Distributed Learning in Noisy-Potential Games for Resource Allocation in D2D Networks

M. Shabbir Ali, P. Coucheney, M. Coupechoux

Abstract—We propose a distributed learning algorithm for the resource allocation problem in Device-to-Device (D2D) wireless networks that takes into account the throughput estimation noise. We first formulate a stochastic optimization problem with the objective of maximizing the generalized alpha fair function of the network. In order to solve it distributively, we then define and use the framework of noisy-potential games. In this context, we propose a distributed Binary Log-linear Learning Algorithm (BLLA) that converges to a Nash Equilibrium of the resource allocation game, which is also an optimal resource allocation for the optimization problem. A key enabler for the analysis of the convergence are the proposed rules for computation of resistance of trees of perturbed Markov chains. The convergence of BLLA is proved for bounded and unbounded noise, with fixed and decreasing temperature parameter. A sufficient number of estimation samples is also provided that guarantees the convergence to an optimal state. At last, we assess the performance of BLLA by extensive simulations by considering both bounded and unbounded noise cases and we show that BLLA achieves higher sum data rate compared to the state-of-the-art.

Index Terms—Distributed Learning, Potential Games, Resource Allocation, Power Control, Interference Management, D2D Networks.

1 Introduction

Ever increasing demand for higher data rates of mobile users and scarcity of wireless frequency spectrum is making efficient utilization of spectrum resources increasingly critical. Device-to-Device (D2D) networks increase the utilization of the spectrum resources by providing spatial spectrum reuse [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15]. D2D networks allow also cellular network offloading, reduction of communication costs and transmit powers among the devices and increased data rates for users [4].

In a D2D network, cellular User Equipments (UEs) communicate on allocated radio resources with a Base Station (BS) while the D2D UEs may reuse in an underlay manner these resources with or without the limited help of the BS. Although it increases cell capacity, channel reuse also creates intra-cell interference that degrades the quality of service (QoS) of the cellular UEs. Therefore, the Resource Allocation Problem (RAP) in underlay D2D networks consists in assigning channels and power to UEs so that some objective function balancing sum data rate and fairness is maximized while maintaining low interference.

RAP in D2D networks is challenging due to the lack of perfect Channel State Information (CSI) at the BS. The estimated CSI of cellular UEs can be fed back to the BS but the CSI of D2D links is difficult to obtain centrally in the network without excessive control signaling. This CSI can moreover be affected by noise so that resource allocation solutions, which ignore this effect may fail to achieve good performance. The estimation noise can arise due to several factors such as randomly varying channel gain, feedback errors, feedback delay errors, and quantization errors [16]. Our goal in this paper is thus to design a low feedback distributed solution to the RAP in D2D networks that takes into account the CSI estimation noise and achieves optimal resource allocation.

- *M. Shabbir Ali and M. Coupechoux are with LTCI, Telecom ParisTech, Université Paris-Saclay.*
Emails: mdshabbirali88@gmail.com, marceau.coupechoux@telecom-paristech.fr.
- *P. Coucheney is with UVSQ, David-Lab, France.*
Email: pierre.coucheney@uvsq.fr

Part of the work of this paper has appeared in [1], [2].

1.1 Literature Survey and Comparison

RAP in wireless networks is a standard problem and it is known to be NP-hard [17]. An extensive survey of RAP for underlay D2D networks can be found in [3]. State-of-the-art solutions are based on dynamic programming [12], graph-theoretical and heuristic solutions [13], [14], [15], based on game theory [5], [6], [7], [8], [9], [10], [11], [18], using linear programming (LP), non-linear programming (NLP), and Markov Random Field [19]. Other generic approaches include neural networks [20], simulated annealing [21], tabu search and genetic algorithms [22]. In the follwoing, we focus on approaches that are closely related to ours.

In [18], RAP is modeled as a Stackelberg game and a distributed stochastic learning algorithm is presented. It is proved that the proposed algorithm converges to a Nash Equilibrium (NE). However, the obtained NE can be suboptimal from a network-wide point of view.

In [12], the authors jointly optimize the mode selection and channel assignment in a cellular network with underlaying D2D communications in order to maximize the weighted sum rate. A dynamic programming (DP) algorithm is proposed but it is exponentially complex. Therefore, a suboptimal greedy algorithm is proposed. However, this approach relies on explicit closed form expressions of sum data rate for different channel fading scenarios.

In [13], a greedy heuristic algorithm is presented that uses the channel gain information of the links. The channel gains are sorted, cellular UEs are alloted channels with highest gain, and D2D UEs are alloted channels with least gains that causes least interference. The drawback of this algorithm lies in the excessive control signaling required to access the perfect CSI of all links.

In [14], a suboptimal graph-theoretical heuristic solution is proposed. The weighted sum signal is maximized using maximum-weighted bipartite matching and the sum interference is minimized using minimum-weighted partitioning. This approach is centralized since the BS uses the partial CSI of all the UEs.

In [15], a heuristic algorithm is proposed for joint mode selection, channel allocation and power allocation in a D2D wireless network. The measured channel gains of the different links is considered to be exact without noise. However, measurement noise affects the link throughput.

Different game-theoretic models such as non-cooperative games, coalition formation games, and auction games have been used to study radio resource allocation problems in D2D networks [5], [6], [7], [8], [9], [10], [11]. In [5], a game-theoretical reverse iterative combinatorial auction is proposed as the allocation mechanism to optimize the system sum rate over the resource sharing of both D2D and cellular modes of the users. In this auction model, the different channels compete for the D2D links. The auction mechanism is known to be strategy proof but requires also a huge overhead for the bid management.

In [7], a mechanism which is a pricing mechanism is proposed to maximize the network throughput under QoS constraint. The base stations transmit a pricing signal to the D2D users. The price increases with the interference from the D2D links. The D2D devices utilize the pricing signal to update their strategies to maximize their data rates. However, the algorithm proposed is a heuristic algorithm whose performance is only evaluated through simulations.

In [8], the uplink resource allocation problem for multiple D2D and cellular users is modeled as a coalition game. The utility function combines different modes, mutual interferences, and resource sharing policy. Convergence to a NE is proved, but it may be sub-optimal and inefficient.

Joint mode selection and spectrum sharing is modeled as a coalition formation game in [9]. A coalition formation algorithm is proposed that converges to a stable though possibly non-optimal coalition. Authors of [10] has a similar approach and evaluate the performance of their algorithm only with simulations.

In [11], author propose a contract-based game theoretic mechanism to incite users to truthfully report their channel quality. Its performance is evaluated through simulations.

1.2 Contributions

A summary of our main contributions is as follows.

- *Novel Approach:* Our approach is to learn the optimal resource allocation in a D2D wireless network using a noisy-potential game that takes into account the estimation noise of the data rate. To the best of our knowledge, this is the first time that distributed learning in noisy environments is considered for resource allocation in D2D networks. We formulate a Stochastic Optimization Problem (SOP) with the objective to maximize the generalized alpha fairness function of the users under the constraints of maximum transmit power and minimum data rate. In order to solve the SOP distributedly, we introduce the notion of noisy-potential game and translate the problem into this framework. We propose a distributed Binary Log-linear Learning Algorithm (BLLA) to achieve a NE of the game, which is also an optimal allocation of the SOP, hence an optimal NE. In contrast, originally, log-linear learning was proposed in [23] to achieve NE in potential games. We extend the results to noisy-potential games and prove that BLLA converges to an optimal NE. We apply the obtained results to obtain the optimal resource allocation in D2D networks.
- *Rules for Resistance:* The convergence of BLLA is analyzed using the resistance of trees of perturbed Markov chains, where the resistance of an edge can be imagined as a cost of playing a suboptimal action. We propose new rules for the computation of resistances that are the key enabler for the analysis of learning algorithms in noisy-potential games.
- *Convergence of BLLA in Presence of Bounded Noise:* In this case, we consider that utilities of the noisy-potential game are corrupted by a bounded (finite support) noise with any distribution. We first prove in Theorem 3 that if a single sample of noisy utility is used then BLLA converges to the global maximum under a strict constraint on the noise range. This extends the convergence results of BLLA to near-potential games using a different proof technique compared to [24]. Then to relax the constraint on the noise range we prove in Theorem 4 that a sufficient number of samples of the noisy utility has to be used to converge to the global maximum.
- *Convergence of BLLA in Presence of Unbounded Noise:* In this case, we consider that utilities of the noisy-potential game are corrupted by unbounded (infinite support) noise with any distribution. In Theorem 5, we give the sufficient number of samples of utilities required for BLLA to converge to the global maximum. A special case of Gaussian noise is presented in Corollary 2.
- *Almost Sure Convergence with Decreasing Temperature:* In Theorem 6, for both bounded and unbounded noise cases we prove the almost sure convergence of BLLA to the global maximum by decreasing the temperature. In [25], the authors has proposed to use BLLA in noisy-potential games and proved the convergence with fixed temperature. In contrast, we extend the results by studying BLLA with both fixed and decreasing temperature.
- *Simulations Results:* Extensive simulations show that BLLA achieves the maximum of the objective function in D2D networks under the constraints. We illustrate the effect of bounded and unbounded noise and show the robustness of BLLA in this uncertain environment. We study the effect of various parameters on the performance of the algorithm and show that BLLA outperforms one of the best known heuristic algorithm of the state-of-the-art [13].

The rest of the paper is organized as follows. The system

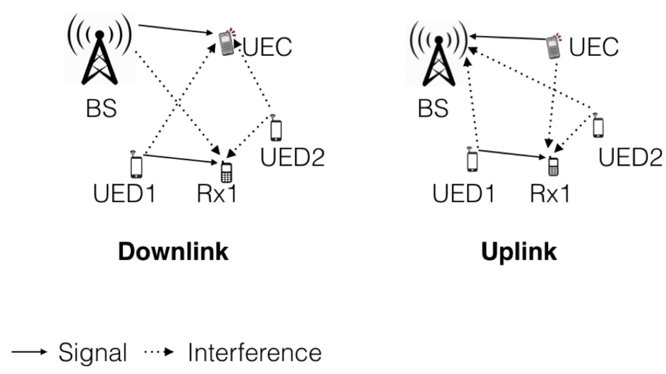


Fig. 1. D2D cellular network layout model.

model and problem formulation are described in Section 2. The noisy-potential game framework and the proposed learning algorithm are described in Section 3. Convergence results of BLLA in various cases are given in Section 4. Simulation results and conclusions are presented in Sections 5 and 6, respectively. Proofs of convergence are in Appendix.

2 D2D Cellular Network Model

In this section, we describe the D2D cellular network model as shown in Fig. 1. This figure shows downlink (DL) and uplink (UL) models.

2.1 System Model

We consider a cell served by a base station (BS)[1]; two types of UEs: (i) cellular UEs (UECs) that communicate with the BS and (ii) D2D UEs (UEDs) that communicate with other UEDs. The set of UEs is denoted as $\mathcal{D}$. We consider a set of orthogonal radio resources or (to make things simple) frequency channels $\mathbb{C}$. A transmit power set $\mathbb{P} = \{0, \dots, p_i^{\max}\}$ is considered where $p_i^{\max}$ is the maximum power allowed for UE i. The maximum transmit power of BS is denote as $p_{\text{BS}}^{\max}$. The UECs are assigned orthogonal channels by the BS, whereas UEDs reuse these channels. We assume that a UE transmits on a single channel. Let $\bar{\boldsymbol{c}} = \left[c_1, \dots, c_{|\mathcal{D}|}\right]$ and $\bar{\boldsymbol{p}} = \left[p_1, \dots, p_{|\mathcal{D}|}\right]$ denote a channel allocation vector and a power allocation vector, respectively. A UE i is allocated the channel $c_i \in \mathbb{C}$ and power $p_i \in \mathbb{P}$. The UEs that transmit on the same channel cause interference that depends on the channel gain.

2.2 Channel Model

We consider a channel model that captures the effect of path-loss, shadowing, and small-scale fading. Formally, the channel power gain g is given by [16]:

$$g = \min\left\{1, K\,|d|^{-\eta}\, e^{\beta y} \hbar\right\}, \tag{1}$$

1. The proposed Algorithm can also be applied to a multiple cell scenario as it requires only the estimated data rates at the base stations. However, for simplicity we consider a single cell model in this paper.

where $K = \left(\frac{\lambda_w}{4\pi d_0}\right)^2$, λ_w is the wavelength, d_0 is the reference distance, d is the distance between the receiver and the transmitter, $\eta \geq 2$ is the path-loss exponent, $e^{\beta y}$ is the shadow fading component, $\beta = \frac{\log 10}{10}$ and y is a Gaussian random variable of zero mean and σ_{sh}^2 variance. The small scale fading component is $\hbar$. Thermal noise power is denoted as P_0.

The data rate of a UE i depends on vectors $\bar{\boldsymbol{c}}$ and $\bar{\boldsymbol{p}}$, and is denoted as $\nu_i\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right)$. Let denote $\mathcal{D}(c)$ as the set of UEs transmitting on channel $c \in \mathbb{C}$. The normalized data rate $\nu_i\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right)$ of UE i on channel c_i is obtained using the classical Shannon capacity formula:

$$\nu_i\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) = \log_2\left(1 + \frac{p_i g_i}{\sum_{j \in \mathcal{D}(c_i)\setminus i} p_j g_{j,i} + P_0}\right). \tag{2}$$

where g_i is the channel power gain between UE i and its receiver, $g_{j,i}$ is the channel power gain between UEs i and j. Rate estimation is done at the receiver and fed back to the transmitter. Let the estimated rate be $\hat{\nu}_i\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right)$, i.e., the rate corrupted by the estimation noise.

2.3 Problem Formulation

We consider a generalized alpha fairness objective function that allows for handling the trade-off between throughput and fairness among the users of the network. Let $\pi_i \geq 0, \forall i \in \mathcal{D}$ denotes the weights that the scheduler associate with users. These weights can be used to prioritize different users; for example, cellular users can be given higher priority. The generalized alpha fair objective function is defined as:

$$\hat{\phi}_\alpha\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) = \sum_{i \in \mathcal{D}} \pi_i \hat{f}_{\alpha,i}\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right), \tag{3}$$

where $\hat{f}_{\alpha,i}\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) = f_{\alpha,i}\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) + \psi_i$, where we assume that the noise ψ_i that originates from the estimated data rate has zero mean and finite variance, and $f_{\alpha,i}$ is defined as:

$$f_{\alpha,i}\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) = \begin{cases} \log \nu_i\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) & \text{if} \quad \alpha = 1, \\ \frac{\nu_i^{1-\alpha}\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right)}{1-\alpha}, & \text{otherwise.} \end{cases} \tag{4}$$

Let $\phi_\alpha\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) = \mathbb{E}\left[\hat{\phi}_\alpha\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right)\right]$ be the expected value over all the randomness. The problem of joint channel and power allocation is formally stated as:

$$\left(\bar{\boldsymbol{c}}^*, \bar{\boldsymbol{p}}^*\right) \in \underset{\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}} \in (\mathbb{C}\times\mathbb{P})^{|\mathcal{D}|}}{\arg\max}\ \phi_\alpha\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right), \tag{5}$$

$$\text{s.t.} \quad 0 \leq p_i \leq p_i^{\max}, \quad \forall i \in \mathcal{D} \tag{6}$$

$$\nu_i\left(\bar{\boldsymbol{c}}, \bar{\boldsymbol{p}}\right) \geq r_i^{\min}, \quad \forall i \in \mathcal{D}. \tag{7}$$

The alpha fair function is well studied in the literature [26] because it captures various trade-offs between data rate and fairness. When $\pi_i = 1$ for all i, the solution of the above problem indeed yields rate maximization for $\alpha = 0$, proportional fairness for $\alpha = 1$, and max-min fairness as $\alpha \to \infty$ [26]. The constraints considered above accounts for the maximum power and minimum quality of service of the users. Particularly, the constraint (6) ensures that the maximum power is not above $p_i^{\max}$. Note that each user can have a different maximum power. Usually the D2D users have lower maximum power so as to limit the interference at the cellular users. The constraint (7) ensures that the data rate of user i is at least $r_i^{\min}$ and thus that its quality of service is met.

We seek to maximize the average value of the alpha fairness function by taking into account the estimation noise of the data rates. Hence, the above problem is a SOP [27]. In the next sections, we develop a general solution for this kind of problem. As we look for a distributed approach, we rely on a game theoretical framework.

3 Noisy Potential Game Framework

In this section, we first reformulate the considered optimization problem as a noisy-potential game and then we describe a distributed learning algorithm that leads to an optimal NE even in presence of noise.

3.1 Game Formulation

As UEs have only access to random estimations of their throughput, the game we first consider is a stochastic game, we call it a RAP game.

Definition 1. [RAP game] A RAP game is defined by the tuple $\hat{\mathcal{G}} := \left\{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \left\{\hat{U}_i\right\}_{i\in\mathcal{D}}\right\}$, where $\mathcal{D}$ is a set of UEs that are the players of the game, $\{X_i\}_{i\in\mathcal{D}}$ are action sets, $\hat{U}_i : X \rightarrow \mathcal{R}$ are utility functions, and $X = X_1 \times X_2 \times \dots X_{|\mathcal{D}|}$.

An action of a player i is a pair of channel c_i and power p_i, i.e., $a_i = (c_i, p_i)$. An action profile is defined as $a := (a_i, a_{-i})$, where $a_i \in X_i$ is the action of player i, $a_{-i} \in X_{-i}$, and X_{-i} is the action set of all the players except player i. The utility functions $\hat{U}_i$ are stochastic as they are functions of random variables.

NE is a classical solution concept for a game. In our case however, a NE may be suboptimal with respect to the SOP we are studying. Our goal is thus to find an optimal NE, i.e., an NE that corresponds to a global maximizer of the problem (5). Potential games are well suited for this purpose provided that the potential function is aligned with the objective function of the optimization problem. In potential games with finite action sets and deterministic utilities indeed, an NE always exists and there exist algorithms that are guaranteed to converge to a NE, which is also a global maximizer of the potential function [28]. In order to take into account the stochastic utilities of our RAP game, we extend this framework to noisy-potential games.

3.2 Near and Noisy-Potential Games

Before defining noisy-potential games, we make a detour through the notion of near-potential game that is required for the proofs of convergence. Based on the notion in [24], [29] a near-potential game is defined as below.

Definition 2. [ζ-potential game] A game $\tilde{\mathcal{G}} := \{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \{U_i\}_{i\in\mathcal{D}}\}$ is a near-potential game or a ζ-potential game if there is a potential function $\Phi : X \rightarrow \mathcal{R}$ such that $\forall i \in \mathcal{D}$, $\forall a_i, a_i' \in X_i$ and $\forall a_{-i} \in X_{-i}$,

$$|U_i(a_i, a_{-i}) - U_i(a_i', a_{-i}) - \Phi(a_i', a_{-i}) - \Phi(a_i, a_{-i})| \leq \zeta. \tag{8}$$

For $\zeta = 0$, it is an exact potential game [28].

Note that the parameter ζ captures the *maximum pairwise difference* (MPD) between an ζ-potential game and an exact potential game with the same potential function as in [24, Definition 2.2].

Definition 3. [Noisy-potential game] Let the expected utility of player i be denoted as $U_i = \mathbb{E}\left[\hat{U}_i\right]$. The game $\hat{\mathcal{G}} := \left\{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \left\{\hat{U}_i\right\}_{i\in\mathcal{D}}\right\}$ is a noisy-potential game if the game $\mathcal{G} := \{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \{U_i\}_{i\in\mathcal{D}}\}$ is a potential game.

We now model the RAP game as a noisy potential game with the potential function given in (3) by carefully designing the utility functions. For that, we consider the following utility functions, which represent the marginal contributions of the players to the global objective function $\phi_\alpha(a)$:

$$\begin{aligned}\hat{U}_i(a_i, a_{-i}) \quad &= \sum_{j\in\mathcal{D}(a_i)} \pi_j \hat{f}_{\alpha,j}(a_i, a_{-i}) \\ &- \sum_{k\in\mathcal{D}(a_i)\setminus i} \pi_k \hat{f}_{\alpha,k}(a_i', a_{-i}),\end{aligned} \tag{9}$$

where $\mathcal{D}(a_i) = \{j \in \mathcal{D} : a_j = a_i\}$ is the set of players using the same channel as i. The additive noise component can be separated as $\hat{U}_i(a_i, a_{-i}) = \mathbb{E}\left[\hat{U}_i(a_i, a_{-i})\right] + Z_i$, where the noise component:

$$Z_i = \sum_{j\in\mathcal{D}(a_i)} \pi_j \psi_j - \sum_{k\in\mathcal{D}(a_i)\setminus i} \pi_k \psi_k \tag{10}$$

is of zero mean and finite variance. In the case of unbounded noise, we assume that the variance of $Z_i, \forall i$ is σ^2. In the case of bounded noise, we assume that $-\frac{\ell}{2} \leq Z_i \leq \frac{\ell}{2}, \forall i$.

Note that the utility $\hat{U}_i$ may have a large variance leading to a noisy potential game with a large deviation from the exact potential game. To decrease the variance of the utility we define a sample mean of the utility function:

$$\hat{U}_i^N = \frac{1}{N} \sum_{k=1}^{N} \hat{U}_i, \tag{11}$$

Higher the value of N lower is the variance of the utility $\hat{U}_i^N$, while for practical reasons, a lower N is desired.

Proposition 1. The RAP game $\hat{\mathcal{G}}^N := \left\{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \left\{\hat{U}_i^N\right\}_{i\in\mathcal{D}}\right\}$ is a noisy potential game with potential function $\phi_\alpha(a)$.

In the rest of the paper, we consider the noisy potential RAP game $\hat{\mathcal{G}}^N$.

3.3 Learning in Noisy-Potential Game

In this subsection, we describe the proposed binary log-linear algorithm (BLLA) for learning in noisy-potential games.

The details of BLLA are described in Algorithm 1 and the steps involved are shown in Fig. 2. Time is divided in time-slots, every slot is itself divided into two phases, and each phase is made of N samples.

- Step 1: The algorithm starts with an action profile a_0 that consist of orthogonal channels for UECs and

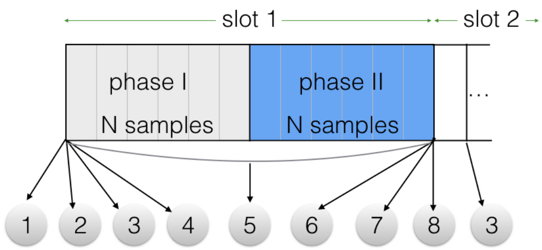


Fig. 2. Time slots, phases, and steps of BLLA are shown in this figure. Steps are represented using circles. During Step 5, measurements are performed by the UEs during the two phases on the channels $c_i(t-1)$ and $\hat{c}_i$.

lowest allowable transmit power. The UEDs are not allocated channels in the beginning. If power control is not implemented then transmit powers are set to their maximum. Then, Steps 2 to 8 repeat at every time-slot t. Steps 1 to 4 are executed at the beginning of Phase I, while steps 6 to 8 are executed at the end of Phase II and step 5 is executed during the two phases.

- Step 2: The temperature parameter τ is set. It can be fixed throughout the algorithm or be a decreasing function of t. It governs the convergence properties of the algorithm.
- Step 3: The BS randomly selects a player i and a trial action $\hat{a}_i \in X_i$ with uniform probability. This player will potentially revise its strategy, while others keep their strategies constant during the time-slot.
- Step 4: The BS asks all the players with actions $a_i(t-1)$ and $\hat{a}_i$ to estimate their data rate during the two phases and feedback the results to the BS at the end of the slot. This will be used to compute the marginal contribution of i.
- Step 5: The player i plays action $a_i(t-1) = (c_i(t-1), p_i)$ and $\hat{a}_i = (\hat{c}, \hat{p}_i)$ during Phase I and Phase II, respectively. All players on channels $c_i(t-1)$ and $\hat{c}_i$ sample their data rate during the two phases.
- Step 6: Let $\hat{f}^k_{\alpha,i}$ be the k^{th} sample of $\hat{f}_{\alpha,i}$. At the end of Phase II, all players j on $a_i(t-1)$ and $\hat{a}_i$ send $\frac{1}{N}\sum_{k=1}^{N} \hat{f}^k_{\alpha,i}\left(a_i(t-1), a_{-i}(t-1)\right)$ and $\frac{1}{N}\sum_{k=1}^{N} \hat{f}^k_{\alpha,i}\left(\hat{a}_i, a_{-i}(t-1)\right)$ to the BS, respectively.
- Step 7: The BS calculates the utility of player i according to (11) and selects an action from the set $\{a_i(t-1), \hat{a}_i\}$ according to (12). If one of the constraints is violated during Phase II then the trial action is not selected.
- Stpe 8: The BS informs player i with the selected action. This feedback requires only one bit.

Note that BLLA is distributed in nature because only a few players have to feedback to the BS.

Algorithm 1 Binary Log-linear Learning Algorithm

[Step:1] Start with an arbitrary action profile a_0.
while $t \geq 1$ **do**
  [Step:2] Set parameter $\tau(t)$.
  [Step:3] BS randomly selects a player i and a trial action $\hat{a}_i \in X_i$ with uniform probability.
  [Step:4] BS asks player i and all the players with actions $a_i(t-1)$ and $\hat{a}_i$ to estimate their sample mean data rates.
  [Step:5] Player i plays action $a_i(t-1)$ and $\hat{a}_i$ during Phase I and Phase II, respectively.
  [Step:6] At the end of Phase II, all players j on $a_i(t-1)$ and $\hat{a}_i$ send $\frac{1}{N}\sum_{k=1}^{N} \hat{f}^k_{\alpha,i}\left(a_i(t-1), a_{-i}(t-1)\right)$ and $\frac{1}{N}\sum_{k=1}^{N} \hat{f}^k_{\alpha,i}\left(\hat{a}_i, a_{-i}(t-1)\right)$ to the BS, respectively.
  if constraints (6) and (7) are satisfied for all players with action $\hat{a}_i$ **then**
    [Step:7] BS calculates $\hat{U}_i^N\left(a(t-1)\right)$, $\hat{U}_i^N\left(\hat{a}_i, a_{-i}(t-1)\right)$, and selects action $\hat{a}_i$ with probability:
$$\left(1 + e^{\Delta_i^N/\tau}\right)^{-1}, \tag{12}$$
    where $\Delta_i^N = \hat{U}_i^N\left(a(t-1)\right) - \hat{U}_i^N\left(\hat{a}_i, a_{-i}(t-1)\right)$.
  end if
  [Step:8] BS informs player i to play the selected action. All the other players repeat their previous actions, i.e., $a_{-i}(t) = a_{-i}(t-1)$.
end while

4 Convergence of the Learning Algorithm

In this section, we present the results of convergence of BLLA for both the cases of bounded and unbounded noise. For $\tau \neq 0$, BLLA generates an irreducible Markov chain over the action space of the RAP game $\hat{\mathcal{G}}^N$. As the parameter τ goes to zero, the stationary distribution concentrates on a few states whose limit probability is strictly positive. These states are called stochastically stable. It is known that for exact potential games the stochastically stable states of BLLA are the maximizers of the potential function [30]. We extend this result to noisy-potential games.

4.1 Preliminaries

We start with preliminaries and present rules required for the analysis. The dynamics of BLLA can be analyzed using the resistance of trees of a perturbed Markov chain. More details on resistance trees of Markov chains can be found in [30], [31]. A perturbed Markov process is characterized by a set $\{P^\tau\}$ of transition matrices over a state space X indexed by a parameter τ, where $\tau \in (0, \tau_h]$ is a parameter that controls the perturbation and τ_h is a constant. P^0_{ab} and P^τ_{ab} denote the transition probabilities from state a to b in the unperturbed and the perturbed Markov chains, respectively. The definition of resistance of transitions and the definition of a regular perturbed Markov process are given below [31].

Definition 4 (Resistance of transition). A perturbed Markov process $\{P^\tau\}$ is regular if it satisfies the following conditions [31]:

1. $\exists \tau_h : \forall \tau \in (0, \tau_h]$, P^τ is aperiodic and irreducible,
2. $\lim_{\tau \to 0} P^\tau_{ab}$ exists and is equal to P^0_{ab},

3. For each P^{τ}_{ab} strictly positive, there exists a non-negative number R_{ab} called the resistance of the transition such that:
$$0 < \lim_{\tau \to 0^+} e^{\frac{R_{\text{ab}}}{\tau}} P^{\tau}_{\text{ab}} < \infty. \tag{13}$$

Note that if $P^0_{ab} > 0$ then $R_{\text{ab}} = 0$.

A tree, T, rooted at a state a, is a set of $|X| - 1$ directed edges such that, from every other state a' in the state space, there is a unique directed path in the tree to a. The resistance of the directed edge $a \to b$ is the resistance of P^{τ}_{ab}, hence R_{ab}. The resistance of a rooted tree, T, is the sum of the resistances on its edges $R(T) = \sum_{a,b \in T} R_{\text{ab}}$. Let $\mathcal{T}(a)$ be defined as the set of trees rooted at the state a. The stochastic potential of the state a is defined as $\gamma(a) = \min_{T \in \mathcal{T}(a)} R(T)$.

We now define a minimum resistance tree and the stochastically stable states of a PMC [31].

Definition 5 (Minimum Resistance Tree). A minimum resistance tree is a tree that has the minimum stochastic potential, that is, any tree T that satisfies:
$$R(T) = \min_{a \in X} \gamma(a). \tag{14}$$

Definition 6 (Stochastically Stable State). Let $\{P^{\tau}\}$ be a regular perturbed Markov process, and for each $\tau > 0$, let μ^{τ} be the unique stationary distribution of P^{τ}. A state a is stochastically stable if:
$$\lim_{\tau \to 0} \mu^{\tau}(a) > 0. \tag{15}$$

Lemma *1 ([31]).* Let P^{τ} be a regular perturbation of P^0 and let μ^{τ} be its stationary distribution. Then $\lim_{\tau \to 0} \mu^{\tau} = \mu^0$ exists and μ^0 is a stationary distribution of P^0. Moreover, $\mu^0_x > 0$ if and only if $\gamma(x) \leq \gamma(y)$ for all $y \in X$.

Due to Lemma 1, the stochastically stable states of BLLA are the roots of the induced minimum resistance tree. Finding the stochastically stable states thus requires the computation of the resistance of trees. This computation is not easy in general, we thus propose in the next section a new definition of the resistance and computation rules that will be used for the analysis of BLLA.

4.2 Resistance Computation Rules

The resistance in Definition 4 can be computed in case the transition probability function can be factorised into a simple function and in case the limit in (13) can be evaluated. However, transition functions can be composite and intricate and may not always be simplified. Moreover, the limit in (13) cannot always be feasible to evaluate. For example, when $P^{\tau}_{ab} = \tau$, the resistance does not exist according to Definition 4. To overcome these limitations of Definition 4 we first give a new generalized definition of resistance that allows us to develop easy rules to compute the resistance of any positive function.

Let $o(.)$ and $\omega(.)$ denote little "o" order and little "ω" order, respectively.

Definition 7. Let f and g be two functions of τ. Then $f(\tau) \in o(g(\tau))$ if $\lim_{\tau \to \infty} \frac{f(\tau)}{g(\tau)} = 0$. And $f(\tau) \in \omega(g(\tau))$ if $\lim_{\tau \to \infty} \left|\frac{f(\tau)}{g(\tau)}\right| = \infty$.

Definition 8 (Resistance of positive function). The resistance of a strictly positive function $f(\tau)$ is $\text{Res}(f)$ if there exists a strictly positive function $g(\tau)$ such that $g \in o\left(e^{k/\tau}\right)$ and $g \in \omega\left(e^{-k/\tau}\right)$ for any $k > 0$, and:
$$\lim_{\tau \to 0} \frac{f(\tau)}{g(\tau) e^{-\frac{\text{Res}(f)}{\tau}}} = 1. \tag{16}$$

Remark Note that Definition 8 includes Definition 4, in which $g(\tau) = \kappa, 0 < \kappa < \infty$ is a constant. Now, we can evaluate the resistance of $P^{\tau}_{ab} = \tau$, i.e., $\text{Res}(\tau) = 0$.

Remark Note that (16) is equivalent to
$$f(\tau) = g(\tau) e^{-\frac{\text{Res}(f)}{\tau}} + h(\tau), \tag{17}$$
where $h(\tau) \in o\left(g(\tau) e^{-\frac{\text{Res}(f)}{\tau}}\right)$.

Remark We call $g(\tau)$ as a *sub-exponential function* if $g \in o\left(e^{k/\tau}\right)$ and $g \in \omega\left(e^{-k/\tau}\right)$ for any $k > 0$. Note that it is equivalent to $|\log g| \in o\left(\frac{1}{\tau}\right)$.

Proposition 2. Let f, f_1 and f_2 be strictly positive functions. Let $\text{Res}(f_1)$ and $\text{Res}(f_2)$ exist. Let κ be a positive constant.

I $f_1(\tau)$ is sub-exponential if and only if $\text{Res}(f_1) = 0$. In particular $\text{Res}(\kappa) = 0$.
II $\text{Res}(e^{-\kappa/\tau}) = \kappa$.
III $\text{Res}(f_1 + f_2) = \min\{\text{Res}(f_1), \text{Res}(f_2)\}$.
IV If $\text{Res}(f_1) < \text{Res}(f_2)$ then $\text{Res}(f_1 - f_2) = \text{Res}(f_1)$.
V $\text{Res}(f_1 f_2) = \text{Res}(f_1) + \text{Res}(f_2)$.
VI $\text{Res}(\frac{1}{f}) = -\text{Res}(f)$.
VII If $\forall \tau, f_1(\tau) \leq f_2(\tau)$ then $\text{Res}(f_2) \leq \text{Res}(f_1)$.
VIII If $\forall \tau, f_1(\tau) \leq f(\tau) \leq f_2(\tau)$ and if $\text{Res}(f_1) = \text{Res}(f_2)$ then $\text{Res}(f)$ exists and $\text{Res}(f) = \text{Res}(f_1)$.

Proof: See proof in [2]. □

Remark In Rule IV, if $\text{Res}(f_1) = \text{Res}(f_2)$ then we cannot compute $\text{Res}(f_1 - f_2)$ because in general the difference of sub-exponential functions may not be a sub-exponential function. For example, choose $f_1(\tau) = 1 + e^{-k/\tau}$ and $f_2(\tau) = 1$ with $k > 0$ then $\text{Res}(f_1) = \text{Res}(f_2) = 0$ but $\text{Res}(f_1 - f_2) = k$.

Remark For Rule VIII, in general if $f_1(\tau) \leq f(\tau) \leq f_2(\tau)$ and $\text{Res}(f_1) \neq \text{Res}(f_2)$ then $\text{Res}(f)$ may not exist. For example, for $f(\tau) = \lambda(\tau) f_1 + (1 - \lambda(\tau)) f_2$, $\lambda(\tau) = \frac{1}{2}\left(\cos\left(\frac{1}{\tau}\right) + 1\right)$ $\text{Res}(f)$ does not exist.

4.3 Main Results on Convergence of BLLA

4.3.1 Bounded Noise

Let ϕ^* and $\phi^{\dagger}$ be the first and second global maxima of a function ϕ.

Theorem *3.* For a noisy-potential game $\hat{\mathcal{G}} := \left\{\mathcal{D}, \{X_i\}_{i \in \mathcal{D}}, \left\{\hat{U}_i\right\}_{i \in \mathcal{D}}\right\}$ with potential ϕ and with bounded noise of interval size ℓ having finite support, the stochastically stable states of BLLA have a potential greater than $\phi^* - 2\ell(|X| - 1)$.

Proof We first show that there exists a near-potential game with MPD ℓ with the same maximum resistance as the noisy potential game $\hat{\mathcal{G}}$. Then the convergence result of BLLA in noisy-potential game is same as in near-potential game [29, Theorem 3]. See Appendix A for more details.

Corollary *1.* For a noisy-potential game $\hat{\mathcal{G}} := \left\{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \left\{\hat{U}_i\right\}_{i\in\mathcal{D}}\right\}$ with potential ϕ the stochastically stable states of BLLA have potential ϕ^* if:

$$\ell < \frac{\phi^* - \phi^\dagger}{2(|X|-1)}. \tag{18}$$

The convergence of BLLA can be obtain by using multiple samples of the noisy utilities, which is given in the below theorem.

Theorem *4.* For a noisy-potential game $\hat{\mathcal{G}}^N$ the stochastically stable states of BLLA are the global maximizers of the potential function $\phi(a)$ if the estimation noise is bounded in an interval of size ℓ and the number of estimation samples verifies:

$$N \geq 2\ell^2 \frac{\left(\log\left(\frac{4}{\xi}\right) + \frac{2}{\tau}\right)}{(1-\xi)^2\,\tau^2}, \tag{19}$$

where $0 < \xi < 1$ is free parameter and can be chosen so as to minimize the number of samples.

Proof: See Appendix C. □

4.3.2 Unbounded Noise

Theorem *5.* Let $0 < \xi < 1$. Let the estimation noise Z of the utility in (9) be unbounded with zero mean and finite moment generating function $M(\theta)$. Let $\theta^* = \arg\max_\theta \theta(1-\xi)\tau - \log(M(\theta)M(-\theta))$. For a noisy-potential game $\hat{\mathcal{G}}^N$ the stochastically stable states of BLLA are the global maximizers of the potential function $\phi(a)$ if the number of samples verifies:

$$N \geq \frac{\log\left(\frac{4}{\xi}\right) + \frac{2}{\tau}}{\log\left(\frac{e^{\theta^*(1-\xi)\tau}}{M(\theta^*)M(-\theta^*)}\right)}. \tag{20}$$

Proof: See Appendix D. □

The corollary below immediately follows in case of Gaussian noise distribution.

Corollary *2.* Let the noise have Gaussian probability distribution with zero mean and σ^2 variance. Then the stochastically stable states of BLLA are the global maximizers of the potential function if:

$$N \geq 4\sigma^2 \frac{\log\left(\frac{4}{\xi}\right) + \frac{2}{\tau}}{\tau^2(1-\xi)^2}. \tag{21}$$

A small N is desired for practical implementations, we thus choose the lowest N that satisfies Theorem 5 in simulations. In Theorem 5, we have a convergence in probability for a fixed parameter τ. In Theorem 6, we now consider the case of a decreasing parameter τ for which we obtain an almost sure convergence to an optimal state as in simulated annealing with cooling schedule [32].

Theorem *6.* Consider BLLA with a decreasing parameter $\tau(t) = 1/\log(1+t)$, and the number of samples $N(\tau)$ is given by Theorems 4 or 5. Then, BLLA converges with probability 1 to a global maximizer of the potential function.

Proof See Appendix E.

TABLE 1
Simulation parameters.

Parameter	Variable	Value
Network radius	$\mathcal{L}$	200 m
Number of orthogonal channels	$\lvert\mathbb{C}\rvert$	5
Number of power values	$\lvert\mathbb{P}\rvert$	10
Number of UECs		5
Number of UEDs		15
Maximum transmit power of BS	$p_{BS}^{\max}$	46 dBm
Maximum transmit power of UE	$p_i^{\max}$	24 dBm
System bandwidth	W	1 MHz
Additive noise power	P_0	-114 dBm
Path-loss exponent	η	3.5
Shadowing standard deviation	σ_{sh}	6 dB
Minimum rate constraint	$r^{\min}$	1 bps/Hz
Bounded noise range	ℓ	5%, 10%
Gaussian noise standard deviation	σ	5%, 10%
BLLA parameter	τ	0.01
Parameter	ζ	0.00001

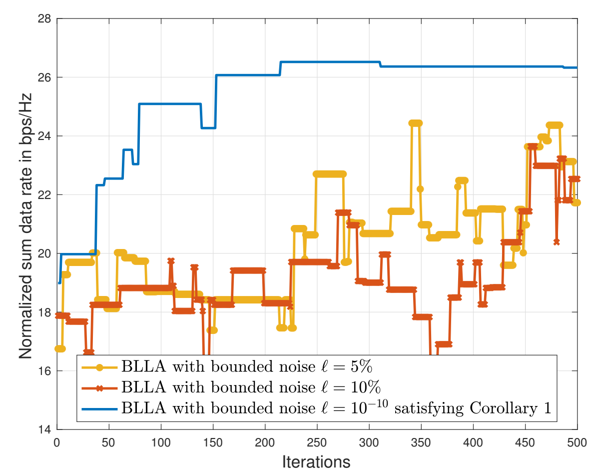


Fig. 3. Normalized sum data rate as a function of the number of BLLA iterations when $N = 1$.

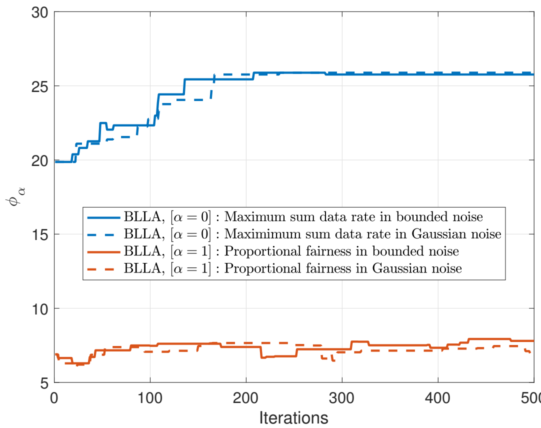
5 Simulations

In this Section, we present simulation results considering standard wireless system parameters[2] shown in Table 1. We consider the downlink of a D2D network with a BS located at the center of the region. UECs and UEDs are located around the BS. The UED receivers are located around their respective UED transmitters uniformly random over a disk of radius 20 m.

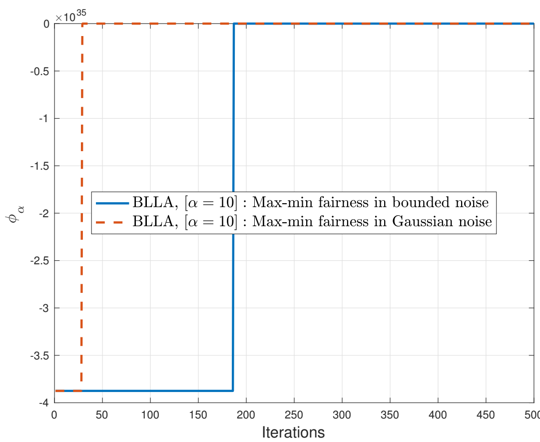
5.1 Effect of Noise on the Convergence of BLLA

In Fig. 3, we show the normalized sum data rate of the system ($\alpha = 0$, $\pi_i = 1\ \forall i$) as a function of the number

2. In all the simulations we consider the noise range ℓ and Gaussian noise variance σ^2 are normalized with maximum potential function value $\phi_{\max}$. We fix the normalized values of ℓ and σ^2 in percentage and use the running maximum value of the potential function in the simulations.



(a) $\alpha = 0$ and $\alpha = 1$.



(b) $\alpha = 10$.

Fig. 4. Objective function as a function of BLLA iterations for different values of α in bounded noise $\ell = 5\%$ and Gaussian noise $\sigma = 5\%$.

of iterations of the learning algorithm and we illustrate the result of Corollary 1. In this case, BLLA is run with a single sample ($N = 1$). When the noise range ℓ does not satisfy the condition of the Corollary ($\ell = 5$ or 10%), we observe large fluctuations of the sum data rate. On the contrary, when ℓ satisfies condition (18), BLLA converges to a stable value. As in practice the noise range cannot be controlled, this highlights the necessity to work with multiple samples.

5.2 BLLA with Sufficient Number of Samples

Fig. 4 shows the convergence of BLLA in presence of bounded noise Fig. 4a and unbounded Gaussian noise Fig. 4b for different values of α. The number of samples for bounded noise and Gaussian noise are set according to Theorem 4 and 5, respectively. We see that BLLA converges to the maximum values of ϕ_α for all different α in both cases of noise.

We now compare the performance of BLLA with that of the Algorithm presented in [13], which is a greedy heuristic algorithm. This algorithm uses the channel gain information of the links to allocate the channel to those UEDs that cause least interference. Fig. 5 shows the performance comparison of BLLA versus this heuristic algorithm for both bounded

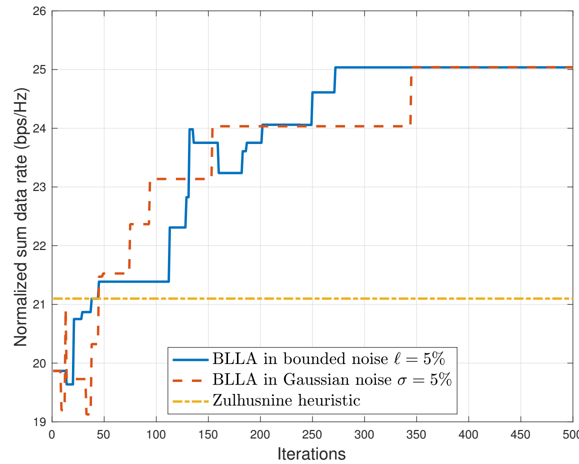


Fig. 5. Normalized sum data rate as a function of the number of BLLA iterations and comparison with the heuristic proposed in [13], for which full and exact channel information is assumed.

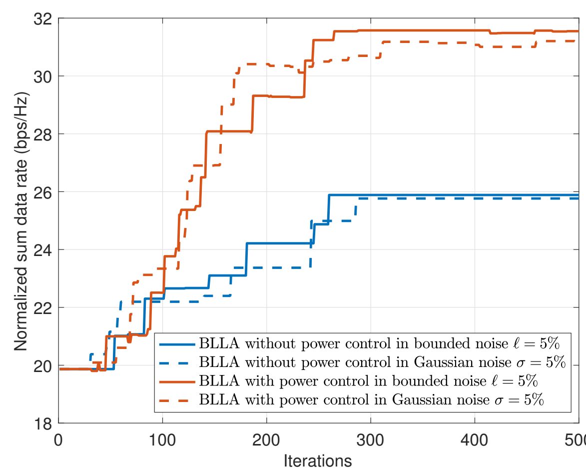


Fig. 6. Normalized sum data rate as a function of the number of BLLA iterations with and without power control.

and unbounded noise. Full and exact channel information is assumed for the heuristic algorithm. On the contrary, BLLA does not need full channel information and uses estimated channel gain feedback from the users. It is clear BLLA converges to the maximum normalized sum data rate of the network for both bounded and unbounded noise, whereas the heuristic algorithm converges to some suboptimal value.

5.3 Effect of Power Control

Fig. 6 shows the normalized sum data rate achieved by BLLA with and without power control under bounded noise and unbounded Gaussian noise. We see how higher data rates are reached by allowing an additional degree of freedom in the resource allocation. Interference is indeed minimized by exploiting channel diversity and by power allocation.

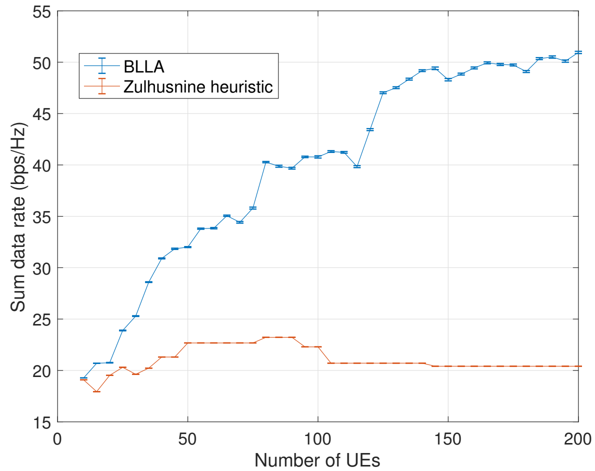


Fig. 7. Normalized sum data rate as a function of the number of D2D users (UED) with BLLA and the heuristic proposed in [13].

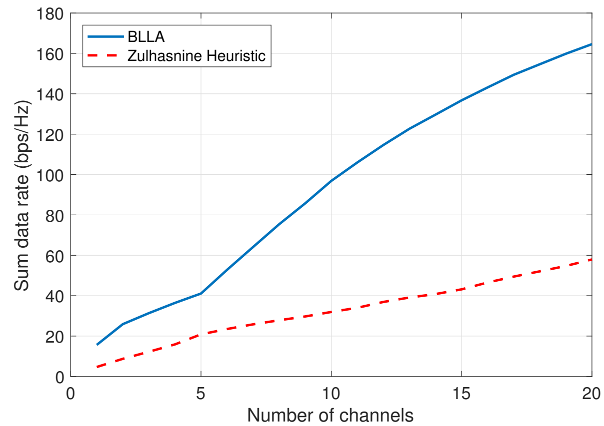


Fig. 8. Normalized sum data rate ($\alpha = 0$) as a function of the number of channels using BLLA and the heuristic proposed in [13]. The number of UECs are 5 and UEDs are 95. For BLLA, bounded noise is assumed with $\ell = 5\%$. For [13], full and exact channel information is assumed.

5.4 Effect of Varying the Number of UEs

We now study in Fig. 7 the effect of increasing the number of UEDs on the sum data rate. We first remark that whatever the number of UEDs, BLLA outperforms the heuristic proposed in [13]. Until approximately 50 UEDs, both BLLA and the heuristic see their sum data rate increasing. This is because the number of UEDs is small compared to the number of channels and the addition of new UEDs increases the channel usage without creating too much interference. After 50 UEDs however, the heuristic is not anymore able to optimally pack the UEDs in the available resources. The amount of interference is then so large that the sum data rate decreases and fluctuates around a low value according to the specific locations of the new UEDs. On the contrary, BLLA manages interference and resource allocation in such a way that the sum data rate continues to grow until a saturation point much higher than the heuristic.

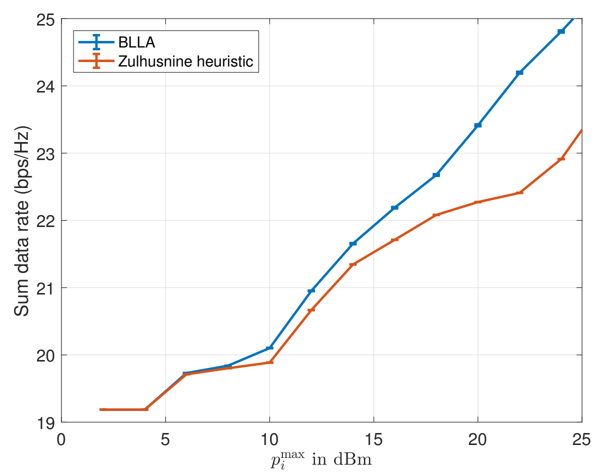


Fig. 9. Normalized sum data rate as a function of the UED power constraint with BLLA and Zulhasnine heuristic [13]. For BLLA, bounded noise with $\ell = 5\%$ is assumed. For the heuristic, full and exact channel information is assumed.

5.5 Effect of Varying the Number of Channels

Fig. 8 shows sum data rate as a function of the number of channels. We obtain a huge gain in sum data rate using BLLA when compared to Zulhusnine heuristic, even though the heuristic algorithm uses the perfect knowledge of the channel state information. This is because BLLA better manages the interference between the links given the number of channels. As the number of channels increases the interference per channel decreases. The decrease in interference per channel obtained by BLLA is much more than that of Zulhusnine heuristic.

5.6 Effect of Power and Rate Constraints

In this section, we study the effect of different constraints on the sum data rate. Note that a UED is not allocated a channel if either of the constraints is not satisfied. Fig. 9 shows the influence of the UED maximum transmit power $p_i^{\max}$ on the sum data rate. As we have assumed unbounded rate function, BLLA benefits from the increase of the UED power constraint and allows close-by UED transmitters and receivers to communicate at a high data rate. The heuristic follows this trend until a certain threshold but results afterwards in a much lower and fluctuating sum data rate. This highlights the fact that the heuristic is less efficient in managing high interference in the network. For lower interference caused due to lower maximum power of UEDs the heuristic gives approximately same data rate as BLLA. This suggests that a simple heuristic works well in lower interference and BLLA works well in both lower and higher interference scenarios.

We now study the effect of the minimum rate constraint $r^{\min}$ in Fig. 10. As the constraint increases, the sum data rate decreases because less communications can be simultaneously sustained. Whatever the constraint, BLLA outperforms the heuristic. The difference is particularly large when there is no constraint ($r^{\min} = 0$) and decreases as the constraint increases. This is explained by the fact that there are

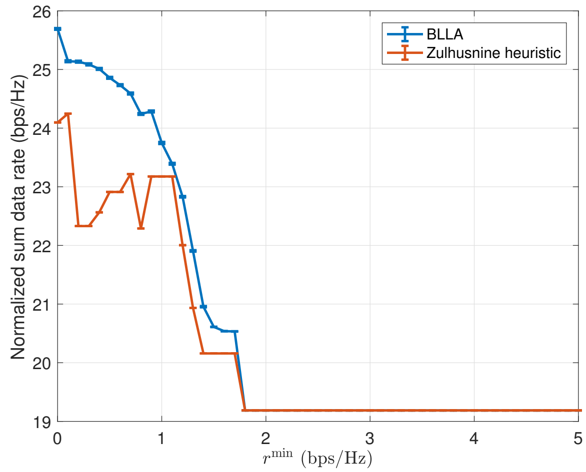


Fig. 10. Normalized sum data rate as a function of the UED rate constraint with BLLA and and Zulhasnine heuristic [13]. For BLLA, bounded noise with $\ell = 5\%$ is assumed. For the heuristic, full and exact channel information is assumed.

more transmissions and thus more interference when $r^{\min}$ is small. As BLLA is more efficient in managing interference, the difference between the two schemes increases.

6 Conclusions

In this paper, we have proposed a distributed learning algorithm for resource allocation in D2D cellular networks in presence of noisy estimates of the data rates. A stochastic optimization problem is first formulated. Then, a noisy-potential game framework is introduced to obtain a distributed solution and to capture the effect of noise. At last, we have proposed a distributed Binary Log-linear Learning Algorithm (BLLA) that achieves the optimal resource allocation, which is also an optimal Nash Equilibrium of the game. Sufficient number of samples are given that guarantee the convergence in case of bounded and unbounded noise. Proofs are supported by a new definition of the resistance of trees in perturbed Markov chains and new computation rules that can be used for any distributed learning algorithms in games. Extensive simulations have shown that BLLA outperforms the state-of-the-art in the field of resource allocation for D2D networks.

References

[1] M. S. Ali, P. Coucheney, and M. Coupechoux, "Optimal distributed channel assignment in D2D networks using learning in noisy potential games," in *Proc. IEEE INFOCOM*, May 2017.

[2] M. S. Ali, P. Coucheney, and M. Coupechoux, "Rules for computing resistance of transitions of learning algorithms in games," in *Proc. GAMENETS*, May 2017.

[3] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, 2014.

[4] M. N. Tehrani, M. Uysal, and H. Yanikomeroglu, "Device-to-device communication in 5g cellular networks: challenges, solutions, and future directions," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 86–92, 2014.

[5] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang, X. Cheng, and B. Jiao, "Efficiency resource allocation for device-to-device underlay communication systems: a reverse iterative combinatorial auction based approach," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 348–358, 2013.

[6] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 136–144, 2014.

[7] Q. Ye, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "Distributed resource allocation in device-to-device enhanced cellular networks," *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 441–454, 2015.

[8] Y. Li, D. Jin, J. Yuan, and Z. Han, "Coalitional games for resource allocation in the device-to-device uplink underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3965–3977, 2014.

[9] H. Chen, D. Wu, and Y. Cai, "Coalition formation game for green resource management in d2d communications," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1395–1398, 2014.

[10] Y. Cai, H. Chen, D. Wu, W. Yang, and L. Zhou, "A distributed resource management scheme for d2d communications based on coalition formation game," in *Proc. ICC*, pp. 355–359, IEEE, 2014.

[11] B.-Y. Huang, S.-T. Su, C.-Y. Wang, C.-W. Yeh, and H.-Y. Wei, "Resource allocation in d2d communication-a game theoretic approach," in *Proc. ICC*, pp. 483–488, IEEE, 2014.

[12] R. Wang, J. Zhang, S. Song, and K. B. Letaief, "Optimal qos-aware channel assignment in d2d communications with partial csi," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7594–7609, 2016.

[13] M. Zulhasnine, C. Huang, and A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying lte network," in *2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications*, pp. 368–375, Oct 2010.

[14] S. Maghsudi and S. Stańczak, "Hybrid centralized–distributed resource allocation for device-to-device communication underlaying cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2481–2495, 2016.

[15] C. Gao, X. Sheng, J. Tang, W. Zhang, S. Zou, and M. Guizani, "Joint mode selection, channel allocation and power assignment for green device-to-device communications," in *Proc. ICC*, pp. 178–183, 2014.

[16] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[17] M. R. Garey and D. S. Johnson, *Computers and intractability*, vol. 29. wh freeman New York, 2002.

[18] S. Dominic and L. Jacob, "Distributed resource allocation for d2d communications underlaying cellular networks in time-varying environment," *IEEE Communications Letters*, vol. 22, pp. 388–391, Feb 2018.

[19] E. Ahmed, A. Gani, S. Abolfazli, L. J. Yao, and S. U. Khan, "Channel assignment algorithms in cognitive radio networks: Taxonomy, open issues, and challenges," *IEEE Commun. Surveys & Tutorials*, vol. 18, no. 1, pp. 795–823, 2016.

[20] P. T. Chan, M. Palaniswami, and D. Everitt, "Neural network-based dynamic channel assignment for cellular mobile communication systems," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 279–288, 1994.

[21] M. Duque-Antón, D. Kunz, and B. Rüber, "Channel assignment for cellular radio using simulated annealing," *IEEE Trans. Veh. Technol.*, vol. 42, no. 1, pp. 14–21, 1993.

[22] G. K. Audhya, K. Sinha, S. C. Ghosh, and B. P. Sinha, "A survey on the channel assignment problem in wireless networks," *Wireless Commun. and Mobile Computing*, vol. 11, no. 5, pp. 583–609, 2011.

[23] L. E. Blume, "The statistical mechanics of strategic interaction," *Games and economic behavior*, vol. 5, no. 3, pp. 387–424, 1993.

[24] O. Candogan, A. Ozdaglar, and P. A. Parrilo, "Learning in near-potential games," in *IEEE Conf. Decision Control and European Control Conference*, pp. 2428–2433, 2011.

[25] D. S. Leslie and J. R. Marden, "Equilibrium selection in potential games with noisy rewards," in *Proc. IEEE Network Games, Control and Optimization (NetGCooP)*, pp. 1–4, 2011.

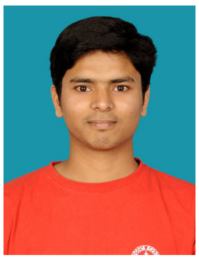
[26] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, 2000.

[27] J. C. Spall, *Introduction to stochastic search and optimization: estimation, simulation, and control*, vol. 65. John Wiley & Sons, 2005.

[28] D. Monderer and L. S. Shapley, "Potential Games," *Games and Economic Behavior*, vol. 14, pp. 124–143, May 1996.

[29] M. S. Ali, P. Coucheney, and M. Coupechoux, "Load balancing in heterogeneous networks based on distributed learning in near-potential games," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 5046–5059, 2016.

[30] J. R. Marden and J. S. Shamma, "Revisiting log-linear learning: Asynchrony, completeness and a payoff-based implementation," *Games and Economic Behaviour*, vol. 75, pp. 788–808, July 2012.
[31] H. P. Young, "The Evolution of Conventions," *Econometrica*, vol. 61, pp. 57–84, Jan. 1993.
[32] B. Hajek, "Cooling schedules for optimal annealing," *Math. of operations research*, vol. 13, no. 2, pp. 311–329, 1988.
[33] S. Anily and A. Federgruen, "Ergodicity in parametric nonstationary markov chains: An application to simulated annealing methods," *Operations Research*, vol. 35, no. 6, pp. 867–874, 1987.
[34] S. Anily and A. Federgruen, "Simulated annealing methods with general acceptance probabilities," *Journal of Applied Probability*, pp. 657–667, 1987.
[35] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, vol. 31. Springer Science & Business Media, 2013.
[36] V. Anantharam and P. Tsoucas, "A proof of the markov chain tree theorem," *Statistics & Probability Letters*, vol. 8, no. 2, pp. 189–192, 1989.



M. Shabbir Ali is a postdoctoral researcher at Orange Labs, Paris. He did his Ph.D. in Computer Science and Networks from Telecom Paristech, Paris in 2017. He received his M.Sc. degree in Electrical Communication Engineering from Indian Institute of Science (IISc), Bangalore in 2014. He did his Bachelor of Engineering degree in Electrical and Electronics Eng. from the MuffakhamJah College of Eng. and Technology (MJCET), Hyderabad, India in 2010. His research interests include game theory, machine learning, heterogeneous cellular networks, cognitive radio networks, resource allocation, load balancing, and interference modeling and management.

Pierre Coucheney is assistant professor in computer science from David-lab, University of Versailles. He graduated from the Master "Optimisation, Jeux, et ModÃ©lisation Economique" from university Paris 6, received his Ph.D. degree from the University of Grenoble in 2011, and achieved a 1-year postdoc at Inria. His thesis took place in a common lab between Inria and Alcatel-Lucent Bell Labs, and he participated in the European project Euro-NF. P.C. is interested in algorithmic aspects of game theory as well as to its applications to the optimization of networks and economics of networks .

Marceau Coupechoux has been working as an Associate Professor at Telecom ParisTech since 2005. He obtained his Masters' degree from Telecom ParisTech in 1999 and from University of Stuttgart, Germany, in 2000, and his Ph.D. from Institut Eurecom, Sophia-Antipolis, France, in 2004. From 2000 to 2005, he was with Alcatel-Lucent Bell Labs. He was a Visiting Scientist at the Indian Institute of Science, Bangalore, India, during 2011–2012. Currently, at the Computer and Network Science department of Telecom ParisTech, he is working on cellular networks, wireless networks, ad hoc networks, cognitive networks, focusing mainly on radio resource management and performance evaluation.

Appendix A
Proof of Theorem 3

Consider a noisy-potential game $\hat{\mathcal{G}} := \left\{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \left\{\hat{U}_i\right\}_{i\in\mathcal{D}}\right\}$ and the corresponding exact potential game $\mathcal{G} := \left\{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \{U_i\}_{i\in\mathcal{D}}\right\}$, where $\mathbb{E}\left[\hat{U}_i\right] = U_i, \forall i$. Utilities $\hat{U}_i$ given in (9) with bounded noise Z_i of range ℓ with finite support have probability distribution in the range $U_i - \frac{l}{2} \leq \hat{U}_i \leq U_i + \frac{l}{2}$. Let $\hat{\Delta}_i = \hat{U}_i(a) - \hat{U}_i(b)$ and $\Delta_i = U_i(a) - U_i(b)$. Then, we have

$$\Delta_i - \ell \leq \hat{\Delta}_i \leq \Delta_i - \ell. \tag{22}$$

Recall that $\hat{\Delta}_i$ has finite support. Let $\{d_1, \ldots, d_n\}$ be a set of values that $\hat{\Delta}_i$ can take. Let $m_i > 0$, denote the probability of choosing player i to revise its action. The transition probability P^τ_{ab} of BLLA in noisy-potential game $\hat{\mathcal{G}}$ is

$$P^\tau_{ab} = \frac{m_i}{|X_i|}\mathbb{E}\left[\left(1 + e^{\hat{\Delta}_i/\tau}\right)^{-1}\right], \tag{23}$$

$$P^\tau_{ab} = \frac{m_i}{|X_i|}\sum_{k=1}^{n}\left(\left(1 + e^{d_k/\tau}\right)^{-1}\Pr\left(\hat{\Delta}_i = d_k\right)\right). \tag{24}$$

Using the proposed rules, we can calculate the resistance as follows.

$$\begin{aligned}
&\mathrm{Res}\left(P^\tau_{ab}\right)\\
&= \mathrm{Res}\left(\frac{m_i}{|X_i|}\right) + \min_{d_k}\left\{\mathrm{Res}\left[\left(1 + e^{d_k/\tau}\right)^{-1}\Pr\left(\hat{\Delta}_i = d_k\right)\right]\right\}, &(25)\\
&= \min_{d_k}\left\{\mathrm{Res}\left[\left(1 + e^{d_k/\tau}\right)^{-1}\right] + \mathrm{Res}\left[\Pr\left(\hat{\Delta}_i = d_k\right)\right]\right\}\\
&+ \mathrm{Res}\left(\frac{m_i}{|X_i|}\right), &(26)\\
&= \min_{d_k}\left\{\mathrm{Res}\left(\left(1 + e^{d_k/\tau}\right)^{-1}\right)\right\}, &(27)\\
&= \min_{d_k}\left\{-\mathrm{Res}\left(1 + e^{d_k/\tau}\right)\right\}, &(28)\\
&= \min_{d_k}\left(-\min\left\{\mathrm{Res}\left(1\right), \mathrm{Res}\left(e^{d_k/\tau}\right)\right\}\right), &(29)\\
&= \min_{d_k}\left(-\min\{0, -d_k\}\right), &(30)\\
&= \min_{d_k}\left(\max\{0, d_k\}\right). &(31)
\end{aligned}$$

where (25) is obtained by using Rule III and V, (26) is by Rule V, (27) is by using $\mathrm{Res}\left(\frac{m_i}{|X_i|}\right) = 0$ and $\mathrm{Res}\left(\Pr\left(\hat{\Delta}_i = d_k\right)\right) = 0$, (28) is by Rule VI, (29) is by Rule III, (30) is by Rule I and II. Since $\min_{d_k}\left(\max\{0, d_k\}\right) \geq \max\{0, \Delta_i - \ell\} \geq \max\{0, \Delta_i\} - \ell$ and $\min_{d_k}\left(\max\{0, d_k\}\right) \leq \max\{0, \Delta_i + \ell\} \leq \max\{0, \Delta_i\} + \ell$. Therefore, we have

$$\left|\mathrm{Res}\left(P^\tau_{ab}\right) - \max\{0, \Delta_i\}\right| \leq \ell, \tag{32}$$

which is the same resistance of BLLA in an ℓ-potential game [29]. Then the convergence proof of BLLA in noisy-potential game follows from the proof of [29, Theorem 3].

Appendix B
Proof of Convergence of BLLA with Fixed τ

Our proof approach is to show that for a particular number of samples the resistance BLLA with estimated utilities is same as that of with the deterministic utilities[3]. This proof idea have often been used in the literature [25], [31].

We first compute the resistance of BLLA in a deterministic potential game using the proposed rules. Then, show that the resistance of BLLA in noisy-potential game is same as that of the exact potential game if a particular number of samples are used. Let consider the exact potential RAP game $\mathcal{G} := \left\{\mathcal{D}, \{X_i\}_{i\in\mathcal{D}}, \{U_i\}_{i\in\mathcal{D}}\right\}$, with expected utilities $U_i = \mathbb{E}\left[\hat{U}^N_i\right]$. BLLA induces a regular Markov process over the action space X of $\mathcal{G}$ [25], [29], [30]. Let denote P^τ as the transition matrix of the regular Markov process.

Let m_i denote the probability of choosing player i to revise its action. In case of deterministic utilities, the transition probability P^τ_{ab} of BLLA is

$$P^\tau_{ab} = m_i\frac{e^{\frac{1}{\tau}U_i(b)}}{e^{\frac{1}{\tau}U_i(a)} + e^{\frac{1}{\tau}U_i(b)}}. \tag{33}$$

Let $\Delta_i = U_i(a) - U_i(b)$.

Using Lemma 2, we have

$$\begin{aligned}
&\mathrm{Res}(P^\tau_{ab})\\
&= \mathrm{Res}\left(m_i\right) + \mathrm{Res}\left(\frac{e^{\frac{1}{\tau}U_i(b)}}{e^{\frac{1}{\tau}U_i(a)} + e^{\frac{1}{\tau}U_i(b)}}\right), &(34)\\
&= \mathrm{Res}\left(e^{\frac{1}{\tau}U_i(b)}\right) - \min\left\{\mathrm{Res}\left(e^{\frac{1}{\tau}U_i(a)}\right), \mathrm{Res}\left(e^{\frac{1}{\tau}U_i(b)}\right)\right\} &(35)\\
&= \Delta^+_i, &(36)
\end{aligned}$$

where $\Delta^+_i = \max\{0, \Delta_i\}$.

In the following, we show that the resistance of BLLA for the noisy-potential CAP game $\hat{\mathcal{G}}^N$ with estimated utilities $\hat{U}^N_i$ is same as in (35). For this, we need the following lemma.

Lemma *2.* Let denote $\Delta^N_i = \hat{U}^N_i(a) - \hat{U}^N_i(b)$, $\Delta_i = U_i(a) - U_i(b)$,

$$p^N_i = \mathbb{E}\left[\left(1 + e^{\Delta^N_i/\tau}\right)^{-1}\right], \tag{37}$$

$$p_i = \left(1 + e^{\Delta_i/\tau}\right)^{-1}, \tag{38}$$

and consider the event $A^\delta = \left\{\left|\Delta^N_i - \Delta_i\right| < \delta\right\}$. Then

$$\left|p^N_i - p_i\right| \leq \delta\tau^{-1}p_i + 2\Pr\left(\bar{A}^\delta\right). \tag{39}$$

Proof: Notice that the probability of transition of BLLA from action a to b in noisy-potential game $\hat{\mathcal{G}}^N$ is $p^N_i = \Pr^N\left(a \to b\right)$ given in (37) and in deterministic potential game is $p_i = \Pr\left(a \to b\right)$ given in (38). Using the law of total probability, we can write

$$p^N_i = \Pr^N\left(a \to b \,\middle|\, A^\delta\right)\Pr\left(A^\delta\right) + \Pr^N\left(a \to b \,\middle|\, \bar{A}^\delta\right)\Pr\left(\bar{A}^\delta\right), \tag{40}$$

3. In all the proofs, the considered utilities are normalized by the maximum potential $\phi_{\max}$.

$$\left|p_i^N - p_i\right| \leq \left| \Pr^N \left(a \to b \middle| A^\delta\right) - p_i\right| \Pr\left(A^\delta\right) + \left| \Pr^N \left(a \to b \middle| \bar{A}^\delta\right) - p_i\right| \Pr\left(\bar{A}^\delta\right). \quad (41)$$

It can be shown that the absolute value of the derivative of p_i with respect to Δ_i is $\tau^{-1} p_i (1 - p_i) \leq \tau^{-1} p_i$. Therefore, we have

$$\left| \Pr^N \left(a \to b \middle| A^\delta\right) - p_i\right| \leq \delta \tau^{-1} p_i. \quad (42)$$

Also, we bound $\left| \Pr^N \left(a \to b \middle| \bar{A}^\delta\right) - p_i\right| \leq 2$. Substituting, this and (42) in (41) we have (39). □

Appendix C
Proof of Theorem 4

Proof: We have $\Delta_i^N - \Delta_i = \frac{1}{N}\sum_{k=1}^N \left(\hat{U}_i(a) - U_i(a) - \left(\hat{U}_i(b) - U_i(b)\right)\right) = \frac{1}{N}\sum_{k=1}^N (Z_i(a) - Z_i(b))$. Since, the noise components $Z_i(a)$, $Z_i(b)$ have the range ℓ their sum $Z_i(a) - Z_i(b)$ has range 2ℓ. We can use Hoefding inequality for bounded independent random variables as below

$$\Pr\left(\bar{A}^\delta\right) = \Pr\left(\frac{1}{N}\sum_{i=1}^N |Z_i(a) - Z_i(b)| > \delta\right) \leq 2\exp\left(-N\frac{\delta^2}{2\ell^2}\right). \quad (43)$$

Substituting (43) in Lemma 2, we have

$$p_i\left(1 - \frac{\delta}{\tau}\right) - 4e^{-N\frac{\delta^2}{2\ell^2}} \leq p_i^N \leq p_i\left(1 + \frac{\delta}{\tau}\right) + 4e^{-N\frac{\delta^2}{2\ell^2}}. \quad (44)$$

Substituting the number of samples N from (19) and $\delta = (1-\xi)\tau$ in above, we have

$$\xi\left(p_i - e^{-\frac{2}{\tau}}\right) \leq p_i^N \leq (2-\xi) p_i + \xi e^{-\frac{2}{\tau}}. \quad (45)$$

As before, the transition probability P_{ab}^τ of BLLA is

$$\xi m_i\left(p_i - e^{-\frac{2}{\tau}}\right) \leq P_{ab}^\tau \leq (2-\xi) m_i p_i + \xi m_i e^{-\frac{2}{\tau}}. \quad (46)$$

In the following, we calculate the resistance of lower and upper bound of the above P_{ab}^τ using Lemma 2. Note that $\text{Res}(p_i) = \Delta_i^+$, $\text{Res}(e^{-\frac{2}{\tau}}) = 2$, and $\Delta_i \leq 2$. The resistance of lower bound of P_{ab}^τ is

$$\begin{aligned}
&\text{Res}\left(\xi m_i\left(p_i - e^{-\frac{2}{\tau}}\right)\right) \\
&= \text{Res}(\xi m_i) + \text{Res}\left(\left(p_i - e^{-\frac{2}{\tau}}\right)\right), && (47)\\
&= \min\left\{\text{Res}(p_i), \text{Res}\left(e^{-\frac{2}{\tau}}\right)\right\}, && (48)\\
&= \text{Res}(p_i). && (49)
\end{aligned}$$

Similarly, the resistance of upper bound of P_{ab}^τ is

$$\begin{aligned}
&\text{Res}\left((2-\xi) m_i p_i + \xi m_i e^{-\frac{2}{\tau}}\right) \\
&= \min\left\{\text{Res}((2-\xi) m_i p_i), \text{Res}\left(\xi m_i e^{-\frac{2}{\tau}}\right)\right\}, && (50)\\
&= \min\left\{\text{Res}(p_i), \text{Res}\left(e^{-\frac{2}{\tau}}\right)\right\}, && (51)\\
&= \text{Res}(p_i). && (52)
\end{aligned}$$

Since both the bounds have the same resistance, by Rule VIII the resistance of P_{ab}^τ exists and is equal to $\text{Res}(p_i)$. Therefore, the resistance of transitions of BLLA with bounded noise is same as in the case of without noise (35). □

Appendix D
Proof of Theorem 5

Proof: In this case, we use Chernoff bound to calculate $\Pr\left(\bar{A}^\delta\right)$ because of the unbounded noise as below. We have $\Delta_i^N - \Delta_i = \frac{1}{N}\sum_{k=1}^N \left(\hat{U}_i(a) - U_i(a) - \left(\hat{U}_i(b) - U_i(b)\right)\right) = \frac{1}{N}\sum_{k=1}^N (Z_i(a) - Z_i(b))$. Since, in this case the noise components $Z_i(a)$, $Z_i(b)$ are independent and having the moment generating function $M(\theta)$.

$$\begin{aligned}
\Pr\left(\bar{A}^\delta\right) &= \Pr\left(\frac{1}{N}\sum_{i=1}^N |Z_i(a) - Z_i(b)| > \delta\right), && (53)\\
&= 2\Pr\left(\frac{1}{N}\sum_{i=1}^N (Z_i(a) - Z_i(b)) > \delta\right), && (54)\\
&\leq 2\exp\left(-N\log\left(\frac{e^{\theta^*\delta}}{M(\theta^*)M(-\theta^*)}\right)\right), && (55)
\end{aligned}$$

where, (53) is obtained by assuming symmetric probability distribution of noise. However, for non-symmetric distribution a more complex expression can be obtained. Also, we used the Chernoff bound for independent and identically distributed random variables to obtain the equation (55).

Substituting (55), $\delta = (1-\xi)\tau$, and N from (20) in Lemma 2, we have

$$\xi p_i - 4\xi e^{-\frac{2}{\tau}} \leq p_i^N \leq (2-\xi) p_i + 4\xi e^{-\frac{2}{\tau}}. \quad (56)$$

Following the same steps as before, we get that the resistance of transitions of BLLA with unbounded noise is same as in the case of without noise (35). □

Appendix E
Proof of Convergence of BLLA with Decreasing $\tau(t)$

We follow the proof approach as in [33] by Anily and Federgruen. The Theorem 1 in [33] cannot be used directly since our TPF does not belong to the class of asymptotically monotone function (CAM) or rationally closed class of bounded variation (RCBV) functions (see [33, Def. 3 and 4]). We give the proof in the case of bounded noise. The proof for unbounded noise can be done similarly.

Proof of Theorem 6:

We check that the assumptions of Theorem 1 in [33] are satisfied for the proof of Theorem 6. In Lemma 5, we prove that BLLA generates a weakly ergodic non-homogeneous Markov chain. In Lemma 6, we show that the stationary distribution $\pi(\tau)$ of the homogeneous Markov chain is a bounded variation function of τ. □

We now give the Lemmas required for the above proof in the following. To simplify the notations we omit to specify the index of player i and particular transition when not needed. For a given parameter τ, we set $N(\tau)$ as in (19), and we consider $p(\tau) = p^{N(\tau)}$. Recall that $p(\tau) = \mathbb{E}\left[f(\Delta^N, \tau)\right]$ with $f(d, \tau) = \left(1 + \exp\left(\frac{d}{\tau}\right)\right)^{-1}$. We denote $\delta = \mathbb{E}[\Delta^N]$.

Lemma *3. For a given τ, function $\dfrac{\partial f(d,\tau)}{\partial \tau}$ is odd, has the sign of d, is bounded in absolute value by k/τ for some $k > 0$, and the maximum is attained (for positive value)* at the point $a^*\tau$, where $a^* > 0$ and it is independent of τ.

Proof We have the function

$$\frac{\partial f(d,\tau)}{\partial \tau} = \frac{d}{\tau^2}\frac{1}{2+\exp(d/\tau)+\exp(-d/\tau)}. \tag{57}$$

This is an odd function in d that has the sign of d. Hence, we just consider the case $d > 0$. Then

$$\frac{\partial^2 f(d,\tau)}{\partial d \partial \tau} = \frac{1}{\tau^2(2+Y+Y^{-1})^2}\left[1 - \frac{d}{\tau}\frac{Y-Y^{-1}}{2+Y+Y^{-1}}\right],$$

with $Y = \exp(d/\tau)$. This is first positive and then negative when d is positive. The maximum is reached when

$$\frac{d}{\tau}\frac{Y-Y^{-1}}{2+Y+Y^{-1}} = 1. \tag{58}$$

We claim that the maximum in d is attained for $d^* = a^*\tau$, with $a^* > 0$ a constant. Indeed, consider $d = a\tau$ with $a > 0$ in (58), which gives

$$2+\exp(a)(1-a)+\exp(-a)(1+a) = 0.$$

Consider the function $g(a) = 2+\exp(a)(1-a)+\exp(-a)(1+a)$. We have $g(0) = 4$, and g tends to $-\infty$ when a goes to ∞. Furthermore, the derivative is $-a(\exp(a)+\exp(-a))$ which is strictly negative, hence there is a unique solution a^* to the equation (58). Replacing d by $d^* = a^*\tau$ in (57) yields:

$$\frac{\partial f(d^*,\tau)}{\partial \tau} = \frac{a^*}{\tau}\frac{1}{2+\exp(a^*)+\exp(-a^*)}.$$

Hence the result follows with $k = \dfrac{a^*}{2+\exp(a^*)+\exp(-a^*)}$.

Lemma *4.* If $\delta > 0$ (resp. $\delta < 0$), then $p(\tau)$ is increasing (resp. decreasing) in the vicinity of $\tau = 0$. Furthermore, $|p'(\tau)|$ has resistance $|\delta|$.

Proof We consider $\delta > 0$. The case $\delta < 0$ is similar.

We will show that the derivative $p'(\tau)$ is positive in the vicinity of 0. Previous lemma shows that $\dfrac{\partial f(d,\tau)}{\partial \tau} \le k/\tau$. Since the constant function k/τ is integrable w.r.t. to the distribution of Δ^N, then

$$p'(\tau) = \mathbb{E}\left[\frac{\partial f(\Delta^N,\tau)}{\partial \tau}\right]. \tag{59}$$

By previous lemma, the point reaching the maximum of $\dfrac{\partial f(d,\tau)}{\partial \tau}$ is $a^*\tau$, then it goes to zero when τ goes to zero, and the function is then decreasing. Hence, for any $\epsilon < \delta$, there is τ small enough such that the minimum (resp. maximum) of the derivative on the interval $[\delta-\epsilon, \delta+\epsilon]$ is attained at $\delta+\epsilon$ (resp. $\delta-\epsilon$). Consider the event

$$A^\epsilon = \left\{\left|\Delta^N - \delta\right| < \epsilon\right\}. \tag{60}$$

In the following, we proof the Lemma by bound the $p'(\tau)$.

$$p'(\tau) = \mathbb{E}\left[\frac{\partial f(\Delta^N,\tau)}{\partial \tau}\right], \tag{61}$$

$$= \mathbb{E}\left[\frac{\partial f(\Delta^N,\tau)}{\partial \tau}|\bar{A}^\epsilon\right]\mathbb{P}[\bar{A}^\epsilon] + \mathbb{E}\left[\frac{\partial f(\Delta^N,\tau)}{\partial \tau}|A^\epsilon\right]\mathbb{P}[A^\epsilon], \tag{62}$$

$$\ge -\frac{k}{\tau}\mathbb{P}[\bar{A}^\epsilon] + \frac{\partial f(\delta+\epsilon,\tau)}{\partial \tau}\mathbb{P}[A^\epsilon], \tag{63}$$

$$\ge -\frac{k\xi}{\tau}\exp(-\frac{2}{\tau}) + 0.5\frac{\partial f(\delta+(1-\xi)\tau,\tau)}{\partial \tau}. \tag{64}$$

In the above (63) is obtained by using Lemma 3 and (64) is obtained by choosing $\epsilon = (1-\xi)\tau$ and N is given by Theorem 4 or Theorem 5. The resistance of the first term $\text{Res}\left(-\frac{k\xi}{\tau}\exp(-\frac{2}{\tau})\right) = 2$ and resistance of second term $\text{Res}\left(0.5\frac{\partial f(\delta+(1-\xi)\tau,\tau)}{\partial \tau}\right) = \delta < 2$. Therefore, the first term is negligible compared to the second term. Hence the lower bound is positive for small enough τ and the resistance of lower bound on $p'(\tau)$ is δ. Hence the derivative is lower bounded by a positive function and then is positive.

The upper bound is obtained with the following inequality:

$$p'(\tau) = \mathbb{E}\left[\frac{\partial f(\Delta^N,\tau)}{\partial \tau}|\bar{A}^\epsilon\right]\mathbb{P}[\bar{A}^\epsilon] + \mathbb{E}\left[\frac{\partial f(\Delta^N,\tau)}{\partial \tau}|A^\epsilon\right]\mathbb{P}[A^\epsilon],$$

$$\le \frac{k}{\tau}\mathbb{P}[\bar{A}^\epsilon] + \frac{\partial f(\delta-\epsilon,\tau)}{\partial \tau}\mathbb{P}[A^\epsilon],$$

$$\le \frac{k\xi}{\tau}\exp(-\frac{2}{\tau}) + 0.5\frac{\partial f(\delta-(1-\xi)\tau,\tau)}{\partial \tau}.$$

This upper bound has the resistance δ. Therefore, from Rule VIII we have that the resistance of $p'(\tau)$ is δ.

Lemma *5.* The non-homogeneous Markov chain generated by the BLLA algorithm with decreasing parameter $\tau(t) = \frac{1}{\log(1+t)}$ is weakly ergodic.

Proof The conditions of validity of Theorem 2 in [34] are checked by Lemma 4, Equation (56) and the classical choice of decreasing parameter τ. More details about weak ergodicity can also be found in [35].

If a real valued function f is defined on the interval $[a,b]$, f is differentiable and its derivative f' is Riemann integrable then its total variation $V_a^b(f)$ is

$$V_a^b(f) = \int_a^b |f'(x)|\,dx. \tag{65}$$

f is a bounded variation function if its total variation is finite i.e., $V_a^b(f) < \infty$. In particular, if the derivative f' is bounded then $V_a^b(f) < \infty$ and f is a bounded variation function.

Let $\pi(\tau)$ be the stationary distribution of the homogeneous Markov chain for a given τ.

Lemma *6.* $\pi(\tau)$ has a bounded derivative.

Proof By the Markov chain tree theorem [36] for every state $c \in X$, we have $\pi_c(\tau) = \frac{u(c)}{\sum_{d\in X} u(d)}$ where

$$u_c(\tau) = \sum_{T\in\mathcal{T}_c}\prod_{e\in T} p_e(\tau), \tag{66}$$

$p_e(\tau)$ (37) is the transition probability to a state e and $\mathcal{T}_c$ is the set of trees rooted at the state c. Then

$$|\pi_c'(\tau)| = \left|\left(\frac{u_c}{\sum_d u_d}\right)'\right| \leq \frac{|u_c'|(\sum_d u_d)}{(\sum_d u_d)^2} + \frac{u_c\,|\sum_d u_d'|}{(\sum_d u_d)^2},$$
$$\leq \frac{|u_c'|}{\sum_d u_d} + \frac{|\sum_d u_d'|}{\sum_d u_d} \leq \frac{|u_c'|}{\sum_d u_d} + \sum_d \frac{|u_d'|}{\sum_{d'} u_{d'}}.$$

Hence it suffices to show that $\dfrac{|u_c'|}{\sum_d u_d}$ is bounded for all states c.

Let $\text{Res}\,(T)$ denotes the total resistance of a tree T and $R_{\min}$ denotes the resistance of the minimal resistance tree. By using Proposition 2, we obtain

$$\text{Res}\left(\sum_d u_d(\tau)\right) = \text{Res}\left(\sum_d \sum_{T\in\mathcal{T}_d} \prod_{e\in T} p_e(\tau)\right), \tag{67}$$
$$= \min_{d\in X} \min_{T\in\mathcal{T}_d} \text{Res}\left(\prod_{e\in T} p_e(\tau)\right), \tag{68}$$
$$= \min_{d\in X, T\in\mathcal{T}_d} \text{Res}\,(T) = R_{\min}. \tag{69}$$

The derivative of transition probability $u_c'(\tau)$ is obtained using (66) as

$$u_c'(\tau) = \sum_{T\in\mathcal{T}_c} \sum_{e\in T} p_e'(\tau) \prod_{d\in T/e} p_d(\tau). \tag{70}$$

The resistance of $u_c'(\tau)$ is

$$\text{Res}\,(u_c') = \min_{T\in\mathcal{T}_c} \min_{e\in T} \left[\text{Res}\,(p_e') + \sum_{d\in T/e} \text{Res}\,(p_d(\tau))\right]. \tag{71}$$

By Lemma 4, $\text{Res}\,(p_e')$ is $|\delta|$. If $\delta > 0$ then $\text{Res}\,(p_e) = \delta$, otherwise $\text{Res}\,(p_e) = 0$, which corresponds to the best response transition. Hence, $\text{Res}\,(p_e') \geq \text{Res}\,(p_e)$ and $\text{Res}\,(p_e')$ is strictly greater than $\text{Res}\,(p_e)$ if $\delta < 0$. In Lemma 7, we prove that the minimal resistance tree must contain a transition with null resistance (which corresponds to the best response). Hence, the state c at which $R_{\min}$ is reached contains at least one transition with $\delta \leq 0$. Therefore, $\text{Res}\,(u_c') > R_{\min}$. Using Proposition 2, we have $\text{Res}\left(\frac{|u_c'|}{\sum_d u_d}\right) = \text{Res}\,(u_c'|) - R_{\min} > 0$. Hence, $\frac{|u_c'|}{\sum_d u_d} \to 0$ as τ goes to zero for all states c. This finally shows that the derivative $|\pi_c'(\tau)|$ is bounded.

Lemma *7.* A minimum resistance tree must contain a transition with zero resistance.

Proof Assume that a minimum resistance tree $T_{\min}$ have all the transitions with non-zero resistance. Let the root of this tree be a state s and let there be a transition from another state s' to s. Let $R_{s'\to s}$ be a non-zero resistance of this transition. Note that the resistance of reverse transition $R_{s\to s'} = 0$ because it corresponds to the best response transition. Construct a new tree T rooted at state s' by adding the transition $s \to s'$ and removing the transition $s' \to s$. The resistance of the tree T is

$$R_T = R_{T_{\min}} - R_{s'\to s} + R_{s\to s'} < R_{T_{\min}}. \tag{72}$$

We arrive at a contradiction. Therefore, a minimum resistance tree must contain a transition with null resistance.