Evaluation of Radio Access Congestion in Heterogeneous Wireless Access Networks

X. Gelabert, J. Pérez-Romero, O. Sallent and R. Agustí
Dept. of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
Email: {xavier.gelabert, jorperez, sallent, ramon}@tsc.upc.edu

***Abstract*—This work takes as a reference the scenario of heterogeneous radio access networks where several Radio Access Technologies (RATs) coexist and cooperate providing a range of services over a specific coverage area. In this context, an appropriate RAT must be selected upon each user request in order to provide the specified Quality of Service (QoS) level for the requested service. Despite of having an efficient RAT selection mechanism, the intrinsic network dynamics may pose some risk on the QoS guarantees in each RAT which may not be met thus leading to a degraded network performance. In such situations, the network undergoes so-called radio access congestion situations. In this work, a statistical characterization of radio access congestion is provided for TDMA and WCDMA based systems. Furthermore, the impact of several representative RAT selection policies is evaluated in terms of their obtained congestion probability. Numerical results conclude that a proper knowledge of congestion information can be used as a relevant input for RAT selection specification.**

***Index Terms*—Beyond 3G, mobile communication systems, congestion control, radio resource management.**

I. Introduction

As new emerging technologies enter the wireless arena, operators and manufacturers are faced with the challenging task of jointly managing, as efficiently as possible, the pool of resources offered by several Radio Access Technologies (RATs) covering a same area. Each of these RATs, in turn, may provide different Quality-of-Service (QoS) levels for each of the supported services, thus adding a higher complexity to the problem of allocating a particular service to the RAT that is best-suited to support it. This network heterogeneity is sometimes referred in the literature as Beyond Third Generation (B3G) systems, and has indeed captured a huge attention in the recent years [1][2]. The set of mechanisms devoted to the efficient and utmost use of radio resources is typically denoted as Radio Resource Management (RRM), term that has been conveniently redefined to Common RRM (CRRM)[1] in order to capture the necessity of managing the total amount of resources offered by the different RATs in a coordinated way rather than considering each RAT as a stand-alone entity. The potential benefits of CRRM have been assessed in a number of papers, see e.g. [3], [4], [5], [6], [7], [8].

A key CRRM functionality is to select the most appropriate RAT according to some specified user/operator criteria. This RAT selection procedure can be carried out at call/session initiation, i.e. initial RAT selection; and/or during the call/session lifetime, which is usually denoted as Vertical HandOver (VHO). To a great extent, the objective of RAT selection and VHO algorithms is to enable the so-called *Always-Best-Connected* paradigm [4], which ensures a user is connected to the best RAT anytime and anywhere.

Radio access congestion arises when QoS failures at the radio interface layer occur due to the intrinsic dynamics of the network (e.g. mobility, interference rise, traffic variability, etc.) [9]. In a B3G scenario with multiple RATs and multiple services, congestion situations in each RAT can be statistically characterized by means of a *congestion probability* given the number of allocated users. Then, provided a set of RAT selection policies, the evaluation of the congestion impact of these policies is of special interest. Under the above framework, this paper intends to define and assess the impact of radio access congestion in RATs with different underlying technologies, such as Time Division Multiple Access (TDMA) and Wideband Code Division Multiple Access (WCDMA). The performance of several RAT selection policies is also evaluated in terms of the experienced congestion probability.

Despite the large amount of work devoted to CRRM algorithms in recent years (specially in what refers to RAT selection, see e.g. [5], [6], [7], [8]), to the best of the author's knowledge, very few contributions have been made to address radio access congestion control strategies. In [10], authors propose several schemes for controlling and distributing the network traffic over two RATs. Cost metrics are assigned to each service in each RAT, and load control is evaluated by means of RAT selection and VHO. In [11], [12], the authors presented a framework for managing congestion situations in heterogeneous networks. In particular, practical methodologies for congestion detection in GSM/EDGE and UMTS networks are provided along with VHO and bit-rate reduction techniques so as to lessen the congestion status of the network.

While aforementioned approaches usually rely on extensive system-level simulations, a theoretical approach to the problem seems to be uncovered. It could then be of great interest to gain insight into the general problem as well as to inspire the definition of practical and efficient congestion control strategies. In this sense, the main contribution of this work is to define a statistical model for the characterization of radio access congestion in B3G scenarios composed by TDMA

This work has been supported by the Spanish Research Council under COGNOS grant (ref. TEC2007-60985).

[1]Also denoted as Joint (JRRM) or Multiple (MRRM) in some works.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

and WCDMA technologies and to evaluate the behavior of representative RAT selection policies.

The rest of the paper is organized as follows. Section II presents the considered resource allocation model for voice and data services in a combined TDMA/WCDMA B3G scenario. Section III presents the statistical congestion model for the case of these technologies. Some results are presented in section IV and, finally, conclusions are sketched in section V.

II. Resource Allocation Model

The problem of multi-service resource allocation in multi-access systems can be approached by means of a multi-dimensional Markov model where each dimension corresponds to the allocation of a particular service type into a given RAT [13]. For illustrative purposes, we assume in the following the allocation of generic voice and data services in a scenario with TDMA and WCDMA RATs[2]. The identification of the state space along with the definition of the state transition rates and the steady state balance equations constitute the focus of this section.

A. Markov State Space

A number of $J = 2$ different traffic classes being supported by a number of $K = 2$ co-sited RATs leads to an $M = K \times J = 4$ state dimension Markov model. The number of admitted users of each service in each RAT is defined by row vector $\boldsymbol{N} = [\boldsymbol{N}_t, \boldsymbol{N}_w]$ with $\boldsymbol{N}_t = \left[N_t^v, N_t^d\right]$ and $\boldsymbol{N}_w = \left[N_w^v, N_w^d\right]$ where elements N_k^j denote the number of service j users in RAT k. Vector $\boldsymbol{N}$ will uniquely identify each state, hereon denoted as $S_{\boldsymbol{N}}$, in the Markov chain model.

Assuming that the capacity of a particular RAT k is given by the maximum number of users it may allocate (i.e. *hard capacity* is assumed), a finite number of states $S_{\boldsymbol{N}}$ is considered. This limit is usually imposed by RAT-specific Call Admission Control (CAC) procedures that determine if a new user should be admitted or not in the system so that minimum QoS requirements of already accepted users are guaranteed. In the following, the feasible states in TDMA and WCDMA are derived in order to obtain the Markov state space.

1) TDMA Case: In TDMA-based systems, a total amount of C channels (or timeslots – TSL) are to be shared among voice and data users within a time frame. It is considered that no service has priority over any other. A voice user is assumed to occupy a whole TSL throughout the duration of a call. As for data users, the same TSL may be shared by up to n_C data transmissions corresponding to different users by means of efficient time scheduling. Moreover, data users are granted q TSLs in order to increase their achievable bit rate (multi-slot capability).

According to the above, the total amount of C TSLs are shared among the simultaneous voice and data users in the system, which leads to the set of feasible number of admitted voice and data users in the uplink (UL):

$$\hat{\mathcal{S}}^t = \left\{\boldsymbol{N}_t : 0 \leq N_t^v/C + N_t^d \hat{\alpha}_d \hat{q}/\hat{n}_C C \leq 1\right\} \qquad (1)$$

with $\hat{\alpha}_d$ the UL activity factor of data users[3]. For the downlink (DL) case, (1) should be particularized accordingly in order to obtain $\check{\mathcal{S}}^t$. Then, the TDMA feasible set of states is given by $\mathcal{S}^t = \hat{\mathcal{S}}^t \cap \check{\mathcal{S}}^t$.

2) WCDMA Case: In WCDMA uplink, the load factor, $\hat{\eta}$, can be used for guaranteeing minimum QoS levels for admitted users by imposing the condition[4] [14]:

$$\hat{\eta}_{\boldsymbol{N}_w} = \frac{\hat{\alpha}_v N_w^v}{W/\hat{\theta}_v \hat{R}_v + 1} + \frac{\hat{\alpha}_d N_w^d}{W/\hat{\theta}_d \hat{R}_d + 1} \leq \hat{\eta}_{max} \qquad (2)$$

with W the chip rate; $\hat{R}_v$ and $\hat{R}_d$ the UL bit rate of voice and data services; $\hat{\theta}_v$ along with $\hat{\theta}_d$ the UL target bit-energy-to-noise-density ratio after de-spreading and decoding for voice and data users and $0 \leq \hat{\eta}_{max} \leq 1$. In this case, the UL WCDMA feasibility set yields:

$$\hat{\mathcal{S}}^w = \{\boldsymbol{N}_w : 0 \leq \hat{\eta}_{\boldsymbol{N}_w}/\hat{\eta}_{max} \leq 1\} \qquad (3)$$

In the DL, it is assumed a CAC algorithm based on the availability of Orthogonal Variable Spreading Factor (OVSF) codes as, e.g., in [14]. Then, considering a same OVSF code can be shared among different data users during inactivity periods, the number of dedicated channels will have to fulfil the relationship [14]:

$$\mathcal{C}_{\boldsymbol{N}_w} = N_w^v/\check{\xi}_v + \check{\alpha}_d N_w^d/\check{\xi}_d \leq \mathcal{C}_{max} \qquad (4)$$

with $\check{\xi}_v$ and $\check{\xi}_d$ the DL spreading factor for voice and data services respectively, and $0 \leq \mathcal{C}_{max} \leq 1$. Then, the feasible set of voice and data users in WCDMA DL is expressed as:

$$\check{\mathcal{S}}^w = \{\boldsymbol{N}_w : 0 \leq \mathcal{C}_{\boldsymbol{N}_w}/\mathcal{C}_{max} \leq 1\} \qquad (5)$$

Accordingly, the resulting WCDMA total feasible set, $\mathcal{S}^w$, must satisfy conditions in (3) and (5), i.e. $\mathcal{S}^w = \hat{\mathcal{S}}^w \cap \check{\mathcal{S}}^w$.

Finally, the global Markov state feasibility space, $\mathcal{S}$, may be defined, from the feasible sets in each RAT as:

$$\mathcal{S} = \left\{\boldsymbol{N} : \boldsymbol{N}_t \in \mathcal{S}^t, \boldsymbol{N}_w \in \mathcal{S}^w\right\} \qquad (6)$$

B. State Transitions

Transitions between states $S_{\boldsymbol{N}} \in \mathcal{S}$ in the resulting 4-dimensional Markov chain happen due to service arrival rates, i.e. $\boldsymbol{\lambda} = [\lambda_v, \lambda_d]$, or due to service departure rates $\boldsymbol{\mu} = [\mu_v, \mu_d]$. It is assumed that arrival rates λ_j are Poisson-distributed and that service times are exponentially distributed with mean service time $1/\mu_j$ [15].

A particular traffic allocation policy, referred to as $\pi_{\boldsymbol{N}}$, is then responsible of determining, in a given state $S_{\boldsymbol{N}} \in \mathcal{S}$, the

[2] Throughout the paper, let index $j = \{v, d\}$ represent voice and data services along with $k = \{t, w\}$ represent TDMA and WCDMA RATs

[3] Throughout this article, let $\hat{x}$ and $\check{x}$ indicate UL and DL directions of parameter x respectively.

[4] Assuming users with same service type to have same requirements in terms of bit-rate and Eb/No.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

specific transition rates of each service type into each of the available RATs, i.e. $\boldsymbol{\lambda}_\pi$, thus defining the following function:

$$\begin{array}{cccc} \pi_{\boldsymbol{N}}: & \mathbb{R}_+^2 & \longrightarrow & \mathbb{R}_+^4 \\ & \boldsymbol{\lambda} & \longrightarrow & \boldsymbol{\lambda}_\pi \end{array} \tag{7}$$

where vector $\boldsymbol{\lambda}_\pi$ contains elements $\lambda_{(\pi,k,j)}$ denoting the arrival transition rate of service $j \in \{v, d\}$ into RAT $k \in \{t, w\}$ due to policy $\pi_{\boldsymbol{N}}$ in state $S_{\boldsymbol{N}}$.

A specific policy π may be characterized by means of a *policy decision function*, $\boldsymbol{\Theta}_{\boldsymbol{N}} \in \mathbb{R}^M$, with elements $\Theta_{(\boldsymbol{N},k,j)} \in [0,1]$ (called *policy actions*) that determine the fraction of traffic j into RAT k at a given state $S_{\boldsymbol{N}}$, i.e.

$$\lambda_{(\pi,k,j)} = \Theta_{(\boldsymbol{N},k,j)} \lambda_{k,j} \tag{8}$$

with $\lambda_{k,j}$ the arrival rate of service j into RAT k.

In the following, some examples are provided for the case of having two services, voice and data, and two RATs, TDMA and WCDMA.

1) Load-Balancing (LB) RAT Selection Policy: This policy intends to allocate a given user demanding a particular traffic class to the RAT that undergoes a lower load level. Naturally, this implies defining appropriate load metrics for each of the considered RATs. In this paper, and without loss of generality, it is assumed that the load metrics used in the LB policy operation for TDMA and WCDMA systems are taken in the UL direction as in [13]. Nevertheless, other configurations, such as e.g. considering the DL or combining the UL with the DL, are possible in the described model.

According to the above, in TDMA-based systems, such as e.g. GSM/EDGE, the *TSL utilization factor*, [16], can be used to measure the load in a given state $S_{\boldsymbol{N}} \in \mathcal{S}$. It is defined as the ratio between the number of occupied TSLs over the number of available TSLs, thus expressed as:

$$L^t_{\boldsymbol{N}_t} = \min\left(C, N_t^v + N_t^d \hat{\alpha}_d \hat{q}\right)/C \tag{9}$$

On the other hand, the load in a WCDMA-based system may be calculated by means of the uplink load factor defined previously in (2), thus

$$L^w_{\boldsymbol{N}_w} = \hat{\eta}_{\boldsymbol{N}_w} \tag{10}$$

At a specified state $S_{\boldsymbol{N}} \in \mathcal{S}$, a service class j user will be allocated to the RAT k that, including this new service class j user, exhibits a lower load. Accordingly, Table I shows the values of $\Theta_{(\boldsymbol{N},k,l)}$ for LB policy where $\boldsymbol{e}_{k,j} \in \mathbb{Z}^2$ is a row vector containing all zeros except for the *j-th* component which is 1.

2) Service-Based (SB) RAT Selection Policy: This policy intends to select a particular RAT k based on the demanding user service type j according to some predefined preference scheme. For brevity purposes, consider a policy, named SB#1, which directs voice users to TDMA and data users to WCDMA. If upon allocation the selected RAT is at full capacity, then the opposite RAT is selected. In addition, consider a policy named SB#2 which acts the opposite to SB#1, i.e. primarily directing voice users to WCDMA and data users to TDMA. Table I shows the values that $\Theta_{(\boldsymbol{N},k,l)}$ takes for SB#1 and SB#2 policies.

[5] Let $\wedge$ and $\vee$ indicate logical AND and OR respectively. Let $\bar{k}$ indicate the opposite RAT to k for the case of two RATs.

TABLE I
POLICY ACTIONS[5]

$\Theta_{(\boldsymbol{N},k,j)}$	Load Balancing (LB) Policy
1	$(L^k_{(\boldsymbol{N}_k+\boldsymbol{e}_{k,j})} < L^{\bar{k}}_{(\boldsymbol{N}_{\bar{k}}+\boldsymbol{e}_{\bar{k},j})}) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^k_{b,j})$
1	$(S_{\boldsymbol{N}} \in \mathcal{S}^{\bar{k}}_{b,j}) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^k_{b,j})$
0.5	$(L^k_{(\boldsymbol{N}_k+\boldsymbol{e}_{k,j})} = L^{\bar{k}}_{(\boldsymbol{N}_{\bar{k}}+\boldsymbol{e}_{\bar{k},j})}) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^k_{b,j}) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^{\bar{k}}_{b,j})$
0	$(L^k_{(\boldsymbol{N}_k+\boldsymbol{e}_{k,j})} > L^{\bar{k}}_{(\boldsymbol{N}_{\bar{k}}+\boldsymbol{e}_{\bar{k},j})}) \vee (S_{\boldsymbol{N}} \in \mathcal{S}^k_{b,j})$
$\Theta_{(\boldsymbol{N},k,j)}$	Service Based #1 (SB#1) Policy
1	$(k = t) \wedge (j = v) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^k_{b,j})$
1	$(k = t) \wedge (j = d) \wedge (S_{\boldsymbol{N}} \in \mathcal{S}^{\bar{k}}_{b,j})$
1	$(k = w) \wedge (j = v) \wedge (S_{\boldsymbol{N}} \in \mathcal{S}^{\bar{k}}_{b,j})$
1	$(k = w) \wedge (j = d) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^k_{b,j})$
0	otherwise
$\Theta_{(\boldsymbol{N},k,j)}$	Service Based #2 (SB#2) Policy
1	$(k = t) \wedge (j = d) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^k_{b,j})$
1	$(k = t) \wedge (j = v) \wedge (S_{\boldsymbol{N}} \in \mathcal{S}^{\bar{k}}_{b,j})$
1	$(k = w) \wedge (j = d) \wedge (S_{\boldsymbol{N}} \in \mathcal{S}^{\bar{k}}_{b,j})$
1	$(k = w) \wedge (j = v) \wedge (S_{\boldsymbol{N}} \notin \mathcal{S}^k_{b,j})$
0	otherwise

C. Steady State Balance Equations (SSBEs)

In equilibrium, the SSBE for state $S_{\boldsymbol{N}} \in \mathcal{S}$ results from equaling the flow rate into state $S_{\boldsymbol{N}}$ to the flow rate out of state $S_{\boldsymbol{N}}$ [15]:

$$\begin{aligned} P_{\boldsymbol{N}} \sum_{k\in\{t,w\}} \sum_{j\in\{v,d\}} \lambda_{k,j} \Theta_{(\boldsymbol{N},k,j)} \delta_{(\boldsymbol{N}+\boldsymbol{a}_k^j)} + N_k^j \mu_j \delta_{(\boldsymbol{N}-\boldsymbol{a}_k^j)} \\ = \sum_{k\in\{t,w\}} \sum_{j\in\{v,d\}} \lambda_{k,j} \Theta_{(\boldsymbol{N}-\boldsymbol{a}_k^j,k,j)} P_{(\boldsymbol{N}-\boldsymbol{a}_{k,j})} \delta_{(\boldsymbol{N}-\boldsymbol{a}_k^j)} \\ + (N_k^j + 1) \mu_j P_{(\boldsymbol{N}+\boldsymbol{a}_k^j)} \delta_{(\boldsymbol{N}+\boldsymbol{a}_k^j)} \end{aligned} \tag{11}$$

where $P_{\boldsymbol{N}}$ is the probability of being in state $S_{\boldsymbol{N}}$, and $\boldsymbol{a}_k^j \in \mathbb{Z}_+^M$ is a row vector containing all zeroes except for the *j-th* service in RAT k element which is 1. In addition, $\delta_{(\boldsymbol{N})}$ is a function which guarantees that non-feasible states are not accounted, i.e. $\delta_{(\boldsymbol{N})} = 0$ if $S_{\boldsymbol{N}} \notin \mathcal{S}$ and $\delta_{(\boldsymbol{N})} = 1$ otherwise.

Once the SSBEs are determined for all states $S_{\boldsymbol{N}} \in \mathcal{S}$, numerical methods can be applied to solve the resulting system of equations given by the SSBEs plus the normalization constraint $\sum_{S_{\boldsymbol{N}} \in S} P_{\boldsymbol{N}} = 1$. The reader is referred to [17] for further details on the numerical solution of Markov chains.

III. CONGESTION PROBABILITY MODEL

The presented Markov model provides the probability, $P_{\boldsymbol{N}}$, of having a particular number of admitted users denoted by vector $\boldsymbol{N}$. Consequently, the congestion probability in each RAT k for a given RAT selection policy π can be formulated as follows:

$$P_c^{k,\pi} = \sum_{S_{\boldsymbol{N}} \in \mathcal{S}} P_c^k(\boldsymbol{N}_k) \cdot P_{\boldsymbol{N}} \quad (12)$$

where $P_c^k(\boldsymbol{N}_k)$, the congestion probability in RAT k given $\boldsymbol{N}_k$ allocated users, is averaged over all state probabilities $P_{\boldsymbol{N}}$ in the state space $\mathcal{S}$. In addition, the congestion probability in RAT k given $\boldsymbol{N}_k$ allocated users, $P_c^k(\boldsymbol{N}_k)$, may arise independently due to congestion measured in the UL and/or due to congestion measured in the DL. Consequently one can express this probability as:

$$P_c^k(\boldsymbol{N}_k) = 1 - [1 - \hat{P}_c^k(\boldsymbol{N}_k)][1 - \check{P}_c^k(\boldsymbol{N}_k)] \quad (13)$$

where $\hat{P}_c^k(\boldsymbol{N}_k)$ and $\check{P}_c^k(\boldsymbol{N}_k)$ indicate UL and DL congestion probability in RAT k respectively.

Provided the congestion probability in RAT k given policy π in (12), the overall congestion probability in the TDMA/WCDMA B3G scenario, i.e. the probability that there exists congestion in at least one RAT, P_c^π, can be then expressed as:

$$P_c^\pi = 1 - (1 - P_c^{t,\pi})(1 - P_c^{w,\pi}) \quad (14)$$

where it is assumed that congestion probabilities arising in each RAT are independent.

In the following subsections we derive the congestion probabilities $\hat{P}_c^t(\boldsymbol{N}_t)$ and $\check{P}_c^t(\boldsymbol{N}_t)$ for TDMA, along with $\hat{P}_c^w(\boldsymbol{N}_w)$ and $\check{P}_c^w(\boldsymbol{N}_w)$ for WCDMA. For convenience, we define the probability of having n simultaneously transmitting users out of N admitted users with activity factor α, as [14]:

$$P_\alpha(n|N) = \binom{N}{n} \alpha^n (1-\alpha)^{N-n} \quad (15)$$

Then, assuming independence between voice and data call/session generation, the probability of having $\boldsymbol{n}_k = (n_k^v, n_k^d)$ simultaneous voice and data users in RAT k when $\boldsymbol{N}_k$ voice and data users are admitted in the system, $P_{\boldsymbol{\alpha}}(\boldsymbol{n}_k|\boldsymbol{N}_k)$, can be formulated as:

$$P_{\boldsymbol{\alpha}}(\boldsymbol{n}_k|\boldsymbol{N}_k) = P_{\alpha_v}(n_k^v|N_k^v) \cdot P_{\alpha_d}(n_k^d|N_k^d) \quad (16)$$

where $\boldsymbol{\alpha} = (\alpha_v, \alpha_d)$ denotes voice and data activity factors which should be chosen conveniently for the UL and DL.

A. WCDMA Case

In WCDMA, radio congestion events may arise independently in the UL and/or in the DL. In the following subsections both cases are addressed.

1) Uplink Case: In WCDMA, congestion situations may be detected in the UL by means of the UL load factor $\hat{\eta}$ whenever its value exceeds a given threshold $\hat{\eta}_c$ during a certain percentage of frames within a period of time [12]. Thus we may define the WCDMA UL congestion probability, $\hat{P}_c^w(\boldsymbol{N}_w)$, as:

$$\hat{P}_c^w(\boldsymbol{N}_w) = \sum_{n_v^w=0}^{N_v^w} \sum_{n_d^w=0}^{N_d^w} \Pr\left(\hat{\eta}_{(\boldsymbol{n}_w)} > \hat{\eta}_c\right) \cdot P_{\hat{\boldsymbol{\alpha}}}(\boldsymbol{n}_w|\boldsymbol{N}_w) \quad (17)$$

where, for two services, we may express the UL load factor in a single-cell scenario when $\boldsymbol{n}_w$ simultaneous voice and data users are in the system, $\hat{\eta}_{(\boldsymbol{n}_w)}$, as [14]:

$$\hat{\eta}_{(\boldsymbol{n}_w)} = \frac{n_w^v}{W/\hat{\theta}_v \hat{R}_v + 1} + \frac{n_w^d}{W/\hat{\theta}_d \hat{R}_d + 1} \quad (18)$$

2) Downlink Case: In WCDMA-based systems, the total DL transmitted power needed to satisfy all $\boldsymbol{n}_w$ simultaneous voice and data users in a single-cell scenario should be [14]:

$$\check{P}_T(\boldsymbol{n}_w) = \frac{P_p + N_T \sum_{i=1}^{n_w^v + n_w^d} \frac{L_i}{\check{A}_i + \rho}}{1 - \check{\eta}} \leq \check{P}_{T,\max} \quad (19)$$

with

$$\check{A}_i = \frac{W}{\check{\theta}_i \check{R}_i} \qquad \check{\eta} = \sum_{i=1}^{n_w^v + n_w^d} \frac{\rho}{\check{A}_i + \rho} \quad (20)$$

Where P_p and N_T are the pilot and thermal noise powers respectively, ρ is the DL orthogonality factor, and $\check{P}_{T,\max}$ is the maximum total DL power available at the base station. W is the WCDMA chip-rate, and user requirements are in the form of requested bit rates $\check{R}_i$ and target bit-energy-to-noise-density ratios $\check{\theta}_i$. The path-loss experienced by each user, L_i, may be characterized by considering macro or micro-cell environments with shadowing effects modeled by means of log-normal variation [14].

Consequently, for $\boldsymbol{n}_w$ simultaneous users, a congestion situation may be detected whenever the total DL power exceeds a given power threshold $\check{P}_{T,c} < \check{P}_{T,\max}$. Then, after some manipulations, the congestion probability in the DL yields:

$$\begin{aligned} &\Pr(\check{P}_T(\boldsymbol{n}_w) > \check{P}_{T,c}) = \\ &= \Pr\left(\sum_{i=1}^{n_w^v + n_w^d} \frac{L_i}{\check{A}_i + \rho} > \frac{\check{P}_{T,c}(1-\check{\eta}) - P_p}{P_N}\right) \\ &= \Pr\left(\gamma_{(\boldsymbol{n}_w)} > \gamma^*_{(\boldsymbol{n}_w)}\right) \end{aligned} \quad (21)$$

which depends on the path-loss distribution and consequently on the user′s geographical location.

Given a number of active voice and data users, the congestion probability due to downlink power availability may be expressed as:

$$\check{P}_c^w(\boldsymbol{N}_w) = \sum_{n_w^v=0}^{N_w^v} \sum_{n_w^d=0}^{N_w^d} \Pr\left(\gamma_{\boldsymbol{n}_w} > \gamma^*_{\boldsymbol{n}_w}\right) P_{\check{\boldsymbol{\alpha}}}(\boldsymbol{n}_w|\boldsymbol{N}_w) \quad (22)$$

B. TDMA Case

As suggested in [12], a possible indicator of congestion in TDMA access technologies can be based on the effect of timeslot sharing among data users. Accordingly, the congestion probability in TDMA, given n_t^v simultaneous voice users are assigned τ_v TSLs and n_t^d simultaneous data users are assigned τ_d data TSL, can be expressed as:

$$\Pr\left(\xi_{(\tau_v,\tau_d)} < \xi_c\right) = \begin{cases} 1 & \text{if } \xi_{(\tau_v,\tau_d)} < \xi_c \\ 0 & \text{if } \xi_{(\tau_v,\tau_d)} \geq \xi_c \end{cases} \quad (23)$$

TABLE II
MAIN SYSTEM PARAMETERS FOR NUMERICAL EVALUATION.

WCDMA Parameter	Value	WCDMA Parameter	Value	TDMA Parameter	Value
N_T	-100 dBm	$\{\check{\xi}_v, \check{\xi}_d\}$	{128,64}	C	23
P_p	30 dBm	$\hat{\eta}_c$	0.7	$\hat{n}_C$	3
ρ	0.6	$\hat{\eta}_{max}$	0.9	$\check{n}_C$	4
$\check{P}_{T,c}$	35 dBm	$\{\hat{\theta}_v, \hat{\theta}_d\}$	{6,5} dB	$\hat{\xi}_c = \check{\xi}_c$	0.35
$\check{P}_{T,max}$	43 dBm	$\mathcal{C}_{max}$	63/64	$\hat{q}$	3
$\{\check{\theta}_v, \check{\theta}_d\}$	{6,7} dB	$\hat{R}_v$	12.2 kbps	$\check{q}$	8
$\check{R}_v$	12.2 kbps	$\hat{R}_d$	32 kbps	$\hat{\alpha}_v = \check{\alpha}_v$	1
$\check{R}_d$	64 kbps	$\hat{\alpha}_v = \hat{\alpha}_v$	0.5	$\hat{\alpha}_d = \check{\alpha}_d$	0.5
W	3.84 Mcps	$\check{\alpha}_d = \check{\alpha}_d$	0.5		

where, if C is the total number of available TSLs,

$$\xi_{(\tau_v,\tau_d)} = \min(C - \tau_v, \tau_d)/\tau_d \quad (24)$$

is the Reduction Factor [16] that accounts for the effect of TSL sharing among data users in a TDMA system, like e.g. GSM/EDGE. It follows from (24) that ξ takes values between 0 and 1, meaning a very saturated network for ξ close to 0 (high TSL sharing), and a low loaded network for ξ close to 1 (low TSL sharing). According to (23), congestion is detected if the Reduction Factor ξ falls below a given threshold ξ_c.

Then, given a number of N_t^v and N_t^d admitted users in TDMA, we may compute the probability of n_t^v voice users and n_t^d data users simultaneously transmitting in TDMA in the same way as (16). These simultaneous users will occupy a single TSL in the case of voice users, i.e. $\tau_v = n_t^v$, and q TSL in the case of data users, i.e. $\tau_d = n_t^d q$. The resulting congestion probability can be defined as:

$$P_c^t(\boldsymbol{N}_t) = \sum_{n_t^v=0}^{N_t^v} \sum_{n_t^d=0}^{N_t^d} \Pr\left(\xi_{(n_t^v, n_t^d q)} < \xi_c\right) \cdot P_{\boldsymbol{\alpha}}(\boldsymbol{n}_t|\boldsymbol{N}_t) \quad (25)$$

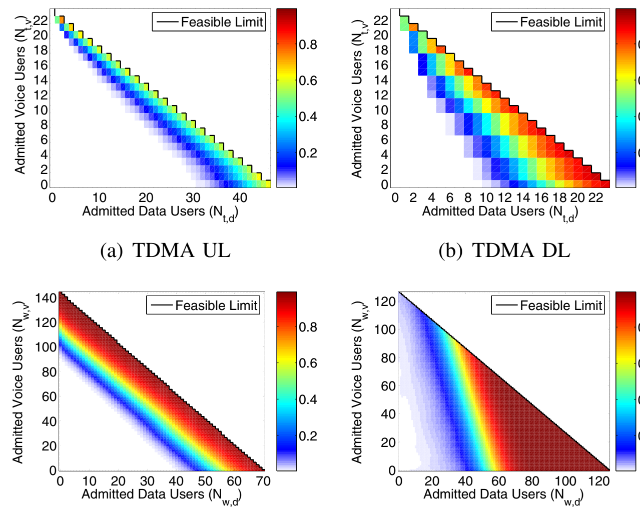
where parameters ξ, ξ_c, q, and α must be particularised in order to obtain the TDMA congestion probability in both links, i.e. $\hat{P}_c^t(\boldsymbol{N}_t)$ and $\check{P}_c^t(\boldsymbol{N}_t)$.

IV. NUMERICAL RESULTS

In this section, congestion probability expressions for both links in TDMA and WCDMA are evaluated for the particular parameter settings given in Table II. In addition, the congestion probability that arises from the application of the presented RAT selection policies, i.e. LB, SB#1 and SB#2, is also evaluated and discussed.

Fig. 1(a) and Fig. 1(b) show the UL and DL congestion probability distribution in TDMA, i.e. $\hat{P}_c^t(\boldsymbol{N}_t)$ and $\check{P}_c^t(\boldsymbol{N}_t)$ respectively[6]. For the considered TDMA system parameters provided in Table II, data users are granted with a higher amount of TSL in DL than in the UL ($\check{q} = 8$ as opposed to $\hat{q} = 3$). This asymmetry in channel allocation causes the congestion probability to be higher in the DL case (Fig. 1(b)), where TSL reuse and thus congestion is potentially higher, than in the UL case (Fig. 1(a)). Recall that congestion in TDMA is due to excessive TSL sharing of data users and therefore these users have higher influence on congestion than voice users.

[6]Congestion probabilities are represented in the form of a color-scale graphs with the color-bar indicating probability values on the right of each graph.



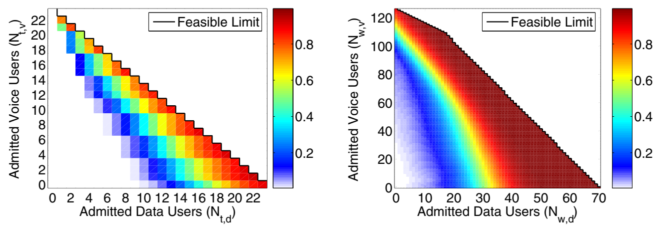
(a) TDMA UL (b) TDMA DL (c) WCDMA UL (d) WCDMA DL

Fig. 1. Congestion probability distributions in TDMA and WCDMA.

Similarly, Fig. 1(c) and Fig. 1(d) show the UL and DL congestion probability distribution in WCDMA, i.e. $\hat{P}_c^w(\boldsymbol{N}_w)$ and $\check{P}_c^w(\boldsymbol{N}_w)$ respectively. In this case, for the considered parameter settings, it can be seen that the strain in the DL congestion probability (Fig. 1(d)) is set on the number of data users in the system due to their higher demands in terms of bit rates, which is translated into higher power demands and, consequently, higher congestion chances. In the UL case (Fig. 1(c)), a more symmetric degradation due to voice and data users is observed.

Considering the above, the congestion probability in TDMA, $P_c^t(\boldsymbol{N}_t)$, and in WCDMA, $P_c^w(\boldsymbol{N}_w)$, (as defined in (13)) is shown in Fig. 2(a) and Fig. 2(b) respectively. Since the congestion probability in TDMA is mainly DL-driven, the congestion probability in TDMA in Fig. 2(a) resembles to that of Fig. 1(b). As for WCDMA, although both UL and DL congestion probabilities influence to some extent the congestion probability in Fig. 2(b), it can be noticed that data users have a higher impact on the overall congestion probability in WCDMA.

In order to capture the influence of different RAT selection policies on the congestion probability as formulated in (12), Fig. 3 shows the congestion probability that arises in TDMA and in WCDMA when RAT selection policies SB#1, SB#2, and LB are used. From the global congestion probability, as defined in (14) and shown in Fig. 3 (top), we observe that the LB strategy achieves an overall lower congestion probability as compared to service-based strategies SB#1 and SB#2, with SB#2 exhibiting the poorest performance. Note that SB#2 mainly allocates data users to TDMA while voice users to



(a) TDMA (b) WCDMA

Fig. 2. Congestion probability distributions in TDMA and WCDMA.

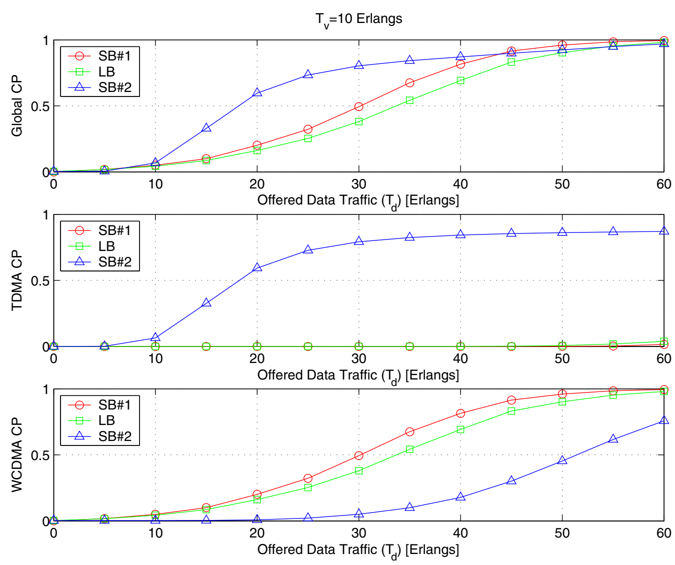


Fig. 3. Global congestion probability (top), congestion probability in TDMA (middle) and in WCDMA (bottom).

WCDMA, thus congestion arises in both RATs with special emphasis in TDMA where data users are forced to share TSLs in excess. As for SB#1, which allocates voice users to TDMA, and LB, which operates in such a way that data users are not forced to share TSLs in TDMA unless no capacity is left in WCDMA [13], they present hardly no congestion in TDMA. In WCDMA, given SB#1 primarily allocates data users to this RAT, this strategy presents the highest overall congestion in WCDMA. LB will allocate both voice and data users in WCDMA, therefore congestion will rise whenever data traffic load increases. Nevertheless, given LB allocates voice and data resources in WCDMA as opposed to SB#1 that allocates only data users, the congestion probability in WCDMA is somewhat lower for LB than for SB#1. Then, the higher flexibility of LB in allocating resources in both RATs causes this strategy to outperform SB#1 and SB#2 in terms of congestion probability.

V. Conclusion

In this paper, a statistical characterization of radio access congestion in multi-RAT environments has been presented for the case of TDMA and WCDMA based technologies. Congestion probabilities have been derived for both the UL and DL as a function of the number of admitted voice and data users in the system. In addition, the evaluation of several RAT selection policies, Load Balancing (LB) and Service-Based (SB), has been carried out and the impact of these policies on the achieved congestion probability assessed. Results indicate a better performance of the LB policy which prevents from data TSL reuse in TDMA along with a higher flexibility in allocating voice and data users in TDMA and WCDMA. Further work will be devoted to use congestion probability information as a RAT selection criteria in order to choose, upon service arrival, the RAT that undergoes a lower congestion probability for such service.

References

[1] Y. Kim, B. J. Jeong, J. Chung, C.-S. Hwang, J. Ryu, K.-H. Kim, and Y. K. Kim, "Beyond 3G: vision, requirements, and enabling technologies," *IEEE Communications Magazine*, vol. 41, no. 3, pp. 120–124, Mar 2003.

[2] S. Uskela, "Key concepts for evolution toward beyond 3G networks," *IEEE Wireless Communications*, vol. 10, no. 1, pp. 43–48, Feb. 2003.

[3] A. Tolli, P. Hakalin, and H. Holma, "Performance evaluation of common radio resource management (CRRM)," in *Proc. IEEE International Conference on Communications (ICC'02)*, vol. 5, New York, NY, USA, 2002, pp. 3429–33.

[4] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Communications*, vol. 10, no. 1, pp. 49 – 55, Feb. 2003.

[5] A. Furuskar and J. Zander, "Multiservice allocation for multiaccess wireless systems," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 174–84, Jan. 2005.

[6] G. Fodor, A. Furuskar, and J. Lundsjo, "On access selection techniques in always best connected networks," in *Proc. of the 16th ITC Specialist Seminar on Performance Evaluation of Wireless and Mobile Systems*, Antwerp, Belgium, 2004.

[7] R. Chakravorty et al., "Performance issues with vertical handovers - experiences from GPRS cellular and WLAN hot-spots integration," in *Proc. Second IEEE Annual Conference on Pervasive Computing and Communications*, Orlando, FL, USA, Mar. 2004, pp. 155 – 64.

[8] J. Perez-Romero, O. Sallent, and R. Agusti, "A generalized framework for multi-RAT scenarios characterisation," in *Proc. of the 2007 IEEE 65th Vehicular Technology Conference*, Dublin, Ireland, Apr. 2007, pp. 980 – 4.

[9] 3GPP TR 25.922 v6.0.1, "Radio resource management strategies (rel. 6)".

[10] F. Malavasi, M. Breveglieri, L. Vignali, P. Leaves, and J. Huschke, "Traffic control algorithms for a multi access network scenario comprising GPRS and UMTS," in *Proc. of the 57th IEEE Semiannual Vehicular Technology Conference. VTC 2003*, vol. 1, Jeju, South Korea, Apr. 2003, pp. 145 – 9.

[11] X. Gelabert, J. Perez-Romero, O. Sallent, and R. Agusti, "Congestion control strategies in multi-access networks," in *Proc. IEEE 3rd International Symposium of Wireless Communication Systems (ISWCS'06)*, Valencia, Spain, Sep. 2006.

[12] ——, "On managing multiple radio access congestion events in B3G scenarios," in *Proc. IEEE 65th Vehicular Technology Conference - VTC2007-Spring*, Dublin, Ireland, 2007, pp. 1137 – 1141.

[13] X. Gelabert, J. Pérez-Romero, O. Sallent, and R. Agustí, "A Markovian Approach to Radio Access Technology Selection in Heterogeneous Multi-Access/Multi-Service Wireless Networks," *Accepted at IEEE Transactions on Mobile Computing*, Apr. 2008.

[14] J. Perez-Romero, O. Sallent, R. Agusti, and M. Diaz-Guerra, *Radio Resource Management Strategies in UMTS*. John Wiley & Sons, 2005.

[15] L. Kleinrock, *Queueing Systems, Volume 1: Theory*. John Wiley & Sons, 1975.

[16] T. Halonen, J. Melero, and J. Romero-Garcia, *GSM, GPRS and EDGE Performance: Evolution Toward 3G/UMTS*. New York, NY, USA: John Wiley & Sons, 2002.

[17] W. J. Stewart, *Introduction to the Numerical Solution of Markov Chains*. Princeton Univ. Press, 1994.