A Competitive Rate Allocation Game

Yanting Wu[1], George Rabanca[2], Bhaskar Krishnamachari[1], and Amotz Bar-Noy[2]

[1] University of Southern California, Los Angeles CA 90089, USA, {yantingw,bkrishna}@usc.edu

[2] The City University of New York, New York, NY 10016, USA grabanca@gc.cuny.edu, amotz@sci.brooklyn.cuny.edu

Abstract. We introduce a competitive rate allocation game in which two receivers compete to forward the data from a transmitter to a destination in exchange for a payment proportional to the amount of forwarded data. At each time slot the channel from the transmitter to each receiver is an independent random variable with two states, high or low, affecting the amount of data that can be transmitted. Receivers make "bids" on the state of their channel and the transmitter allocates rate accordingly. Receivers are rewarded for successful transmissions and penalized for unsuccessful transmissions. The goal of the transmitter is to set the penalties in such a way that even if the receivers are selfish, the data forwarded is close to the optimal transmission rate. We first model this problem as a single shot game in which the receivers know the channel probability distributions but the transmitter does not, and show that it is possible for the transmitter to set penalties so as to ensure that both receivers have a dominant strategy and the corresponding Price of Anarchy is bounded by 2. We show, moreover, that this is in a sense the best possible bound. We next consider the case when receivers have incomplete information on the distributions, and numerically evaluate the performance of a distributed online learning algorithm based on the well-known UCB1 policy for this case. From simulations, we find that use of UCB1 policy yields a performance close to the dominant strategy.

Keywords: competitive rate allocation game, Nash equilibrium, online learning.

1 Introduction

Optimizing throughput is one of the central problems in wireless networking research. To make good use of the available wireless channels, the transmitter must allocate rate efficiently. We study in this paper a simple yet fundamental rate allocation problem in which the transmitter does not precisely know the state of the channels, and the corresponding receivers are selfish.

In this problem, there is one transmitter that must allocate rates to two different receivers to forward data on its behalf to a given destination (see illustration in figure 1). The two channels from the transmitter to each receiver are independent channels with two states: high or low. The channel states are assumed to

V. Krishnamurthy et al. (Eds.): GameNets 2012, LNICST 105, pp. 16–30, 2012.

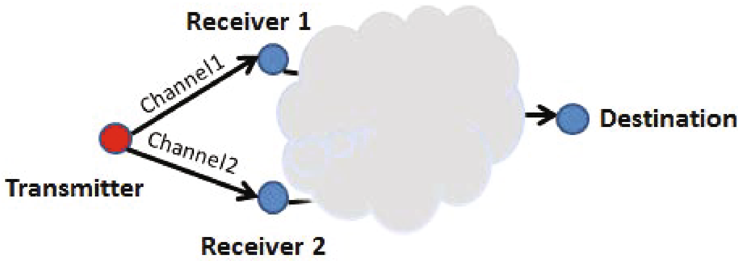

Fig. 1. Illustration of problem

be i.i.d. Bernoulli random variables. Initially we assume that the receivers both know each others channel parameters, but the transmitter does not. At each time, the receivers communicate to the transmitter a binary bid corresponding to the possible state of their respective channels. The transmitter responds to these bids by deciding whether to send aggressively (at a high, or very high rate) or conservatively (at a low rate) on one or both channels. Specifically when both receivers bid low, the transmitter sends data at a low rate R_1 over both channels. And when both receivers bid high, the transmitter splits its power to send data at a high rate R_2 over both channels. When one of the receivers bids low and the other bids high, the transmitter sends data at a very high rate R_3 over the latter channel. When the sender sends data at a high or very high rate, we assume that there is a failure and nothing gets sent if the transmission channel actually turns out to be bad. In this case, the sender levies a penalty on the receivers. But whenever data is successfully sent, it pays the receiver a fee proportional to the rate obtained.

There are two roles in this setting: the receivers and the transmitter. The receivers want to get as much reward as possible, avoiding the penalties. Since the transmitter's rate allocation is a competitive resource that directly affects the receivers' utilities, the setting can be modeled as a two player, non-cooperative game. On the other hand, the transmitter is the game designer: it can choose the penalties in order to influence how the receivers play the game. The goal of the transmitter is to transmit as much data as possible and, without knowledge of the receiver's channel states, to guarantee that the total transmission is not much worse than the optimal. In this paper we prove that there is a way to set the penalty terms such that both receivers have dominant strategies, and the data forwarded by two receivers is at least $1/2$ of the optimal, in other words, that the Price of Anarchy from the transmitter's point of view is at most 2.

If the underlying channels' states are known we can assume that the two receivers will play their dominant strategies if they have one. However, if the underlying channel status is unknown, the receivers need to learn which action is more beneficial. Assuming that the underlying channel state is drawn from an unknown underlying distribution at each time slot, we show that modeling each payers' choice of action as a multi-armed bandit leads to desirable results.

In this paper we adapt UCB1 algorithm [1], which there are two arms for each receiver, each arm corresponding to an action: bidding high or bidding low. From the simulations, we find that the UCB1 algorithm gives a performance which is close to the dominant strategy, and, when both receivers use UCB1 to choose their strategies, it can give even better payoffs than playing the dominant strategy.

Related Work: Game theory, which is a mathematical tool for analyzing the interaction of two or more decision makers, has been used in wireless communications by many authors [2], [3]. While we are not aware of other papers that have addressed exactly the same formulation as discussed here, other researchers have explored related game theoretic problems pertaining to power allocation over multiple channels. For instance, the authors of [4] formulate a multiuser power control problem as a noncooperative game, show the existence and uniqueness of a Nash equilibrium for a two-player version game and propose a water-filling algoirhtm which reaches the Nash equilibrium efficiently; and the authors of [5] study a power allocation game for orthogonal multiple access channels, prove that there exists a unique equilibrium of this game when the channels are static and show that a simple distributed learning schema based on the replicator dynamics converges to equilibrium exponentially fast. Unlike most of the prior works, our formulation and analysis is not focused on optimizing the power allocation per se, but rather on issues of information asymmetry between the transmitter and receivers and the design of appropriate penalties levied by the transmitter to ensure that the receiver's selfishness do not hurt performance too much. Somewhat related to the formulation in this paper are two recent papers on non-game-theoretic formulations for a transmitter to decide on the whether to send conservatively or aggressively over a single (known or unknown) Markovian channel [6] [7]. Although we consider a simpler Bernoulli channels here in which case the transmitter's decisions would be simplified, our formulation focuses on strategic interactions between two receivers. In the case of unknown stochastic payoffs, we consider the use of a multi-armed bandit-based learning algorithm. Relatively little is known about the performance of such online learning algorithms in game formulations, though it has been shown that they do not always converge to Nash equilibria [8].

2 Problem Formulation

In the rate allocation game we consider two receivers and one transmitter. The transmitter uses the two receivers to forward data to the destination. The channel from the transmitter to each receiver has one of the two states at each time slot: low (L or 0) or high (H or 1). The two channels are independent with each other and their state comes from an i.i.d. distribution. We denote p_i $(i = 0, 1)$ as the probability that channel i is in state high at any time. Before transmitting, neither the receivers nor the transmitter know the state of the channel. At the beginning of each time slot, each receiver makes a "bid" (high or low). The transmitter allocates rate to the receivers according to the bids sent. At the end of the time slot both receivers observe whether or not their transmission was successful. A transmission is not successful if the respective channel is in a low state but has been assumed to be in a high state.

Since the channel state is unknown in advance, the receivers' bid may lead to an unsuccessful transmission. If the the transmission is successful, the receiver is paid an amount proportional to the transmission rate. Otherwise, it will get a penalty (negative reward). Table 1 shows the reward functions for each receiver.

Table 1.

Bid	Actual State	Other Channel Bid	Reward
L	L	L	R_1
		H	0
L	H	L	R_1
		H	0
H	L	L	$-C$
		H	$-D$
H	H	L	R_3
		H	R_2

Throughout the rest of the paper we will assume that $R_1 < R_2 < R_3 < 2R_2$. C and D are the penalties that the receivers get for making a high bid when their channel state is low.

There are two roles in this game setting: the transmitter and the receivers. The transmitter wants to carry as much data as possible to the destination. It is not interested in penalizing the receivers, but only uses the penalty to give incentive to the receivers to make good guesses. The receivers are only interested in the reward and they don't lose any utility from transmitting more data.

3 Parameters Known Cases - Receivers' Perspective

Table 2 shows the relationship between the expected rewards for the two receivers as a normal form game. In each cell, the first value corresponds to the reward for receiver 1, and the second value corresponds to the reward for receiver 2.

Table 2.

Receiver 1 \ Receiver 2	L	H
L	(R_1, R_1)	$(0, p_2R_3 - (1-p_2)C)$
H	$(p_1R_3 - (1-p_1)C, 0)$	$(p_1R_2 - (1-p_1)D, p_2R_2 - (1-p_2)D)$

3.1 Mixed Nash Equilibrium

We denote by XY_i the expected reward for receiver i ($i = 1, 2$) when receiver 1 bids X, receiver 2 bids Y (where X and Y are high or low).

$$\begin{aligned} LL_1 &= R_1, \\ LL_2 &= R_1, \\ LH_1 &= 0, \\ LH_2 &= p_2 R_3 - (1 - p_2)C, \\ HL_1 &= p_1 R_3 - (1 - p_1)C, \\ HL_2 &= 0, \\ HH_1 &= p_1 R_2 - (1 - p_1)D, \\ HH_2 &= p_2 R_2 - (1 - p_2)D. \end{aligned} \tag{1}$$

Let receiver 1 bid high with probability q_1, and receiver 2 bid high with probability q_2. At Nash equilibrium, receiver 1 selects the probability such that the utility function for receiver 2 is the same for both bidding high and bidding low. Therefore we have:

$$(1 - q_1)LL_2 = (1 - q_1)LH_2 + q_1 HH_2. \tag{2}$$

Similarly for receiver 2:

$$(1 - q_2) \times LL_1 = (1 - q_2)HL_1 + q_2 HH_1. \tag{3}$$

Solving 2 and 3 we get :

$$q_1 = \frac{-C + Cp_2 - R_1 + p_2 R_3}{-C + D + Cp_2 - Dp_2 - R_1 - p_2 R_2 + p_2 R_3}, \tag{4}$$

$$q_2 = \frac{-C + Cp_1 - R_1 + p_1 R_3}{-C + D + Cp_1 - Dp_1 - R_1 - p_1 R_2 + p_1 R_3}. \tag{5}$$

Setting q_1 and q_2 to be 0 or 1, we can find a relationship between the values of p_1 and p_2 and the existence of a pure Nash equilibrium.

$$\begin{array}{l} \text{If } q_1 = 0 \text{ and } q_2 = 0, \text{then } p_1 = \frac{C+R_1}{C+R_3}\ p_2 = \frac{C+R_1}{C+R_3}. \\ \text{If } q_1 = 0 \text{ and } q_2 = 1, \text{then } p_1 = \frac{D}{D+R_2}\ p_2 = \frac{C+R_1}{C+R_3}. \\ \text{If } q_1 = 1 \text{ and } q_2 = 0, \text{then } p_1 = \frac{C+R_1}{C+R_3}\ p_2 = \frac{D}{D+R_2}. \\ \text{If } q_1 = 1 \text{ and } q_2 = 1, \text{then } p_1 = \frac{D}{D+R_2}\ p_2 = \frac{D}{D+R_2}. \end{array} \tag{6}$$

Denote

$$b_1 = \min\{\frac{C + R_1}{C + R_3}, \frac{D}{D + R_2}\}, \tag{7}$$

$$b_2 = \max\{\frac{C + R_1}{C + R_3}, \frac{D}{D + R_2}\}. \tag{8}$$

Theorem 1. *If* $p_1 \notin [b_1, b_2]$ *or* $p_2 \notin [b_1, b_2]$*, then there exists a unique pure Nash equilibrium.*

Proof. Let $p_1 < b_1$, thus $p_1 < \frac{C+R_1}{C+R_3}$ and $p_1 < \frac{D}{D+R_2}$

$$HL_1 = p_1 R_3 - (1 - p_1)C < b_1(R_3 + C) - C \leq R_1,$$
$$HH_1 = p_1 R_2 - (1 - p_1)D < b_1(R_2 + D) - D \leq 0.$$

Thus receiver 1 has a dominating strategy for bidding low. When receiver 1 bids low, the optimal action for receiver 2 is bidding low if $LL_2 > LH_2$, and high otherwise.

Let $p_1 > b_2$, thus $p_1 > \frac{C+R_1}{C+R_3}$ and $p_1 > \frac{D}{D+R_2}$

$$HL_1 = p_1 R_3 - (1 - p_1)C > b_2(R_3 + C) - C \geq R_1,$$
$$HH_1 = p_1 R_2 - (1 - p_1)D > b_2(R_2 + D) - D \geq 0.$$

Thus the dominating strategy for receiver 1 is bidding high. When receiver 1 bids high, the optimal action for receiver 2 is bidding high if $HL_2 < HH_2$ and low otherwise.

Similarly for $p_2 \notin [b_1, b_2]$.

Lemma 1. *If* $p_1 \in (b_1, b_2)$ *and* $p_2 \in (b_1, b_2)$*, there exists more than one Nash equilibrium.*

Proof. Let $p_1 \in (b_1, b_2)$ and $p_2 \in (b_1, b_2)$, then there are two possible scenarios:

Scenario 1: $b_1 = \frac{C+R_1}{C+R_3}$, $b_2 = \frac{D}{D+R_2}$, then

$$LH_2 = p_2 R_3 - (1 - p_2)C = p_2(R_3 + C) - C > b_1(R_3 + C) - C = R_1.$$

Similarly, $HL_1 > R_1$.

$$HH_1 = p_1 R_2 - (1 - p_1)D = p_1(R_2 + D) - D < b_2(R_2 + D) = 0.$$

Similarly, $HH_2 < 0$.

The payoff matrix for receivers will become as Table 3 shown:

Table 3.

Receiver 1 \ Receiver 2	L	H
L	(R_1, R_1)	$(0, > R_1)$
H	$(> R_1, 0)$	$(< 0, < 0)$

There are two Nash equilibrium: when one receiver bids high, the other receiver bids low.

Scenario 2: $b_1 = \frac{D}{D+R_2}$, $b_2 = \frac{C+R_1}{C+R_3}$, then

$$LH_2 = p_2 R_3 - (1 - p_2)C = p_2(R_3 + C) - C < b_2(R_3 + C) - C = R_1.$$

Similarly, $HL_1 < R_1$.

$$HH_1 = p_1R_2 - (1-p_1)D = p_1(R_2 + D) - D > b_2(R_2 + D) = 0.$$

Similarly, $HH_2 > 0$.

The payoff matrix for receivers will become as Table 5 shown:

Table 4.

Receiver 1 \ Receiver 2	L	H
L	(R_1, R_1)	$(0, < R_1)$
H	$(< R_1, 0)$	$(> 0, > 0)$

There are two Nash equilibrium: both bid high, or both bid low.

In the range of $(b1, b2) \times (b1, b2)$, if both receivers play mixed Nash equilibrium, their utility could become much worse than they play pure Nash equilibrium.

If the mixed Nash equilibrium is used. The expected total utility function for each receiver are:

$$\begin{aligned} U_1 = (1-q_1)(1-q_2)R_1 + q_1(1-p_1)(1-q_2)(-C), \\ +q_1(1-p_1)q_2(-D) + q_1p_1(1-q_2)R_3 + q_1p_1q_2R_2. \end{aligned} \tag{9}$$

$$\begin{aligned} U_2 = (1-q_2)(1-q_1)R_1 + q_2(1-p_2)(1-q_1)(-C), \\ +q_2(1-p_2)q_1(-D) + q_2p_2(1-q_1)R_3 + q_2p2q_1R_2. \end{aligned} \tag{10}$$

In cases where $b_1 = \frac{D}{D+R_2}, b_2 = \frac{C+R_1}{C+R_3}$, when $p_1 \to b_1+$, $p_2 \to b_1+$, we have $q_1 \to 1$ and $q_2 \to 1$. Substituting in Eq. (10) and Eq. (11), we can get $U_1 \to 0$ and $U_2 \to 0$, which is much worse than they just play LL Nash equilibrium. Both receivers suffer if they play mixed Nash equilibrium.

For simplicity, we want to set C and D such that we only have pure Nash equilibrium, independent of the probability distributions p_1 and p_2.

Lemma 2. *Given C, there exists a D, such that there only exist pure Nash equilibrium.*

Proof. When $D = \frac{-CR_2 - R_1R_2}{R_1 - R_3}$, we can get $b_1 = b_2$, there only exists pure Nash equilibrium region.

Lemma 3. *If there only exists pure Nash equilibrium both receivers have a dominant strategy.*

Proof. If we only have pure Nash equilibrium then we must have $b_1 = b_2 = p$. There are four possible scenarios:

Scenarios 1: $p_1 < p$ and $p_2 < p$,

The payoff matrix for receivers will become as Table 5 shown:

Table 5.

Receiver 1 \ Receiver 2	L	H
L	(R_1, R_1)	$(0, < R_1)$
H	$(< R_1, 0)$	$(< 0, < 0)$

The dominant strategies for both receivers are bidding low.

Similarly, we have

Scenario 2: $p_1 < p$ and $p_2 > p$, dominant strategy for receiver 1 is bidding low and dominant strategy for receiver 2 is bidding high.

Scenario 3: $p_1 > p$ and $p_2 < p$, dominant strategy for receiver 1 is bidding high and dominant strategy for receiver 2 is bidding low.

Scenario 4: $p_1 > p$ and $p_2 > p$, dominant strategy for both receivers is bidding high.

4 Parameters Known Cases - Transmitter' Perspective

In this section, we consider the amount of data which can be sent by the two receivers. Think the transmitter asks the two receivers to forward its data. What the transmitter really cares about is how much data is sent. In this case, when sending fails, we consider the data sent is 0. The penalty term C and D are to let the receivers adjust their bidding, but for transmitter, it does not get such a penalty.

Table 6 represents the expected rewards table got from the transmitter's view:

Table 6.

Receiver 1 \ Receiver 2	L	H
L	(R_1, R_1)	$(0, p_2 R_3)$
H	$(p_1 R_3, 0)$	$(p_1 R_2, p_2 R_2)$

Utility functions from the transmitter's point of view:

$$V_{LL} = R_1 + R_1, \quad (11)$$

$$V_{HL} = p_1 R_3, \quad (12)$$

$$V_{LH} = p_2 R_3, \quad (13)$$

$$V_{HH} = p_1 R_2 + p_2 R_2. \quad (14)$$

$$(15)$$

Price of Anarchy (PoA):

$$PoA = \frac{\max_{s \in S} V(s)}{\min_{s \in NE} V(s)}. \tag{16}$$

where S is the strategy set, NE is the Nash equilibrium set, and $V(s) = \{V_{LL}, V_{HL}, V_{LH}, V_{HH}\}$.

Theorem 2. *If $C = \frac{R_1R_3 - R_1R_2}{R_2 - R_1}$ and $D = \frac{R_1R_2}{R_2 - R_1}$, then $PoA < 2$.*

Proof. If $C = \frac{R_1R_3 - R_1R_2}{R_2 - R_1}$ and $D = \frac{R_1R_2}{R_2 - R_1}$, then $b_1 = b_2 = \frac{R_1}{R_2}$. There only exists pure Nash equilibrium.

Let $p = \frac{R_1}{R_2}$,

If $p_1 < p$ and $p_2 < p$,

$$V_{LL} = 2R_1, \tag{17}$$
$$V_{HL} = p_1R_3 < \frac{R_1R_3}{R_2} < 2R_1, \tag{18}$$
$$V_{LH} = p_2R_3 < \frac{R_1R_3}{R_2} < 2R_1, \tag{19}$$
$$V_{HH} = p_1R_2 + p_2R_2 < 2R_1. \tag{20}$$
$$\tag{21}$$

The optimal is LL. The Nash equilibrium is also LL. Thus $PoA = 1$.

If $p_1 < p$ and $p_2 > p$,

$$V_{LL} = 2R_1 < 2p_2R_2 < 2p_2R_3, \tag{22}$$
$$V_{HL} = p_1R_3 < p_2R_3, \tag{23}$$
$$V_{LH} = p_2R_3, \tag{24}$$
$$V_{HH} = p_1R_2 + p_2R_2 < 2p_2R_2 < 2p_2R_3. \tag{25}$$
$$\tag{26}$$

The optimal is at most $2p_2R_3$. The Nash equilibrium is LH. Thus $PoA < 2$.

If $p_1 > p$ and $p_2 < p$, similar to the $p_1 < p$ and $p_2 > p$ case.

If $p_1 > p$ and $p_2 > p$,

$$V_{LL} = 2R_1 < 2p_1R_2, \tag{27}$$
$$V_{HL} = p_1R_3 < 2p_1R_2, \tag{28}$$
$$V_{LH} = p_2R_3 < 2p_2R_2, \tag{29}$$
$$V_{HH} = p_1R_2 + p_2R_2. \tag{30}$$
$$\tag{31}$$

The optimal is at most $2(p_1R_2 + p_2R_2)$. Nash equilibrium is HH. Thus $PoA < 2$

Lemma 4. *In the rate allocation game, for any fixed penalties C and D, there exist p_1 and p_2 such that the PoA is at least $\frac{2R_1}{R_3}$.*

Proof. Assume that $p_1 = 0$ and $p_2 = 1$. Then the table 7 shows the receivers' payoff matrices and table 8 shows the transmitter's payoff.

Table 7. Receivers' payoff for $p_1 = 0$ and $p_2 = 1$

Receiver 1 \ Receiver 2	L	H
L	(R_1, R_1)	$(0, R_3)$
H	$(0, 0)$	$(-C, R_2)$

Table 8. Transmitter's payoff for $p_1 = 0$ and $p_2 = 1$

Receiver 1 \ Receiver 2	L	H
L	$(2R_1)$	(R_3)
H	(0)	(R_2)

Since $R3 > R1$, then the only Nash equilibrium in this instance of the game is (L, H) for a transmitter utility of R_3. If $2R_1 > R_3$ then the optimal solution from the transmitter perspective is (L, L) for an utility of $2R_1$.

The Price of Anarchy is at least $\frac{2R_1}{R3}$.

Corollary 1. *The Price of Anarchy for the rate allocation game over all instances can be arbitrarily close to 2 for any C and D.*

Proof. Setting $R_1 = R_2 + \epsilon = R_3 + 2\epsilon$ ($\epsilon \to 0+$) in the lemma 4 leads to a $PoA \to 2$.

This corollary implies that our result in Theorem 2 showing how that the PoA can be bounded by 2 is essentially tight in the sense that no better guarantee could be provided that applies to all problem parameters.

5 Online Learning Using Multi-armed Bandits Algorithms

When the channels' status are known, and C and D are set as described in Theorem 2, both receivers have dominant strategies. However, when the channels' status are unknown, the receivers need to try both actions: sending with high data rate or sending with low data rate. The underlying channels are stochastic, even to each receiver, the probability that the channel will be good is unknown. Multi-armed bandits are handy tool to tackle the stochastic channel problems, so we adopt the well known UCB1 algorithm [1] to figure out the optimal strategies for each receiver. The arms correspond to actions: bidding high or low, each receiver only records the average rewards and number of plays and play by the UCB1 algorithm in a distributed manner without taking into account the other receiver's actions.

We recap the UCB1 algorithm in Alg. 1, normalizing the rewards in our case to lie between 0 and 1.

Algorithm 1. Online learning using UCB1

There are two arms corresponding to each receiver: bidding high or bidding low. Let x_l be the rewards which represents the average reward gained by each receiver by playing arm l (l = H, L), n_l represents how many times the arm l is played.

Initialization: Initially, playing each arm once, store the initial rewards in x_l, and set $n_l = 1$.

for time slot $n = 1, 2, \cdots$ **do**

Select the arm with highest value of $\frac{\bar{x}_l+D}{R_3+D} + \sqrt{\frac{2ln(n)}{n_l}}$. Play the selected arm for a time slot. Update the average reward of the selected arm as well as n_l of the selected arm.

end for

6 Simulations

In this section we present some simulation results showing that the UCB1 learning algorithm performs well. In all simulations we fix the penalties C and D as in theorem 2 which leaves each receiver with a dominant strategy, but which is not usually known by the receivers. In the figures below we show how the UCB1 learning algorithm compares with playing the dominant strategy (if the receiver knew it) and determine that using UCB1 does not lose much utility in average, and sometimes is better than the dominant strategy.

First, in figure 2, we assume that receiver 2 knows its probability for the state of the channel being high, and plays its own dominant strategy. In this case receiver 1 would be better off if it knew the probability of its state and would play the dominant strategy. However, playing UCB1 does not lose much utility in average. Figure 2 shows for each R_1 as a fraction of R_3, the average payoff over multiple games in which R_2, p_1 and p_2 are distributed over their entire domain.

In figure 3 we show the average payoff over multiple choices of R_2, p_1 and p_2, when receiver 1 plays either the dominant strategy or the UCB1 strategy, and receiver 2 plays the UCB1 strategy. We can see here that the dominant strategy is only better in average for large values of R_1 and for small values of R_1 playing UCB1 brings better payoff.

Figure 4 and 5 show the same scenarios from the transmitter's perspective. Figure 4 compares the optimal average utility the transmitter could get from each game to the average utility the transmitter gets from receiver 1 using UCB1 or receiver 1 using its dominant strategy, when receiver 2 plays its dominant strategy. We notice that both strategies give almost the same payoff to the transmitter, especially when the value of R_1 is much smaller compared to R_3. This happens because when receiver 1 uses UCB1 against a player that uses its dominant strategy then receiver 1 will quickly learn to play its dominant strategy as well. Figure 5 shows how the transmitter optimal payoff compares to the transmitter payoff when the receiver 2 uses the UCB1. When both receivers use the UCB1 algorithm to choose their strategies, the transmitter payoff is better than when one receiver uses the dominant strategy and the other receiver

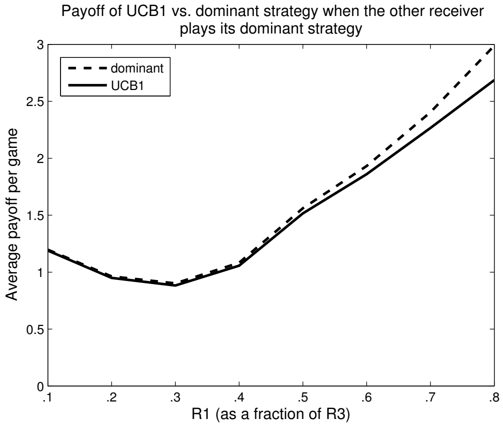

Fig. 2. Receiver 1 payoff against receiver 2 using dominant strategy

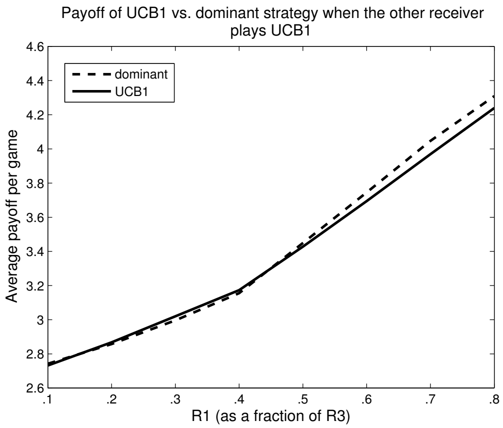

Fig. 3. Receiver 1 payoff against receiver 2 using UCB1 strategy

uses the UCB1 learning algorithm. When both receivers are using the UCB1 learning algorithm the receivers don't play the Nash equilibrium when that is much worse than cooperating. This is why the UCB1 sometimes performs better than the dominant strategy.

Finally, figure 6 shows how the transmission rate varies when receivers use the UCB1 learning algorithm, compared to the optimal transmission rate. In this simulation we vary the actual probabilities of the two channels while keeping the rewards unchanged, and we observe that when the two channels are equally good the UCB1 algorithm obtains almost optimal transmission rate.

We now consider two specific problem instances to illustrate the performance when UCB1 is adopted by both receivers. In both cases, we assume the following parameters hold:

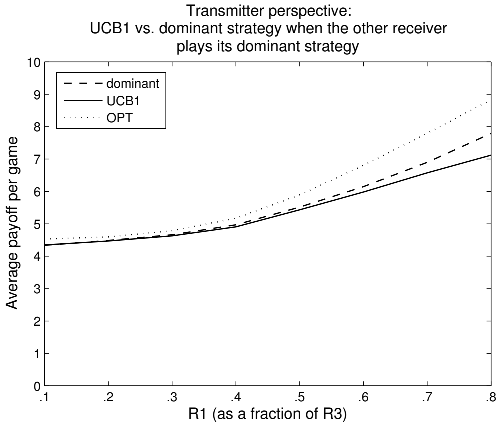

Fig. 4. Transmitter payoff when one receiver uses dominant strategy

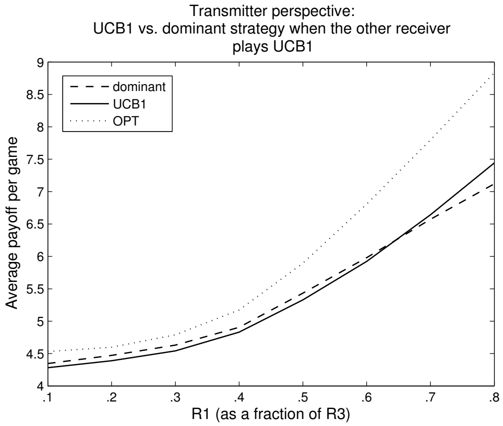

Fig. 5. Transmitter payoff when one receiver uses UCB1 strategy

$$R_1 = 40,\ R_2 = 45,\ R_3 = 60,\ C = 120,\ D = 360,\ T = 10^5,\ b_1 = b_2 = 8/9.$$

Example 1: Probability parameters $p_1 = 6/9,\ p_2 = 7.9/9$

In this case, the payoff matrix from the receivers' point of view is shown in table 10:

The optimal action (from the transmitter's perspective) is both receivers bidding low. When both receivers apply UCB1, we find that for receiver 1, the number of times out of 100,000 that it bids high is 657, the number of times it bids low is 99343; for receiver 2, the number of times it bids high is 39814, and the number of times it bids low is 60186.

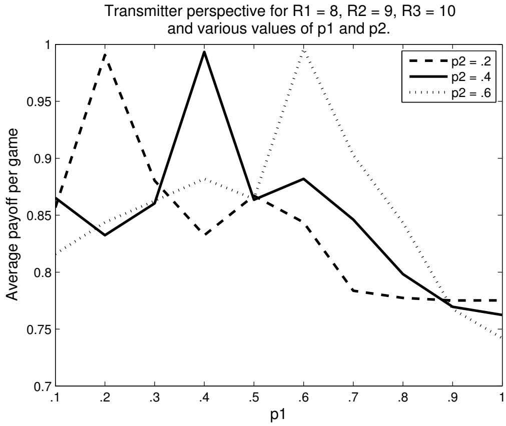

Fig. 6. Normalized transmitter payoff with respect to optimum when both play UCB1 as a function of the two channel parameters

Table 9.

Receiver 2 / Receiver 1	L	H
L	$(40, 40)$	$(0, 38)$
H	$(0, 0)$	$(-90, -4.5)$

Example 2: Probability parameters: $p_1 = 6/9,\ p_2 = 8.1/9$
The payoff matrix from the receivers' point of view is shown in table 10:

Table 10.

Receiver 2 / Receiver 1	L	H
L	$(40, 40)$	$(0, 42)$
H	$(0, 0)$	$(-90, 4.5)$

In this case, the optimal action (from the transmitter's perspective) is receiver 1 bidding low, receiver 2 bidding high. for Receiver 1, the number of times out of 100,000 that it bids high is 622, the number of times it bids low is 99378; for Receiver 2, the number of times it bids high is 62706, and the number of times it bids low is 62706.

These examples illustrate how the distributed learning algorithm is sensitive to the underlying channel parameters and learns to play the right bid over a sufficient period of time, although as expected, the regret is higher when the channel parameter is close to b_1.

7 Conclusion

We have presented and investigated a competitive rate allocation game in which two selfish receivers compete to forward the data from a transmitter to a destination for a rate-proportional fee. We showed that even if the transmitter is unaware of the stochastic parameters of the two channels, it can set penalties for failures in such a way that the two receivers' strategic bids yield a total rate that is not less than half of the best possible rate it could achieve if it had knowledge of the channel parameters. We have also studied the challenging case when the underlying channel is unknown, resulting in a game with unknown stochastic payoffs. For this game, we numerically evaluated the use of the well-known UCB1 strategy for multi-armed bandits, and showed that it gives performance close to the dominant strategies (in the case the payoffs are known) or sometimes even better. In future work, we would like to obtain more rigorous results for the game with unknown stochastic payoffs.

Acknowledgment. This research was supported by the U.S. Army Research Laboratory under the Network Science Collaborative Technology Alliance, Agreement Number W911NF-09-2-0053.

References

1. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine Learning 47, 235–256 (2002)
2. MacKenzii, A., DaSilva, L.: Game Theory for Wireless Engineers. Morgan and Claypool Publishers (2006)
3. Altman, E., Boulogne, T., El-Azouzi, R., Jimenez, T., Wynter, L.: A survey on networking games in telecommunications. Computer and Operations Research 33, 286–311 (2006)
4. Yu, W., Ginis, G., Cioffi, J.M.: Distributed Multiuser Power Control for Digital Subscriber Lines. IEEE Jour. on Selected Areas in Communications 20(5), 1105–1115 (2002)
5. Mertikopoulos, P., Belmega, E.V., Moustakas, A.L., Lasaulce, S.: Distributed Learning Policies for Power Allocation in Multiple Access Channels. IEEE Journal on Selected Areas in Communications 30(1), 96–106 (2012)
6. Laourine, A., Tong, L.: Betting on Gilbert-Elliott Channels. IEEE Transactions on Wireless Communications 50(3), 484–494 (2010)
7. Wu, Y., Krishnamachari, B.: Online Learning to Optimize Transmission over an Unknown Gilbert-Elliott Channel. WiOpt (May 2012)
8. Daskalakis, C., Frongillo, R., Papadimitriou, C.H., Pierrakos, G., Valiant, G.: On Learning Algorithms for Nash Equilibria. In: Kontogiannis, S., Koutsoupias, E., Spirakis, P.G. (eds.) SAGT 2010. LNCS, vol. 6386, pp. 114–125. Springer, Heidelberg (2010)