

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/4302710>

# Constrained Stochastic Games in Wireless Networks

Conference Paper · December 2007

DOI: 10.1109/GLOCOM.2007.66 · Source: IEEE Xplore

CITATIONS

25

READS

60

6 authors, including:



**K. E. Avrachenkov**

National Institute for Research in Computer Science and Control

316 PUBLICATIONS 4,449 CITATIONS

[SEE PROFILE](#)



**mérrouane Debbah**

École Supérieure d'Electricité

822 PUBLICATIONS 27,674 CITATIONS

[SEE PROFILE](#)



**Rachid El-Azouzi**

Université d'Avignon et des Pays du Vaucluse

255 PUBLICATIONS 3,421 CITATIONS

[SEE PROFILE](#)



**Daniel S. Menasché**

Federal University of Rio de Janeiro

115 PUBLICATIONS 996 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Stochastic Geometry Modelling and Analysis of Wireless Edge Caching [View project](#)



Intelligence Surface [View project](#)

# Constrained Stochastic Games in Wireless Networks

Eitan Altman<sup>\*</sup>, Konstantin Avrachenkov<sup>\*</sup>, Nicolas Bonneau<sup>\*</sup>,  
Mérrouane Debbah<sup>†</sup>, Rachid El-Azouzi<sup>‡</sup>, Daniel Menasché<sup>\*</sup>

<sup>\*</sup> INRIA, Centre Sophia-Antipolis, 2004 Route des Lucioles, B.P. 93, 06902 Sophia-Antipolis Cedex, France

<sup>†</sup> Mobile Communications Group, Institut Eurécom, 2229, Route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, France

<sup>‡</sup> Univesite d'Avignon, 339, Chemin des Meinajaries, Agroparc BP 1228, 84911 Avignon Cedex 9, France

**Abstract**—We consider the situation where  $N$  nodes share a common access point. With each node  $i$  there is an associated buffer and channel state that change in time. Node  $i$  dynamically chooses both the power and the admission control to be adopted so as to maximize the expected throughput, which depends on the actions and states of all the players, given its power and delay constraints. The information structure that we consider is such that each player knows the state of its own buffer and channel and its own actions. It does not know the states of, and the actions taken by other players. Using Markov Decision Processes we analyze the single player optimal policies under different model parameters. In the context of the stochastic games we study the equilibria of the  $N$  player scenario.

## I. INTRODUCTION

We consider a set of mobiles who transmit packets through a wireless network and who share a common access point (AP). Each mobile must dynamically choose its power and packet admission control based on the information that it has about its queue size (number of backlogged packets) and on its current channel state.

Although the large majority of wireless networks operating today do not delegate these decisions to the users, the possibility of leaving the decisions to the mobiles is an important alternative worth studying since it scales much better with increasing traffic demand, throughputs and the number of subscribers.

In the absence of queues, in the case that each mobile always has a packet to transmit (this is the saturated regime), the problem can be described as a game with a random environment. We call the random environment the state of the system. The state is given by a set of Markov chains, one for each user. The users control the power (and hence the instantaneous costs or rewards) but have no control over the system state. This situation was studied in Altman et al. [1], [2]. In contrast, in this paper the control of a user has an impact not only on the instantaneous costs or rewards but also on the state transition probabilities, as the state of player  $i$  now includes also the (controlled) queue size (while [1], [2] assumes a saturated system).

Adding the queues allows us to reduce losses at the cost of introducing queueing delays within the network. In formulating the optimal control problem we shall take these delays into account in the optimization objectives in the form of additional constraints.

The problem of maximizing the throughput of some traffic subject to constraints on its delay has received extensive attention, starting with the pioneering work [3] by Lazar with its extension to a game setting in [4]. The research on power control in cellular networks is also rich [5]. Our work is the first to consider both admission control and discrete power control to maximize the throughput given certain power and delay constraints. We study this situation for the case of a single mobile, as well as for the case of several players. When there are several players, we consider both the non-cooperative and the cooperative (team problem) formulation, using game theory. The team problem occurs when there is more than one mobile but a common objective function is optimized by all mobiles. On the contrary, in the non-cooperative game, the objective of each player is to selfishly maximize its own payoff. We consider both the Nash equilibrium concept as well as that of correlated equilibrium.

The notion of correlated equilibrium was introduced by R. Aumann in [6] as a generalization of Nash equilibrium. The important feature of correlated games is the presence of an *arbitrator*. An arbitrator needs not have any intelligence or knowledge of the game, it needs only send random (private or public) signals to the players that are independent of all other data in the game. In the context of non-cooperative games, each player has the possibility not to consider the signal(s) it receives. Coordination between players turns out to be useful also in the case of cooperative optimization. The signals enable joint randomization between the strategies of the players, possibly resulting in equilibria with higher payoffs. The concept of correlated games was recently introduced in a networking context in [7], where the authors consider a simple ALOHA setting. To our knowledge, our paper is the first to consider correlated equilibria of stochastic games with constraints.

This paper is structured as follows. In Sec. II we present the proposed model. In Sec. III we analyze using Markov Decision Processes the optimal policies when a single controller is involved. In Sec. IV we provide results about the equilibria in the scenario with multiple users with completely independent state information. Finally, in Sec. V we assume that there is a common device that emits signals to all the users and we present the correlated equilibria that emerges from this system.

The scripts used to generate the pictures presented in this paper and the solution techniques adopted [Lemke linear complementarity program (LCP) and successive best response]

as well as an extended version of this work are available online at [8].

## II. THE MODEL

We consider the situation where  $N$  nodes share a common AP. Time is divided into slots. We assume that packets are generated at each node by application(s). At each time slot  $t$ , each mobile  $i$  transmits data with power level  $p_i^t$  chosen among a finite set  $P = \{0, 1, \dots, P_{\max}\}$  containing  $P_{\max} + 1$  power levels. The action  $p_i = 0$  means back-off and the action  $p_i \geq 1$  means transmission with power level  $p_i$ .

We propose a rate control mechanism where we assume that node  $i$  has elastic traffic and uses an admission control mechanism to select the input rates to the buffer with finite capacity  $B_i$ . A node receiving a new packet can decide to drop or accept it. The admission control actions available to each mobile are in the set  $C = \{0, 1\}$ , where 1 (resp. 0) means that mobile  $i$  accepts (resp. drops) a new packet from the upper layer.

We assume that the channel between mobile  $i$  and the base station can be modeled as an ergodic finite Markov Chain taking values in a set  $E_i = \{1, 2, \dots, L_i\}$  of  $L_i$  states with transition probabilities  $\lambda_{xy}^i$  (i.e., the entry  $(x, y)$  of the transition matrix  $\Lambda^i$  is equal to  $\lambda_{xy}^i$ ,  $\sum_{y=1}^{L_i} \lambda_{xy}^i = 1$ ). We shall denote by  $\pi^i = [\pi_1^i, \pi_2^i, \dots, \pi_{L_i}^i]$  the steady state probability distribution of  $\Lambda^i$ . We assume throughout that  $\Lambda^i$  is irreducible and aperiodic for all  $i = 1, 2, \dots, N$ .

At each node packets arrive from the upper layer according to the arrival process  $\{\gamma^t, t = 1, 2, \dots\}$  (an independent and identically distributed process).

Knowing the number of packets in its buffer and the channel state between itself and the access point mobile  $i$  must choose (a)  $p_i \in P$ , i.e., the transmission power level and (b)  $c_i \in C$ , i.e., to accept or reject new packets which may arrive from the upper layer. Only local information is available to each mobile and there is no coordination in their actions (except when considering correlated equilibria). Let  $a_i^t = (p_i^t, c_i^t)$  where  $p_i^t \in P$  and  $c_i^t \in C$ .

Let  $q_i^t$  be the queue size of node  $i$  at the beginning of slot  $t$ . In a given time slot we assume that all the arrivals from the upper layer occur after transmission of packets to the network. Therefore, the dynamics of the buffer size at node  $i$  are given by

$$q_i^{t+1} = \min([q_i^t + c_i^t \gamma_i^t - w_i^t]^+, B_i) \quad (1)$$

where  $\gamma_i^t$  is the realization of a random variable  $\Gamma_i^t$  indicating the number of arrivals in slot  $t$  and  $w_i^t$  is the realization of a r.v.  $W_i^t$  indicating the number of departures in slot  $t$ . In general we adopt the convention of using capital letters for r.v.s and small letters for their realization.

We assume that the mobiles will never retransmit packets. In each time slot a mobile can transmit at most one packet ( $w_i^t \in \{0, 1\}$ ). When  $q_i^t > 0$ ,  $w_i^t$  is equal to 0 (resp. 1) if  $p_i^t = 0$  (resp.  $p_i^t > 0$ ). When  $q_i^t = 0$ ,  $w_i^t = 0$ .

The state of mobile  $i$  at time  $t$  is described by  $x_i^t = (y_i^t, q_i^t)$ , where  $y_i^t$  is the state of the channel in slot  $t$  and  $q_i^t$  is the

number of packets in the queue at the beginning of slot  $t$ . If the state of mobile  $i$  at time  $t$  is  $x_i$  and action  $a_i$  is adopted the next state will be  $z_i$  with probability  $P_{x_i a_i z_i}$ .

**Mobile strategies**  $u_i((p_i, c_i)|x_i)$  is the probability that mobile  $i$  chooses the transmission power level  $p_i \in P$  and the action  $c_i \in C$  when its state is  $x_i$ . The class of decentralized policies of mobile  $i$  is denoted by  $U_i$ . Define  $U = \prod_{i=1}^N U_i$ .

**Performance measures** The signal to interference ratio  $SIR_i$  of mobile  $i$  at the base station when the power level choices of the mobiles are  $\mathbf{p} = (p_1, \dots, p_N)$  and the mobile states are  $\mathbf{x} = (x_1, \dots, x_N)$  is given by

$$\begin{aligned} SIR_i((\mathbf{y}, \mathbf{q}), \mathbf{p}) &= \frac{g_i(y_i)p_i}{N_0 + \sum_{j \neq i, q_j > 0} c_{ij} g_j(y_j)p_j} \text{ if } q_i > 0 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

where  $c_{ij}$  are the coding orthogonality coefficients and  $N_0$  the thermal noise in the medium. We consider the following instantaneous utility of mobile  $i$ :

$$r_i(\mathbf{x}, (\mathbf{p}, \mathbf{c})) = r_i(\mathbf{x}, \mathbf{a}) = \log_2(1 + SIR_i(\mathbf{x}, \mathbf{p})) \quad (2)$$

Notice that we optimize an upper bound on the throughput. In the sequel we refer to it simply as capacity.

For a given policy  $u_i$  and a given initial state  $x_i^{t=0} = \beta_i$  we now define the constraints of each mobile.

**Power constraints:** Let mobile  $i$ 's power level at time  $t$  be  $p_i^t \in P$ . The exact power of the signal of mobile  $i$  is  $P_i = P_0 p_i$  where  $P_0$  is some base value of the power (in this paper,  $P_0 = 1$ ). For each mobile  $i$  let its expected average power consumption be constrained by the following bound:

$$\text{Power}_{\beta_i, u_i}^i = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E_{\beta_i}^{u_i} d_i(X_i^t, A_i^t) \leq P_{\text{const}}^i \quad (3)$$

where  $d_i(x_i, a_i) = P_0 p_i$  if  $q_i > 0$  and 0 otherwise. Note that the power constraints of a mobile don't depend on the decisions of the other mobiles.

**Expected buffer length constraints:** Some mobiles (applications) might have stringent delay constraints. Let the expected average buffer length be constrained by the following bound:

$$\text{Buf}_{\beta_i, u_i}^i = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E_{\beta_i}^{u_i} Q_i^t \leq Q_{\text{const}}^i \quad (4)$$

**Problem statement:** The objective of mobile  $i$  is to maximize the mean throughput. Given  $\mathbf{x}^{t=0} = \beta$ , the controller's problem is thus to obtain the optimal strategy  $u_i$  that maximizes (5) subject to (3) and (4).

$$\text{Thp}_{\beta, \mathbf{u}}^i = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} E_{\beta}^{\mathbf{u}} r_i(\mathbf{X}^t, \mathbf{A}^t) \quad (5)$$

## III. THE SINGLE PLAYER CASE

In this section we analyze the single player case in the context of the Markov Decision Processes (MDP). The notation is the same as the one used in Table I although we don't write the indexes  $i$  for notational convenience.

TABLE I  
TABLE OF NOTATION FOR PLAYER  $i$

$q_i^t$	size of the queue at the beginning of slot $t$
$y_i^t$	channel state during slot $t$
$w_i^t$	number of departures at time slot $t$
$\gamma_i^t$	number of arrivals at time slot $t$
$x_i^t = (y_i^t, q_i^t)$	state vector (channel state, buffer size)
$a_i^t = (p_i^t, c_i^t)$	action vector (power control, admission control)
$g_i(y_i)$	attenuation at channel state $y_i$
$d_i(p_i) = p_i$	power consumption related to action $p_i$
$d_i(x_i, a_i)$	$d_i(p_i)$ if $q_i > 0$ (0 otherwise)
$r_i^t(\mathbf{x}^t, \mathbf{a}^t)$	instantaneous throughput in slot $t$

### A. The Linear Program

The single player problem can be solved using a linear program [9]. Let  $\mathcal{K} = \{(x, a) : x \in \mathbf{X}, a \in \mathbf{A}\}$ .

**LP1 :**

Find  $z^* := \{z^*(x, a)\}_{x,a}$ , where  $(x, a) \in \mathcal{K}$ , that minimizes

$$\sum_{x \in \mathbf{X}} \sum_{a \in \mathbf{A}} r(x, a) z(x, a) \quad \text{subject to:} \quad (6)$$

$$\sum_{x \in \mathbf{X}} \sum_{a \in \mathbf{A}} z(x, a) [\delta_r(x) - \mathcal{P}_{yar}] = 0, \quad \forall r \in \mathbf{X}, \quad (7)$$

$$\sum_{x=(y,q) \in \mathbf{X}} \sum_{a \in \mathbf{A}} qz(x, a) \leq Q_{\text{const}} \quad (8)$$

$$\sum_{x \in \mathbf{X}} \sum_{a \in \mathbf{A}} d(x, a) z(x, a) \leq P_{\text{const}} \quad (9)$$

$$z(x, a) \geq 0, \quad \forall (x, a) \in \mathcal{K} \quad \sum_{(x,a) \in \mathcal{K}} z(x, a) = 1 \quad (10)$$

$z(x, a)$  is an occupation measure. It gives the steady state probability that the system is in state  $x \in \mathbf{X}$  and action  $a \in \mathbf{A}$  is chosen. For any policy  $u$  there is one and only one corresponding occupation measure  $z$  [9]. Solving the linear program above we find  $z^*$  and the optimal policy is the one that chooses action  $a$  at state  $x$  with probability

$$u(a|x) = \frac{z^*(x, a)}{\sum_{a' \in \mathbf{A}} z^*(x, a')}. \quad (11)$$

### B. Model Parameters and the Presentation of the Results

Here we illustrate how the parameters of the model will be set and how the optimal policies will be presented throughout this paper. To this end we consider the scenario shown in the first column of Table II.

The parameter  $L$  specifies the size of the Markov Chain  $\Lambda$  that characterizes the channel. The precise definition of  $\Lambda$  assumed throughout this paper is analog to the BF-FSMC model [10]:  $\lambda_{0,0} = 1/2$ ,  $\lambda_{0,1} = 1/2$ ;  $\lambda_{L-1,L-1} = 1/2$ ,  $\lambda_{L-1,L-2} = 1/2$ ;  $\lambda_{i,i} = 1/3$ ,  $\lambda_{i,i-1} = 1/3$ ,  $\lambda_{i,i+1} = 1/3$  ( $2 \leq i \leq L-2$ ). This means that at each time slot the channel preserves its state or changes by one unit. The power gain  $g(y)$  when the MC is in state  $y$  is given by  $y/(L-1)$ .

Recall that the parameter  $P_{\text{max}}$  specifies the maximum power level available. The power action set is  $P = \{0, 1, 2, \dots, P_{\text{max}}\}$ .

We associate to each state and to each of the controller's actions an index, as shown below (an analog indexing scheme is used in the other examples):

state index	1	2	3	4	5	6	7	8	...	24
$q$	0	0	0	0	1	1	1	1	...	5
$y$	0	1	2	3	0	1	2	3	...	3

action index	0	1	2	3	4	5	6	7
$c$	0	0	0	0	1	1	1	1
$p$	0	1	2	3	0	1	2	3

In order to fully characterize the model we have to show how to assign  $\mathcal{P}_{xar}$ , the probability of moving from state  $x$  to  $r$  given that action  $a$  is used. Let  $\mathcal{P}_{xar} = \mathcal{P}_{(y_0, q_0)(p, c)(y_1, q_1)}$ . We assume that the transitions in the channel state are independent of the buffer state, hence  $\mathcal{P}_{(y_0, q_0)(p, c)(y_1, q_1)} = \lambda_{y_0, y_1} \mathcal{P}'_{q_0(p, c)q_1}$  where  $\mathcal{P}'_{q_0(p, c)q_1}$  depends on the arrival process and is given as follow. Let  $w$  be the number of packets served during the time slot  $t$ , where  $w = 0$  if  $p = 0$  and  $w = \min(q_0, 1)$  otherwise.

Let  $\Gamma$  be the r.v. that characterizes the number of arrivals in a time slot,  $q_1$  is equal to  $\min(q_0 - w + c\gamma, B)$  with probability  $P\{\Gamma = \gamma\}$ .  $\mathcal{P}'_{q_0(p, c)q_1} = P\{\Gamma = \gamma\}$  where  $q_1 = \min(q_0 - 1_{p>0} \min(q_0, 1) + c\gamma, B)$ .

Once the parameters of the linear program are fully determined we solve it using the simplex method. The optimal policy obtained from the occupation measures  $z(x, a)$  is shown in Fig. 1 (in the sequel of this paper, instead of 3d pictures we will only present the support of the optimal policies).

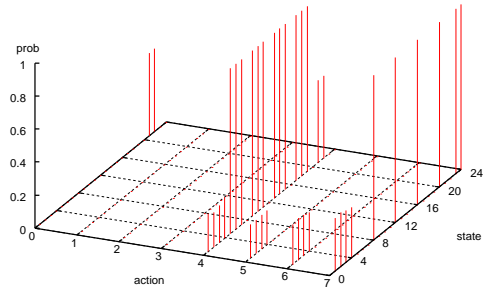


Fig. 1. Optimal policy.

Note that when the buffer is empty (states 0 to 4) the optimal policy consists of randomizing freely between actions 4 to 7. There is no difference between the outcome generated by these actions since when there are no packets to be transmitted the power level chosen is irrelevant.

When the amount of packets in the buffer varies between 1 and 4 (states 5 up to 20) the optimal policy is very restrictive: a packet is transmitted using maximum power just if the channel condition is perfect (states 8, 12, 16 and 20). However, when the buffer is full the optimal policy is less restrictive: packets are transmitted even if the channel condition is not perfect

(states 23 and 24). The rationale is that the optimal policy avoids as much as possible to empty the buffer (states 1-4) because when the buffer is empty the controller loses the opportunity to transmit packets and to increase its throughput. Therefore, the lower the amount of packets in the buffer the more restrictive the optimal policy. *One of our major conclusions is that the controller becomes more careless as the amount of packets in the buffer increases.* This comment applies to all the policies presented in this paper.

Finally, notice that in states 21 and 22 the optimal policy consists of randomizing between actions 0 and 4. This occurs because when the buffer is full and the controller decides not to transmit packets it is irrelevant if newly arrived packets that come from the upper layer will be accepted or not – the buffer will remain full independently of the adopted admission control action.

### C. Even Without Constraints, ‘Always Transmit’ is Not Optimal!

Consider now what happens when there are neither delay constraints nor power constraints. One could guess that in this scenario the optimal strategy would be to always transmit with maximum power. However, since we are considering a non saturated system, that’s not the case: the optimal policy consists of transmitting with maximum power when the network state is good, *but not transmitting when the network state is bad. When the queue empties the controller may lose opportunities to send packets so it is important to avoid empty queues.*

Considering the parameters shown in the second column of Table II we obtain the optimal policy in Fig. 2. When the number of packets in the buffer is small (states 12-55) the optimal policy states that packets should be transmitted if the channel condition is greater than or equal to 2 (a channel condition of 0 represents the worst possible scenario). When the number of packets in the buffer is large (states 56-110) the optimal policy is less restrictive: packets are not transmitted only if the network condition is 0.

A more prominent example illustrating that even without constraints the policy ‘always transmit’ is not optimal is obtained as follows. We start from the scenario in the second column of Table II and change the arrival probability of the packets from 0.9 to 0.1 (ceteris paribus). (Please, refer to [8] for the optimal policy in that case) The new optimal policy has the same structure but is much more restrictive than the one shown in Fig. 2. The optimal throughput obtained is 0.30 while the throughput obtained with the policy ‘always transmit’ is 0.25. All the numerical examples that we conducted had this property of leading to optimal policies that become more restrictive as the arrival probability decreases. This happens because the lower the arrival probability the greater the drift towards the states where the buffer is empty. Hence, the optimal policies must be more restrictive to compensate this drift and to avoid empty buffers.

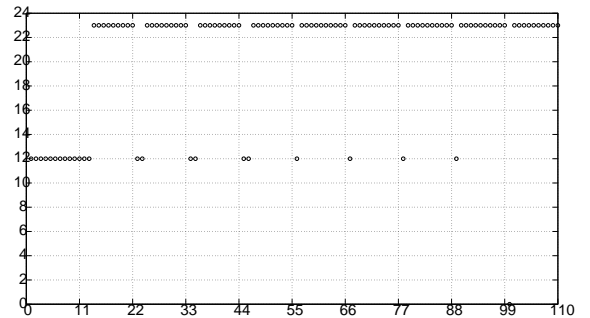


Fig. 2. No constraints. (x axis - states, y axis - actions)

### D. Coping with Power and Delay Constraints

We now analyze what happens when we have to cope with delay constraints. Starting from the scenario of the second column of Table II we add a power constraint of 5 and a delay constraint of 4. The optimal policy is depicted in Fig. 3.

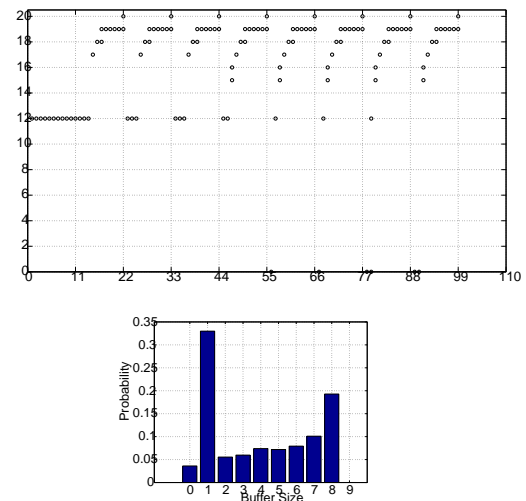


Fig. 3. Optimal policy with power and delay constraints (x axis - states, y axis - actions). Buffer distribution (down).

The optimal policies described up to this point did not involve the rejection of packets. The admission control action adopted could always be set to ‘accept’ without leading to a decrease in the mean throughput. However, due to the delay constraints the optimal policy shown in Fig. 3 involves necessarily the rejection of some packets that arrive from the upper layer when the system is in states 56, 67, 78, 79, 89 or 90. In Fig. 3 we see the pmf of the buffer size: the probability of a full buffer is indeed almost zero and the mean buffer size is equal to 4. Notice also that it is beneficial to have only one packet in the buffer but not the empty buffer.

## IV. MULTIPLE PLAYERS WITH INDEPENDENT STATE INFORMATION

In this section we study stochastic games where each player has independent state information. In particular, we seek for

TABLE II  
MODEL PARAMETERS

Parameter	Sect 1B	Sect 1C	Non cooperative and Cooperative	Correlated Cooperative (Sec IV-C)
$L$	4	11	4	4
$B$	5	9	7	7
$P_{\max}$	3	11	5	5
$P_{\text{const}}$	5	-	2	1.8
$Q_{\text{const}}$	-	-	none	none
arr. prob.	0.1	0.9	0.5	0.5

the Nash equilibria in different settings. The existence of those Nash equilibria among stationary policies is guaranteed, and the proof is presented in a subsequent paper [11].

#### A. Linear Program

Next we present a Linear Program (LP) for computing the set of all optimal responses for player  $i$  against a stationary policy  $\mathbf{u}_{-i}$ . Let  $\mathcal{K}_i = \{(x, a) : x \in \mathbf{X}_i, a \in \mathbf{A}_i(x)\}$ .

$\mathbf{LP}(i, \mathbf{u}) :$

Find  $z_i^* := \{z_i^*(x, a)\}_{x,a}$ , where  $(x, a) \in \mathcal{K}_i$ , that maximizes

$$\sum_{x \in \mathbf{X}_i} \sum_{a \in \mathbf{A}_i(y)} R_i^u(x, a) z_i(x, a) \quad \text{subject to:} \quad (12)$$

$$\sum_{x \in \mathbf{X}_i} \sum_{a \in \mathbf{A}_i(x)} z_i(x, a) [\delta_r(x) - \mathcal{P}_{xar}^i] = 0, \quad \forall r \in \mathbf{X}_i, \quad (13)$$

$$\sum_{x=(y,q) \in \mathbf{X}_i} \sum_{a \in \mathbf{A}_i} q z_i(x, a) \leq Q_{\text{const}} \quad (14)$$

$$\sum_{x \in \mathbf{X}_i} \sum_{a \in \mathbf{A}_i} d(x, a) z_i(x, a) \leq P_{\text{const}} \quad (15)$$

$$z_i(x, a) \geq 0, \quad \forall (x, a) \in \mathcal{K}_i \quad \sum_{(x,a) \in \mathcal{K}_i} z_i(x, a) = 1 \quad (16)$$

$R_i^u(\cdot)$  denotes the immediate cost induced by players other than  $i$ , when player  $i$  uses action  $a$  and the other players use a stationary multi policy  $\mathbf{u}_{-i}$ :

$$R_i^u(x, a) := \sum_{\mathbf{x}_{-i} \in \mathbf{X}_{-i}} \sum_{\mathbf{a}_{-i} \in \mathbf{A}_{-i}} \left[ \prod_{l \neq i} u_l(a_l | x_l) \pi_l^u(x_l) \right] t_i(\mathbf{x}, \mathbf{a}) \quad (17)$$

where  $\pi_l^u(x_l)$  is the steady state probability of user  $l$  being in state  $x_l$  given that multi policy  $\mathbf{u}$  is adopted (note that  $\pi_l^u(x_l)$  depends on  $\mathbf{u}$  only through  $u_l$ ).  $t_i(\mathbf{x}, \mathbf{a})$  is equal to  $r_i(\mathbf{x}, \mathbf{a})$  defined by (2) (unless we explicitly consider something different).

#### B. Non Zero Sum Games

We now consider two decentralized problems. In the first one each mobile maximizes its own throughput while in the second one each mobile maximizes the aggregated throughput. Our aim is to compare the equilibrium policies and the throughput obtained in the equilibrium of these two scenarios.

1) *Non-Cooperative Game*: Assuming that there are two mobiles sharing a channel and that the parameters of the model are those described in the fourth column of Table II, the optimal policy for both players is the one shown in Fig. 4. The throughput obtained by each player in the equilibrium is 0.69616 (the aggregated throughput is 1.39231).

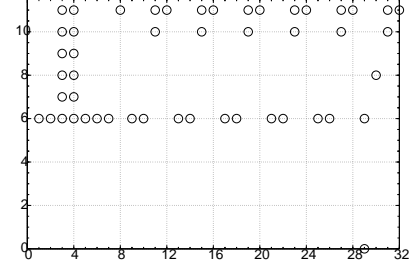


Fig. 4. Non cooperative equilibrium (x axis, states; y axis, actions).

2) *Cooperative Game (Team Problem)*: We consider here the same parameters as in the decentralized non-cooperative case. However, now we assume that the two players have the same objective function, which is the aggregated throughput. Player  $i$  seeks the policy  $u_i$  that maximizes the objective function

$$R(u_i, \mathbf{u}_{-i}^*) = R_i(u_i, \mathbf{u}_{-i}^*) + R_{-i}(u_i, \mathbf{u}_{-i}^*) \quad (18)$$

where  $\mathbf{u}_{-i}^*$  is the strategy of the other player. For the two player case, this is equivalent to setting in equation (17) the quantity  $t_1(\mathbf{y}, \mathbf{a}) = t_2(\mathbf{y}, \mathbf{a}) = r_1(\mathbf{y}, \mathbf{a}) + r_2(\mathbf{y}, \mathbf{a})$  where  $r_i$ 's were defined in (2).

Due to space restrictions, please refer to [8] for the optimal policy in this scenario. The aggregated throughput is 1.39726.

## V. CORRELATED EQUILIBRIA

In this section, we consider the simple case when there are only two players involved. Results can be straightforwardly extended to the case of multiple players.

We wish to introduce a coordination mechanism in this model. The simplest and most intuitive coordination mechanism is given by a common signal which both users as well as the base station overhear before each transmission. This signal is a random boolean variable in  $\{0, 1\}$ , and is independent of all other random variables in the setting. Instead of using the regular matched filter as in (2), the base station will use Successive Interference Cancellation (SIC) matched filter [12]. If the signal is "0", the base station will decode first user 1 then user 2; if the signal is "1", the base station will decode them in the reverse order, first user 2 then user 1. We assume that users will always be decoded successfully, which is critical in SIC. Thus each time a signal is decoded its contribution to the interference is perfectly subtracted. This removes some of the inter-user interference and therefore increases the SNR of the following decoded user. This rule of decoding according to the coordination signal is known by the users, which allows them to jointly optimize their strategies. In this section we assume

that there is a device that emits signals to the mobiles at every time slot. This means that the players may coordinate through the emitted signals and we seek for correlated equilibria [6], [7].

### A. Linear Program

The LP presented in Section IV-A can be adapted to find correlated equilibria as follows. First, the system state of each mobile  $i$  is now characterized by  $x_i = (y_i, q_i, s_i)$  where the extra variable  $s_i$  is the correlation signal. Second, we adapt  $t_i$ .  $t_i(\mathbf{x}, \mathbf{a}) = t_i((y_1, q_1, s_1), \dots, (y_N, q_N, s_N), \mathbf{a})$  is equal to zero if it is not the case that  $s_1 = s_2 = \dots = s_N$  and  $t_i$  is greater than or equal to 0 otherwise (the exact definition will depend on the scenario considered, as we will see next). Finally, we adjust  $\mathcal{P}_{xar}^i$ . Assuming that the correlation signal alternates between two values, given  $x = (y_1, q_1, s_1)$  and  $r = (y_2, q_2, s_2)$ ,  $\mathcal{P}_{xar}^i \geq 0$  if  $s_1 \neq s_2$  and  $\mathcal{P}_{xar}^i = 0$  otherwise.

### B. Non-Cooperative Game

We assume that the AP alternates in a round robin fashion between serving first the packets from player 1 and then those from player 2. The parameters of the game are the same as those considered in the last section (last column of Table II). At each time slot the AP sends a signal (which alternates between 1 and 2) to both mobiles. When the signal is 2 the packets from mobile 2 are decoded before those of mobile 1 [5]. Thus the throughput perceived by mobile 1 is

$$t_1(((y_1, q_1, 2), (y_2, q_2, 2)), \mathbf{a}) = \log_2 \left( 1 + \frac{g_1(x_1)d_1(y_1, a_1)}{N_0} \right) \quad (19)$$

and the throughput obtained by mobile 2 is

$$t_2(((y_1, q_1, 2), (y_2, q_2, 2)), \mathbf{a}) = \log_2 \left( 1 + \frac{g_2(x_2)d_2(y_2, a_2)}{N_0 + g_1(x_1)d_1(y_1, a_1)} \right) \quad (20)$$

Similar formulas apply when the signal is equal to 1.

Notice that the state space has now size 64 (twice the size of the state space in the last section). The first (last) 32 states correspond to a signal equal to 1 (2).

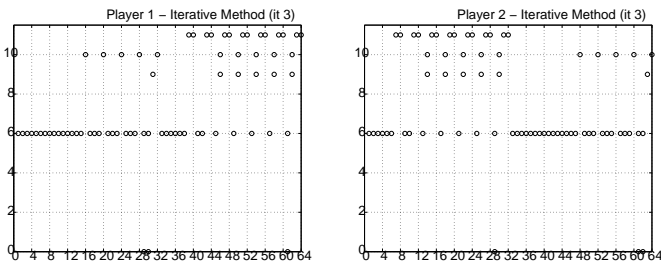


Fig. 5. Correlated equilibrium: optimal strategy for players 1 (left) and 2 (right) (x axis, states; y axis, actions).

The throughput obtained by each player in the equilibrium is 0.8168 (greater than the 0.69616 obtained in Section IV-B1 for the decentralized non-cooperative case). The optimal strategies obtained using the iterative method for player 1 and 2 are shown in Fig. 5. Player 1 invests most of his power in states

33-64 when he has an advantage over player 2, while player 2 invests more power in states 1-32 than in states 33-64.

### C. Cooperative Game (Team Problem)

In this section we analyze the correlated equilibrium in the decentralized cooperative scenario (all the mobiles wish to maximize the aggregated throughput). The correlation mechanism considered here is not the same as the one considered in the last section. Here we assume the simplest possible of those mechanisms: at each time slot the AP sends a signal which alternates between 1 and 2 to the mobiles. The AP does not take different actions based on the signal sent.

Consider the parameters shown in the last column of Table II. The optimal policy obtained using the LCP for players 1 and 2 is shown in Fig. 6. It is interesting to notice that the equilibrium has a time division multiplexing structure: the mobiles never transmit at the same time. The joint throughput in this case is equal to 1.49294.

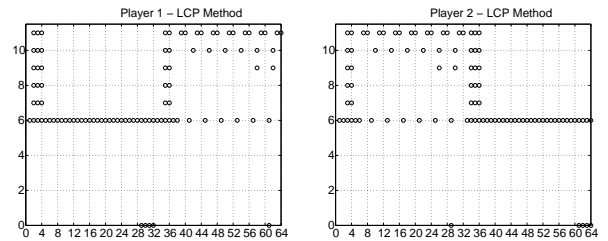


Fig. 6. Optimal strategy for players 1 (left) and 2 (right) (x axis, states; y axis, actions).

**Acknowledgement:** This work was supported by the Bionets European project.

### REFERENCES

- [1] E. Altman, K. Avrachenkov, R. Marquez, and G. Miller, "Zero-sum constrained stochastic games with independent state processes," *Math. Meth. Oper. Res.*, vol. 62, pp. 375–386, 2005.
- [2] E. Altman, K. Avrachenkov, G. Miller, and B. Prabhu, "Uplink dynamic discrete power control in cellular networks," *INRIA Technical Report 5818*, 2006.
- [3] A. Lazar, "optimal flow control of a class of queuing networks in equilibrium," *IEEE Transactions on Automatic Control*, vol. 28, pp. 1001–1007, 1983.
- [4] M. T. Hsiao and A. A. Lazar, "Optimal decentralized flow control of Markovian queuing networks with multiple controllers," *Performance Evaluation*, vol. 13, pp. 181–204, 1991.
- [5] L. Lai and H. E. Gamal, "The water-filling game in fading multiple access channels," *arXiv*, 2005, arXiv:cs.IT/0512013 v1 3 Dec 2005e-.
- [6] R. Aumann, "Subjectivity and Correlation in Randomized Strategies," *Journal of Mathematical Economics*, vol. 1, pp. 67–96, 1974.
- [7] E. Altman, N. Bonneau, and M. Debbah, "Correlated equilibrium in access control for wireless communications," in *Networking*, 2006.
- [8] E. Altman, K. Avrachenkov, N. Bonneau, M. Debbah, R. El-Azouzi, and D. Menasché, "Constrained stochastic games: Dynamic control in wireless networks," *Tech Report*, 2007, www-net.cs.umass.edu/~sadorc/mdp.
- [9] E. Altman, *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.
- [10] H. Wang and N. B. Mandayam, "Dynamic power control under energy and delay constraints," in *Globecom*, 2001.
- [11] E. Altman, "Constrained cost coupled stochastic games with independent state information," *to be published*, 2006.
- [12] R. Müller and S. Verdú, "Design and Analysis of Low-Complexity Interference Mitigation on Vector Channels," pp. 1429–1441, Aug. 2001.