Joint uplink and downlink admission control to both streaming and elastic flows in CDMA/HSDPA systems

T. Chahed [a], E. Altman [b], S.E. Elayoubi [c]

[a] *IT-TELECOM SudParis*
UMR CNRS 5157
9 rue C. Fourier
91011 Evry CEDEX, France
tijani.chahed@it-sudparis.eu

[b] *INRIA-Sophia Antipolis*
2004 route des Lucioles - BP 93
06902 Sophia Antipolis, France
eitan.altman@sophia.inria.fr

[c] *Orange Labs*
38, rue du Général Leclerc
92794 Issy les Moulieaux, France
salaheddine.elayoubi@orange-ftgroup.com

Abstract

TCP-based data flows generate packets and ACKs in two directions, be it in the wireline or wireless networks. In the latter case, packets are typically found in the downlink whereas ACKs are in the uplink. Those two links are asymmetric in the case of CDMA-based High Data Rate (HDR)/High-Speed Downlink Packet Access (HSDPA) systems, the uplink being much slower than the downlink and thus, in some cases, restrictive in terms of the achievable throughput of the TCP flow. The aim of this work is to evaluate the performance of such a setting, in the presence of both streaming and elastic traffic, under a dynamic scenario where users arrive to the system and leave it after completion of their service. We specifically quantify the impact of the uplink on the overall performance of TCP and study the model variations as a function of several parameters such as load, file size and radio conditions [1].

Key words: Capacity, joint uplink and downlink, HSDPA/HDR, TCP.

[1] This work was jointly supported by INRIA ARC IFANY and Euro-NGI/FGI specific project CELLULAR

1 Introduction

The choice that has been made in CDMA-based systems to accompany the arrival of data-oriented applications in the third generation mobile systems is to offer larger bit rates in the downlink where most of these applications are expected to be. And so the overall system became asymmetric, with lower rate uplink, mostly based on Dedicated CHannels (DCHs), suitable for streaming applications such as voice, and higher bit rate downlink, implementing High Data Rate (HDR) [5]/High-Speed Downlink Packet Access (HSDPA) [7], and which offers packet-based applications a high-rate shared medium.

Elastic flows however typically use TCP in the transport layer, with packets flowing on the higher-rate downlink, with instantaneous rates up to some 10Mbps, the ACKs, as a flow on their own, returning on the slower rate uplink, with rates which can be as small as 16Kbps. The asymmetry in architecture may result in cases where the uplink is restrictive and does not let data flows take full advantage of all the available capacity in the downlink.

So far, to the best of our knowledge, works in CDMA systems considered capacity separately in the uplink [3] and downlink [8]. The only work that we are aware of that considers the joint capacity in the uplink and downlink is contained in Reference [12]. It however considers dedicated channels only and not HSDPA and the data flows are considered in one direction only, i.e., it does not take into account the flow of ACKs in the opposite direction.

Other works considered asymmetry in links and quantified their impact on TCP performance (for instance Reference [14] and references cited therein). These works however consider solely the wireline context, are mostly carried out at the packet level and do not consider a mixture of streaming and elastic flows.

In this work, we investigate the simultaneous joint capacity of the uplink and downlink in the presence of both streaming and elastic flows, the latter governed by TCP at the transport layer, with packets in the downlink and ACKs returning back in the uplink. The main difference between those types of flows is that streaming flows require some constant bit rate and their service duration is independent of the quantity of resources they get. This is not the case of data flows which have the ability to share resources in a fair manner among themselves and would leave the system sooner if they get more resources.

Please note that the case of packets in the uplink and corresponding ACKs in the downlink is less typical and less restrictive with respect to our problem statement; our model and analysis are however applicable to that case and to the case with packets and ACKs in both links as well.

Please note also that we also apply, in this work, admission control to both types of flows, streaming and elastic. This results in a finite Quasi-Birth-Death (QBD) model. Have elastic flows been left without admission control, this results in an infinite QBD model as shown in our previous work contained in Reference [4].

The remainder of this work is organized as follows. In section 2, we develop analytical models for the capacity and throughputs in a UMTS/HSDPA system. In Section 3, we show our model for both the uplink and downlink and detail the arrival and departure processes and mean rates of both streaming and elastic flows. In the next section we show our analysis based on the Quasi-Birth Death (QBD) process with a matrix-geometric solution to the steady-state probabilities of the system. In Section 5, we show some performance evaluation results to illustrate the impact of the uplink on the overall TCP performance and this with respect to several system parameters, namely load, file size and radio conditions. Section 6 contains some analysis on how to make the uplink less restrictive by means of squeezing the streaming flows present in the uplink. Section 7 eventually concludes the paper.

2 System

2.1 Description

The system we focus on in this work is as shown in Figure 1. It is mainly composed of the UMTS/HSDPA network, itself subdivided into a CDMA-based radio link and a Core Network (CN), possibly some wireline Internet cloud and a mobile equipment. The downlink implements HSDPA and the uplink is composed of classical UMTS dedicated channels for streaming flows and one common channel for elastic ones, ACKs in our case.

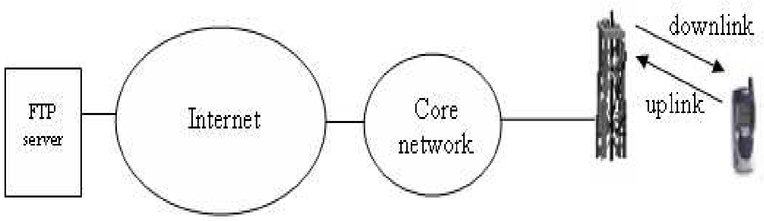

Fig. 1. End-to-end system

We consider a system with two types of calls: i. streaming, e.g., voice, carried on both the uplink and downlink and ii. elastic, e.g. FTP, where a file is for instance retrieved from a server and is then carried, using TCP, to the user on the downlink only. The downlink elastic flows generate TCP ACKs that are sent on the uplink.

We next model the capacity of the UMTS/HSDPA system for both uplink and downlink. These results will be used in the subsequent section where analysis and performance evaluation is performed.

2.2 Downlink model

2.2.1 Streaming calls

In the downlink, if cell 0 contains X_d active users, the Signal-to-Interference plus Noise Ratio (SINR) achieved for user k, situated at distance d_k from its own base station, is given by:

$$SINR_k = \frac{P_{k,0}/q_{k,0}}{I_{inter,k} + I_{intra,k} + N_0} S_k$$

where $P_{k,0}$ is the power used by the base station of cell 0 towards user k, $q_{k,0}$ is the path loss between the target base station and user k and which depends on the distance d_k, I is the interference (both intra- and inter-cell), N_0 is the noise term and S_k is the spreading factor.

To analyze the interference, let us first note that the intra-cell interference originates from the common channels and from other users can be written as:

$$I_{intra}(d_0) = \alpha(P_{tot,0} - P_{k,0})/q_{k,0}$$

α being the orthogonality factor and $P_{tot,l}$ the total transmitted power by base station l. For the inter-cell interference, it is given by:

$$I_{inter,k} = \sum_{l \neq 0} P_{tot,l}/q_{k,l}$$

with $P_{tot,l} = \chi_d P_{max}$, where χ_d is the average load in a typical cell of the system, defined as the ratio between the used and total powers, and P_{max} is the maximal transmission power (please refer to Table 1 for a collection of the main notations used throughout the text).

$SINR_k$	Signal to Interference plus Noise Ratio of user k
$(SINR)_u^s$	SINR of streaming user in the uplink
$P_{k,0}$	power emitted by BS 0 towards mobile k
P_{max}	maximal transmission power
P_{CCH}	power of CCH channels
$P_d^{s,e}$	power of streaming or elastic users in the downlink
P_{SCCH}	power of SCCH channel
P_u^s	power of streaming user in the uplink
$P_u^{i,j}$	power emitter by user i in cell j
$P_{tot,l}$	total power emitter by BS l
P_{CPCH}	power for CPCH channel
I_{tot}	overall power received by BS
$\bar{P}$	average received power at BS
$q_{k,0}$	path loss between BS 0 and mobile k
$q_k^{i,j}$	path loss between user i in cell j and cell 0
I	interference
I_{intra}	intra-cell interference
$I_{inter,k}$	inter-cell interference
d_k	distance between BS and user k
N_0	noise term
S_k	spreading factor for user k
α	orthogonality factor
χ_d	average load
$\chi_{u,max}$	maximum load in the uplink
F_k	F-Factor
$E^{s,e}$	bit energy of streaming or elastic flows
W/R_u^s	processing gain

Table 1
Main notations used in the model

On the other hand, the value $\sum_{l\neq 0} q_{k,0}/q_{k,l}$ is the well-known F-factor F_k [11] [17]. The SINR is then equal to:

$$SINR_k = \frac{P_{k,0}}{\alpha(P_{tot,0} - P_{k,0}) + \chi_d P_{max} F_k + N_0 q_{k,0}} S_k$$

leading to the expression:

$$\beta_k = \frac{P_{k,0}}{\alpha P_{tot,0} + \chi_d P_{max} F_k + N_0 q_{k,0}}$$

where we define

$$\beta_k = \frac{SINR_k}{S_k + \alpha.SINR_k} \tag{1}$$

If X_d^s streaming users, with no HSDPA users, were present in the downlink, this leads to the power expression:

$$P_{tot,0} = \frac{P_{CCH} + \beta^s \sum_k (\chi_d P_{max} F_k + N_0 q_{k,0})}{1 - \alpha X_d^s \beta^s}$$

where $\beta^s = \frac{SINR^s}{S^s + \alpha.SINR^s}$, P_{CCH} is the power associated with the Common CHannels (CCHs), $SINR^s$ is the target SINR for streaming flows and S^s is their spreading factor.

Considering the average values over the cell, we have:

$$P_{tot,0} = \frac{P_{CCH} + \beta^s (\chi_d P_{max} \bar{F} + N_0 \bar{q}) X_d^s}{1 - \alpha \beta^s X_d^s}$$

now considering the constraint on the maximal transmission power ($P_{tot,0} \leq P_{max}$), the constraint on the number of users becomes:

$$\beta^s (\chi_d P_{max} \bar{F} + N_0 \bar{q} + \alpha P_{max}) X_d^s \leq P_{max} - P_{CCH}$$

The capacity of the system in the downlink C_d is thus equal to:

$$C_d = \frac{P_{max} - P_{CCH}}{\beta^s (\chi_d P_{max} \bar{F} + N_0 \bar{q} + \alpha P_{max})}$$

2.2.2 Elastic calls

Data calls are handled on the HSDPA High Speed Downlink Shared Channel (HS-DSCH). In the presence of at least one HSDPA call, the whole available power will be used, i.e. $P_{tot,0} = P_{max}$, and the power used by a streaming call becomes:

$$P_d^s = \frac{\alpha \beta^s}{1 - \alpha \beta^s} P_{CCH} + \frac{\beta^s}{1 - \alpha \beta^s} (\alpha P_{max} + \chi_d P_{max} \bar{F} + N_0 \bar{q}).$$

For an HSDPA call, the achieved value of β^e is equal to:

$$\beta^e = \frac{P_d^e}{(\alpha P_{max} + \bar{\chi_d} P_{max} \bar{F_i} + N_0 \bar{q_i})}, \tag{2}$$

where P_d^e is the power received by a HSDPA call. Since a HSDPA user utilizes the whole available power while receiving, its power is:

$$P_d^e = P_{max} - P_{CCH} - P_{SCCH} - P_d^s X_d^s \tag{3}$$

with P_{SCCH} the power associated with the Shared Control Channel (SCCH). Therefore, the SINR of one HSDPA user can be calculated using the definition of β in Eqn. (1).

Knowing this SINR of HSDPA users, the throughput is based on link level curves function of the SINR; let t denote such a function. Thus, the overall HSDPA throughput for data flows in the cell is given by:

$$T_d^e(X_d^s) = t\left(\frac{P_{max} - P_{CCH} - P_{SCCH} - \beta^s(\alpha P_{max} + \chi_d P_{max}\bar{F} + N_0\bar{q})X_d^s}{\chi_d P_{max}\bar{F} + N_0\bar{q} + \alpha(P_{CCH} + P_{SCCH}) + \alpha\beta^s(\alpha P_{max} + \chi_d P_{max}\bar{F} + N_0\bar{q})X_d^s}\right) \tag{4}$$

and the throughput of one HSDPA user is given by $\frac{T_d(X_d^s)}{X_d^e}$ where X_d^e is the number of elastic flows in the downlink. Note that this throughput depends on the radio conditions of data users in the cell (through the mean path loss and F-factor).

Owing to the presence of streaming calls in the downlink, the condition:

$$P_{max} - P_{CCH} - P_{SCCH} - P_d^s X_d^s > 0 \tag{5}$$

has to be imposed in order to guarantee that there exists some power dedicated to this purpose. Thus, Eqn. (5) provides the maximal number of UMTS calls:

$$N_d = \left\lfloor \frac{P_{max} - P_{CCH} - P_{SCCH}}{\beta^s(\chi_d P_{max}\bar{F} + N_0\bar{q} + \alpha P_{max})} \right\rfloor \tag{6}$$

when there exists a nonzero number of HSDPA calls.

2.3 Uplink model

2.3.1 Streaming calls

In the uplink, the $(SINR)_u^s$ received from a streaming mobile of a given cell 0 must be greater than a given constant to guarantee the reception of the signal at the Base Station (BS):

$$(SINR)_u^s = \frac{P_u^s}{I_{intra,0} + I_{inter,0} + N_0} \geq \tilde{\Delta}^s = \frac{E^s}{N_0}\frac{R_u^s}{W}$$

E^s/N_0 is the minimum allowed ratio between the bit energy and the interference plus noise density, which guarantees the target quality of service in terms of bit error probability; W/R_u^s is the processing gain, i.e., the ratio between the chip rate and the source bit rate, N_0 is the background noise, $I_{intra,0}$ and $I_{inter,0}$ are the total powers received from other mobiles within the considered cell and all its neighbours and P_u^s is the power of a streaming user in the uplink.

$$I_{intra,0} = X_u^s P_u^s$$

$$I_{inter,0} = \sum_{j \neq 0} \sum_{i=1}^{X_{u,j}^s} P_u^{i,j}$$

where $P_u^{i,j}$ is the power emitted by mobile i in cell j and $X_{u,j}^s$ is the number of streaming users in the uplink of cell j. We introduce the factor $f = E[\sum_{j\neq 0}\sum_{i=1}^{X_{u,j}^s} \frac{q_0^{i,j}}{q_j^{i,j}}]$, where $q_k^{i,j}$ is the path loss between a user i in cell j and the base station of cell k. If $\bar{P}$ is the average received power by a base station of the system, and considering the minimal power that can achieve the target SIR, we obtain:

$$\tilde{\Delta}^s = \frac{P_u^s}{X_u^s P_u^s + \bar{P} f + N_0 - P_u^s}$$

Defining

$$\Delta^s = \frac{\tilde{\Delta}^s}{1 + \tilde{\Delta}^s} \tag{7}$$

we obtain:

$$P_u^s = \Delta^s (X_u^s P_u^s + \bar{P} f + N_0)$$

leading to:

$$P_u^s = \frac{(\bar{P}f + N_0)\Delta^s}{1 - X_u^s \Delta^s} \tag{8}$$

To ensure a good operation of the system, we must fix a constraint on the load of the cell. Note that this load is defined by:

$$\chi_u = \frac{I_{tot}}{I_{tot} + N_0}$$

where I_{tot} is the overall power received by the base station:

$$I_{tot} = X_u^s P_u^s + \bar{P}f$$

And the uplink constraint is

$$X_u^s \Delta^s \leq \frac{(\bar{P}f + N_0)\chi_{u,max} - \bar{P}f}{N_0}$$

where $\chi_{u,max}$ is the maximum load in the uplink. Eventually, the capacity of the system in the uplink is given by:

$$C_u = \left\lfloor \frac{(\bar{P}f + N_0)\chi_{u,max} - \bar{P}f}{N_0 \Delta^s} \right\rfloor$$

Remark 1. The definition of capacity we adopt in this work is the *effective bandwidth* that appears in our previous work contained in Reference [9] and is indeed a bound on the sum of the target SIR as shown in Reference [13].

2.3.2 TCP ACKs

For TCP ACKs, and in the absence of high Speed Uplink Packet Access (HSUPA), they may be carried either on Dedicated (DCH) or on Common Packet CHannels (CPCHs) [11]. DCHs are not suitable for these kinds of transmissions, as the code resources will be taken for the whole communication, CPCH is the most suitable channel. In the presence of X_u^s streaming calls in the uplink, the available power that can be accepted for the reception of CPCH is:

$$P_{CPCH}(X_u^s) = \frac{\chi_{u,max} N_0}{1 - \chi_{u,max}} - (X_u^s P_u^s + \bar{P}f)$$

Using Eqn. (8), this leads to the value:

$$\Delta_{CPCH}(X_u^s) = P_{CPCH}(X_u^s)\frac{1 - X_u^s\Delta^s}{(\bar{P}f + N_0)}$$

and using Eqn. (7), we obtain the throughput of data calls in the uplink:

$$T_u^e(X_u^s) = \frac{W}{E^e/N_0}\frac{\Delta_{CPCH}(X_u^s)}{1 - \Delta_{CPCH}(X_u^s)} \tag{9}$$

where E^e/N_0 is the minimum allowed ratio between the bit energy and the interference plus noise density, required to decode a communication on the CPCH channel.

2.3.3 Capacity reservation

In order to insure a minimal throughput for sending TCP ACKs in the uplink, a bandwidth reservation is possible, e.g., by fixing a maximal load value to the streaming calls $\chi_{u,max}^s < \chi_{u,max}^s$. The maximal equivalent number of streaming users is thus:

$$N_u = \frac{(\bar{P}f + N_0)\chi_{u,max}^s - \bar{P}f}{N_0\Delta^s}$$

And the corresponding minimal throughput allocated to ACKs is equal to:

$$T_{u,min}^e = \frac{W}{E^e/N_0}\frac{\Delta_{CPCH}(X_{u,max}^s)}{1 - \Delta_{CPCH}(X_{u,max}^s)}$$

3 Capacity modeling

Let the uplink be modeled as a server with capacity C_u and let the downlink be modeled as a server with capacity C_d, as derived above.

Let the arrival of streaming flows be Poisson with mean arrival rate λ_u^s in the uplink and λ_d^s in the downlink. These flows are assumed to have a service exponentially distributed with mean duration equal to T^s or equivalently a mean service rate $\mu^s = 1/T^s$. Each streaming flow is transported over a dedicated link with rate R_u^s, for instance 16Kbps. Again, the service duration is independent of the amount of capacity granted to this type of flow.

We consider for the time being that those two sets of streaming flows, uplink and downlink, are independent and that the maximum number of such flows in the uplink is N_u and in the downlink N_d; $N_u \leq C_u$ and $N_d \leq C_d$. If we are to model interactive streaming traffic, such as telephony, the number of flows of this type of traffic should be the same in both directions.

Let λ^e denote the arrival rate for data flows in the downlink. We adopt an admission scheme that gives priority to voice flows over data ones. Based on this, data flows share (fairly) the available capacity left over by streaming ones. Once in the system, we assume that this data packet flow generates instantaneously a corresponding stream of ACKs in the uplink. Now this overall data flow, packets and ACKs, shall take the minimum capacity between the bandwidth left over by streaming flows in the downlink to process data packets and the bandwidth left over by streaming flows in the uplink to process smaller size ACKs. Recall that we assume that ACKs are transported over shared channels in the uplink too.

And so, in total, each data flow obtains on average, some R^e throughput given by:

$$R^e = \min(\frac{T_d^e(X_d^s(t))}{X^e(t)}, \frac{T_u^e(X_u^s(t))}{X^e(t)} \frac{s_p b}{s_a}) \text{ for } X^e > 0 \tag{10}$$

where X^e is the number of concurrent data flows in the system, X_u^s is the number of streaming flows in the uplink, X_d^s is the number of streaming flows in the downlink, b is the number of packets acknowledged by a cumulative ACK, s_p is the packet size and s_a is the ACK size. T_u^e and T_d^e are the throughputs achieved by data calls in the uplink and downlink, respectively, and are calculated in the previous sections (Eqns. (4) and (9)).

The departure rate μ^e for data flows is given by:

$$\mu^e = \frac{R^e}{E[Z]} X^e \tag{11}$$

where $E[Z]$ is the mean file size.

Remark 2. In the above, we assumed that data flows are managed through some round robin scheduling algorithm in the downlink. It is however common to use some kind of opportunistic scheduling in HSDPA-based systems, such as Proportional Fair Scheduling (PFS). In this case, a gain function representing the radio conditions of a given user [6] must be added to the throughput shown in the above-mentioned equations.

Remark 3. It should be clear that the resources used by data flows are function of their number as well as the number of streaming calls in progress in the system, i.e., we should have written $R^e(X)$ and $\mu^e(X)$ where X is a vector denoting the number of streaming and elastic flows X^s and X^e respectively. As will be seen in the last section of this paper, even R^s, the throughput of streaming flows, will be made dependent on X. We however drop (X) for the sake of notational convenience.

It is important to note that the standard approach to compute the system's capacity so far has been to identify which direction seems to be the bottleneck, and then to compute the capacity in that direction assuming that there are no bandwidth limitation on the opposite direction. If one does not know a-priori which of the directions is the bottleneck then one would compute the uplink capacity ignoring the downlink limitation then compute the downlink capacity ignoring the uplink limitations and take the minimum.

To see that this approach gives the wrong capacity, we note from Eq. (10) that the total expected throughput C^e available to all data connections is $E[\min(T^e_d(X^s_d), T^e_u(X^s_u)\frac{s_p b}{s_a}]$. The approach that would decouple the up and down directions amounts to modeling the uplink expected available throughput as $E[T^e_u(X^s_u)\frac{s_p b}{s_a}]$ and that of the downlink as $E[T^e_d(X^s_d)]$. The available throughput of the system $C^e_{decouple}$ would then be the minimum of the two expressions. Now since $\min(x, y) = \min(x - y, 0) + y$ and since $\min(z, 0)$ is concave, if follows from Jensen's inequality that $C^e \leq C^e_{decoupled}$. One can easily find parameters where this inequality is strict. In that case one can have an unstable system for some arrival rate of data sessions while the standard approach would predict that the system is stable. One can show in fact that the number of data sessions in the decoupled model is stochastically smaller than that of the real system.

4 Analysis

We assume that both types of flows are subject to admission control. The CAC ensures in this case that the capacities in both uplink and downlink are not exceeded, i.e., $X^s_i \leq N_i$, $i = u, d$ and $X^e \leq N^e_{max}$, where N^e_{max} is the maximum number of elastic users to be admitted in the system.

Our system of asymmetric uplink and downlink transporting streaming and elastic flows modeled above can be solved as follows [2]. The number of streaming flows in progress $X^s_i(t)$, $i = u, d$, is a birth-death process with parameters λ^s_i and μ^s. The steady state probabilities $\pi(.)$ are given by the Erlang formula as:

$$\pi(X_i^s = x) = \frac{1}{\sum_{k=0}^{N_i} \frac{(\rho_i^s)^k}{k!}} \frac{(\rho_i^s)^x}{x!} \tag{12}$$

where $\rho_i^s = \lambda_i^s/\mu^s$.

The blocking probability B_i^s, $i = u, d$, of streaming flows is given by:

$$B_i^s = \frac{1}{\sum_{k=0}^{N_i} \frac{(\rho_i^s)^k}{k!}} \frac{(\rho_i^s)^{N_i}}{N_i!} \tag{13}$$

The process $(X_u^s(t), X_d^s(t), X^e(t))$ referring to the number of streaming flows in the uplink and downlink as well as the number of data flows, respectively, is a finite homogeneous Quasi-Birth and Death (QBD) process with infinitesimal generator Q given by [10]:

$$\mathbf{Q} = \begin{bmatrix} B_1 & A_0 & 0 & 0 & ... \\ A_2 & A_1 & A_0 & 0 & ... \\ 0 & A_2 & A_1 & A_0 & ... \\ 0 & . & . & . & 0 \\ ... & 0 & A_2 & A_1 & A_0 \\ ... & ... & 0 & A_2 & B_2 \end{bmatrix}$$

where A_0, A_1 and A_2 are square matrices of size $(N_u + 1)(N_d + 1)$ which we denote by N (recall that N_u and N_d are the maximum number of streaming flows that are admitted to the system in the uplink and downlink respectively). A_0 represents the data flows arrivals, with arrival rates λ^e at the diagonal, i.e., $A_0[i, i] = \lambda^e$, A_2 represents their departures, with mean departure rates μ^e given by Eqn. (11) at the diagonal too, and hence $A_2[i, i] = \mu^e$. A_1 corresponds to the arrival and departure of voice flows. It is a tri-diagonal matrix, with mean arrival rates at the upper diagonal, or, $A_1[i, i + 1] = \lambda^s$, and departure rates at the lower one, or, $A_1[i, i - 1] = i\mu^s$. The diagonal entries are simply the negative sum of all other entries at the same row, which are arrival and departure rates of voice and data flows, that is, $A_1[i, i] = -\lambda^s - i\mu^s - \lambda^e - \mu^e$, so as to make the sum of the elements of the row of Q equal to zero. Eventually, $B_1 = A_1 + A_2$ and $B_2 = A_0 + A_1$.

The steady-state equations for streaming flows are given by:

$$\pi(0)B_1 + \pi(1)A_2 = 0 \tag{14}$$

and

$$\pi(i-1)A_0 + \pi(1)A_1 + \pi(i+1)A_2 = 0; i \geq 0$$

The stationary probability of having i data calls in the system is now obtained using the matrix-geometric solution [1]:

$$\pi^e(i) = v_1 S_1^i + v_2 S_2^{N^e - i}; 0 \leq i \leq N^e$$

where S_1 is the solution to:

$$A_0 + S_1 A_1 + S_1^2 A_2 \tag{15}$$

and can be solved recursively as follows:

$$S_1 = (A_0 + S_1 T + S_1^2 A_2) D^{-1}$$

starting from $S_1 = 0$. Matrices T and D are such that $A_1 = T - D$, T having zero diagonal and D a diagonal matrix, positive and invertible.

S_2 is the solution to the dual equation [10]:

$$A_2 + S_2 A_1 + S^2 A_0 \tag{16}$$

Vectors v_1 and v_2 are obtained from the following boundary equation:

$$[v_1 v_2] \begin{bmatrix} B_1 + S_1 A_2 & S_1^{N^e-1}(S_1 B_2 + A_0) \\ S_2^{N^e-1}(S_2 B_2 + A_2) & B_2 + R_2 A_0 \end{bmatrix} = [00]$$

and normalization condition:

$$[v_1 v_2] \begin{bmatrix} \sum_{j=0}^{N^e} S_1^j \\ \sum_{j=0}^{N^e} S_2^j \end{bmatrix} e = 1$$

The mean number of data flows $\bar{N}^e$ as well as their mean transfer time T^e includes their blocking probability B^e given by:

$$B^e = Prob(X^e = N^e_{max}) = \pi^e(N^e_{max})e = v_1 R_1^{N^e} e + v_2 e \tag{17}$$

with:

$$\bar{N^e} = \sum_{j=0}^{N^e} j\pi^e(j)e = \sum_{j=0}^{N^e} j(v_1 R_1^j + v_2 R_2^{N^e-1})e \tag{18}$$

and:

$$T^e = \frac{\bar{N^e}}{\lambda^e(1 - B^e)} \tag{19}$$

5 Performance evaluation

We now evaluate numerically the performance of our system. Let the total capacity of the uplink C_u be equal to 320 Kbps, i.e., 20 voice calls with 16Kbps each, and let the capacity of the downlink C_d be equal to 10 Mbps.

Unless otherwise stated, let us consider that users experience the same radio conditions on average and are at some 200m distance from the base station (Figure 9 shows the achievable throughput in this case). Let the file size be equal to 100 packets of size 1500 bytes each. Let the mean arrival rates for streaming flows in both the uplink and downlink λ_i^s, $i = u, d$, be taken as 1 flow/sec, and let the mean service rate $\mu^s = 0.1$ and the mean arrival rate for elastic flows be equal to 2 flows/sec.

5.1 Impact of uplink

We first vary the capacity given to data flows in the uplink by decreasing the threshold on the maximum number of streaming ones N_u from 20 to 10.

Figure 2 shows the mean transfer time of a TCP file for different values of the uplink rate. We observe that below a certain value, 140Kbps in this case, the uplink is restrictive (and should thus be taken into account into acceptance constraints). Beyond that value, it is not.

The blocking probabilities for both elastic and streaming flows vary accordingly, as shown in Figures 3 and 4, respectively. The former decreases as more room is reserved to data flows and the latter increases as less resources are devoted to streaming flows.

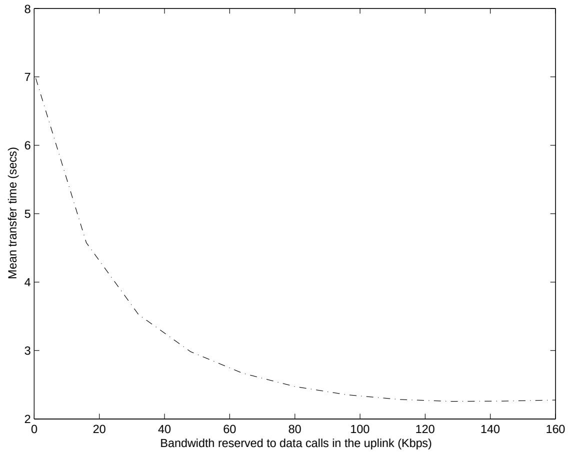

Fig. 2. Mean transfer time of data flows

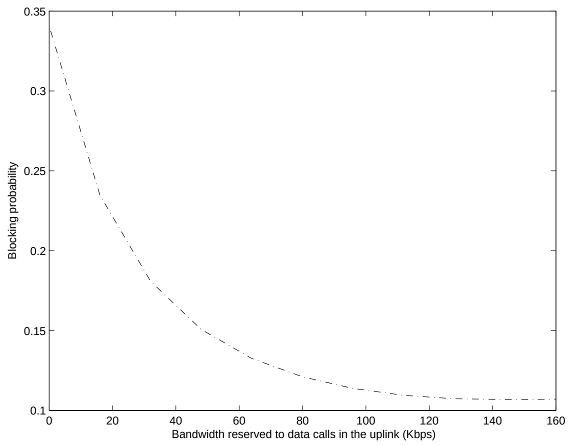

Fig. 3. Blocking of elastic flows

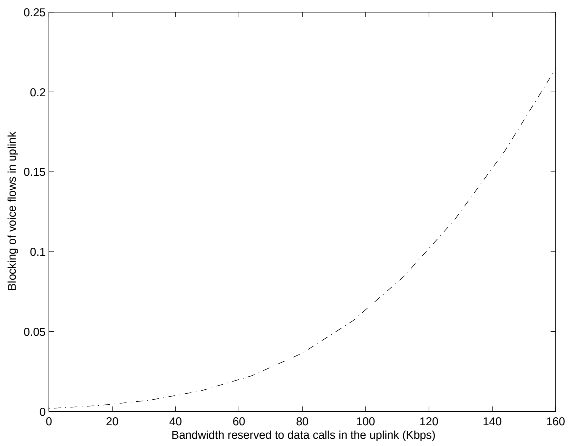

Fig. 4. Blocking of streaming flows in uplink

5.2 Impact of load

We now increase the mean arrival rate of elastic flows from 2 to 4 flows/sec and show in Figures 5 and 6 the mean transfer time and blocking rate respectively.

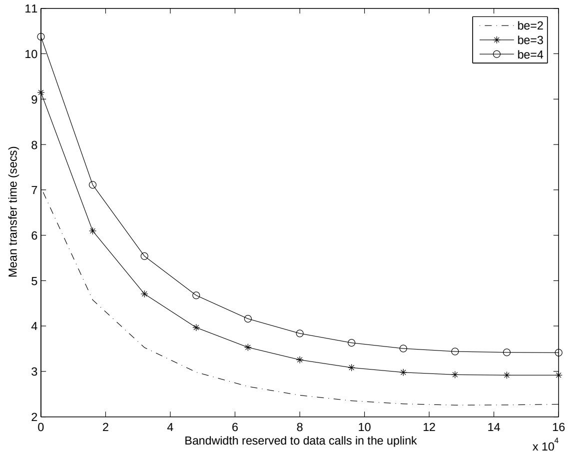

Fig. 5. Mean transfer time of data flows - *be* stands for mean arrival rate of elastic flows in units of flows per second

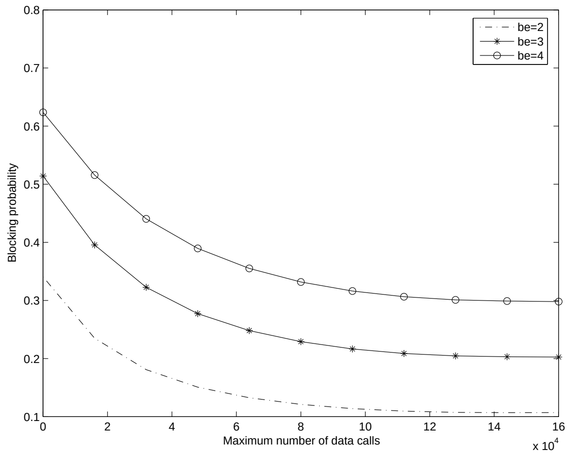

Fig. 6. Blocking of data flows - *be* stands for mean arrival rate of elastic flows in units of flows per second

5.3 Impact of file size

We now change the file size, from 50 to 100 packets, and plot in Figures 7 and 8 the corresponding mean transfer time and blocking probability for different values of reserved bandwidth in the uplink. The mean performance of TCP flows degrades as the file to be transferred gets larger.

5.4 Impact of radio conditions

We now see how the uplink is restrictive in terms of radio conditions. We change the latter from good to bad by increasing the distance from 100 to 200 and 250 m. Indeed, as can be seen from Eqn. (4), the throughput achieved for

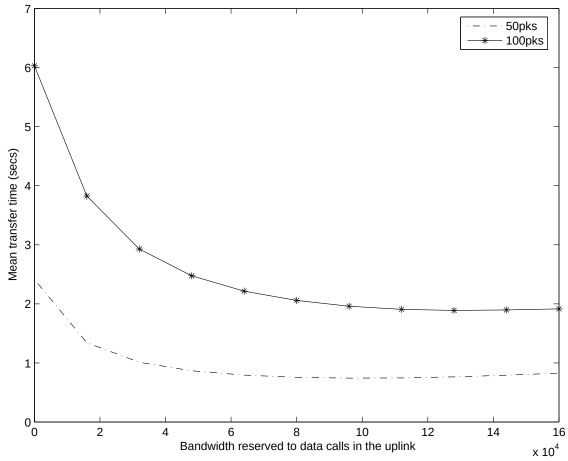

Fig. 7. Mean transfer time of data flows

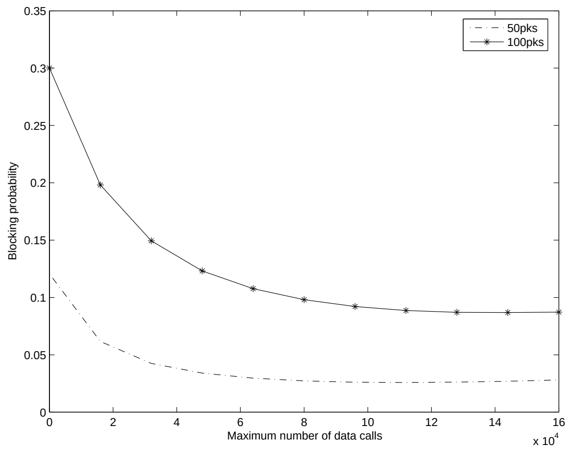

Fig. 8. Blocking of data flows

HSDPA depends on the position in the cell. This is illustrated in Figure 9, where we plot the throughput achieved for a cell that contains only HSDPA users (i.e. emits on the DSCH channel with maximal power), as a function of the position in the cell.

We show in Figures 10 and 11 the corresponding mean transfer time and blocking probabilities for different values of the uplink rate respectively. We observe that when the uplink is most restrictive, all types of users are penalized in the same way. The discrepancy between users changes as the uplink becomes less restrictive, and in this case, the radio conditions play a role too, i.e., users far from the base station have a degraded performance not because of a lesser available capacity but because they cannot take full advantage of it.

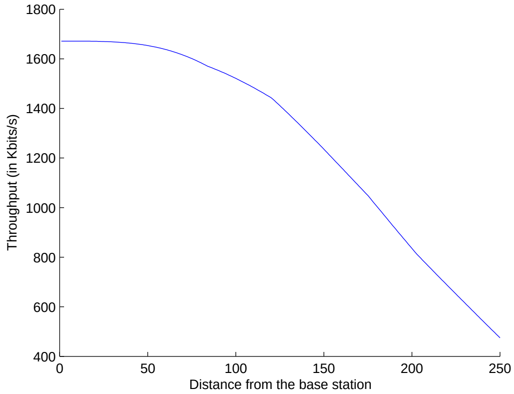

Fig. 9. Achieved throughput as a function of the distance

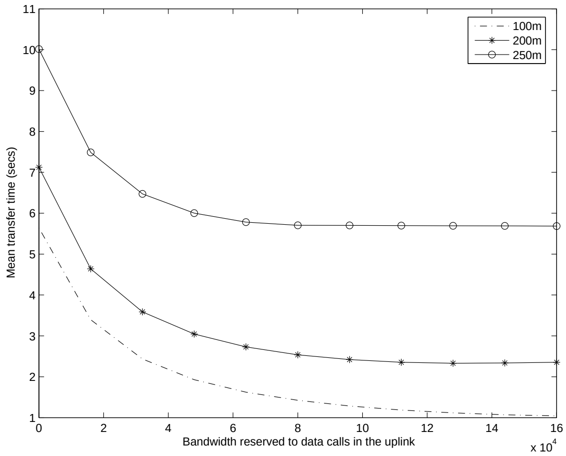

Fig. 10. Mean transfer time of data flows

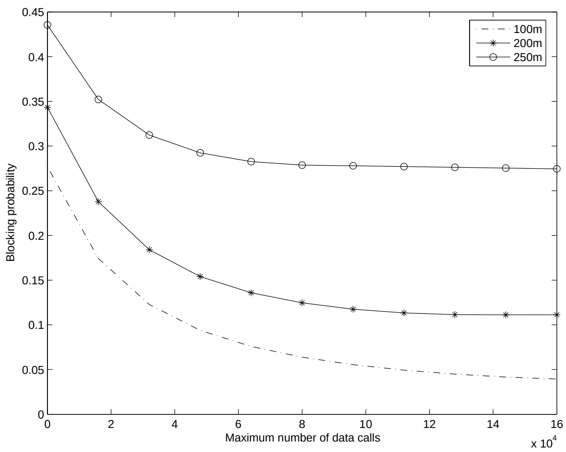

Fig. 11. Blocking of data flows

5.5 Impact of cumulative ACK

We eventually show how the uplink is restrictive in terms of different values of cumulative ACK, namely $b = 1$ and 2. Figures 12 and 13 show the corresponding mean transfer time and blocking probabilities for different values of the uplink and we can observe that a larger value of b limits the problem of the uplink but does not solve it.

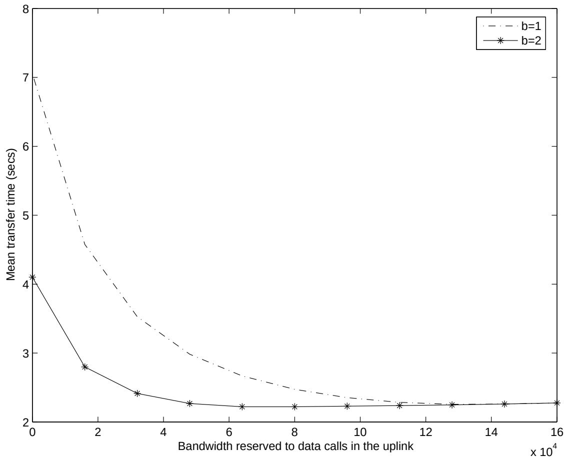

Fig. 12. Mean transfer time of data flows

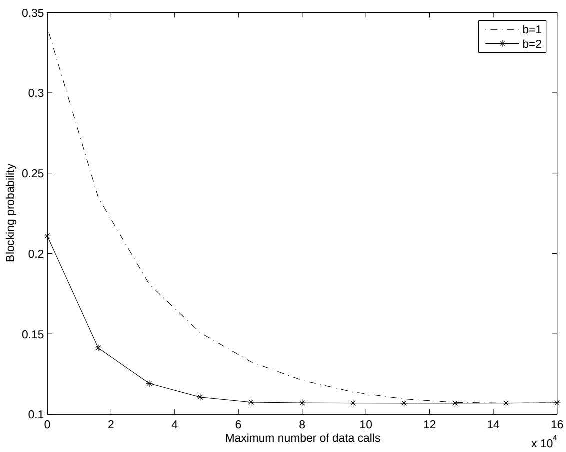

Fig. 13. Blocking of data flows

6 Extending the Erlang capacity region

The previous section illustrated the regions where the uplink is restrictive and how this restriction changes with load, file size and radio conditions. The question we pose now is how to remedy to such a situation. One answer goes as follows. Whenever the uplink is restrictive, we can actually take advantage of the fact that CDMA uses an Adaptive Multi-Rate (AMR) codec for streaming applications and which allows for eight different transmission rates $R^s_{u,i}$, $i = 1, ...8$ [11] ($R^s_{u,j} > R^s_{u,k}$ for $j < k$). In this case, we can squeeze the rate of transmission of each streaming flow in the uplink and make hence more room for ACKs to accompany the higher rate of packets in the downlink and make thus the uplink less restrictive.

The need for squeezing is obtained whenever the capacity left over by streaming flows in the uplink is less than that of the downlink, or:

$$\frac{T^e_d(X^s_d(t))}{X^e(t)} > \frac{T^e_u(X^s_u(t))}{X^e(t)} \frac{s_p b}{s_a} \tag{20}$$

where, again, X^e is the number of concurrent data flows in the system, X^s_u is the number of voice flows in the uplink, X^s_d is the number of data flows in the downlink, b is the number of packets acknowledged by a cumulative ACK, s_p is the packet size, s_a is the ACK size and T^e_u and T^e_u are the throughputs achieved by data calls in the uplink and downlink, respectively.

The difference, when positive, between the two terms for uplink and downlink capacities mentioned in Eqn. (20), hereby denoted by δ, indicates how much capacity is not being used and consequently how much squeezing is needed. This translates evenly among on-going streaming calls which will now have a rate $R^s_u(x)$ equal to $R^s_u - \delta/X^s_u$ rounded up to the nearest value of $R^s_{u,i}$, $i = 2, ...8$.

The overall utility $E[R]$ perceived by streaming flows undergoing both phases, squeezed and not squeezed, is the mean of the individual utility of each phase. And is given by [10]:

$$E[R] = \frac{\sum_{i=1}^{N_u} Pr(X^s_u(t) = x) x R^s_u(x)}{\sum_{i=1}^{N_u} Pr(X^s_u(t) = x) x} \tag{21}$$

The analysis follows the same QBD process as above.

Remark 4. Please note that it is not possible to squeeze streaming flows so as to accommodate other streaming flows. Indeed, squeezing degrades performance and is only acceptable for short periods of times, such as above, to make more room for short-sized ACKs. Long periods of squeezing, which will result in the case of admitting other streaming flows, causes the dropping of the squeezed streaming flows.

7 Concluding remarks

We investigated in this work the joint capacity of the uplink and downlink in CDMA-based system where both directions of flows are asymmetric due to the implementation of HDR/HSDPA in the downlink which makes it capable to offer higher bit rates than the classical uplink. The notable impact of such an asymmetry is on the transport of TCP-based data flows which use ACKs, typically in the uplink, whereas data packets enjoy higher bit rates in the downlink. In this case, the uplink may turn out to be restrictive. This restriction is exacerbated by higher load, larger file size and worst radio conditions.

In this work, we considered that ACKs are carried over shared links in the uplink. The case of dedicated links for ACKs is a little bit different. In this case, the maximum number of elastic flows that can be admitted to the system is limited by the uplink, as the latter imposes a lower limit on the resources given to each flow of ACKs. This is not the case of the downlink where data flows can share the leftover capacity with no lower bound on the resulting individual share. The analysis would follow a Level-Dependent QBD (LDQBD) process and can be carried out using numerical methods similar to the ones used in Reference [10].

In this case, one interesting way to extend the capacity of the system beyond its limits is to suppress ACKs. This is possible if one adopts an open-loop reliable transport, such as the one based on Luby-Transform (LT) codes [15]. This is an issue of future work perspective.

References

[1] N. Akar and K. Sohraby, Finite and Infinite QBD Chains: A Simple and Unifying Algorithmic Approach, IEEE INFOCOM'97, Kobe, April 1997.

[2] G. Alpan, E. Altman, H. Magroun, D. Kofman, Call admission control in the presence of point-to-multipoint best-effort connections, ICC'99.

[3] E. Altman, Capacity of multi-service cellular networks with transmission-rate control: : a queueing analysis, ACM Mobicom'02, Atlanta, September 2002.

[4] E. Altman, T. Chahed, S-E. Elayoubi, Joint uplink and downlink capacity considerations in admission control in multiservice CDMA/HSDPA systems, Valuetools, October 2007.

[5] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, A. Viterbi, CDMA/HDR: A Bandwidth-Efficient High-Speed Wireless Data Service for Nomadic Users, IEEE Communications Magazine, July 2000.

[6] T. Bonald, A. Proutiere, Wireless downlink data channels : User performance and cell dimensionning, ACM Mobicom, 2003.

[7] T. Chahed, M. Dirani, Cross-layer modeling of capacity in UMTS/HSDPA networks under dynamic user setting, VTC'2006-Fall, Montreal, September 2006.

[8] S-E. Elayoubi, T. Chahed, G. Hbuterne, Mobility-aware admission control schemes in the downlink of third generation wireless systems, IEEE Transactions on Vehicular Technology, January 2007.

[9] S-E. Elayoubi, T. Chahed and G. Hbuterne, Connection Admission Control in UMTS in the Presence of Shared Channels, Computer Communications, volume 27, issue 11, June 2004.

[10] N. Hegde, E. Altman, Capacity of Multiservice WCDMA Networks with Variable GoS, Wireless Networks, 12(2), April 2006.

[11] H. Holma, A. Toskala, WCDMA for UMTS, John Wiley and Sons Ltd., 2001.

[12] W. Jeon and D. Jeong, Call Admission Control for Mobile Multimedia Communications with Traffic Asymmetry between Uplink and Downlink, IEEE Transactions on Vehicular Technology, Volume : 50 Issue: 1 , January 2001.

[13] Kiran and D. Tse, Effective Bandwidths and Effective Interference for Linear Multiuser Receivers in Asynchronous Systems, IEEE Transactions on Information Theory, vol 46(4), July 2000, pp. 1426-1447.

[14] T. V. Lakshman, U. Madhow, B. Suter, TCP/IP performance with random loss and bidirectional congestion. IEEE/ACM Trans. Netw. 8(5): 541-555, 2000.

[15] M. Luby, LT codes, IEEE-FOCS, pp. 271-282, 2002.

[16] M. F. Neuts, Matrix-Geometric Solutions in Stochastic Models, an Algorithmic Approach, Dover publications, 1994.

[17] K. Sipila, Z. Honkasalo, J. Laiho-Steffens and A. Wackr, Estimation of capacity and required transmission power of WCDMA downlink based on a downlink pole equation, IEEE VTC 2000.