Reinforcement Learning Algorithm for Load Balancing in Self-Organizing Networks

Mohammed S. Ali[1], Pierre Coucheney[2], and Marceau Coupechoux[1]

[1] *LTCI/LINCS, Telecom Paris, Institut Polytechnique de Paris, Palaiseau, France*
[2] *David-lab, UVSQ, Versailles, France*

Introduction

Due to the increase of traffic and data rate requirements, 5G consists of many more base stations (BS) than previous generation networks: traditional macro cells are coexisting with many small cells in dense urban areas. With the network densification, conventional cellular networks are thus becoming heterogeneous. In this context, it is crucial for operators that want to reduce their operation expenditures to implement self-organized functions for their network. The goal is to take advantage of the huge amount of data generated by the network in order to predict potential issues, plan solutions, take decisions, and optimize in a fully autonomous way. The 3GPP standards provide architectures and protocols to implement Self-Organized Network (SON) functions; they, however, do not define SON algorithms. In this article, we thus propose a reinforcement learning approach to solve the specific case of load balancing in 5G heterogeneous networks.

Heterogeneous networks consist of macro BSs and small BSs that transmit with high and low power, respectively. Conventional user-association rule is such that the users select a BS that provides the highest received power, so that the macro BSs associate with more users. The resulting imbalance between BSs traffic creates overload situations at macro BSs, whereas some resources are underutilized in small BSs. Load balancing therefore aims at a better user association so that network resources are efficiently utilized, and the load is evenly shared among the cells. Two important Intercell Interference Coordination (ICIC) parameters are key tools for load balancing algorithms, namely the Cell Range Expansion (CRE) and the Almost Blank Subframes (ABS). The former is a bias applied to small cell received power and used at user association to offload macro BSs. The latter is a ratio of subframes left without transmission at the macro BSs to control the interference on small cells.

Our approach to solve the load balancing problem is based on game theory and reinforcement learning, and leverages CRE and ABS. More precisely, our idea is to distributively minimize an α–fairness objective function using distributed learning algorithms in near-potential games with load and outage constraints. First, we model the load balancing problem as a nonconvex constrained optimization. Then, we adapt log-linear

Wiley 5G Ref. Edited by Rahim Tafazolli, Chin-Liang Wang and Periklis Chatzimisios.

DOI: 10.1002/9781119471509.w5GRef137

learning algorithms (LLLA) to outage and load constraints, and we achieve the global minimum of the objective function or an approximation of it. By running extensive simulations, we show that the proposed algorithms converge within few hundreds of iterations to the global minimum. We also show that outage can be controlled without ABS, but at the price of undermining the interest of using CRE technique. The introduction of ABS allows for low outage together with better load balancing.

Self-Organizing Networks

Definitions

SON is a set of functions that allow the cellular network to self-organize in a fully autonomous way, i.e. without human intervention. It is a central feature to make 5G networks more efficient in 3GPP Release 16. It is also crucial for mobile network operators that intend to reduce their operation expenditures. Indeed, due to the increasing densification of the network, the time and cost of deployment of new network elements should be reduced. The complexity of the network scales exponentially with the size of the network, so that traditional and manual techniques to manage and optimize these networks become impractical. Therefore, SON aims to enhance automation capabilities of the network and to reduce human dependencies to increase the performance of the network.

We usually classify SON functions in three groups: self-configuration, self-optimization, and self-healing (Moysen and Giupponi 2018; Ramiro and Hamied 2011). Self-configuration is related to the plug-and-play capability of the network elements and interfaces. The preparation, installation, authentication, and sending status report are included in this class of SON. Self-optimization aims at improving the performance of the network in terms of coverage, capacity, handover, or interference. Adaptive network optimization is a key requirement of SON formulated by Next-Generation Mobile Networks (NGMN) (Alliance 2006). At last, self-healing includes keeping the network operational and tracking of events potentially harmful for the network and automatically acting to avoid any such kind of events.

3GPP has defined in 3GPP (2019) three different types of SON architectures, depending on the location of the SON function. (i) In the centralized SON (C-SON) solution, the SON function is located in the management system. It makes use of the information feedback from the Network Functions (NFs) and runs centralized policies to configure network parameters. (ii) When SON is distributed SON (D-SON), the SON function is localized in the NFs. The NFs retrieve data, analyze them, and take decisions. The management and the control of the D-SON functions is performed in the central management system. (iii) Hybrid SON (H-SON) is a combination of centralized and D-SON. The SON function is partially located in the management system and in the NFs and all work in a coordinated way. H-SON results from the trade-off of available centralized information and desired network performance.

SON benefits from the Services-Based Architecture introduced in 5G (3GPP 2018a), where NFs enable other authorized NFs to access their services. In particular, the Network Data Analytics Function (NWDAF) may retrieve data from management and NF and provide data analytics to SON functions (3GPP 2018b).

SON Functions in 3GPP

Several SON functions have been introduced by 3GPP in its different releases (Jorguseski et al. 2014):

1) Self-configuration: standard parameters of new base station can be set using 3GPP automatic radio configuration data-handling function (ARCF), which may include preconfigured parameters such as cell identity, base station neighbors, antenna configuration, and transmit power level.
2) Automated Neighbor Relation (ANR): the ANR function manages the cell neighbor relations. It retrieves the neighbor information from the User Equipments (UEs) and builds the neighbor set of the BS.
3) Automatic cell identity management: this feature makes use of the reporting of the UEs to manage the physical cell identity of the cell to avoid any confusion and collision.
4) Random access optimization: Random access parameters are subject to optimization. The BS can assist the access by acquiring the access attempt report from the UEs.
5) Mobility robust management: this feature manages the mobility connection of the UEs. It minimizes the connection failure rate due to mobility and also avoids unnecessary handovers.
6) Mobility load balancing: this feature manages the uneven load distribution across the BSs by offloading traffic to neighboring BSs.
7) Energy saving: this feature is important for cost saving by reducing unnecessary usage of the equipment.
8) Minimization of drive tests: acquiring radio environment information of the BS is costly and time-consuming. With this function, the reports from UEs can be used to better assess the radio environment, especially in indoor.

Load Balancing in the Literature

Load balancing has been extensively studied in the literature using various approaches. An overview can be found in Andrews et al. (2014), Tonguz and Yanmaz (2008). These can be broadly classified as centralized, e.g. in Son et al. (2009), Rengarajan and De Veciana (2011), Stevens-Navarro et al. (2008), Elayoubi et al. (2010), and decentralized optimization approaches, see, e.g. Kim et al. (2012), Xu et al. (2010), Niyato and Hossain (2009), Aryafar et al. (2013), Kudo and Ohtsuki (2013). However, centralized solutions are computationally extensive, require huge information exchange overhead, and are thus not scalable. To overcome these limitations, decentralized approaches have been proposed. There are several decentralized proposals that are user centric, i.e. such that the users decide to which BS they associate, see, e.g. Kim et al. (2012), Kudo and Ohtsuki (2013). However, 3GPP cellular networks are network centric, in the sense that this is the network that controls the load balancing.

In a network-centric approach, BSs take decisions and users follow a predefined association rule called CRE. With CRE, users associate with a BS that provides the maximum biased received power. A CRE bias is broadcast by every BS and is typically higher for small BSs than for macro BSs. This results in an increase of the small cell coverage and thereby of the number of users associated to them. CRE technique has the drawback of increasing outage probability at the cell edge (Damnjanovic et al. 2011); it is, therefore, often deployed in conjunction with ABS at the macro BS (Oh and Han 2012). During these subframes that represent a fixed ratio of the radio frame, macro BSs drastically lower their transmit power, so that small BSs cell edge users can experience less interference, when scheduled during these periods.

The challenge we intend to tackle here is to jointly determine in a distributed way the optimal CRE bias values and ABS ratios for a required optimal performance of the network. Several papers try to achieve a similar goal. Early papers Wang et al. (2011), Kikuchi and Otsuka (2012), Al-Rawi (2012) propose heuristics without any goal of achieving some kind of optimality. In Ye et al. (2013), the authors formulate an integer programming problem but ignore ABS and related trade-offs. In Tall et al. (2014), two independent algorithms are presented for optimal CRE bias and ABS ratio, respectively. Optimal parameters are obtained for a given number of active users, and the outage constraint is ignored. In Liu et al. (2015), Liu et al. model the system as a potential game and use best response algorithm to reach a Nash Equilibrium (NE) that is also a local optimizer of a proportional fairness objective function. A static *full buffer* traffic model is assumed, whereas a dynamic model is more realistic.

System Model

Network Model

We consider the downlink of a cellular network consisting of a set $\mathcal{B}_e$ of macro BSs with transmit power P_{macro} and a set $\mathcal{B}_s$ of small BSs with transmit power P_{small} in a two-dimensional region $\mathcal{L}$. The set of all stations is denoted $\mathcal{S} \triangleq \mathcal{B}_e \cup \mathcal{B}_s$. There are special subframes called ABS, during which a macro BS transmits with reduced power P_{ABS}. The proportion of ABS subframes is denoted $\theta_i \in [0; 1]$ for BS i. Let $\overline{\boldsymbol{\theta}} = [\theta_1, \theta_2, \dots, \theta_{|\mathcal{S}|}]$ be the ABS ratio vector. We assume that small BSs do not implement ABS technique, i.e. $\theta_i = 0$ for $i \in \mathcal{B}_s$. Every small BS i maintains a parameter $c_i \in [1; c_{\max}]$ called CRE bias. The CRE bias vector is denoted $\overline{\boldsymbol{c}} = [c_1, c_2, \dots, c_{|\mathcal{S}|}]$. The CRE biases for macro BSs are fixed to unity, i.e. $c_k = 1, \forall k \in \mathcal{B}_e$.

Channel Model

The received power at location x from BS i is $P_i g_i(x)$, where P_i is the transmit power and $g_i(x)$ is the channel gain, which captures the effect of path loss and shadowing.

Formally, the channel gain model considered is Goldsmith (2005)

$$g_i(x) = \min\{1, K|x - x_i|^{-\eta} e^{\beta y_i(x)}\} \tag{1}$$

where $K = \left(\frac{\lambda_w}{4\pi d_0}\right)^2$, λ_w is the wavelength, d_0 is the reference distance, x_i is the location of the BS i, $\eta \geq 2$ is the path-loss exponent, and $e^{\beta y_i(x)}$ is the shadowing component, where $\beta = \frac{\log 10}{10}$ and $y_i(x)$ is a realization of Gaussian random process of zero mean and covariance function $C_{y_i}(\Delta x)$ (Gudmundson 1991):

$$C_{y_i}(\Delta x) = \sigma_{sh}^2 e^{-\frac{\Delta x}{D_c}} \tag{2}$$

where σ_{sh}^2 is the variance, Δx is the displacement, and D_c is the decorrelation distance (Goldsmith 2005). A constant cross-correlation between the $y_i(x)$ and $y_j(x)$ is considered as in Fraile et al. (2008).

Let M be the number of subframes in a given radio frame. For every allowed value of θ, we assume that there is a fixed ABS pattern $\Upsilon(\theta)$, i.e. a set of subframes during which a BS transmits at lower power. Notice that $\Upsilon(0) = \emptyset$. Then, the SINR $\gamma_i^f(x, \overline{\boldsymbol{\theta}})$ of a user at location x in a subframe f is given as

$$\gamma_i^f(x, \overline{\boldsymbol{\theta}}) = \frac{P_i^f(\theta_i) g_i(x)}{\sum\limits_{j \in \mathcal{S}} P_j^f(\theta_j) g_j(x) + N_0} \tag{3}$$

where

$$P_i^f(\theta_i) = \begin{cases} P_{\mathrm{ABS}} & \text{if } \; f \in \Upsilon(\theta_i), \; i \in \mathcal{B}_{\mathrm{e}} \\ P_{\mathrm{macro}} & \text{if } \; f \notin \Upsilon(\theta_i), \; i \in \mathcal{B}_{\mathrm{e}} \\ P_{\mathrm{small}} & \text{otherwise} \end{cases} \tag{4}$$

and $N_0 = -174 + 10 \log W$ is the thermal noise power in dBm, and W is system bandwidth in Hz.

CRE User-Association Rule

According to the CRE rule, a user located at x is associated to the BS i that provides the highest biased received power. The set of locations $\mathcal{D}_i(\overline{\boldsymbol{c}})$ associated to BS i is defined as

$$\mathcal{D}_i(\overline{\boldsymbol{c}}) = \{x | \forall j \in \mathcal{S}, P_i g_i(x) c_i \geq P_j g_j(x) c_j\} \tag{5}$$

where $P_i = P_{\mathrm{macro}}$ if $i \in \mathcal{B}_{\mathrm{e}}$, and P_{small} otherwise.

Physical Data Rate

The physical data rate received by a user at x in a subframe f when it is served by BS i is denoted $\tilde{v}_i^f(x, \overline{\boldsymbol{\theta}})$. The user average data rate over a radio frame is $v_i(x, \overline{\boldsymbol{\theta}}) = \frac{1}{M} \sum_{f=1}^{M} \tilde{v}_i^f(x, \overline{\boldsymbol{\theta}})$. This should be understood as the throughput achievable by the user alone in its cell. The function $\tilde{v}_i^f(x, \overline{\boldsymbol{\theta}})$ is a nonnegative and nondecreasing function of the SINR $\gamma_i^f(x, \overline{\boldsymbol{\theta}})$. For SINR below minimum threshold $\gamma_{\min}$, the user is not served, and $\tilde{v}_i^f(x, \overline{\boldsymbol{\theta}}) = 0$.

Traffic Model

Users are assumed to arrive in the system according to a spatial random process, download a file of random size, and leave the system when the download is over. This is referred to as elastic traffic. All users are scheduled in all subframes. At location x, the arrival rate is denoted $\lambda(x)$ [arrivals s ^ ?1 m$^?$2] and the average file size is $1/\mu(x)$ [bits]. Following Kim et al. (2012), we model every BS i as an M/G/1/PS queue of load:

$$\rho_i(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}}) = \int_{\mathcal{D}_i(\overline{\boldsymbol{c}})} \frac{\lambda(x)}{\mu(x) v_i(x, \overline{\boldsymbol{\theta}})} 1_{\{\max_f \gamma_i^f(x, \overline{\boldsymbol{\theta}}) \geq \gamma_{\min}\}} \mathrm{d}x \tag{6}$$

BS i is stable if and only if $0 \leq \rho_i < 1$. In this work, only stable network configurations are considered.

Outage Probability is defined as the fraction of users that are not served. Recall that a user is not served if its SINR is below minimum threshold $\gamma_{\min}$. Formally, the outage probability O_i observed by BS i is given by

$$O_i(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}}) = \frac{\int_{D_i(\overline{\boldsymbol{c}})} \lambda(x) 1_{\{\max_f \gamma_i^f(x, \overline{\boldsymbol{\theta}}) < \gamma_{\min}\}} \mathrm{d}x}{\int_{D_i(\overline{\boldsymbol{c}})} \lambda(x) \mathrm{d}x} \tag{7}$$

In this definition, as soon as there is at least one subframe during which the SINR is above the threshold, the user is supposed to be served.

Problem Formulation and Objective Function

Following Kim et al. (2012), we intend to minimize an $\alpha-$fairness function $\phi_\alpha(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})$ over a feasible set $\mathcal{F}$, which are defined as

$$\phi_\alpha(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}}) = \begin{cases} \sum_{i \in \mathcal{S}} \frac{(1-\rho_i(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}}))^{1-\alpha}}{\alpha - 1}, & \alpha \geq 0, \alpha \neq 1 \\ -\sum_{i \in \mathcal{S}} \log(1 - \rho_i(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})), & \alpha = 1 \end{cases} \tag{8}$$

$$\mathcal{F} = \{\{\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}}\} | \forall i \in \mathcal{S}, \rho_i(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}}) < 1, O_i(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}}) < \overline{O}_i, \} \tag{9}$$

where $\overline{O}_i$ is the maximum outage probability for BS i. The function $\phi_\alpha(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})$ is in general nonconvex, and even if it is convex, the set $\mathcal{F}$ is nonconvex because $\overline{\boldsymbol{c}}$ takes discrete values. The function $\phi_\alpha(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})$ captures various aspects of fairness and performance for the network, depending on the choice of α.

($\alpha = 0$) **Min-sum-load policy:** Minimizing $\phi_0(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})$ minimizes the sum of BS loads in general. In the particular case, where $\overline{\boldsymbol{\theta}} = 0$, it results in a rate-optimal policy. ($\alpha = 1$) **Proportional fair policy:** Minimizing $\phi_1(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})$ is equivalent to achieving proportional fairness between BSs (Mo and Walrand 2000). ($\alpha = 2$) **Delay-optimal policy:** It can be shown that minimizing $\phi_2(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})$ is equivalent to minimizing the average delay of the network. ($\alpha \to \infty$) **Minmax policy:** As $\alpha \to \infty$, the minimizer of $\phi_\alpha(\overline{\boldsymbol{c}}, \overline{\boldsymbol{\theta}})$ tends to the min–max load vector (Mo and Walrand 2000; Bonald and Massoulié 2001; Kim et al. 2012).

In Figure 1, we show an example of $\mathcal{F}$ set obtained with two BSs having different transmit powers located on a two-dimensional region. It is clear from the figure that even if the CRE set were continuous, $\mathcal{F}$ would not be convex. We also show the optimal loads obtained for different α values. All the optimal load points are located on the Pareto frontier. The point for $\alpha \geq 200$ in Figure 1 is the min–max load point because a point of equal coordinates on the Pareto frontier is the min–max point.

Near-Potential Game Framework

In this section, we present an approach using near-potential game framework for distributed optimization of the objective function.[1] We do not intend to describe and analyze the selfish nature of BSs that aim to minimize their costs. Rather, our goal is to achieve the global objective of load balancing by prescribing a cost function to the BSs. For this context, potential games provide a good

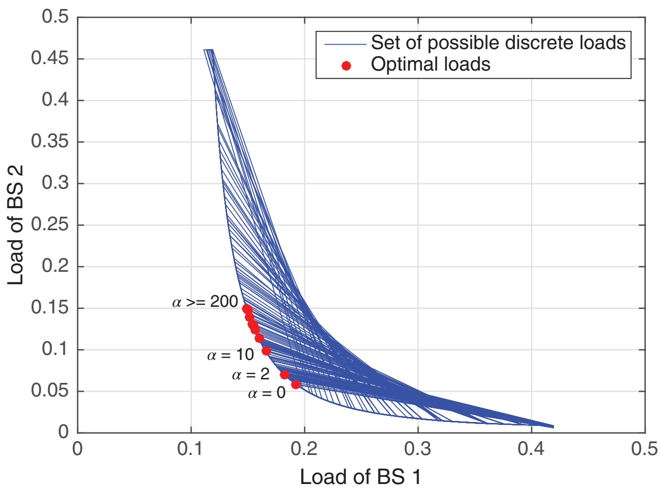

Figure 1 Feasible set $\mathcal{F}$ for two BSs.

framework because players of such a game distributively optimize a potential function.

We model the problem as a *user-association game*, where the BSs are players and allowed CRE bias, and ABS ratio values are their strategies. BSs play the user-association game with the objective of minimizing their costs. An ϵ-NE of the game is reached when no player can benefit more than ϵ by changing its strategy unilaterally.

Definition 1 [ϵ-Nash equilibrium] Let $\mathcal{G} = \{\mathcal{S}, \{X_i\}_{i\in\mathcal{S}}, \{U_i\}_{i\in\mathcal{S}}\}$ be a game, where $\mathcal{S}$ is the set of players, $\{X_i\}_{i\in\mathcal{S}}$ are the strategy sets, and $\{U_i\}_{i\in\mathcal{S}}$ are the cost functions. Let a_i be a strategy profile of player i and a_{-i} be a strategy profile of all players except for player i. A strategy profile (a_i^*, a_{-i}^*) is an ϵ-NE if

$$U_i(a_i^*, a_{-i}^*) - U_i(a_i, a_{-i}^*) \leq \epsilon, \quad \forall a_i \in X_i, \forall i \in \mathcal{S} \tag{10}$$

If $\epsilon = 0$, then it is a pure Nash equilibrium (PNE).

Based on the notion developed in Candogan et al. (2011), we now define the ξ-potential game.

Definition 2 [ξ-potential game] A game $\mathcal{G} = \{\mathcal{S}, \{X_i\}_{i\in\mathcal{S}}, \{U_i\}_{i\in\mathcal{S}}\}$ is an ξ-potential game if there is a potential function $h : X \to \mathcal{R}$ such that $\forall i \in \mathcal{S}, \forall a_i, a_i^{'} \in X_i$, and $\forall a_{-i} \in X_{-i}$,

$$|U_i(a_i, a_{-i}) - U_i(a_i', a_{-i}) + h(a_i', a_{-i}) - h(a_i, a_{-i})| \leq \xi \tag{11}$$

For $\xi = 0$, it is an exact potential game (Monderer and Shapley 1996).

An exact potential game has at least one PNE, and the local optimizers of the potential function are PNEs (Monderer and Shapley 1996). In the following lemma, we provide the relationship between the PNEs of a potential game and a near-potential game with the same potential.

Lemma 1 Let $\mathcal{G} = \{\mathcal{S}, \{X_i\}_{i\in\mathcal{S}}, \{U_i\}_{i\in\mathcal{S}}\}$ and $\mathcal{G}' = \{\mathcal{S}, \{X_i\}_{i\in\mathcal{S}}, \{U_i'\}_{i\in\mathcal{S}}\}$ be an exact potential game and an ϵ-potential game, respectively, sharing a common potential function. If a^* is a PNE for $\mathcal{G}$, then it is an ξ-NE for $\mathcal{G}'$.

In our problem, we seek that the objective function (8) is turned is turned into a potential function of the user-association game. The issue is in designing the cost functions of the BSs to obtain an ξ-potential game, where ξ represents a trade-off between the quality of the solution and the distributed nature of the algorithm. We consider a simple cost structure for BS i, which takes into account only the effects of its neighbors. The cost functions of the individual BSs are defined as

$$U_i^{\varpi}(a_i, a_{-i}) = \sum_{j\in N_i^{\varpi}} \frac{(1-\rho_j(a_i, a_{-i}))^{1-\alpha}}{\alpha - 1} \tag{12}$$

where N_i^{ϖ} is the neighborhood of BS i, ϖ is a parameter to control its size, and $\rho_j(a_i, a_{-i})$ is the load of BS j given in (6). With this cost function, we now formally define the user-association game.

Definition 3 [User-Association Game] It is defined by the tuple $\Gamma^{\varpi} = \{\mathcal{S}, \{X_i\}_{i\in\mathcal{S}}, \{U_i^{\varpi}\}_{i\in\mathcal{S}}\}$, where $\mathcal{S}$ is a set of BSs, X_i is a set of strategies of BS i, and U_i^{ϖ} is given in (12). Strategy set X_i is a discrete set of CRE bias values for small BS $i \in \mathcal{B}_{\mathrm{s}}$, and X_i is a discrete set of ABS ratios for macro BS $i \in \mathcal{B}_{\mathrm{e}}$.

In the following sections, we first show the construction of N_i^{ϖ}, and then in Proposition 1 we prove that the user-association game Γ^{ϖ} is an ξ-potential game.

Base Station Neighborhood

We start with the definition of the neighbor set N_x of small BSs at location x:

$$N_x = \{j \in \mathcal{S} \mid \max_{c_j} P_j g_j(x) c_j \geq \max_{k\in\mathcal{S}} \min_{c_k} P_k g_k(x) c_k\} \tag{13}$$

BS j is in N_x if it is likely to serve the user at x for some CRE vector. Take the example of Figure 2, which shows the bias received power range at a given location x for all BSs. The BSs whose biased received power ranges intersect with the line that passes through the max–min-biased received power are the neighbor BSs. In Figure 2, BS 1 is a macro BS and has a single possible CRE bias. It also has the max–min bias received power. This means a user at x will receive at least this bias power. The bias received power from BS 5 can exceed this max–min, so that BS 5 is likely to serve the user for some CRE bias and is thus included in N_x. In the same way, BS 7 is also included in N_x. On the other hand, the bias received power from BS 2 will never exceed that of BS 1, and thus BS 2 will never serve the user at x.

We now construct the neighborhood of small BSs based on sets N_x. We assume that the users located at x calculate N_x and report it to their serving BSs, which multicast this information to all the BSs in N_x. BS j is considered to be a neighbor of BS i if the proportion of reports where BS i and BS j are in N_x is at least a threshold ϖ. This constraint aims at excluding from the neighborhood BSs that have insignificant influence on load. Otherwise, due to the infinite support of shadowing in our model, all BSs in the network can be potentially neighbors. Formally, the neighbor set of small BS i is defined as

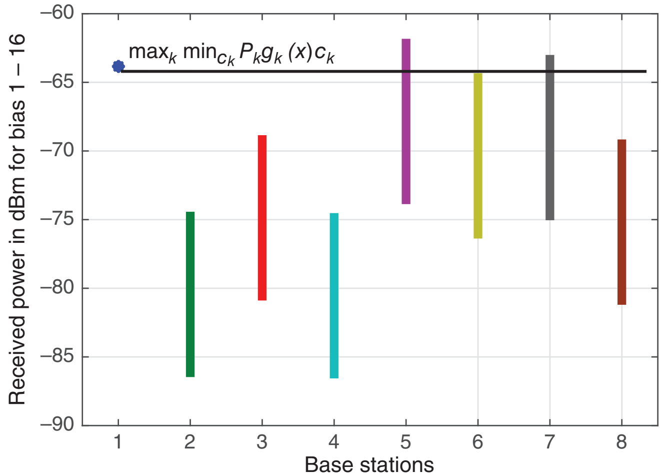

Figure 2 Illustration of the neighbor set.

$$N_i^{\varpi} = \left\{ j \in \mathcal{S} | \frac{\int_{x\in\mathcal{L}} \lambda(x) 1_{i,j\in N_x} \mathrm{d}x}{\int_{x\in\mathcal{L}} \lambda(x) 1_{i\in N_x} \mathrm{d}x} \geq \varpi \right\} \tag{14}$$

For the threshold $\varpi = 0$, the neighbor set N_i^0 boils down to $N_i^0 = \bigcup_{x:i\in N_x} N_x$. The neighbor set N_i^{ϖ} is empty for $\varpi > 1$. For $0 < \varpi \leq 1$, N_i^{ϖ} is a decreasing sequence of sets.

Note that a change in the ABS ratio theoretically affects the load of all BSs of the network through interference. We thus assume that a macro BS neighborhood is made of all BSs. In practice, however, the neighborhood of a macro BS is finite because interference power decreases with distance. Macro BS neighborhood can be constructed similarly to the small BS neighborhood.

Proposition 1 The user-association game Γ^{ϖ} with the potential function (8) is an ξ-potential game, where

$$\xi = \max_{a_i, a_i' \in X_i, a_{-i} \in X_{-i}, i\in\mathcal{S}} \Big| \sum_{j\in N_i^0 \setminus N_i^{\varpi}} \frac{(1-\rho_j(a_i,a_{-i}))^{1-\alpha}}{\alpha-1} - \sum_{j\in N_i^0 \setminus N_i^{\varpi}} \frac{(1-\rho_j(a_i',a_{-i}))^{1-\alpha}}{\alpha-1} \Big| \tag{15}$$

Corollary 1 The game Γ^0 is an exact potential game.

Distributed Learning Algorithms

In this section, we introduce distributed learning algorithms that are used to find the optimal PNE of the user-association game. First, we present the BR algorithm and the LLLA for the complete information setting. Next, the Binary Log-Linear Learning Algorithm (BLLLA) for the partial information setting is described.

Best Response Algorithm

Best response algorithm is an asynchronous algorithm where at any given time only a single BS updates its strategy. Set ϖ and assume a time-varying random process with which a BS is chosen to revise its strategy. The selected BS computes its cost $U_i(a_i, a_{-i}(t-1))$ for all $a_i \in X_i$ and sets $U_i(a_i, a_{-i}(t-1)) = \infty$ if $\rho_j \geq 1$ or $O_j \geq \overline{O}_j$ for $j \in N_i^{\varpi}$. Then, the BS chooses a strategy $a_i \in X_i$ that minimizes its cost, given the strategies $a_{-i} \in X_{-i}$ of other players. In other words, BS i chooses a strategy from its best response set B_i:

$$B_i(a_{-i}) = \text{argmin}_{a_i} U_i(a_i, a_{-i}) \tag{16}$$

Note that the BR algorithm requires complete information, i.e. the effects of choosing all the other strategies are supposed to be known. Moreover, BR algorithm is not guaranteed to converge to the optimal PNE even in exact potential game Γ^0 because the potential function may have multiple suboptima (Monderer and Shapley 1996). For the ξ-potential game Γ^{ϖ} ($\varpi \neq 0$) and $\xi \neq 0$, a PNE may not even exist.

Log-Linear Learning Algorithm

The LLLA is a classical asynchronous algorithm that guarantees the convergence to the optimal PNE of an exact potential game (Marden and Shamma 2012). This algorithm is similar to BR but allows deviations from the best response with a small probability. It is summarized in Algorithm 1. However, for this algorithm, the BSs require again complete information, and with this select a strategy to play according to a probability distribution.

Algorithm 1 Log-Linear Learning Algorithm

1: **Initialization:** Start with arbitrary action profile a.
2: Set parameter τ and ϖ.
3: **While** $t \geq 1$ **do**
4: Randomly select a player i.
5: Compute cost $U_i\left(a_i, a_{-i}(t-1)\right)$ for all $a_i \in X_i$.
6: For any $a_i \in X_i$, set $U_i\left(a_i, a_{-i}(t-1)\right) = \infty$ if $\rho_j \geq 1$ or $O_j \geq \overline{O}_j$ for $j \in i^{\varpi}$.
7: Take action $a_i(t)$ from X_i with probability $p_i^{a_i}(t)$,

$$p_i^{a_i}(t) = \frac{\exp\left(-\frac{1}{\tau}U_i\left(a_i, a_{-i}(t-1)\right)\right)}{\sum_{a_i' \in X_i} \exp\left(-\frac{1}{\tau}U_i\left(a_i', a_{-i}(t-1)\right)\right)} \tag{17}$$

8: All the other players must repeat their previous actions, i.e. $a_{-i}(t) = a_{-i}(t-1)$.

Binary Log-Linear Learning Algorithm

The BLLLA converges to the optimal PNE of an exact potential game even if only partial information about the game is available to the players (Marden and Shamma 2012).

Partial information is the information that a player has about its current strategy. Unlike complete information, the effect of choosing any other strategy is not known to the player. As LLLA, the BLLLA is also an asynchronous algorithm. In this algorithm, whenever the BS updates its strategy, it does it in two steps. In the first step, the BS tries a strategy from its strategy set to obtain its payoff. In the second step, the BS randomly chooses among the two strategies (present strategy and trial strategy), as summarized in Algorithm 2.

Algorithm 2 Binary Log-linear Learning Algorithm

1: **Initialization:** Start with arbitrary action profile a.
2: Set parameter τ and ϖ.
3: **While** $t \geq 1$ **do**
4: Randomly select a BS i.
5: Select a trial action $\hat{a}_i \in X_i$ with uniform probability.
6: Compute cost $U_i\left(\hat{a}_i, a_{-i}(t-1)\right)$.
7: Set $U_i\left(\hat{a}_i, a_{-i}(t-1)\right) = \infty$ if $\rho_j \geq 1$ or $O_j \geq \overline{O}_j$ for $j \in i^{\varpi}$.
8: Play action $a_i(t) \in \left\{a_i(t-1), \hat{a}_i\right\}$ as given below.

$$a_i(t) = \begin{cases} a_i(t-1), & \text{w.p. } \frac{e^{-\frac{1}{\tau}U_i(a(t-1))}}{e^{-\frac{1}{\tau}U_i(a(t-1))}+e^{-\frac{1}{\tau}U_i\left(\hat{a}_i, a_{-i}(t-1)\right)}} \\ \hat{a}_i, & \text{w.p. } \frac{e^{-\frac{1}{\tau}U_i\left(\hat{a}_i, a_{-i}(t-1)\right)}}{e^{-\frac{1}{\tau}U_i(a(t-1))}+e^{-\frac{1}{\tau}U_i\left(\hat{a}_i, a_{-i}(t-1)\right)}} \end{cases} \tag{18}$$

9: All the other players must repeat their previous actions, i.e. $a_{-i}(t) = a_{-i}(t-1)$.

Note that in all the above algorithms the actions that are not in the feasible set $\mathcal{F}$ have infinite cost.

Convergence of LLLA and BLLLA to Optimal PNE

The proof of convergence of LLLA and BLLLA to optimal PNE for an exact potential game Γ^0 is given in Marden and Shamma (2012). The conditions of convergence of LLLA and BLLLA to the global minimum of the potential function of a near-potential game Γ^{ϖ}, $\varpi > 0$, are given in the following theorem. As the underlying Markov chain is ergodic, each state has a positive probability to be chosen throughout the iterations of the algorithm. We say that the algorithm *converges* to a state if that probability is nonzero when parameter τ goes to zero.

Let ϕ_α^* and $\phi_\alpha^\dagger$ be the first minimum and second minimum values of the potential.

Theorem 1 For any $\epsilon > 0$ and any ξ-potential game Γ^{ϖ} with ξ given in (15) and $\xi < \frac{\epsilon}{2(|X|-1)}$, LLLA and BLLLA converge to a set of ξ-NEs with potential less than $\phi_\alpha^* + \epsilon$.

Corollary 2 For the game Γ^{ϖ}, if $\xi < \frac{\phi_\alpha^\dagger - \phi_\alpha^*}{2(|X|-1)}$, then both the LLLA and BLLLA converge to a set of PNEs whose potential value is ϕ_α^*.

We now define the neighborhood of every BS so that the condition of Corollary 2 is met. The following theorem gives an upper bound on ϖ that guarantees LLLA and BLLLA to converge to an optimal PNE.

Theorem 2 The constraint in Theorem 2 is satisfied if

$$\varpi \leq \epsilon Q(1-\rho_{\max})^{\alpha} \tag{19}$$

where $\rho_{\max}$ is the maximum possible load of a BS, $Q = \frac{\max_{x,\overline{\theta},j\in S} \frac{1}{v_j(x,\overline{\theta})}}{4|S||X|\lambda_m \max_x \left\{\frac{1}{\mu(x)}\right\}}$, and λ_m is an upper bound for the sum arrival rate in a cell.

Corollary 3 The constraint in corollary 2 is satisfied if

$$\varpi \leq Q(\phi_{\alpha}^{\dagger} - \phi_{\alpha}^{*})(1-\rho_{\max})^{\alpha} \tag{20}$$

Simulation Results

In this section, we show simulation results considering standard parameters as adopted in 3GPP (3GPP 2010). These parameters are listed in Table 1. We consider eight BSs located in a two-dimensional region $\mathcal{L}$. BS 1 is a macro BS and the rest are small BSs. There are two hotspots where the traffic is five times the average traffic, which can be seen in Figure 3. We consider shadow fading with a standard deviation of $\sigma_{sh} = 8$ dB and a decorrelation distance of $D_c = 20$ m. Cross-correlation between the shadowing components at a location is considered to be 0.5. We use the classical Shannon formula for calculating channel capacity $\tilde{v}_i^f(x,\overline{\theta}) = W\log_2(1+\gamma_i^f(x,\overline{\theta}))$.

Table 1 Simulation parameters.

Parameter	Variable	Value
Number of BSs	N_s	8
Macro BS during NS	P_{macro}	46 dBm
Macro BS during ABS	P_{ABS}	0 dBm
Small BS	P_{small}	24 dBm
Average file size	$\frac{1}{\mu}$	0.5 Mbytes
Average traffic load density	$\frac{\lambda}{\mu}$	64 bits s^1 m^2
System bandwidth	W	20 MHz
Carrier frequency	f_c	2.6 GHz
Noise power	N	$-174 + 10\log(\text{W})$ dBm
Minimum SINR	$\gamma_{\min}$	-10 dB
Path-loss exponent	η	3.5
Reference distance	d_0	10 m
CRE bias set	c_i	$\{1, 1.1, 1.2, \ldots, 16\}$
ABS ratio	θ_i	$\{0, 0.01, 0.02, \ldots, 1\}$

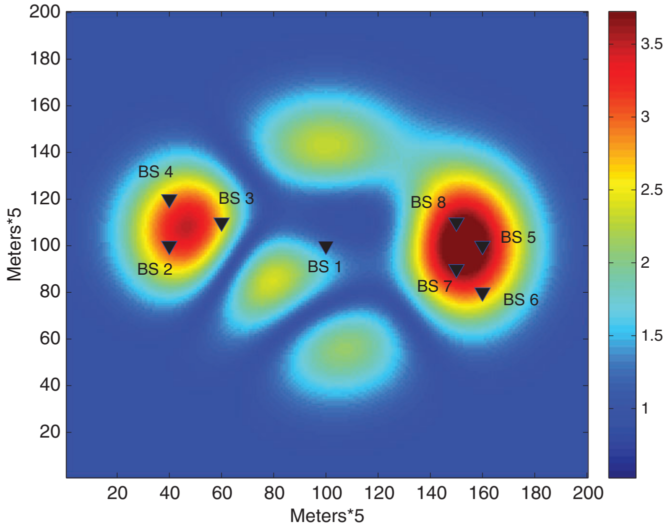

Figure 3 Normalized traffic variations.

Performance Comparison of Algorithms

We first focus on the CRE optimization and assume in this section that the macro BS does not implement ABS ($\theta_1 = 0$). We consider a square region $\mathcal{L}$ of side 1000 m (Figure 4).

The convergence of LLLA and BLLLA to the global minimum of the objective function is shown in Figure 5. We observe that in all cases proposed algorithms converge within few tens and sometimes few hundreds of iterations. Due to the complete information available with LLLA, this algorithms converges faster than BLLLA. Note, however, that BLLLA does not lose so much in terms of convergence speed; this is an interesting conclusion for a practical implementation.

We also compare in Figure 6 LLLA and BLLLA with Pradelski and Young learning algorithm (PYLA), which is completely uncoupled and a variant of trial-and-error algorithm (Pradelski and Young 2012). It guarantees asymptotic convergence to the optimal NE for any finite game that possesses at least one NE. LLLA and BLLLA converge quickly, whereas PYLA oscillates between fast search and slow search phases. Therefore, the performance of LLLA and BLLLA for load balancing is much superior to that of PYLA.

Effect of ϖ

The effect of threshold ϖ on the convergence of LLLA is shown in Figure 7a. For $\varpi = 0$, all the BSs are neighbors so that our framework is an exact potential game and LLLA converges to an optimal PNE. The threshold parameter $\varpi = 10^{-22}$ results in an ϵ-potential game and satisfies the sufficient condition of Corollary 3. Therefore, LLLA also converges to the global minimizer of the objective function. If we now further increase ϖ to a value that violates the condition of Corollary 3 ($\varpi = 0.9$), neighborhoods are further reduced. LLLA is, however, not anymore guaranteed to

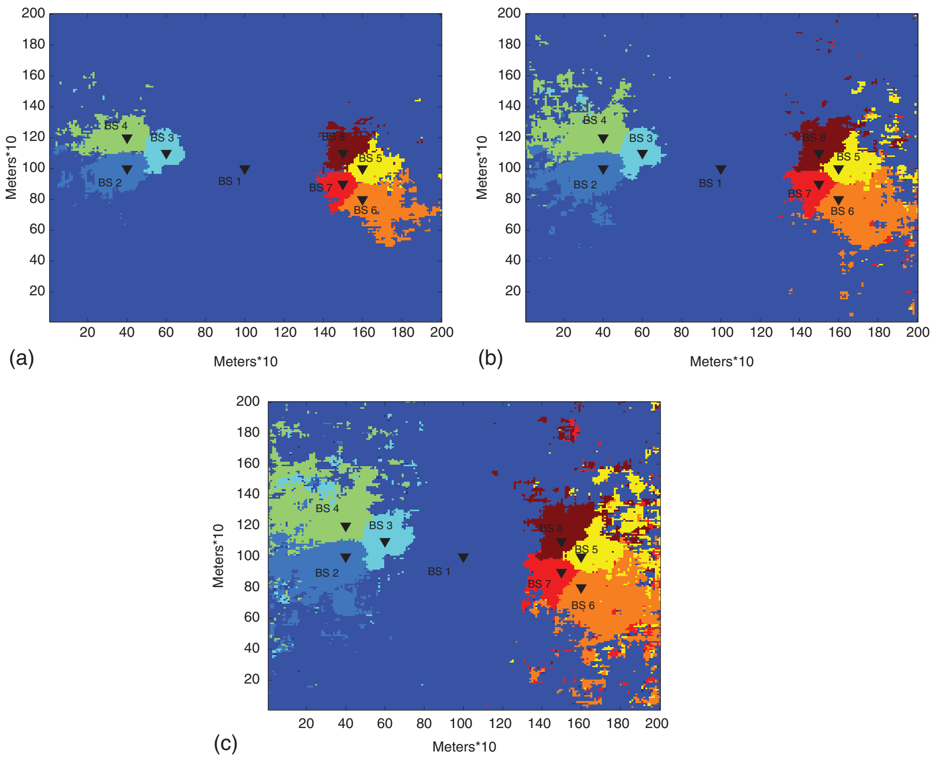

Figure 4 The variations of the coverage regions of BSs obtained using the optimal CRE for different α. (a) ($\alpha = 0$) Rate-optimal policy. (b) ($\alpha = 2$) Delay-optimal policy. (c) ($\alpha = 50$) Min–max policy.

converge to an optimal PNE as seen from the figure. The threshold ϖ therefore strikes a balance between the size of the neighborhood and the optimality of the solution.

Effect of τ

As shown in Figure 7b and Figure 7c, there is a trade-off in the choice of τ. LLLA and BLLLA converge with high probability to the global minimum of the objective function for $\tau \in (0, \infty)$ under the conditions of Corollary 2. This means that, asymptotically, the probability that the algorithm is at the global minimum approaches to 1 as τ goes to 0.

For high values of τ (e.g. $\tau = 0.05$ in the figure), LLLA and BLLLA result into oscillations. This is due to the fact that the algorithms converge fastly in probability to the uniform distribution. As a matter of fact, it does not spend much time in optimal states, which is not practically desirable. For small values of τ (e.g. $\tau = 0.01$ or 0.001), asymptotically, the algorithms will spend most of the time in the global optimal. However, convergence is slow in probability. This explains that the system can take long time to escape from suboptimal states. Contrary to best response, however, the proposed algorithms will not get stuck into these suboptimal states.

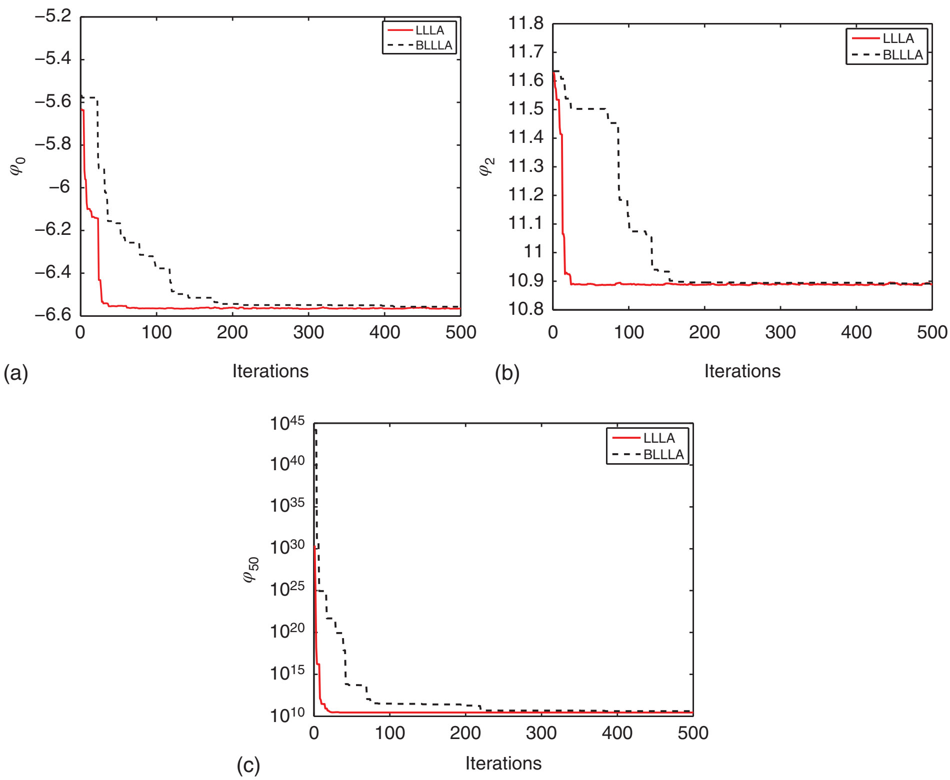

Figure 5 Convergence of LLLA and BLLLA ($\tau = 0.001$, $\varpi = 10^{-22}$). (a) ($\alpha = 0$) Rate-optimal policy. (b) ($\alpha = 2$) Delay-optimal policy. (c) ($\alpha = 50$) Min–max policy.

Effect of α

We now compare in Table 2 the optimal bias values and BS loads. The corresponding coverage regions are shown in Figure 4. With $\alpha = 0$, every user is served by the BS that provides the maximum data rate, which is obtained for bias values equal to one. This corresponds to the classical best signal association rule that results in heavy load imbalance between stations: the load of the macro BS reaches 92%, while small BSs have loads less than 11%. As α increases to 2, the coverage regions of all small BSs expand and that of the macro BS shrink. The load of the macro BS is decreased to 61%, and concurrently, the utilization of small BSs is increased (up to 21%). Min–max policy is approximated with a value of $\alpha = 50$. The load of the macro BS is further reduced to 45%, and the load dispersion is decreased.

This phenomenon can also be observed in Figures 8 and 9, where optimal CRE and loads are shown as functions of α. We see in this figure how the CREs of small BSs are gradually increased and how the load dispersion is reduced.

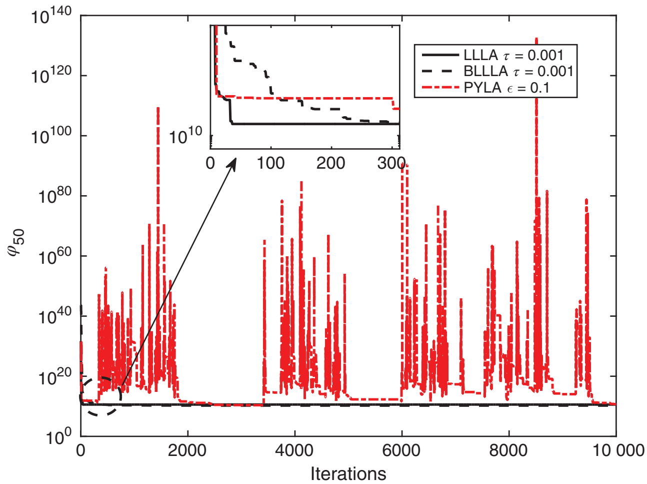

Figure 6 Comparison of LLLA, BLLLA, and PYLA for $\alpha = 50$, $\varpi = 10^{-22}$, and $\epsilon = 0.1$.

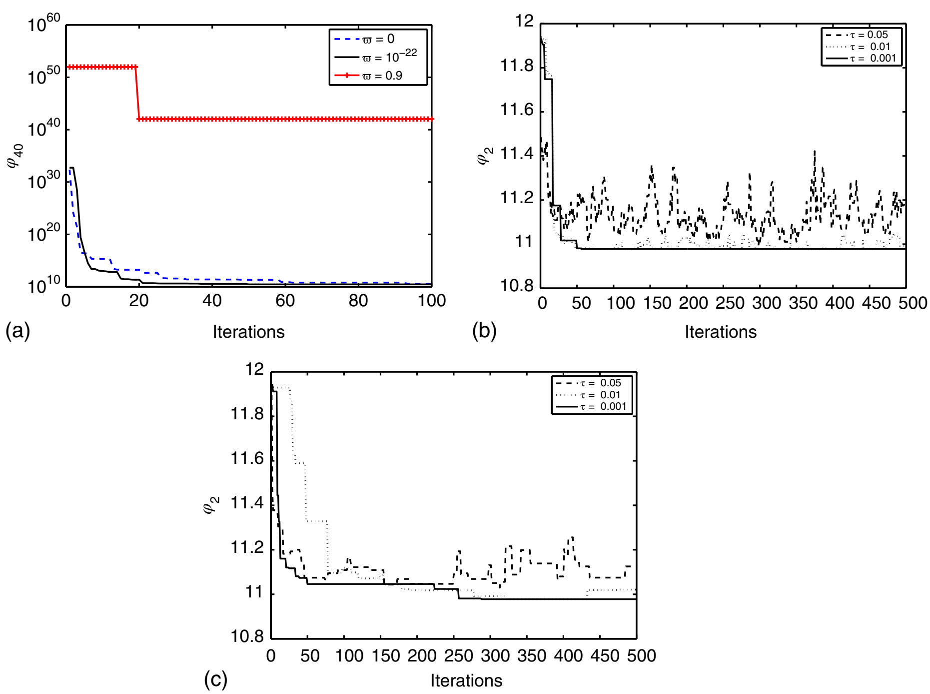

Figure 7 Effect of ϖ and τ on the convergence of LLLA and BLLLA. (a) Effect of ϖ on LLLA. (b) Effect of τ on LLLA. (c) Effect of τ on BLLLA.

Table 2 Comparison of optimal CRE, optimal loads of BSs for different α ($\tau = 10^{-3}$, $\varpi = 10^{-22}$)

	$\alpha = 0$		$\alpha = 2$		$\alpha \to \infty$	
BS i	c_i^*	$\rho_i^*\%$	c_i^*	$\rho_i^*\%$	c_i^*	$\rho_i^*\%$
1	1	92	1	61	1	45
2	1	7	3	20	8	42
3	1	4	3	9	9	23
4	1	9	3	18	8	37
5	1	11	3	21	7	37
6	1	8	3	20	7	43
7	1	5	3	11	8	30
8	1	7	3	19	6	37

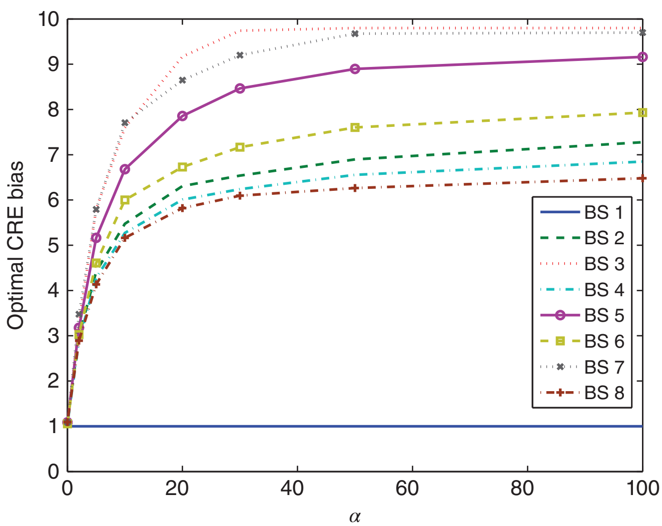

Figure 8 Evolution of optimal CRE bias.

Fairness-Outage Trade-off Using ABS

In this section, we allow the macro BS to implement ABS and study the effect of this technique on outage and fairness. For the sake of simplicity, we set $\alpha = 50$, $\tau = 10^{-3}$, and $\varpi = 10^{-22}$ and use LLLA. We also consider a square region $\mathcal{L}$ of side 2000 m in the simulations. Three cases may be compared to evaluate the interest of using ABS:

1) *No outage constraint no ABS*: In this case, the macro BS does not implement ABS ($\theta_1 = 0$), and we do not impose outage constraint. This serves as a benchmark to compare with other two cases.

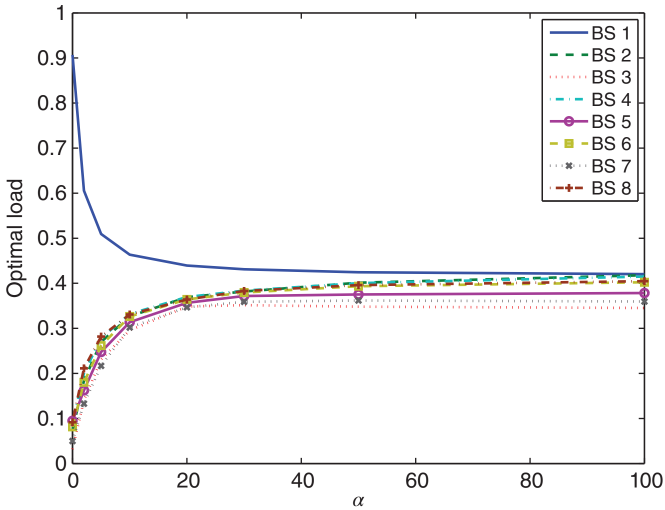

Figure 9 Evolution of optimal loads.

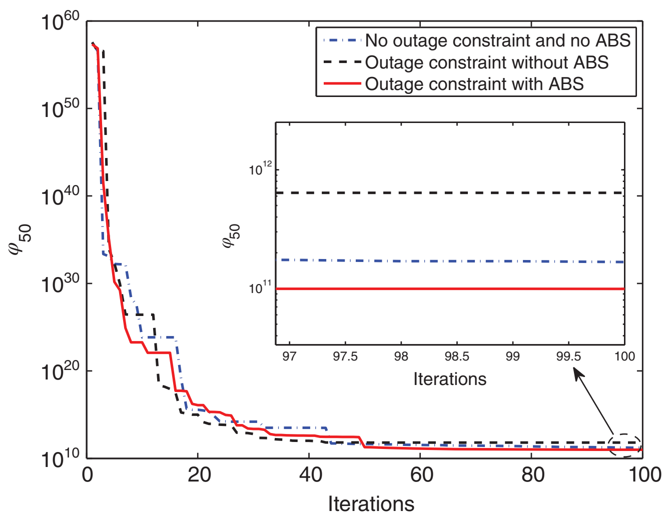

Figure 10 Effect of outage constraint and ABS (LLLA, ($\alpha = 50$), $\tau = 10^{-3}$, $\varpi = 10^{-22}$, and $\overline{O}_i = 2\%$).

2) *Outage constraint without ABS*: We introduce here an outage constraint ($\overline{O}_i = 2\%$ for all i) but still do not allow ABS ($\theta_1 = 0$).
3) *Outage constraint with ABS*: We impose an outage constraint and allow ABS at the macro BS.

In Figure 10, we show the cost function ϕ_α for the three considered cases. In the considered scenario, outage probabilities exceed the threshold in the first case. When outage constraint is introduced in the second case, the cost increases because the feasible set shrinks (some actions are not anymore available). When ABS is introduced in the third

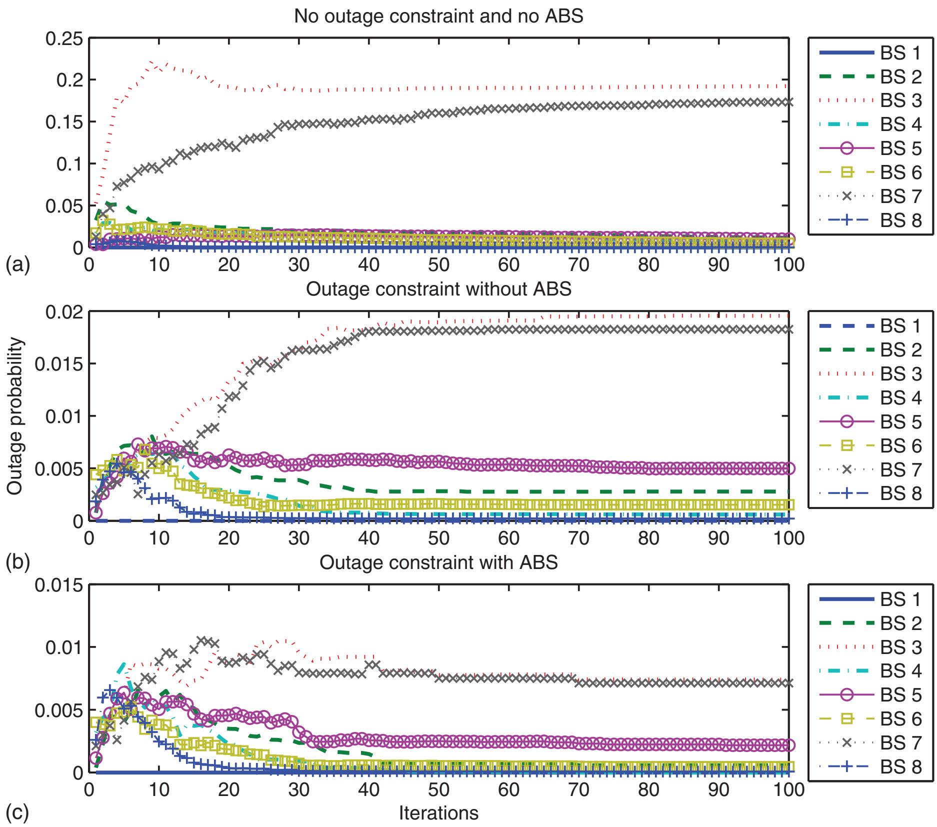

Figure 11 Outage probability comparisons of (a) no outage constraint no ABS, (b) outage constraint without ABS, and (c) outage constraint with ABS (LLLA, $\alpha = 50$, $\tau = 10^{-3}$, $\varpi = 10^{-22}$, and $\overline{O}_i = 2\%$).

case, the feasible set expands again and so the optimal cost of the system decreases, and fairness is improved.

Figure 11 shows the evolution of the outage probabilities as the algorithm iterates. With no ABS and no constraint (Figure 11a), the outage probabilities of BS 3 and 7 considerably exceed the threshold of 2%. The reason is that small BSs increase their CRE to achieve optimality without taking care of the users in outage, so that the users at cell edge may experience a very bad signal quality. Therefore, fairness is achieved at the cost of an unacceptable outage. Imposing an outage constraint without using ABS is sufficient to achieve a good quality of service (Figure 11b). The function ϕ_α, however, converges to a higher value. ABS is a good means to both meet the outage constraint and achieve fairness (Figure 11c). The reason is that small BS cell edge users experience a better signal quality during ABS subframes, and the ABS ratio also offers the macro BS an additional degree of freedom for adapting its load and achieving fairness. This can also be seen in Figure 12, where we have plotted average loads over 50 realizations after LLLA has converged. From the first to the second case, the load vector expands because of the smaller feasible set and then shrinks in the third case, thanks to ABS.

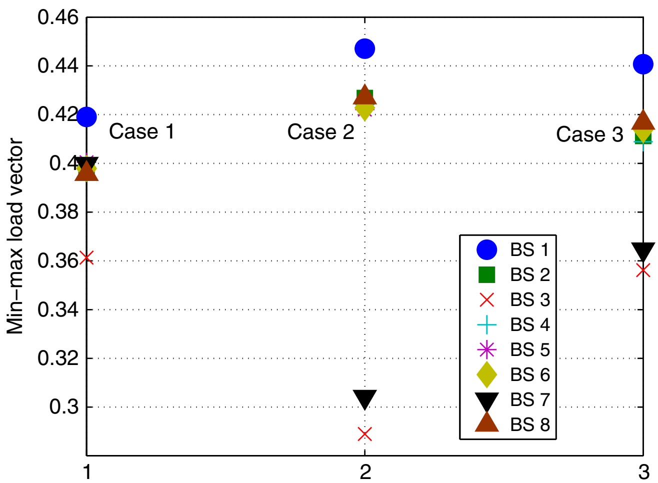

Figure 12 Comparison of min–max load vector of cases: (Case 1) no outage constraint no ABS, (Case 2) outage constraint without ABS, and (Case 3) outage constraint with ABS (LLLA, $\alpha = 50$, $\tau = 10^{-3}$, $\varpi = 10^{-22}$, and $\overline{O}_i = 2\%$).

Conclusions

In this article, a novel reinforcement learning approach for load balancing SON using CRE association technique and ABS interference management technique is presented. Our approach exploits the near-potential game structure and distributed learning algorithms. We show that the load balancing problem can be solved distributively by restricting the number of neighbors, and we provide theoretical guarantees of convergence for our algorithms. By running extensive simulations in two settings, which are complete and partial information settings, we show that the proposed algorithms converge within a few tens of iterations to the optimal PNE, which is also a minimizer of a $\alpha-$fairness function of the network. The convergence speed of the BLLLA that uses partial information is comparable to the LLLA that uses complete information, meaning that partial information is sufficient in practical implementations. Simulations show that for load balancing, LLLA and BLLLA perform better than a variant of trial-and-error algorithm. Finally, we show that by introducing ABS, the outages can be reduced and a better load balancing can be achieved.

Related Articles

Interference Management
User Association in UDN
Radio Resource Management in Small Cell Networks
Energy Efficient Radio Resource Management
AI-Enabled Radio Access Networks
5G Development: Autonomic Networking in RAN (from C-RAN to small cells)

Endnote

1 Proofs are provided in Ali et al. (2016).

References

3GPP (2010). Technical specification group radio access network; evolved universal terrestrial radio access (eutra); further advancements for e-utra physical layer aspects. TR 36.814, 3GPP.

3GPP (2018a). Technical Specification Group Services and System Aspects; System Architecture for the 5G System; Stage 2 (Release 15). TR 23.501, 3GPP.

3GPP (2018b). Technical Specification Group Services and System Aspects; Study of Enablers for Network Automation for 5G (Release 16). TR 23.791, 3GPP.

3GPP (2019). Technical Specification Group Services and System Aspects; Telecommunication Management; Study on the Self-organizing Networks (SON) for 5G Networks (Release 16). TR 28.861, 3GPP.

Ali, M.S., Coucheney, P., and Coupechoux, M. (2016). Load balancing in heterogeneous networks based on distributed learning in near-potential games. *IEEE Transactions on Wireless Communications* 15 (7): 5046–5059.

Alliance, N. (2006). Next generation mobile networks beyond HSPA & EVDO. White Paper, Next Generation Mobile Networks Ltd.

Al-Rawi, M. (2012). A dynamic approach for cell range expansion in interference coordinated LTE-advanced heterogeneous networks. Proc. ICCS, 533–537.

Andrews, J., Singh, S., Ye, Q. et al. (2014). An overview of load balancing in hetnets: old myths and open problems. *IEEE Wireless Communications* 21: 18–25.

Aryafar, E., Keshavarz-Haddad, A., Wang, M., and Chiang, M. (2013). RAT selection games in HetNets. Proc. INFOCOM, 998–1006.

Bonald, T. and Massoulié, L. (2001). Impact of fairness on internet performance. ACM SIGMETRICS Performance Evaluation Review, 29, 82–91.

Candogan, O., Ozdaglar, A., and Parrilo, P.A. (2011). Learning in near-potential games. IEEE Conf. Decision Control and European Control Conference, 2428–2433.

Damnjanovic, A., Montojo, J., Wei, Y. et al. (2011). A survey on 3GPP heterogeneous networks. *IEEE Wireless Communications* 18 (3): 10–21.

Elayoubi, S.-E., Altman, E., Haddad, M., and Altman, Z. (2010). A hybrid decision approach for the association problem in heterogeneous networks. Proc. INFOCOM, 1–5.

Fraile, R., Monserrat, J.F., Gozálvez, J., and Cardona, N. (2008). Mobile radio bi-dimensional large-scale fading modelling with site-to-site cross-correlation. *European Transactions on Telecommunications* 19 (1): 101–106.

Goldsmith, A. (2005). *Wireless Communications*. Cambridge, UK: Cambridge university press.

Gudmundson, M. (1991). Correlation model for shadow fading in mobile radio systems. *Electronics Letters* 27 (23): 2145–2146.

Jorguseski, L., Pais, A., Gunnarsson, F. et al. (2014). Self-organizing networks in 3GPP: standardization and future trends. *IEEE Communications Magazine* 52 (12): 28–34.

Kikuchi, K. and Otsuka, H. (2012). Proposal of adaptive control CRE in heterogeneous networks. Proc. PIMRC, 910–914.

Kim, H., de Veciana, G., Yang, X., and Venkatachalam, M. (2012). Distributed alpha-optimal user association and cell load balancing in wireless networks. *IEEE/ACM Transactions on Networking* 20: 177–190.

Kudo, T. and Ohtsuki, T. (2013). Cell range expansion using distributed Q-learning in heterogeneous networks. *EURASIP Journal on Wireless Communications and Networking* 2013: 1–10.

Liu, Y., Chen, C.S., and Sung, C.W. (2015). Joint optimization on inter-cell interference management and user attachment in LTE-A hetnets. International Workshop on Resource Allocation, Cooperation and Competition in Wireless Networks (RAWNET). IEEE.

Marden, J.R. and Shamma, J.S. (2012). Revisiting log-linear learning: asynchrony, completeness and a payoff-based implementation. *Games and Economic Behaviour* 75: 788–808.

Mo, J. and Walrand, J. (2000). Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking* 8 (5): 556–567.

Monderer, D. and Shapley, L.S. (1996). Potential games. *Games and Economic Behavior* 14: 124–143.

Moysen, J. and Giupponi, L. (2018). From 4G to 5G: self-organized network management meets machine learning. *Computer Communications* 129: 248–268.

Niyato, D. and Hossain, E. (2009). Dynamics of network selection in heterogeneous wireless networks: an evolutionary game approach. *IEEE Transactions on Vehicular Technology* 58: 2008–2017.

Oh, J. and Han, Y. (2012). Cell selection for range expansion with almost blank subframe in heterogeneous networks. Proc. PIMRC, 653–657.

Pradelski, B.S. and Young, H.P. (2012). Learning efficient nash equilibria in distributed systems. *Games and Economic Behavior* 75 (2): 882–897.

Ramiro, J. and Hamied, K. (2011). *Self-organizing Networks: Self-planning, Self-optimization and Self-healing for GSM, UMTS and LTE*. Hoboken, NJ: John Wiley & Sons.

Rengarajan, B. and De Veciana, G. (2011). Architecture and abstractions for environment and traffic-aware system-level coordination of wireless networks. *IEEE/ACM Transactions on Networking* 19 (3): 721–734.

Son, K., Chong, S., and Veciana, G. (2009). Dynamic association for load balancing and interference avoidance in multi-cell networks. *IEEE Transactions on Wireless Communications* 8 (7): 3566–3576.

Stevens-Navarro, E., Lin, Y., and Wong, V.W. (2008). An MDP-based vertical handoff decision algorithm for heterogeneous wireless networks. *IEEE Transactions on Vehicular Technology* 57 (2): 1243–1254.

Tall, A., Altman, Z., and Altman, E. (2014). Self organizing strategies for enhanced ICIC (eICIC). International Symp. Modeling and Optimisation in Mobile, Ad Hoc, and Wireless Netw., 318–325. IEEE.

Tonguz, O.K. and Yanmaz, E. (2008). The mathematical theory of dynamic load balancing in cellular networks. *IEEE Transactions on Mobile Computing* 7 (12): 1504–1518.

Wang, J., Liu, J., Wang, D. et al. (2011). Optimized fairness cell selection for 3GPP LTE-A macro-pico HetNets. Proc. VTC (Fall), 1–5.

Xu, F., Tan, C.C., Li, Q. et al. (2010). Designing a practical access point association protocol. Proc. INFOCOM, 1–9.

Ye, Q., Rong, B., Chen, Y. et al. (2013). User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications* 12: 2706–2716.

Further Reading

Ali, M.S., Coucheney, P., and Coupechoux, M. (2016). Load balancing in heterogeneous networks based on distributed learning in near-potential games. *IEEE Transactions on Wireless Communications* 15(7): 5046–5059.

Ali, M.S., Coucheney, P., and Coupechoux, M. (2017). Rules for computing resistance of transitions of learning algorithms in games, EAI Gamenets, 97–107.

Ali, M.S., Coucheney, P., and Coupechoux, M. (2019). Distributed learning in noisy-potential games for resource allocation in D2D networks. *IEEE Transactions on Mobile Computing*, 1–18. doi: 10.1109/TMC.2019.2936345.