

On design of TDD for joint uplink and downlink resource allocation in OFDMA-based WiMax

Tijani Chahed
GET/INT - UMR CNRS 5157
9 rue C. Fourier
91011 Evry CEDEX - France
tijani.chahed@int-edu.eu

Salah-Eddine Elayoubi
Orange Labs
38 rue du Général Leclerc
92794 Issy les Moulineaux - France
salaheddine.elayoubi@orange-ftgroup.com

Eitan Altman
INRIA-Sophia Antipolis
2004 route des Lucioles - BP 93
06902 Sophia Antipolis - France
altman@sophia.inria.fr

Abstract— In this paper, we study the joint design of uplink and downlink resources in OFDMA-based systems. We first analyze the interactions between uplink and downlink, due essentially to the usage of Time Division Duplexing (TDD) and the TCP ACK/NACKs. This analysis is based on analytical models for interference and throughputs that we develop for both directions. We next study the capacity of the system using a Markovian analysis and matrix geometric solutions. Our performance analysis study allows us to determine the optimal proportion of resources that has to be allocated to the uplink; In addition to that, we show that an admission control policy reserving some uplink resources to TCP ACKs may improve the overall capacity of the system.

I. INTRODUCTION

The IEEE 802.16e standard [1] is promoted by the World-wide Interoperability for Microwave Access organization (WiMAX) as the system offering broadband wireless access for mobile users. Indeed, cellular WiMAX networks based on Orthogonal Frequency Division Multiple Access (OFDMA) are able to provide high data rates for non-line-of-sight applications. There is thus an increasing interest in designing efficient Radio Resource Management (RRM) schemes for these networks.

In this work, we show how to design RRM schemes that jointly optimize usage of uplink (UL) and downlink (DL) resources. We first develop analytical models for interference and throughputs in WiMAX, taking into account the uplink and downlink specificities, such as uplink power control and adaptive modulation and coding. We then 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 one direction and ACKs returning back in the other direction. The main difference between those types of flows is that streaming flows require some constant bit rate but their service duration is independent of the quantity of resources they get. This is not the case of data flows which first have the ability to share resources in a fair manner among themselves and second would leave the system sooner if they get more resources. Another source of interactions between both links is caused by the TDD nature of the transmission as the same frequency band is used for uplink and downlink transmissions. In order to analyze the system under the different RRM policies and taking into account the UL/DL interactions, we

build the underlying Markov process and show how, using matrix geometric solutions [6], we can obtain the steady-state probabilities and the performance measures.

Our design is performed within two steps. First, we determine, for a given mix of traffic, the optimal TDD factor, i.e. the proportion of resources that we have to give to UL in order to minimize blocking. The second step is to design, knowing this TDD factor, efficient RRM schemes that insure that there neither of the two links is not a bottleneck for data transmissions. We thus show how, by reservation some of the UL resources to ACK, we are able to enhance the performance of elastic flows without degrading the overall performance.

The remainder of this extended abstract is organized as follows. In Section II, we describe the results given by our model for uplink and downlink capacity of OFDMA-based WiMax. In Section III, we detail the interactions between UL and DL caused by TDD and TCP ACKs. 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 V, we show some performance evaluation results. Section VI eventually concludes the paper.

II. SYSTEM

We consider a WiMAX homogeneous cellular system with cells numbered from 0 to infinity and focus on the performance of cell 0. The frequency band of N subcarriers is partitioned into C subchannels, each containing $M = N/C$ subcarriers. We consider K interfering cells. If we note by \mathbf{Y} a vector or zeros and ones whose dimension is equal to the number of interfering cells and whose elements correspond each to an interfering cell, with the value 1 meaning that collision occurs with the corresponding cell, it is easy to show that:

$$Pr(\mathbf{Y}) = \chi^{\mathbf{Y} \cdot \mathbf{Y}} (1 - \chi)^{\mathbf{Y} \cdot \mathbf{Y}} \quad (1)$$

$\mathbf{X} \cdot \mathbf{X}$ being the scalar product of \mathbf{Y} with itself, and χ the load of a typical interfering cell, defined by the proportion of occupied subchannels in it.

A. Downlink

To evaluate the performance of the system, we must first characterize the arrival and departure rates. Let D be the

instantaneous throughput of a subchannel. This throughput depends, in addition to the offered bandwidth by subcarrier W , on the efficiency of the used modulation and the Block Error Rate (BLER). This relationship is given by:

$$D = M \times W \times e \times (1 - BLER) \quad (2)$$

where e is the efficiency of the used modulation (e.g. e is equal to 1 bit/symbol for QPSK 1/2 and to 5 bits/symbol for 64 QAM 5/6). The BLER depends on the physical layer characteristics (used modulation and path loss) and on the amount of interference.

In WiMAX systems, Adaptive Modulation and Coding (AMC) will be used. The choice of the modulation depends on the value of Signal to Interference plus Noise Ratio (SINR, also called C/I) through the perceived BLER: The most efficient modulation that achieves a BLER larger than 10^{-1} is used. For each SINR value, this leads to a couple of values ($e, BLER$), determined by link level curves $e(C/I)$ and $BLER(C/I)$, available in the literature. This gives:

$$e \times (1 - BLER) = e(C/I) \times (1 - BLER(C/I))$$

We then define the resulting function $B(C/I) = e(C/I) \times (1 - BLER(C/I))$ and use it in the remainder of the paper for throughput calculations:

$$D(C/I) = M \times W \times B(C/I) \quad (3)$$

In the downlink, a base station emits, for each subchannel, a constant power P . The SINR in cell 0 for a call situated at distance r_0 from the base station is:

$$SINR^D(\mathbf{Y}, r_0) = \frac{P/q_0^D}{\sum_{i=1}^n X_i \frac{P}{q_i^D} + N_0} \quad (4)$$

where \mathbf{Y} is the vector of collisions, N_0 is the background noise and q_i^D is the path loss between interfering base station i and the corresponding receiver. $q_i^D = r_i^\alpha 10^{\frac{\xi_i}{10}}$, with r_i the distance from base station i to the receiver, ξ_i a normal random variable due to shadowing and $\alpha \in [2, 4]$ is a constant. The corresponding throughput of the subchannel is then given by:

$$D^D(\mathbf{Y}, r_0) = MW B(SINR^D(\mathbf{Y}, r_0)) \quad (5)$$

The subchannel being subject to different interfering configurations, and possibly allocated to users in several position of the cell, its average throughput is given by:

$$\bar{D}^D = E_{r_0} \left[\sum_{\mathbf{Y}} D^D(\mathbf{Y}, r_0) Pr(\mathbf{Y}) \right] \quad (6)$$

The expectation $E_{r_0}[\cdot]$ means that the value of the integral is averaged over the surface of cell 0.

B. Uplink

In the uplink of WiMAX, power control will be used in order to optimize the battery consumption. The power control algorithm being not yet standardized, we propose the following algorithm: As Adaptive Modulation and Coding (AMC) is used, the best achievable modulation and coding scheme is

chosen. The lowest transmitted power P_e is thus calculated by the following algorithm:

- 1) Calculate the SINR that can be achieved using the maximal transmission power P_{max} .
- 2) Obtain the modulation and coding scheme that corresponds to this SINR.
- 3) Calculate the transmission power P_e as the lowest power that achieves the above obtained modulation and coding scheme.

Taking into account this power control, we calculate the SINR, the average emitted power and the SINR. Note that this calculation is more complicated in the uplink than in the downlink as the interference depends on the position of all interfering mobiles in adjacent cells. We thus use an integration of interfering cell surfaces. Because of the lack of space, we do not give the details of the calculations that give the average UL rate \bar{D}^U by an iterative calculation.

III. UPLINK/DOWNLINK INTERACTIONS

The obtained models for the throughputs can be used in order to dimension each link independently. However, this task is not so easy as there is a strong interaction between UL and DL. These interactions are due to two properties. The first is related to the nature of WiMAX, where UL and DL connections share the same bandwidth, and the second is related to the nature of the elastic traffic that generates TCP ACK/NACKs in the opposite direction.

A. TDD

In WiMAX, TDD (Time Division Duplex) is proposed as a duplexing scheme that requires only one channel for transmitting downlink and uplink sub-frames at two distinct time slots. TDD therefore has higher spectral efficiency than FDD. TDD can flexibly handle both symmetric and asymmetric broadband traffic. Moreover, the TDD downlink to uplink ratio can be adjusted dynamically depending on the load of both directions. This dynamic adjustment will be studied later.

B. Interactions due to TCP ACKs

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 . The term capacity refers to the net capacity, in number of subchannels.

Let the arrival of streaming flows be Poisson with mean arrival rate λ_s^U in the uplink and λ_s^D 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 assumed to use some rate R_s , for instance 64Kbps. Naturally, 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_s^U and in the downlink N_s^D ; $N_s^U \leq C^U$ and $N_s^D \leq C^D$. 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 elastic flows in the downlink. We assume that they are also subject to admission control and that the maximum number of data flows is N_e ; $N_e^D \leq C^D$.

Once in the system, we assume that a data packet flow generates instantaneously a corresponding stream of ACKs in the opposite direction. Now this overall data flow, packets and ACKs, shall take the minimum capacity between the bandwidth left over by streaming flows to process data packets in one direction and the bandwidth left over by streaming flows in the opposite direction to process smaller size ACKs.

We adopt a scheduling 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. Each data flow obtains on average in the downlink, some R_e throughput given by:

$$R_e = \min\left(\frac{T_e^D(X_s^D(t))}{X_e^D(t)}, \frac{T_e^U(X_s^U(t))}{X_e^U(t)} \frac{s_p b}{s_a}\right) \quad (7)$$

where X_e is the number of concurrent data flows in the downlink, X_s^U (resp. X_s^D) is the number of streaming flows in the uplink (resp. 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_e^U and T_e^D are the throughputs achieved by elastic calls in the uplink and downlink, respectively, calculated from the previous section:

$$T_e^U(X_s^U) = (C^U - X_s^U) \bar{D}^U$$

$$T_e^D(X_s^D) = (C^D - X_s^D) \bar{D}^D$$

Remark 1. 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 $R_e(\mathbf{X})$ where \mathbf{X} is a vector denoting the number of streaming and elastic flows. We however drop \mathbf{X} for the sake of notational convenience.

We assume that data flows are also subject to admission control. Let N_e be the maximum number of elastic flows.

IV. ANALYSIS

Now, our model can be solved as follows [5]. 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 $\pi(\cdot)$ are given by the Erlang formula as:

$$\pi(X_s^i = x) = \frac{1}{\sum_{k=0}^{N_s^i} \frac{(\rho_s^i)^k}{k!}} \frac{(\rho_s^i)^x}{x!} \quad (8)$$

where $\rho_s^i = \lambda_s^i / \mu_s$.

The blocking probability B_s^i , $i = U, D$, of streaming flows is given by:

$$B_s^i = \frac{1}{\sum_{k=0}^{N_s^i} \frac{(\rho_s^i)^k}{k!}} \frac{(\rho_s^i)^{N_s^i}}{N_s^i!} \quad (9)$$

Now, define the composed variable $X_s^{UD} = X_s^U + (1 + N_s^U)X_s^D$, and consider the process: $\mathbf{X} = (X_s^{UD}, X_e)$. The process $\mathbf{X}(t)$ is a finite homogeneous Quasi-Birth and Death (QBD) process with infinitesimal generator Q , of size $(N_s^U + 1)(N_s^D + 1)(N_e + 1)$, given by:

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

where A_0 , A_1 and A_2 are square matrices of size $(N_s^U + 1)(N_s^D + 1)$ which we denote by N (Recall that N_s^U and N_s^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 at the diagonal, A_2 represents their departures, with mean departure rates, given by the achieved rates and divided by the mean file size, at the diagonal too. A_1 corresponds to the arrival and departure of voice flows. It is tri-diagonal, with mean arrival rates at the upper diagonal and departure rates at the lower one. 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, so as to make the sum of the elements of the row of Q equal to zero. $B_1 = A_1 + A_2$ and $B_2 = A_0 + A_1$.

The stationary probability of having $X_e = i$ is now obtained using the modified matrix-geometric solution [7]:

$$\pi_e(i) = v_1 S_1^i + v_2 S_2^{N_e - i}; 0 \leq i \leq N_e$$

where S_1 is the solution to:

$$A_0 + S_1 A_1 + S_1^2 A_2 \quad (10)$$

and S_2 is the solution to the dual equation:

$$A_2 + S_2 A_1 + S_2^2 A_0 \quad (11)$$

Vectors v_1 and v_2 are obtained from the following boundary:

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

and normalization condition:

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

where $N_e = (N_e^U + 1)(N_e^D + 1)$.

Solving for S_i , $i = 1, 2$, can be done recursively as follows:

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

starting from $S = 0$.

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

The blocking probabilities for both uplink and downlink for elastic flows is given by:

$$B_e = \pi_e(N_e) \quad (12)$$

The mean number of data flows is given by:

$$\bar{X}_e = \sum_{j=0}^{N_e} j \pi_e(j) \quad (13)$$

and eventually the mean transfer time is:

$$T_e = \frac{\bar{X}_e}{\lambda_e(1 - B_e)} \quad (14)$$

V. PERFORMANCE EVALUATION

In this section, we consider a WiMAX system subject to the following mix of traffic:

- 1) Streaming traffic in both uplink and downlink with arrival rates $\lambda_s^U = \lambda_s^D = 0.01$ calls per second, each call having an average duration of 100 seconds.
- 2) Elastic traffic in the downlink with an arrival rate $\lambda_s^U = \lambda_s^D = 0.01$ calls per second. Each connection generates a flow of ACKs that shares the capacity with streaming calls in the uplink.

A. Optimizing the TDD UL/DL factor

We first consider a system where the proportion of resources given to UL, i.e. the ratio $\frac{C_u}{C_u + C_d}$, is variable. We plot in Figure 1 the blocking rates for elastic, UL streaming and DL streaming calls, in addition to the overall blocking rate. We can see that, even if the minimal blocking for UL and DL streaming is attained when the resources are equally shared (as the streaming traffic is equal in both directions), the overall blocking is minimal for another capacity share (30% of resources allocated to uplink) due to the presence of the best effort traffic.

On the other hand, Figure 2 plots the average file transfer time function of the UL/DL ratio. We can see that this download time has also a minimal value. In fact, for a small downlink capacity, the throughput is limited by the lack of resources, while for small uplink capacity, the ACKs cannot be sent leading to a slowed transmission.

B. Capacity reservation

In this section, we consider the optimal UL/DL ratio found in the previous section and try to find a way to further enhance the performance. To do so, we adopt a capacity reservation scheme where a part of the uplink capacity is reserved to TCP ACKs. We plot in Figure 3 the average download time function of the proportion of uplink resources reserved to ACKs. We can see that, when this reserved capacity increases, the download time decreases dramatically, before being stabilized. However, this capacity reservation has a drawback, illustrated in Figure 4, where the blocking rate of uplink streaming calls is shown to increase. A good compromise is thus by reserving a small percentage of the capacity to ACKs (e.g. 25%). This will result in a large decrease of download times, with an almost constant overall blocking probability.

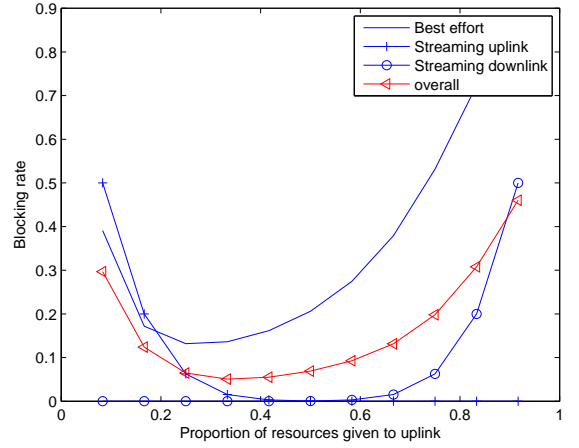


Fig. 1. Blocking rates function of the TDD UL/DL factor.

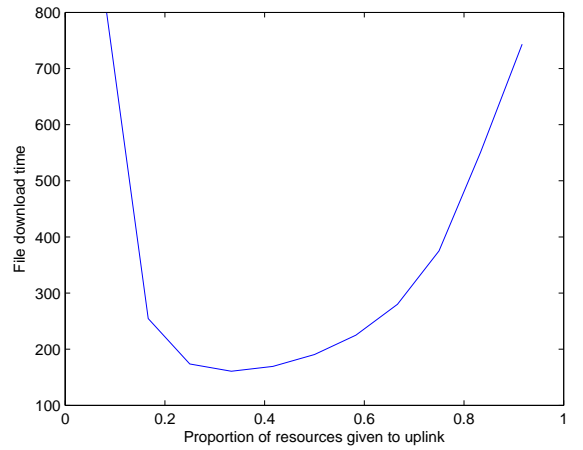


Fig. 2. Average download time function of the TDD UL/DL factor.

VI. CONCLUDING REMARKS

In this paper, we studied the global capacity of a WiMAX system taking into account interactions between uplink and downlink. These interactions may result from the TDD nature of the transmission (both directions share the same bandwidth), or else from the TCP ACKs generated by downlink elastic flows on the opposite side. We developed models for throughputs and capacity in both directions of a WiMAX system, taking into account the inter-cell interference and the power control. We then analyzed the system, using these models and matrix geometric solutions to find the steady state probabilities of the underlying Markov chain. This analysis has led to a performance evaluation study where we showed how, for a given traffic mix, how to optimize the proportion of capacity given to uplink and, within these uplink resources, the subchannels reserved to TCP ACKs in order to insure good download times.

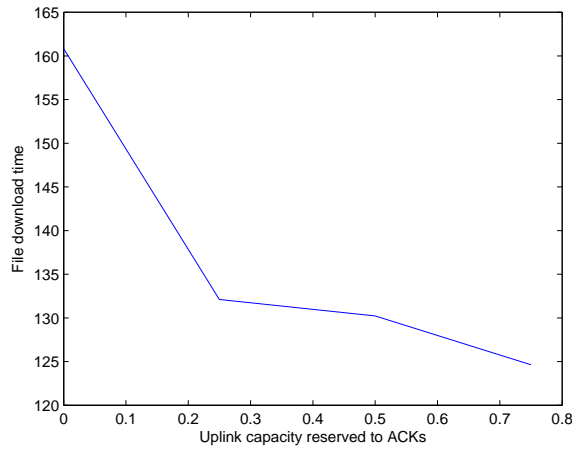


Fig. 3. Average download time function of the capacity reserved to ACKs in the uplink.

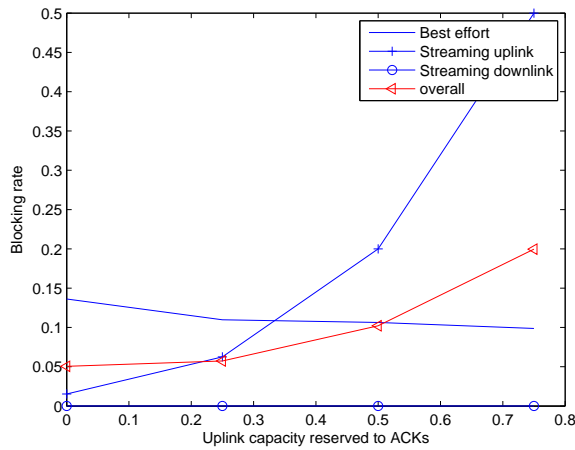


Fig. 4. Blocking rates function of the capacity reserved to ACKs in the uplink.

REFERENCES

- [1] Draft 802.16e/D9, *Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE Standard for Local and Metropolitan Area Networks", June 2005.
- [2] 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.
- [3] 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.
- [4] N. Hegde, E. Altman, Capacity of Multiservice WCDMA Networks with Variable GoS, *Wireless Networks*, 12(2), April 2006.
- [5] G. Alpan, E. Altman, H. Magrour, D. Kofman, Call admission control in the presence of point-to-multipoint best-effort connections, *ICC'99*.
- [6] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models*, an Algorithmic Approach, Dover publications, 1994.
- [7] N. Akar and K. Sohraby, Finite and Infinite QBD Chains: A Simple and Unifying Algorithmic Approach, *IEEE INFOCOM'97*, Kobe, April 1997.