S base stations. The minimum power received at a mobile k from its base station l call is determined by a condition concerning the signal to interference ratio, which should be larger than some constant

$$(C/I)_k = \frac{E_s}{N_o}\frac{R_s}{W}\Gamma, \tag{1}$$

where E_s is the energy per transmitted bit of type s, N_o is the thermal noise density, W is the WCDMA modulation bandwidth, R_s is the transmission rate of the type s call, and Γ is a constant is related to the shadow fading and imperfect power control [3]. Let $P_{k,l}$ be the power received at mobile k from the base station l. Assume that there are M mobiles in a cell l; the base station of that cell transmits at a total power $P_{tot,l}$ given by $P_{tot,l} = \sum_{j=1}^{M} P_{j,l} + P_{CCH}$ where P_{CCH} corresponds to the power transmitted for the orthogonal Common Channels. Note that these two terms are not power controlled and are assumed not to depend on l. Due to the multipath propagation, a fraction α of the received own cell power is experienced as intracell interference. Let $g_{k,l}$ be the attenuation between base station l and mobile k. Denoting by $I_{k,inter}$ and $I_{k,intra}$ the intercell and intracell interferences, respectively, we have

$$\left.\frac{C}{I}\right|_k = \frac{P_{k,l}/g_{k,l}}{I_{k,inter} + I_{k,intra} + N}$$

where N is the receiver noise floor (assumed not to depend on k), $I_{k,intra} = \alpha(P_{CCH} + \sum_{j\neq k} P_{j,l})/g_{k,l}$ and $I_{k,inter} = \sum_{j=1,j\neq l}^{S} P_{tot,j}/g_{k,j}$. Define $F_{k,l} = \frac{\sum_{j=1,j\neq l}^{S} P_{tot,j}/g_{k,j}}{P_{tot,l}/g_{k,l}}$. It then follows that

$$\beta_k = \frac{P_{k,l}/g_{k,l}}{(F_{k,l}+\alpha)P_{tot,l}/g_{k,l} + N} \qquad \text{where } \beta_k = \frac{(C/I)_k}{1+\alpha(C/I)_k}. \tag{2}$$

We next consider two service classes (that will correspond to real time (RT) and non-real time (NRT) traffic, respectively). Let $(C/I)_s$ be the target SIR ratio for mobiles of service class s and let β_s be the corresponding value in (2). Let there be in a given cell M_s mobiles of class s. We shall use the following

[4] UMTS uses the Adaptive Multi-Rate (AMR) codec that offers eight different transmission rates of voice varying between 4.75 kbps to 12.2 kbps.

approximations. First we replace $F_{k,l}$ by a constant (e.g. its average value, as in [4]; this is a standard approximation, see [6]). Secondly, we approximate $g_{k,l}$ by their averages. More precisely we define G_s to be the average of $g_{l,k}$ over all mobiles k belonging to class s, $s = 1, 2$. With these approximations (2) gives the following value for $P_{tot,l}$ (we shall omit the index l):

$$P_{tot} = \frac{P_{CCH} + N \sum_s \beta_s M_s G_s}{1 - (\alpha + F) \sum_s \beta_s M_s}. \tag{3}$$

Further assuming that the system is designed so as to have $P_{CCH} = \psi P_{tot}$ and defining the downlink load as $Y_{DL} = \sum_s \beta_s M_s$, this gives

$$P_{tot} = \frac{N \sum_s \beta_s M_s G_s}{Z_2} \text{ where } Z_2 = (1 - \psi) - (\alpha + F) Y_{DL}. \tag{4}$$

In practice, to avoid instabilities and due to power limitation of the base stations, one wishes to avoid that Z_2 becomes close to zero, thus one poses the constraint $Z_2 \geq \epsilon$ for some $\epsilon > 0$. Define the system's capacity as $\Theta_\epsilon = 1 - \psi - \epsilon$ and the capacity required by a call as

$$\Delta(s) := (\alpha + F)\beta_s. \tag{5}$$

We note that β_s will allow to depend on $M_\sigma, \sigma = 1, 2$. Combining this with (1) and with (2) we get

$$R_s = \frac{\Delta(s)}{\alpha + F - \alpha \Delta(s)} \times \frac{N_o W}{E_s \Gamma}. \tag{6}$$

3 Downlink with Macro diversity

Our approach is inspired by [4] who considered the single service case. A mobile i in macrodiversity is connected to two base stations, b and l. b is defined to be the station with larger SIR. Following [4] we assume that the Maximum Ratio Combining is used and hence the power control tries to maintain $\gamma_i = \frac{C}{I}\big|_i = \frac{C}{I}\big|_{i,b} + \frac{C}{I}\big|_{i,l}$ where γ_k is given by the constant in (1). We have $\Omega_i \leq 1$ where $\Omega_i := \frac{C/I|_{i,l}}{C/I|_{i,b}}$. ($\Omega_i$ is defined to be 0 for a mobile i that is not in macrodiversity.) This gives for the combined C/I [4]:

$$\frac{C}{I}\bigg|_i = \frac{(1 + \Omega_i) P_{i,b} / g_{bi}}{\alpha_b (P_{tot,b} - P_{i,b}) / g_{bi} + F_{i,b} P_{tot,b} / g_{bi} + N}$$

The transmission power becomes $P_{i,b} = \kappa_i (\alpha P_{tot,b} + F_{i,b} P_{tot,b} + g_{bi} N)$, where

$$\kappa_i = \frac{(C/I)_i}{1 + \Omega_i + \alpha (C/I)_i} \tag{7}$$

Let there be M mobiles in a cell b (we shall omit this index) of which a fraction μ is in macrodiversity. Then as in [4], $P_{tot} = \sum_{i=1}^{(1-\mu)M} P_i + \sum_{j=1}^{2\mu M} P_j + P_{CCH}$.

We now consider two classes of services $s = 1, 2$ corresponding to RT and NRT mobiles. We make the following approximations, similar to [4]: For a given service class $s = 1, 2$ Ω_i is replaced by a constant Ω_s (its average over all mobiles of the same service as i); we also replace $F_{i,b}$ by one of four constants F_s^{NMD} and F_s^{MD}, $s = 1, 2$, where F_s^{NMD} (resp. F_s^{MD}) corresponds to an average value of $F_{i,b}$ over mobiles in service s which are not in macrodiversity (and which are in macrodiversity, resp.). Finally, we replace g_{bi} by one of the four constants G_s^{NMD} and G_s^{MD}, $s = 1, 2$, where G_s^{NMD} (resp. G_s^{MD}) corresponds to an average value of g_{ib} over mobiles in service s which are not in macrodiversity (and which are in macrodiversity, resp.). This gives the total power of a base station b: $P_{tot} = Z_1/Z_2$ as long as Z_2 is strictly positive, where $Z_1 := (1-\mu)\sum_{s=1,2} M_s\kappa_s G_s^{NMD} N + 2\mu \sum_{s=1,2} M_s\kappa_s G_s^{MD} N$, $Z_2 := (1-\psi) - (1-\mu)\sum_{s=1,2} M_s\kappa_s(\alpha + F_s^{NMD}) - 2\mu\sum_{s=1,2} M_s\kappa_s(\alpha + F_s^{MD})$. In practice one wishes to avoid that Z_2 becomes close to zero, thus we pose the constraint $Z_2 \geq \epsilon$ for some $\epsilon > 0$. Define the system's capacity as $\Theta_\epsilon = 1 - \psi - \epsilon$, and the capacity required by a call of type s as $\Delta(s) = \kappa_s\left[(1-\mu)(\alpha + F_s^{NMD}) + 2\mu(\alpha + F_s^{MD})\right]$. Combining this with (1) and (7), we get

$$R_s = \frac{\Delta(s)(1 + \Omega_s)}{(1-\mu)(\alpha + F_s^{NMD}) + 2\mu(\alpha + F_s^{MD}) - \alpha\Delta(s)} \cdot \frac{N_o W}{E_s \Gamma}, \quad s = 1, 2. \tag{8}$$

4 Uplink

We recall the capacity notions from the case of uplink from [3]. Define

$$\tilde{\Delta}_s = \frac{E_s}{N_o}\frac{R_s}{W}\Gamma, \text{ and } \Delta'(s) = \frac{\tilde{\Delta}(s)}{1 + \tilde{\Delta}(s)}, \qquad s = 1, 2. \tag{9}$$

The power that should be received at a base station originating from a type s mobile in order to meet the QoS constraints is Z_1/Z_2 [3] where $Z_1 = N\Delta'(s)$ and $Z_2 = 1 - (1+f)\sum_{s=1,2} M_s\Delta'(s)$ (N is the background noise power at the base station, f is a constant describing the average ratio between inter and intra cell interference, and M_s is the number of mobiles of type s in the cell). To avoid instability one requires that $Z_2 \geq \epsilon$ for some $\epsilon > 0$. Define the system's capacity as $\Theta_\epsilon = 1 - \epsilon$, and the capacity required by a type s call as $\Delta(s) = (1+f)\Delta'(s)$. Combining this with (9) we get

$$R_s = \frac{\Delta(s)}{1 + f - \Delta(s)} \times \frac{N_o W}{E_s \Gamma}, \quad s = 1, 2. \tag{10}$$

5 Admission and rate control

We assume that either the uplink or the downlink are the bottleneck in terms of capacity, so we can focus only on the more restrictive direction when accepting calls. All the notations will be understood to relate to that direction.

Capacity reservation. We assume that there exists a capacity L_{NRT} reserved for NRT traffic. RT traffic can use up to a capacity of $L_{RT} := \Theta_\epsilon - L_{NRT}$.

GoS control of RT traffic. UMTS will use the Adaptive Multi-Rate (AMR) codec that offers eight different transmission rates of voice that vary between 4.75 kb/s to 12.2 kb/s, and that can be dynamically changed every 20 ms. The lower the rate is, the larger the amount of compression is, and we say that the GoS is lower. For simplicity we shall assume that the set of available transmission rates of RT traffic has the form $[R^{\min}, R^{\max}]$. We note that $\Delta(RT)$ is increasing with the transmission rate. Hence the achievable capacity set per RT mobile has the form $[\Delta^{\min}, \Delta^{\max}]$. Note that the maximum number of RT calls that can be accepted is $M_{RT}^{\max} = \lfloor \Theta_\epsilon / \Delta^{\min} \rfloor$. We assign full rate $R^{\max}$ (and thus the maximum capacity $\Delta^{\max}$) for each RT mobile as long as $M_{RT} \leq N_{RT}$ where $N_{RT} = \lfloor L_{RT} / \Delta_{RT}^{\max} \rfloor$. For $N_{RT} < M_{RT} \leq M_{RT}^{\max}$ the capacity of each present RT call is reduced to $\Delta_{MR} = L_{RT}/M_{RT}$ and the rate is reduced accordingly (e.g. by combining (1), (2) and (5) for the case of downlink).

Rate control for NRT traffic. The capacity $C(M_{RT})$ unused by the RT traffic is fully assigned to one NRT mobile, and the mobile to which it is assigned is time multiplexed rapidly so that the throughput is shared equally between the present NRT mobiles. The available capacity for NRT mobiles is

$$C(M_{RT}) = \begin{cases} \Theta_\epsilon - M_{RT}\Delta^{\max} & \text{if } M_{RT} \leq N_{RT}, \\ L_{NRT} & \text{otherwise.} \end{cases}$$

The total transmission rate R_{NRT}^{tot} of NRT traffic for the downlink and uplink is then given by

$$R_{NRT}^{tot}(M_{RT}) = \frac{C(M_{RT})}{Z_3} \times \frac{N_o W}{E_s \Gamma},$$

where $Z_3 = \alpha + F - \alpha C(M_{RT})$ for the downlink and $Z_3 = 1 + f - C(M_{RT})$ for the uplink. The expression for downlink with macrodiversity is derived similarly.

6 Statistical model and the QBD approach

Model. RT and NRT calls arrive according to independent Poisson processes with rates λ_{RT} and λ_{NRT}, respectively. The duration of a RT call is exponentially distributed with parameter μ_{RT}. The size of a NRT file is exponentially distributed with parameter μ_{NRT}. Interarrival times, RT call durations and NRT file sizes are all independent. The departure rate of NRT calls depends on the current number of RT calls: $\nu(M_{RT}) = \mu_{NRT} R_{NRT}^{tot}(M_{RT})$

QBD approach. The number of active sessions in all three models (downlink with and without macrodiversity and uplink) can be described as a QBD process, and we denote by Q its generator. We shall assume that the system is stable. The stationary distribution of this system, π, is calculated by solving: $\pi Q = 0$, with the normalization condition $\pi e = 1$ where e is a vector of ones of proper dimension. π represents the steady-state probability of the two-dimensional process lexicographically. We may thus partition π as $[\pi(0), \pi(1), \ldots]$ with the vector $\pi(i)$ for level i, where the levels correspond to the number of

NRT calls in the system. We may further partition each level into the number of RT calls, $\pi(i) = [\pi(i,0), \pi(i,1), \ldots, \pi(i, M_{RT}^{\max})]$, for $i \geq 0$. Q is given by

$$Q = \begin{bmatrix} B & A_0 & 0 & 0 & \cdots \\ A_2 & A_1 & A_0 & 0 & \cdots \\ 0 & A_2 & A_1 & A_0 & \cdots \\ \\ 0 & 0 & \ddots & \ddots & \ddots \end{bmatrix} \tag{11}$$

where the matrices B, A_0, A_1, and A_2 are square matrices of size $(M_{RT}^{\max} + 1)$. The matrix A_0 corresponds to a NRT connection arrival, given by $A_0 = \mathrm{diag}(\lambda_{NRT})$. The matrix A_2 corresponds to a departure of a NRT call and is given by $A_2 = \mathrm{diag}(\nu(i); 0 \leq i \leq M_{NRT}^{\max})$. The matrix A_1 corresponds to the arrival and departure processes of RT calls. A_1 is tri-diagonal as follows: $A_1[i, i+1] = \lambda_{RT}, A_1[i, i-1] = i\mu_{RT}, A_1[i,i] = -\lambda_{RT} - i\mu_{RT} - \lambda_{NRT} - \nu(i)$. We also have $B = A_1 + A_2$. π is given by [3,9] $\pi(i) = \pi(0)\mathbf{R}^i$ where $\mathbf{R}$ is the minimal non-negative solution to the equation: $A_0 + \mathbf{R}A_1 + \mathbf{R}^2 A_2 = 0$. $\pi(0)$ is obtained by solving $\pi(0)(I - \mathbf{R})^{-1}e = 1$ [3]. Note that the evolution of number of RT calls is not affected by the process of NRT calls and the Erlang formula can be used to compute their steady state probability. The blocking probability of a RT call is:

$$P_B^{RT} = \frac{(\rho_{RT})^{M_{RT}^{\max}} / M_{MRT}^{\max}!}{\sum_{i=1}^{M_{RT}^{\max}} (\rho_{RT})^i / i!}$$

where $\rho_{RT} = \lambda_{RT}/\mu_{RT}$. This is the main performance measure for the RT traffic. For NRT calls the important performance measure is expected sojourn time which is given by Little's law as $T_{NRT} = E[M_{NRT}]/\lambda_{NRT}$.

Conditional expected sojourn times. The performance measures so far are similar to those already obtained in the uplink case in [3]. We wish however to present more refined performance measures concerning NRT calls: the expected sojourn times conditioned on the file size and the state upon the arrival of the call. We follow [10] and introduce a non-homogeneous QBD process with the following generator Q^* and the corresponding steady state probabilities π^*:

$$Q^* = \begin{bmatrix} B & A_0 & 0 & 0 & \cdots & & \\ (1/2)A_2 & A_1^* & A_0 & 0 & 0 & \cdots & \\ 0 & (2/3)A_2 & A_1^* & A_0 & 0 & \cdots & \\ 0 & 0 & (3/4)A_2 & A_1^* & A_0 & \cdots & \\ \\ 0 & 0 & \ddots & & \ddots & \ddots & \ddots \end{bmatrix} \tag{12}$$

where the matrices A_0, A_2, B are the same as introduced before, and A_1^* is the same as A_1 defined before except that the diagonal element is chosen to be minus the sum of the off-diagonal elements of Q^*, i.e. $A_1^*[k,k] = -\lambda_{RT} - k\mu_{RT} - \lambda_{NRT} - \frac{k-1}{k}\nu(k)$. The conditional expected sojourn time of a NRT mobile given that its size is v, that there are i RT mobiles and $k-1$ NRT mobiles upon it's arrival, is obtained from [10]:

$$T_{k,i}(v) = \frac{v/R_{NRT}^{tot}(0)}{R^* - \rho^*} + \overline{1_{k,i}} \left[I - \exp\left(\frac{v}{R_{NRT}(0)} \mathcal{R}^{-1} Q^* \right) \right] \overline{w} \quad \text{where} \tag{13}$$

$$R^* := \sum_{k,i} \pi^*(k,i) \frac{R_{NRT}^{tot}(i)}{R_{NRT}^{tot}(0)}, \quad \mathcal{R} = diag\left[\frac{1}{k} \frac{R_{NRT}^{tot}(i)}{R_{NRT}^{tot}(0)} \right], \quad \rho^* := \frac{\lambda_{NRT}}{\mu_{NRT} R_{NRT}^{tot}(0)},$$

$\overline{1_{k,i}}$ is a vector whose entries are all zero except for the (k,i)th entry whose value is 1, and $\overline{w}$ is the solution [10] of $Q^*\overline{w} = \frac{1}{R^*-\rho^*}\mathcal{R} \cdot \overline{1_{k,i}} - \overline{1_{k,i}}$.

Remark 1. Suppose that the number of RT sessions stays fixed throughout the time (this can be used as an approximation when the average duration of RT sessions is very large). Then with $R^* = R_{NRT}^{tot}(i)/R_{NRT}^{tot}(0)$, (13) becomes

$$T_k(v) = \frac{v/R_{NRT}^{tot}(0)}{R^* - \rho^*} + \frac{1}{\mu_{NRT} R_{NRT}^{tot}(i) - \lambda_{NRT}} \times \left(k - \frac{1}{R^* - \rho^*}\right) \times \left(1 - \exp\left(-\frac{v}{R_{NRT}^{tot}(0)}(\mu_{NRT} R_{NRT}^{tot}(0) - \lambda_{NRT})\right)\right) \tag{14}$$

In Fig. 1 we use (14) to compute the maximum number k of NRT calls present at the arrival instant of an NRT call (we include in this number the arriving call) such that the expected sojourn time of the connection, conditional on its size in kb and on k, is below 1 sec. For example, if the size of the file is 100 kb then its conditional expected sojourn time will be smaller than 1 sec as long as the number of mobiles upon arrival (including itself) does not exceed 12. This figure and the next one are obtained with $R_{NRT}^{tot}(0) = 1000kbps$, $\lambda_{NRT} = 1$, $\mu_{NRT}^{-1} = 160$ kbits so that $\rho^* = 0.16$ (We took no RT calls, i.e. $i = 0$).

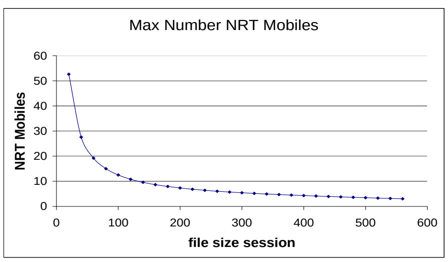

Fig. 1. Max number of NRT calls upon arrival s.t. the conditional expected sojourn time is below 1 sec

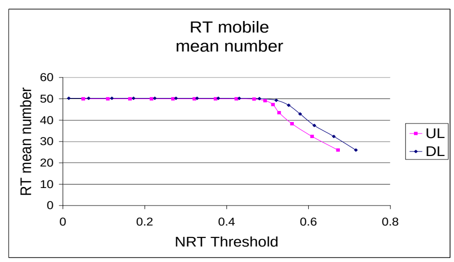

Fig. 2. Mean number of RT calls in a cell as a function of the reservation level for RT traffic

7 Numerical results

We consider the following setting. Unless stated otherwize, the data are for both down and uplink. Transmission rates of RT mobiles: maximum rate is 12.2kbps, minimum rate 4.75kbps. NRT uplink data rate: 144 kbps. NRT downlink data rate: 384 kbps. E_{RT}/N_o (at 12.2 kbps) is 7.9 dB (UL) and 11.0 dB (DL). E_{NRT}/N_o is 4.5 dB (for 144 kbps UL) and 4.8 dB (for 384 kbps DL). Total bandwidth $W = 3.84MHz$. Mean NRT session size 160 kbits, arrival rate of NRT calls: $\lambda_{NRT} = 0.4$. Average duration of RT call: 125 s, arrival rate of RT calls: $\lambda_{RT} = 0.4$. Uplink interference factor is 0.73. Downlink interference factor is 0.55. $\alpha = 0.64$, $\psi = 0.2, \epsilon = 10^{-5}$. Γ is computed so as to guarantee that the probability of exceeding the target C/I ratio is 0.99. It corresponds to a standard deviation constant $\sigma = 0.5$ (see [3]).

Influence of NRT reservation on RT traffic In Fig. 2 we depict the average cell capacity in terms of the average number of RT mobiles for both uplink and downlink as a function of the reservation threshold for NRT traffic. We see that it remains almost constant (50 mobiles per cell) for up to 50% of the load. In Fig. 3 we present the blocking rate of RT traffic. At a reservation L_{NRT} of 50% of the maximum load, the dropping rate is still lower than 1%.

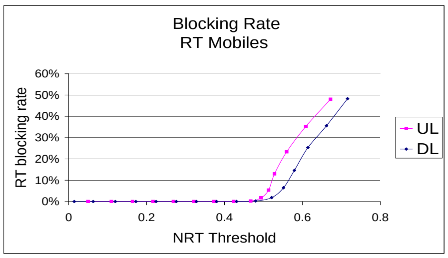

Fig. 3. Blocking rate for RT calls as a function of the reservation level for NRT traffic

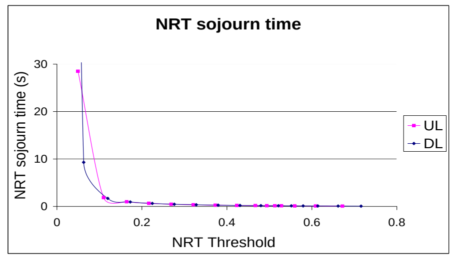

Fig. 4. Expected sojourn times of NRT traffic as a function of the NRT reservation

Influence of NRT reservation on NRT traffic Fig. 4 shows the impact of the reservation threshold L_{NRT} on the expected sojourn time of NRT calls both on uplink and downlink. We see that the expected sojourn times becomes very large as we decrease L_{NRT} below 0.15% of the load. This demonstrates the need for the reservation. In the whole region of loads between 0.16 to 0.5 the NRT expected sojourn time is low and at the same time, as we saw before, the rejection rate of RT calls is very small. This is thus a good operating region for both RT and NRT traffic.

Conditional expected sojourn time Below, the reservation limit L_{NRT} is 0.27. In Fig. 5 we depict the expected sojourn time conditioned on the number of NRT and RT calls found upon the arrival of the call both being k and on the file being of the size of 100 kbits. k is varied in this figure.

Fig. 6 depicts for various file sizes, the maximum number k such that the conditional expected sojourn time of that file with the given size is below 1 sec. k is defined to be the total number of RT calls as well as the total number of NRT calls (including the call we consider) in the cell. We thus assume (as in the previous example) that the number of NRT and of RT calls is the same, and seek for the largest such number satisfying the limit on the expected sojourn time.

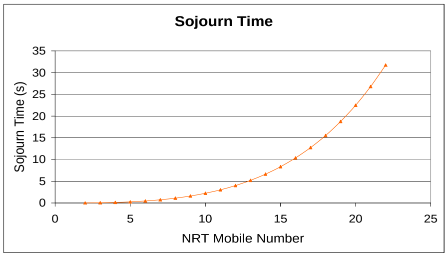

Fig. 5. Conditional expected Sojourn time of an NRT mobile as a function of the number of mobiles in the cell

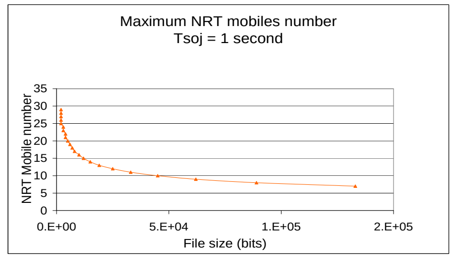

Fig. 6. Max number of NRT calls upon arrival such that the conditional expected sojourn time is below 1 sec

8 Extension to handover calls

Till now we did not differentiate between arrivals of new and of handover calls. We now wish to differentiate between these calls. We assume that RT new calls (resp. NRT new calls) arrive with a rate of λ_{RT}^{New} (resp. λ_{NRT}^{New}) where as the handover calls arrive at rate h_{RT} (resp. h_{NRT}). We assume that RT calls remain at a cell during an exponentially distributed duration with parameter μ_{RT}. Avoiding blocking of handover calls is considered more important than avoiding blocking of new ones. So we define a new threshold $\overline{M}_{RT}^{New} < M_{RT}^{\max}$. Now new RT calls are accepted as long as $M_{RT} \leq \overline{M}_{RT}^{New}$ where as handover RT calls are accepted as long as $M_{RT} \leq \overline{M}_{RT}^{New}$. The behavior of NRT calls is as before. Define $\rho_{RT} = \lambda_{RT}/\mu_{RT}$ and $\rho_{RT}^{HO} = h_{RT}/\mu_{RT}$. Let $p_{RT}(i)$ denote the number of RT mobiles in steady state. It is given by

$$p_{RT}(i) = \begin{cases} \frac{(\rho_{RT})^i}{i!} p_{RT}(0) & \text{if } 0 \leq i \leq \overline{M}_{RT}^{New} \\ \frac{(\rho_{RT})^{\overline{M}_{RT}^{New}} (\rho_{RT}^{HO})^{i-\overline{M}_{RT}^{New}}}{i!} p_{RT}(0) & \text{if } \overline{M}_{RT}^{New} \leq i \leq M_{RT}^{\max} \end{cases}$$

$$\text{where} \quad p_{RT}(0) = \left(\sum_{i=0}^{\overline{M}_{RT}^{New}} \frac{(\rho_{RT})^i}{i!} + \sum_{i=\overline{M}_{RT}^{New}}^{M_{RT}^{\max}} \frac{(\rho_{RT})^{\overline{M}_{RT}^{New}} (\rho_{RT}^{HO})^{i-\overline{M}_{RT}^{New}}}{i!} \right)^{-1}$$

The QBD approach can be directly applied again to compute the joint distribution of RT and NRT calls and performance measures.

A numerical example The data we consider are as before, except that now a fraction of 30% of arriving RT calls are due to handovers. In Fig. 7 we present the impact of the choice of the NRT threshold on the blocking rate of RT mobiles. We also illustrate the impact of the differentiation between New and Handover calls. The middle UL curve is obtained with no differentiation. The total dropping rate of the model with handover (the HO curve) is larger, but among the rate of dropping rate of calls already in the system (that arrive through a handover) is drastically diminished (the curve called "Dropping").

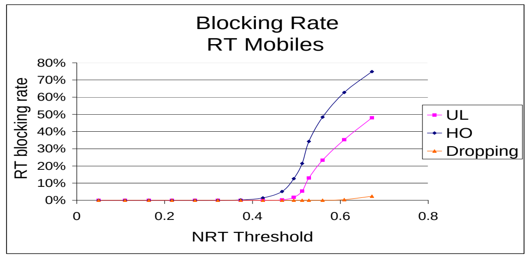

Fig. 7. RT dropping probabilities

9 Conclusions

We have analyzed the performance of a CAC combined with a GoS control in a WCDMA environment used by RT and NRT traffic. RT traffic has dedicated resources where as NRT traffic obtains through time sharing the unused capacity leftover by RT traffic. We illustrated the importance of adding reserved capacity L_{NRT} for NRT traffic only and demonstrated that this can be done in a way not to harm RT traffic.

References

1. P. Bender et al, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users", *IEEE Communications Magazine*, 70–77, July 2000.
2. Y. Fang and Y. Zhang, "Call admission control schemes and performance analysis in wireless mobile networks", *IEEE Trans. Vehicular Tech.* **51**(2), 371-382, 2002.
3. N. Hegde and E. Altman, "Capacity of multiservice WCDMA Networks with variable GoS", Proc. of IEEE WCNC, New Orleans, Louisiana, USA, March 2003.
4. K. Hiltunen and R. De Bernardi, "WCDMA downlink capacity estimation", VTC'2000, 992-996, 2000.
5. I. Koo, J. Ahn, H. A. Lee, K. Kim, "Analysis of Erlang capacity for the multimedia DS-CDMA systems", IEICE Trans. Fundamentals, Vol E82-A No. 5, 849-855, 1999.
6. J. Laiho and A. Wacker, "Radio network planning process and methods for WCDMA", *Ann. Telecommun.*, **56**, No. 5-6, 2001.
7. C. W. Leong and W. Zhuang, "Call admission control for voice and data traffic in wireless communications", Computer Communications 25 No. 10, (2002) 972-979.
8. B. Li, L. Li, B. Li, X.-R. Cao "On handoff performance for an integrated voice/data cellular system" Wireless Networks, Vol. 9, Issue 4 pp. 393 – 402, July 2003.
9. M. F. Neuts. *Matrix-geometric solutions in stochastic models: an algorithmic approach.* The John Hopkins University Press, 1981.
10. R. Núñez Queija and O. J. Boxma. Analysis of a multi-server queueing model of ABR. *J. Appl. Math. Stoch. Anal.*, 11(3), 1998.
11. S. Parkvall, E. Dahlman, P. Frenger, P. Beming and M. Persson, "The high speed packet data evolution of WCDMA", *Proc. of the 12th IEEE PIMRC*, 2001.
12. X. Tang and A. Goldsmith "Admission control and adaptive CDMA for integrated voice and data systems", available on http://citeseer.ist.psu.edu/336427.html