Distributed Learning for Utility Maximization over CSMA-based Wireless Multihop Networks

Hyeryung Jang[†], Se-Young Yun[‡], Jinwoo Shin[†], and Yung Yi[†]

***Abstract*—Game-theoretic modeling and equilibrium analysis have provided valuable insights into the design of robust local control rules for the individual agents in multi-agent systems, *e.g.*, Internet congestion control, road transportation networks, etc. In this paper, we introduce a non-cooperative MAC (Medium Access Control) game for wireless networks and propose new fully-distributed CSMA (Carrier Sense Multiple Access) learning algorithms that are probably optimal in the sense that their long-term throughputs converge to the optimal solution of a utility maximization problem over the maximum throughput region. The most significant part of our approach lies in introducing a novel cost function in agents' utilities so that the proposed game admits an ordinal potential function with (asymptotically) no price-of-anarchy. The game formulation naturally leads to known game-based learning rules to find a Nash equilibrium, but they are computationally inefficient and often require global information. Towards our goal of fully-distributed operation, we propose new fully-distributed learning algorithms by utilizing a unique property of CSMA that enables each link to estimate its temporary link throughput without message passing for the applied CSMA parameters. The proposed algorithms can be thought as 'stochastic approximations' to the standard learning rules, which is a new feature in our work, not prevalent in other traditional game-theoretic approaches. We show that they converge to a Nash equilibrium, which is a utility-optimal point, numerically evaluate their performance to support our theoretical findings and further examine various features such as convergence speed and its tradeoff with efficiency.**

I. INTRODUCTION

In many engineering systems, we often observe trade-offs between efficiency and complexity, where optimal algorithms require heavy computational challenges or extensive message passing, but light-weight approximate algorithms incur efficiency degradation. MAC (Medium Access Control) in wireless networks is no exception. The seminal work is done by Tassiulas and Ephremides [1], referred to as Max-Weight scheduling, which is centralized and computationally intractable (for a large-scale network). The high complexity in Max-Weight stems from the fact that an NP-hard problem (maximum weight independent set problem) has to be solved repeatedly over time. Since then, various subsequent papers based on many principles, *e.g.*, random access [2], [3], pick-and-compare [4]–[7], and maximal/greedy [8]–[10], have been published, and most of them more or less show that the tradeoff between efficiency and complexity indeed exists, *e.g.*, see [11], [12] for surveys.

†: Department of Electrical Engineering, KAIST, South Korea, e-mails: hrjang@lanada.kaist.ac.kr, jinwoos@kaist.ac.kr, yiyung@kaist.edu. ‡: KTH Royal Institute of Technology, Sweden, e-mail:seyoung@kth.se

978-1-4799-3360-0/14/$31.00

In this paper, we aim at developing fully-distributed MAC algorithms without message passing that are utility-optimal over the maximum throughput region (as achieved by Max-Weight). To that end, we adopt CSMA as a base-line MAC, where we smartly update CSMA's parameters, so that the long-term throughput over links is the optimal solution of a utility maximization problem. We take a game-theoretic learning framework—by formulating a non-cooperative CSMA game. Game theory has been emerged as a powerful tool for the design and distributed control of multi-agent systems, *e.g.*, [13], where agents just optimize their local objectives and react to limited network information, yet their local decisions often result in system-wide efficient behaviors. This paper inherits such a philosophy of using a game for distributed optimization. To achieve our goal, we design a novel cost function for each link, characterize the existence, uniqueness, efficiency of NE (Nash Equilibrium), and propose dynamic algorithms (strategy evolutions) to achieve the NE in a fully-distributed manner without any message passing. In the game, each link uses its own CSMA parameter as a strategy, and the payoff function is designed to reflect both (i) the net-utility from the network, *e.g.*, some function of the stationary throughput, and (ii) the cost measured by the harmful effects on other links. We prove that the game is an ordinal potential game and has the unique (non-trivial) NE which is equivalent to the socially optimal point (*i.e.*, no price-of-anarchy).

A game formulation naturally leads to popular learning dynamics in classical game theory, but it turns out that they are computationally inefficient and require global information in our game. In particular, they are interactive: each player's learning process correlates and affects what has to be learned by every other player over time. To resolve this issue, we exploit a feature of CSMA that temporary link throughput is naturally locally-observable without message passing, and design three learning rules, called ***SA-BRD*** (Stochastically Approximated-Best Response Dynamics), ***SA-JD*** (SA-Jacobi Dynamics) and ***SA-GD*** (SA-Gradient Dynamics), that update the link access intensities in a fully-distributed manner: each player can adjust its behavior only in response to its own realized throughput without knowledge of the game structure, without observing the behaviors and/or throughputs of the others. We prove that ***SA-BRD***, ***SA-JD*** and ***SA-GD*** converge to the unique non-trivial NE. They can be thought as 'stochastic approximations' to the standard learning rules, which is a new feature, not popular in other traditional game-theoretic approaches. It has been shown in Hart and Mas-Colell [14] that for a broad class of games, there is no general algorithm which allows the players' period-by-period behavior (even not fully distributed) to converge to a NE. Also, there exists a distributed game-based learning in [15], which, however, provides only probabilistic convergence guarantee under highly strict conditions such as finiteness of a game.

To connect CSMA (with utility maximization being a goal for saturated traffic) to the angles from machine learning or statistical physics, it can be regarded as a problem of finding the parameters of the hard-core graphical model[1] [16] (in a distributed manner) leading to the marginal distribution (or link throughput in networking) that is the optimal solution of a utility maximization problem. The hard-core model there corresponds to the interference graph in multi-hop wireless network and the parameters are the link access intensities in CSMA. Technical challenges lie in complex inter-plays between the parameter-update and the underlying dynamics of the hard-core graphical models, where non-trivial time-scale issues also exist, *e.g.*, the parameter-update before the underlying Markov chain for a given parameter reaches stationary regime. The proposed dynamics in this paper can be also interpreted as variants of the *contrastive divergence learning* [17] in hard-core graphical models, which is of intellectually independent interest in other fields such as machine learning and statistical physics.

Recently, there have been works that fully-distributed MAC algorithms based on CSMA (Carrier Sense Multiple Access) can achieve optimality in both throughput and utility, *e.g.*, see [18]–[22]. Our work takes a different angle (*i.e.*, game-based one) toward fully-distributed algorithms, which reverse-engineer existing algorithms as well as propose new types of algorithms. Also, as summarized in Section II, our paper also differs from other MAC game papers in that our work leads to fully-distributed utility-optimal learning dynamics over the maximum throughput region, whereas most of other papers are based on the smaller throughput region.

II. Related Work

D1. Single-hop ALOHA-based MAC: There exist game-theoretic studies on ALOHA or Slotted-ALOHA with selfish users [23]–[28]. To summarize some of those papers, the authors in [23] considered a multipacket reception model for selfish users and analyzed a Nash equilibrium and its stability region with the assumption of perfect information. The case with partial information has been studied in [24]. In [25], non-cooperative two-player Aloha game was shown to have two different Nash equilibria though only one was locally asymptotically stable. The authors in [27], [28] studied the impact of channel-state information. All these papers considered a single-hop wireless network, *e.g.*, WLAN (Wireless LAN).

D2. Single-hop 802.11 or CSMA/CA: In [29], it has been studied how selfish users can cheat those who obey the standard CSMA/CA. The authors in [30] abstract 802.11 DCF by focusing on 802.11's average behavior and connecting its window-based access and backoff to transmission probability. Then, the stability of 802.11 has been studied when heterogeneous selfish users exist, where each user dynamically changes its contention window size based on its disutility in terms of contention degree.

D3. Multi-hop random access: The authors in [31] reverse-engineered exponential back-off based contention resolution mechanism which can be modeled by a non-cooperative game with a player's strategy being access probability (*i.e.*, Aloha-like MAC). They also showed that the resulting NE is not generally socially optimal. This motivates the work [32] which forward-engineer utility-optimal contention resolution algorithms using a standard optimization decomposition approach. The authors in [33], [34] take a similar medium access model based on access probability and study how cost function in each player's payoff function should be designed to achieve a good NE and propose a dynamic access probability update rule converging to the NE. The authors in [35], [36] propose distributed MAC algorithms which have provable convergence, optimality, and robustness under a wider range of utility functions with single message passing for each node in [35] for general topologies, and without message passing for fully interfered topologies in [36].

D4. Achieving optimality based on CSMA: As mentioned in Section I, CSMA has recently been studied from an optimization based framework to achieve optimality in throughput and/or fairness, *e.g.*, see [18]–[22], [37], [38]. The main intuition underlying these results is that links dynamically adjust their CSMA parameters, *backoff* and *channel holding* times, using local information such as queue-length so that they solve a certain network-wide optimization problem for the desired high performance. These analytical work has been transferred to practical implementations [39]–[41]. This research has been regarded as an exciting progress to achieve both simplicity and optimality in the area of wireless cross-layer design.

Major difference from prior work. Our work differs from *D1* and *D2* in that a multi-hop wireless network is considered, which seems to generate more challenging scenarios than a single-hop case. Our work also enhances the research in *D3* in the sense that (i) we propose a game whose NE is optimal over the maximum throughput region (throughput region achieved by Max-Weight), whereas the studies in *D3* are utility-optimal over a throughput region from Slotted-Aloha (a much smaller region than the maximum throughput region), (ii) their dynamic update algorithms require message passing. CSMA's utility-optimality has been studied by the researches in *D4* mainly from an optimization-perspective, but our work starts from a game, followed by the resulting NE's efficiency (asymptotically no PoA) and proposes diverse dynamic algorithms not revealed by the work in *D4*.

III. Model, Objective, and Challenges

A. System Model

Network, interference, and traffic. In a wireless network, links share the common wireless medium where they may interfere in their transmissions. As a popular model for such wireless interference networks, a network topology can be represented

[1] This corresponds to a graphical model that neighboring nodes cannot be active simultaneously.

as an undirected graph $G = (V, E)$, called *interference graph,* where n links correspond to vertices V, *i.e.*, $|V| = n$, and undirected edges E are generated among interfering links. In other words, we assume that the interference is symmetric, captured by undirected edges in the graph, *i.e.*, $(i, j) \in E$ if and only if links i and j interfere with each other. We are interested in single-hop link-level traffic[2], and assume that the network is saturated, *i.e.*, each link has infinite backlog to transmit.

Schedule and throughput region. We consider a continuous time framework, where our primary interest is to track which links transmit over time. Let $\sigma_i(t) \in \{0, 1\}$ denote whether link i is transmitting at time t or not, where $\sigma_i(t) = 1$ means that the transmission at link i is active (*i.e.*, transmitting) at time t and 0 otherwise. We also denote by $\boldsymbol{\sigma}(t) = [\sigma_i(t)]_{i\in V}$ a *schedule* for links at time t. A scheduling algorithm is regarded as a policy that chooses a sequence of schedules $\{\boldsymbol{\sigma}(t)\}_{t=0}^{\infty}$ over time. Since interfering links cannot successfully transmit packets simultaneously, a schedule $\boldsymbol{\sigma}$ is called *feasible* (*i.e.*, no collision) unless there exists $(i, j) \in E$ such that both σ_i and σ_j are 1. Thus, the set of all feasible schedules $\mathcal{I}$(G) is defined as:

$$\mathcal{I}(G) \triangleq \{\boldsymbol{\sigma} \in \{0, 1\}^n : \sigma_i + \sigma_j \le 1, \forall (i, j) \in E\}.$$

We now define the maximum throughput region (or simply throughput region) Λ of a given network, which is the convex hull of $\mathcal{I}(G)$, *i.e.*,

$$\Lambda \triangleq \left\{ \sum_{\boldsymbol{\rho}\in\mathcal{I}(G)} \alpha_{\boldsymbol{\rho}} \boldsymbol{\rho} : \sum_{\boldsymbol{\rho}\in\mathcal{I}(G)} \alpha_{\boldsymbol{\rho}} = 1, \alpha_{\boldsymbol{\rho}} \ge 0, \forall \boldsymbol{\rho} \in \mathcal{I}(G) \right\}.$$

The intuition behind this notion of throughput region comes from the fact that any scheduling algorithm has to choose a schedule from $\mathcal{I}(G)$ at each time and hence the time average of the service rate in each link induced by any scheduling algorithm must belong to Λ.

CSMA (Carrier Sense Multiple Access). As mentioned in Section I, our interest lies in simple, fully-distributed CSMA scheduling algorithms to avoid interferences efficiently in wireless networks. Under a CSMA algorithm, prior to trying to transmit a packet, links first check whether the medium is busy or idle, and then transmit the packet only when the medium is sensed idle, *i.e.*, no interfering link is transmitting. To control the aggressiveness of such medium access, each link maintains a backoff timer, which is reset to a random value when it expires. The timer runs only when the medium is idle, and stops otherwise. With the backoff timer, links try to avoid collisions by the following procedure:

- Each link does not start transmission immediately when the medium is sensed idle, but keeps silent until its backoff time expires.
- After a link grabs the channel (or medium), the link holds the channel for some duration, called the holding time.

[2]However, our analysis of this paper can be readily extended to multi-hop flows if a classical combination of back-pressure routing and source congestion control [12] is inserted to our analysis.

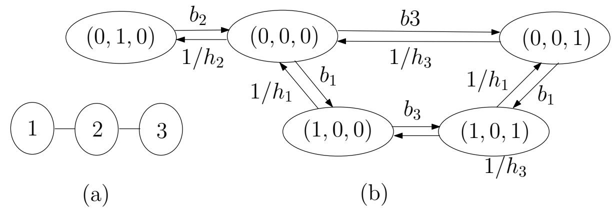


Fig. 1. (a) 3-link interference graph, where links 1 and 3 do not interfere with each other, but interfere only with link 2. (b) The resulting CSMA Markov process for fixed intensities, where for example the state $(1, 0, 1)$ means that links 1 and 3 are active and link 2 is inactive.

We assume that backoff and holding times of link i follow exponential distributions with means $1/b_i$ and h_i, respectively, for some positive real numbers b_i and h_i. Then, it is not hard to see that a CSMA algorithm induces a sequence of schedules $\{\boldsymbol{\sigma}(t)\}_{t=0}^{\infty}$, which is a time-reversible Markov process. To illustrate, consider a simple three-link interference graph and its resulting Markov process in Fig. 1.

For fixed CSMA parameters $\boldsymbol{b} = [b_i]_{i\in V}$ and $\boldsymbol{h} = [h_i]_{i\in V}$, using the time-reversibility of Markov process $\{\boldsymbol{\sigma}(t)\}_{t=0}^{\infty}$, its stationary distribution $\pi^{\boldsymbol{b},\boldsymbol{h}} = [\pi^{\boldsymbol{b},\boldsymbol{h}}_{\boldsymbol{\sigma}}]_{\boldsymbol{\sigma}\in\mathcal{I}(G)}$ can be characterized as follows:

$$\pi^{\boldsymbol{b},\boldsymbol{h}}_{\boldsymbol{\sigma}} = \frac{\prod_{i\in V}(b_i h_i)^{\sigma_i}}{\sum_{\boldsymbol{\sigma'}\in\mathcal{I}(G)} \prod_{i\in V}(b_i h_i)^{\sigma'_i}}. \quad (1)$$

Namely, the stationary distribution of a schedule depends only on the product of $\boldsymbol{b}$ and $\boldsymbol{h}$ of links. For simple presentation, we let $r_i = \log(b_i h_i)$ and $\boldsymbol{r} = [r_i]_{i\in V}$, and call r_i the *intensity* of link i, intuitively meaning the transmission aggressiveness of the link. Hence, we also use $\pi^{\boldsymbol{r}}$ instead of $\pi^{\boldsymbol{b},\boldsymbol{h}}$.

Given the links' fixed intensities $\boldsymbol{r}$, the ergodicity of the Markov process implies that the marginal probability $s_i(\boldsymbol{r})$ that link i is scheduled under the stationary distribution $\pi^{\boldsymbol{r}}$ becomes the link i's long-term (average) throughput or service rate, *i.e.*, $\lim_{t\to\infty} \frac{1}{t}\int_0^t \sigma_i(s)\, ds$. Furthermore, using the reversibility of the Markov process, the marginal probability can be characterized as:

$$\begin{aligned} s_i(\boldsymbol{r}) &= \mathbb{E}_{\pi^{\boldsymbol{r}}}[\sigma_i] = \sum_{\boldsymbol{\sigma}\in\mathcal{I}(G):\sigma_i=1} \pi^{\boldsymbol{r}}_{\boldsymbol{\sigma}} \\ &= \frac{\sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)|\sigma_i=1} \exp(\sum_{i\in V}\sigma_i r_i)}{\sum_{\boldsymbol{\sigma'}\in\mathcal{I}(G)} \exp(\sum_{i\in V}\sigma'_i r_i)}. \end{aligned} \quad (2)$$

B. Problem: Utility Maximization

We aim at developing a CSMA algorithm that controls the intensity of each link so as to make the long-term service rate close to some fairness point of the boundary of Λ. Specifically, each link i adaptively changes its intensity r_i (*i.e.*, CSMA's backoff and holding time parameters b_i and h_i) over time, so that the long-term service rates over links form a solution of the following utility maximization problem:

$$\textbf{(OPT)} \qquad \max_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in V} U_i(\lambda_i), \quad (3)$$

where $\boldsymbol{\lambda}^*$ denotes the solution of the maximization. In the above, each link i has a concave, increasing, and (twice) differentiable utility function, $U_i : [0, 1] \to \mathbb{R}$, where its value represents the utility when rate $\lambda_i \in [0, 1]$ is allocated at link

i. As is well-known, the various forms of utility functions enforce different concepts of fairness, *e.g.*, famous α-fairness.

To achieve the desired utility maximization using a CSMA algorithm, the core question is how each link i chooses transmission intensity r_i so that $s_i(\boldsymbol{r})$ is the solution of (3). To this end, we take a game-theoretic approach, where a smart design of payoff (also called net-utility and price) functions for links is necessary to have the desired property such that the (Nash) equilibrium of the game corresponds to the solution of (3). Under such a game design, we will consider various dynamic learning algorithms that provably converge to the equilibrium, where major technical challenges for proving the convergence lie in handling a non-trivial coupling between CSMA Markov process and CSMA parameter updates.

IV. Game Design and Equilibrium Analysis

A. O-CSMA Game: oCSMA(β)

We first describe our non-cooperative game, denoted by oCSMA(β), which is parameterized by β, by presenting its components. We then explain the rationale behind our game design.

---

oCSMA(β)

(i) ***Players.*** Each link $i \in V$ (*i.e.*, a node in the interference graph $G(V,E)$) acts as a player.

(ii) ***Strategy.*** Each player i chooses an intensity $r_i \in (-\infty,\infty)$ as its own strategy, which determines how aggressively i accesses the medium. We use the conventional notation that the strategy vector for all links except i is $r_{-i} := (r_1, r_2, \cdots, r_{i-1}, r_{i+1}, \cdots, r_n)$ and write a strategy profile $\boldsymbol{r} = (r_i, r_{-i})$.

(iii) ***Payoff function.*** The payoff function $\Phi_i(r_i, r_{-i})$ of player i is designed to be utility U_i subtracted by an incurring price C_i, *i.e.*,

$$\Phi_i(r_i, r_{-i}) = U_i(s_i(\boldsymbol{r})) - \frac{1}{\beta} C_i(r_i, r_{-i}),$$

where

$$C_i(r_i, r_{-i}) = \int_{-\infty}^{r_i} x s_i'(x, r_{-i}) dx. \quad (4)$$

Note that a payoff of player i is determined by how aggressively other links access the medium as well as how itself does. The parameter β quantifies 'price level' in the players' payoffs, and we realize that it balances the tradeoff between the quality (*i.e.*, price-of-anarchy, see Section IV-C) of equilibria in the game and the convergence speed to the equilibria under the learning dynamics (see Section V).

B. Role of Price Function

To have nice properties of our game, *e.g.*, good equilibria or provable transfer to fully-distributed dynamics (converging to a good equilibrium), the choice of price function is of critical importance. This subsection is devoted to explaining how such nice properties can be obtained from our price function (4) which has two following design features ***P1*** and ***P2***.

P1. Appropriate measure of link's contention: We smartly choose a price function, so that it appropriately measures each link's contention impact on other links' throughput. This choice differs from other price function choices, *e.g.*, one in Aloha systems, which is the key to probably have almost no price-of-anarchy (see Section IV-C). Specifically, a simple algebra gives us the following expression of our price function:

$$C_i(\boldsymbol{r}) \quad = \quad s_i(r_i, r_{-i}) \cdot \int_{-\infty}^{r_i} x \frac{s_i'(x, r_{-i})}{s_i(r_i, r_{-i})} \mathrm{d}x.$$

To interpret the above price design, let $R_i \in [-\infty, r_i]$ be a continuous random variable with the density $f_{R_i}(\cdot)$:

$$f_{R_i}(x) \quad = \quad \frac{1}{s_i(r_i, r_{-i})} \frac{\partial s_i(x, r_{-i})}{\partial x}.$$

Then, our price function can be regarded as the product of the service rate and R_i's expectation:

$$C_i(\boldsymbol{r}) = s_i(r_i, r_{-i}) \cdot \mathbb{E}[R_i]. \quad (5)$$

In other words, our design of price function considers the *relative increasing speed of the service rate* in the interval $(-\infty, r_i)$.

<u>Remark</u> 1 A popular choice of price function is $C_i(\boldsymbol{r}) = r_i s_i(r_i, r_{-i})$. The intuition behind such a choice is that the price to be paid by a link's access intensity is proportional to the aggressiveness in media usage multiplied by its achieved gain (*i.e.*, throughput). This type of chice has already been made in other works, *e.g.*, [34] which studies Aloha-based MAC. However, it is unclear that this price function provides a provable framework for no price-of-anarchy.

P2. Indirect coupling of players' strategies: Our price function also leads to a form that is a function of self-strategy and its marginal distribution of the given strategy vector, not the individual strategy values of others. This feature enables us to develop a fully-distributed dynamics that works based only on throughput measurements (see Section V). To explain this in details, we first re-express the price function to better understand how it is structured in terms of the local strategy and other players' strategies. The price function can be re-expressed as:

$$\begin{aligned} C_i(r_i, r_{-i}) &= \int_{-\infty}^{r_i} x s_i'(x, r_{-i}) \mathrm{d}x \\ &= \Big[x s_i(x, r_{-i}) \Big]_{-\infty}^{r_i} - \int_{-\infty}^{r_i} s_i(x, r_{-i}) \mathrm{d}x \\ &= r_i s_i(\boldsymbol{r}) + \ln\big(1 - s_i(\boldsymbol{r})\big), \end{aligned} \quad (6)$$

where for the second term we use: first,

$$\begin{aligned} s_i(x, r_{-i}) &= \frac{\sum_{\boldsymbol{\sigma} \in \mathcal{I}(G): \sigma_i = 1} \exp\big(\sum_{j \in V} r_j \sigma_j\big)}{\sum_{\boldsymbol{\sigma} \in \mathcal{I}(G)} \exp\big(\sum_{j \in V} r_j \sigma_j\big)} \\ &= \frac{\sum_{\boldsymbol{\sigma} \in \mathcal{I}(G): \sigma_i = 1} \exp\big(\sum_{j \in V \setminus \{i\}} r_j \sigma_j\big) \exp(x)}{\sum_{\boldsymbol{\sigma} \in \mathcal{I}(G)} \exp\big(\sum_{j \in V \setminus \{i\}} r_j \sigma_j\big) \exp(x \sigma_i)} \\ &= \frac{B \exp(x)}{A + B \exp(x)}, \end{aligned}$$

where

$$A \equiv \sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)|\sigma_i=0} \exp(\sum_{j\in V\setminus\{i\}} r_j\sigma_j),$$
$$B \equiv \sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)|\sigma_i=1} \exp(\sum_{j\in V\setminus\{i\}} r_j\sigma_j),$$

and second,

$$\int_{-\infty}^{r_i} s_i(x, r_{-i})dx = \int_{-\infty}^{r_i} \frac{B\exp(x)}{A+B\exp(x)} = \ln\frac{A+B\exp(r_i)}{A} = -\ln\Big(1-s_i(\boldsymbol{r})\Big).$$

From (6), the payoff function reads:

$$\Phi_i(\boldsymbol{r}) = U_i(s_i(\boldsymbol{r})) - \frac{1}{\beta}\Big(r_i s_i(\boldsymbol{r}) + \ln\big(1-s_i(\boldsymbol{r})\big)\Big).$$

It is important to see that the payoff function depends only on the local strategy r_i and local service rate $s_i(r_i, r_{-i})$, not directly on the individual strategies of other players. This *indirect coupling*, which is a unique feature in our game, is highly convenient to develop a fully-distributed dynamic algorithm, because $s_i(\cdot)$ can be measured in the midst of playing a player's own strategy, thus not requiring message passing.

C. Equilibrium Analysis: Existence, Uniqueness and Price-of-Anarchy

We now analyze the equilibrium of oCSMA(β) using the popular notion of Nash equilibrium, whose definition is presented as follows:

Definition 1 A strategy profile $\boldsymbol{r}^{\text{NE}}$ is a *Nash equilibrium* (NE) if

$$\Phi_i(r_i^{\text{NE}}, r_{-i}^{\text{NE}}) \geq \Phi_i(r_i, r_{-i}^{\text{NE}}), \quad \forall r_i \in \mathbb{R}, \forall i \in V.$$

Furthermore, we say that a NE $\boldsymbol{r}^{\text{NE}}$ (if exists) in the game is *non-trivial*, if each player i's service rate at equilibrium $s_i(\boldsymbol{r}^{\text{NE}})$ is positive for all players $i \in V$, and *trivial* otherwise. We now present our main results on the equilibrium analysis in the following theorem.

Theorem 1 (Uniqueness and no PoA) In the oCSMA(β), for any $\beta > 0$,

(i) *Existence and uniqueness.* There exists a unique non-trivial NE $\boldsymbol{r}^{\text{NE}}$.

(ii) *Price-of-anarchy.* Furthermore, at the non-trivial NE $\boldsymbol{r}^{\text{NE}}$,

$$\sum_{i\in V} U_i\Big(s_i(\boldsymbol{r}^{\text{NE}})\Big) \geq \sum_{i\in V} U_i\Big(s_i(\boldsymbol{r}^*)\Big) - \frac{\log|\mathcal{I}(G)|}{\beta}, \quad (7)$$

where $\boldsymbol{r}^*$ represents a strategy profile such that the service rate vector $[s_i(\boldsymbol{r}^*)]_{i\in V}$ is the solution of the optimization problem in (3), *i.e.*, $[s_i(\boldsymbol{r}^*)]_{i\in V} = \boldsymbol{\lambda}^*$.

Theorem 1 implies that there is almost no PoA (Price-of-Anarchy) in our game, *i.e.*, the aggregate utility at the unique non-trivial NE can be arbitrarily close to the social optimum by choosing β sufficiently large. Namely, PoA can become arbitrarily small.

Proof: (i) *Existence and uniqueness.* We first prove the existence and the uniqueness of non-trivial NE using a potential game approach. Consider the following function $P(\boldsymbol{r})$ on the space $\mathcal{R}^+ = \{\boldsymbol{r}|s_i(\boldsymbol{r}) > 0, \text{ for all } i \in V\}$ (the set of strategies producing "non-trivial" service rates), defined by:

$$P(\boldsymbol{r}) \triangleq -\sup_{\boldsymbol{\lambda}\in[0,1]^n, \boldsymbol{\mu}\in\mathcal{P}} L(\boldsymbol{\lambda}, \boldsymbol{\mu}; \frac{\boldsymbol{r}}{\beta})$$

where $\mathcal{P}$ is the set of all probability measures over the set of all feasible schedules $\mathcal{I}(G)$, and

$$L(\boldsymbol{\lambda}, \boldsymbol{\mu}; \frac{\boldsymbol{r}}{\beta}) \triangleq \sum_{i\in V} U_i(\lambda_i) - \frac{1}{\beta}\sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)} \mu_{\boldsymbol{\sigma}}\log\mu_{\boldsymbol{\sigma}} + \sum_{i\in V}\frac{r_i}{\beta}\Bigg(\sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)}\mu_{\boldsymbol{\sigma}}\sigma_i - \lambda_i\Bigg).$$

It is easy to check that $P(\boldsymbol{r})$ is strictly concave in $\boldsymbol{r}$, since $P(\cdot)$ is the infimum of $-L(\cdot)$ which is a family of affine functions in $\boldsymbol{r}$. We now show that oCSMA(β) is an *ordinal potential game* [42] with the potential function $P(\boldsymbol{r})$, *i.e.*, $\text{sgn}\frac{\partial\Phi_i(\boldsymbol{r})}{\partial r_i} = \text{sgn}\frac{\partial P(\boldsymbol{r})}{\partial r_i}$, for all $i \in V$. To show this, we first have:

$$\begin{aligned}\frac{\partial\Phi_i(\boldsymbol{r})}{\partial r_i} &= \frac{\partial}{\partial r_i}\Big(U_i(s_i(\boldsymbol{r})) - \frac{1}{\beta}\int_{-\infty}^{r_i} x s_i'(x, r_{-i})dx\Big)\\ &= \frac{\partial s_i(\boldsymbol{r})}{\partial r_i}\Big(U_i'(s_i(\boldsymbol{r})) - \frac{r_i}{\beta}\Big)\\ &= s_i(\boldsymbol{r})\Big(1-s_i(\boldsymbol{r})\Big)\Big(U_i'(s_i(\boldsymbol{r})) - \frac{r_i}{\beta}\Big), \quad (8)\end{aligned}$$

where the last equality comes from a simple calculation:

$$\frac{\partial s_i(r_i, r_{-i})}{\partial r_i} = s_i(r_i, r_{-i})\Big(1 - s_i(r_i, r_{-i})\Big), \quad (9)$$

and second:

$$\frac{\partial P(\boldsymbol{r})}{\partial r_i} = \frac{1}{\beta}\Big(U_i'^{-1}(r_i/\beta) - s_i(\boldsymbol{r})\Big).$$

Thus on the space $\{\boldsymbol{r}|s_i(\boldsymbol{r}) > 0, \text{ for all } i \in V\}$, $\text{sgn}\frac{\partial\Phi_i(\boldsymbol{r})}{\partial r_i} = \text{sgn}\frac{\partial P(\boldsymbol{r})}{\partial r_i}$. From the standard results in potential game theory and strict concavity of $P(\cdot)$, the solution that maximizes $P(\cdot)$ is a NE $\boldsymbol{r}^{\text{NE}}$, where each player's strategy is a *best response* to the others' strategies at NE, and is non-trivial and unique.

(ii) *Price-of-anarchy.* Consider an approximated problem **A-OPT** of **OPT**, given by:

$$\textbf{(A-OPT)} \quad \max_{\boldsymbol{\lambda}\in[0,1]^n, \boldsymbol{\mu}\in\mathcal{P}} \sum_{i\in V} U_i(\lambda_i) - \frac{1}{\beta}\sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)}\mu_{\boldsymbol{\sigma}}\log\mu_{\boldsymbol{\sigma}}$$
$$\text{subject to} \quad \lambda_i \leq \sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)}\mu_{\boldsymbol{\sigma}}\sigma_i, \quad \forall i \in V. (10)$$

Since the objective function is concave and the entropy follows $-\sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)}\mu_{\boldsymbol{\sigma}}\log\mu_{\boldsymbol{\sigma}} \leq \log|\mathcal{I}(G)|$, **A-OPT** problem has a unique solution $(\boldsymbol{\lambda}^o, \boldsymbol{\mu}^o)$ and the solution $\boldsymbol{\lambda}^o$ satisfies that

$$\sum_{i\in V} U_i(\lambda_i^o) \geq \max_{\boldsymbol{\lambda}\in\Lambda} \sum_{i\in V} U_i(\lambda_i) - \frac{\log|\mathcal{I}(G)|}{\beta}. \quad (11)$$

We now consider the Lagrangian L' of **A-OPT** with dual variables $\boldsymbol{k} = [k_i]_{i\in V}$:

$$L'(\boldsymbol{\lambda}, \boldsymbol{\mu}; \boldsymbol{k}) = \sum_{i\in V} U_i(\lambda_i) - \frac{1}{\beta}\sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)}\mu_{\boldsymbol{\sigma}}\log\mu_{\boldsymbol{\sigma}}$$

$$+\sum_{i\in V} k_i\Big(\sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)} \mu_{\boldsymbol{\sigma}}\sigma_i - \lambda_i\Big)$$
$$= \frac{1}{\beta}\left(\sum_{i\in V}\beta k_i\cdot\mathbb{E}_{\boldsymbol{\mu}}[\sigma_i] - \sum_{\boldsymbol{\sigma}\in\mathcal{I}(G)}\mu_{\boldsymbol{\sigma}}\log\mu_{\boldsymbol{\sigma}}\right) + \sum_{i\in V}\left(U_i(\lambda_i) - k_i\lambda_i\right),$$

where $\mathbb{E}_{\boldsymbol{\mu}}[\cdot]$ denotes the expectation for distribution $\boldsymbol{\mu}$. The solution of **A-OPT** is the minimum point of the dual function, which is given by

$$D(\boldsymbol{k}) = \sup_{\boldsymbol{\lambda}\in[0,1]^n,\boldsymbol{\mu}} L'(\boldsymbol{\lambda},\boldsymbol{\mu};\boldsymbol{k}),$$

where $L'(\boldsymbol{\lambda},\boldsymbol{\mu};\cdot)$ is maximized when $\mu^o_{\boldsymbol{\sigma}} = \pi^{\boldsymbol{r}}_{\boldsymbol{\sigma}}$ with $\boldsymbol{r} = \beta\boldsymbol{k}$ (element-wise), and $\lambda_i^o = U_i'^{-1}(k_i)$. Let $\boldsymbol{r}^o$ be the strategy vector such that $\boldsymbol{r}^o = \beta\boldsymbol{k}^o$, where $\boldsymbol{k}^o$ is the solution that minimizes $D(\boldsymbol{k})$, thus

$$s_i(\boldsymbol{r}^o) = U_i'^{-1}(r_i^o/\beta) = \lambda_i^o \quad \text{for all } i\in V. \tag{12}$$

This completes the proof, because the strategy vector $\boldsymbol{r}^{\text{NE}}$ maximizing the potential function $P(\boldsymbol{r})$ coincides with the strategy vector $\boldsymbol{r}^o = \beta\boldsymbol{k}^o$ that minimizes $D(\boldsymbol{k})$, and thus (7) holds from (11) and (12). ■

V. Distributed Learning Dynamics

A. Challenges and Approaches

In Section IV, we show that our game has desirable equilibrium properties, such as uniqueness and (asymptotically) no price-of-anarchy (thus asymptotic utility optimality). We now aim at developing dynamic learning algorithms that operate in a fully-distributed manner, yet converge to the unique non-trivial equilibrium. By "fully-distributed", we mean that users (or players) update their strategies without any message passing among them, relying only on pure local information and local observations.

In achieving our goal, major challenges as well as our approaches to overcome them are summarized in what follows:

- *Hardness of convergence to NE.* It has been known that it is generally hard for a fully-distributed algorithm to converge to a NE. As discussed in [14] that for a broad class of games, there exists no generalized algorithm which operates even in a "partially"-distributed manner (*i.e.*, based on observation of other players' payoff, and thus with message passing), converging to a NE. We overcome this challenge by using the unique property of our game that the payoff function depends only on the local strategy and the marginal distribution (*i.e.*, service rates for a given strategy profile), not directly on other players' strategies. The advantage of this *indirect coupling* allows each user to exploit the *locally-observed service rate* rather than other players' strategies or experienced utilities (as done in classical fully-distributed game-based learning) to update its strategy over time.
- *Long convergence time for classical dynamics.* In updating strategies over time, the locally-observed service rate is not the actual marginal distribution, because after a strategy (*i.e.*, an intensity) is played, it takes long time to reach the stationary regime. In other words, the observed service rates may be far from the 'stationary' service rates. This time-scale issue incurs additional challenges of extremely long convergence times for classical game dynamics, because a certain amount of time (formally called mixing time) to reach the stationary regime is required for each strategy update, and for convergence, long strategy update cycles are necessary. This challenge prevents us from applying the classical dynamics, *e.g.*, best response dynamics. We tackle this challenge by adopting a special learning dynamics, called *stochastically-approximated dynamics* that utilize the time-aggregated service rates in the strategy updates.

B. Three Stochastically-Approximated Dynamics

We provide three stochastically-approximated (SA) dynamic algorithms, all of which provably converge to the unique non-trivial NE: *(i)* ***SA-BRD*** (SA-Best Response Dynamic), *(ii)* ***SA-JD*** (SA-Jacobi Dynamic), and *(iii)* ***SA-GD*** (SA-Gradient Dynamic). In all three dynamics, time is divided into discrete frames $t = 0, 1, \ldots$, where the frame duration is fixed by, say the time to transmit MAC packets of a fixed size. We first let $\bar{s}_i(t)$ and $\hat{s}_i(t)$ be the *aggregate* and *instantaneous* service rate of player i until and at frame t, respectively, *i.e.*,

$$\bar{s}_i(t) = \frac{1}{t}\sum_{n=0}^{t-1}\hat{s}_i(n),$$

where $\hat{s}_i(t)$ denotes the number of transmitted packets at link (or player) i over frame t. Hence, $\bar{s}_i(t)$ can be locally maintained.

(i) SA-BRD. A simple learning dynamic is the best response that each player chooses the best strategy given strategy vector (at the previous frame) of other players:

$$r_i(t+1) = \text{BR}_i(r_{-i}(t)) := \arg\max_{r_i\in\mathbb{R}} \Phi_i(r_i, r_{-i}(t)),$$

which leads to a fixed point of following function:

$$r_i(t+1) = \beta U_i'\Big(s_i(r_i(t+1), r_{-i}(t))\Big). \tag{13}$$

As we mentioned earlier, measuring the 'stationary' service rate $s_i(r_i(t+1), r_{-i}(t))$ directly might incur the long convergence issue (*i.e.*, it takes the mixing time of the underlying CSMA Markov chain). Hence, we study a variant, called ***SA-BRD***, which replaces $s_i(r_i(t+1), r_{-i}(t))$ by the aggregate service rate $\bar{s}_i(t)$:

$$r_i(t+1) = \Big[\beta U_i'\Big(\bar{s}_i(t)\Big)\Big]_{r^{\min}}^{r^{\max}}.$$

For the above and following algorithms, where $[\cdot]_a^b \equiv \max(b, \min(a,\cdot))$, we assume that $r^{\min}$ and $r^{\max}$ are the parameters such that $\boldsymbol{r}^{\text{NE}}$ is in $[r^{\min}, r^{\max}]^n$. The explicit values of $r^{\min}$ and $r^{\max}$ can be also computable [38].

(ii) SA-JD. In the standard Jacobi dynamic, each player gradually adjusts its current strategy towards the best response strategy:

$$r_i(t+1) = r_i(t) + \alpha\Big(\mathrm{BR}_i(r_{-i}(t)) - r_i(t)\Big),$$

where $\alpha \in (0,1]$ is called a smoothing parameter[3]. Similarly as ***SA-BRD***, we suggest the following learning dynamics, called ***SA-JD***:

$$r_i(t+1) \quad = \left[r_i(t) + \alpha\Big(\beta U_i'\Big(\bar{s}_i(t)\Big) - r_i(t)\Big)\right]_{r^{\min}}^{r^{\max}},$$

where $\alpha \in (0,1]$ captures how aggressively the dynamic follows the best response dynamic, where $\alpha = 1$ corresponds to ***SA-BRD***.

(iii) SA-GD. The third dynamic is the gradient play [44], which can be viewed as a "better response" dynamic. In the gradient dynamic, each player gradually updates its intensity in a gradient direction:

$$r_i(t+1) = r_i(t) + \alpha \cdot \nabla\Phi_i(r_i(t)),$$

where we consider a smoothing parameter $\alpha \in (0,1]$ and

$$\nabla\Phi_i(\boldsymbol{r}) = \frac{\partial\Phi_i(\boldsymbol{r})}{\partial r_i} = \frac{\partial s_i(\boldsymbol{r})}{\partial r_i}\left(\frac{\beta}{s_i(\boldsymbol{r})} - r_i\right).$$

A nice economic interpretation of the gradient dynamic is that if the marginal utility exceeds the marginal cost, *i.e.*, $\nabla\Phi_i(\boldsymbol{r}) > 0$, link i's intensity is increased and vice versa. Similarly as before, we design the following variant, called ***SA-GD***:

$$r_i(t+1) = \left[r_i(t) + \alpha\frac{\partial s_i(\boldsymbol{r}(t))}{\partial r_i(t)}\Big(\beta U_i'\Big(\bar{s}_i(t)\Big) - r_i(t)\Big)\right]_{r^{\min}}^{r^{\max}}.$$

The following theorem states that all of three dynamics converge to the unique non-trivial NE, which is in turn asymptotically equal to the socially optimum, as discussed in Section IV-C.

Theorem 2 (Convergence) In all of ***SA-BRD***, ***SA-JD***, and ***SA-GD***, $\boldsymbol{r}(t)$ converges to the unique non-trivial NE $\boldsymbol{r}^{\mathrm{NE}}$ in the sense that

$$\lim_{t\to\infty} \boldsymbol{r}(t) = \boldsymbol{r}^{\mathrm{NE}}, \quad \text{component-wise, almost surely.}$$

The proof of Theorem 2 is presented in Appendix. We remark that the convergence of ***SA-BRD*** has been already studied in [37], but from a different perspective. Specifically, the authors propose a steepest ascent algorithm which turns out to be ***SA-BRD***, whereas we derive it via a game-theoretic approach. Therefore, we provide the proof only for ***SA-JD*** and ***SA-GD***. The additional technical challenge dealing with later two dynamics (not existing for ***SA-BRD***) is that they have higher-order temporal dependencies in their updating rules, *i.e.*, use the current strategy $r_i(t)$ for obtaining the next strategy $r_i(t+1)$. To handle the issue, we define a 'virtual' process (see $\nu_i(t)$ in Appendix) and argue its convergence under the relation to that of the original process $\{r_i(t), t \in \mathbb{Z}_{\geq 0}\}$.

[3] Jacobi dynamics generally achieves a smoother move than best response does in case of non-supermodular games which has unique equlibrium. The small smoothing parameter plays the role of compensating for the instability of the best response dynamics, see [43].

VI. Numerical Results

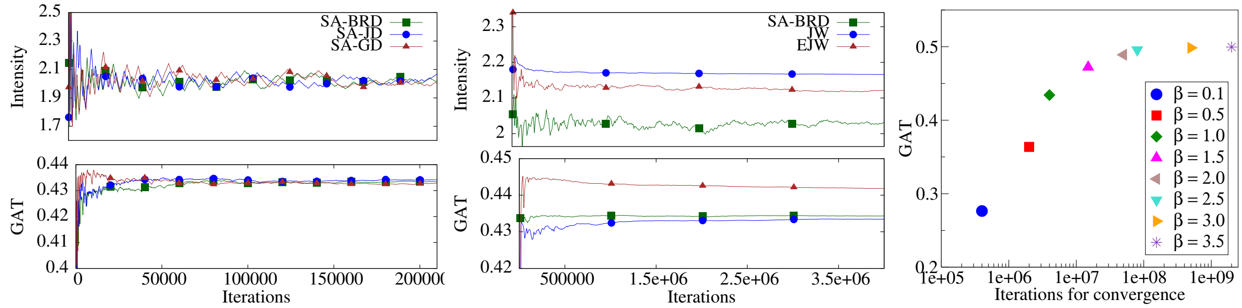
We now provide numerical results that demonstrate our analytical findings. For numerical experiments, we consider proportional fairness across users, *e.g.*, $U_i(\cdot) = \log(\cdot)$ for all users i. We first plot convergence speeds of the learning dynamics designed from our game, in terms of intensity and network utility to support Theorem 2. Second, we compare the convergence speed of the proposed game dynamics and other utility optimal CSMA algorithms in [38] and [19], which we denote by ***JW*** and ***EJW***, respectively. Both compared algorithms iteratively update intensities based on the gradient of the dual problem of **A-OPT** with small and decreasing step size $\alpha_i(t) = 1/t$, and common utility function $U(\cdot)$ as specified by:

$$r_i(t+1) = r_i(t) + \alpha_i(t)\left(U'^{-1}\left(\frac{r_i(t)}{\beta}\right) - \hat{s}_i(t)\right). \tag{14}$$

Note that to guarantee the convergence, the update intervals of ***JW*** increase exponentially so that the $\hat{s}_i(t)$ becomes close to $s_i(\boldsymbol{r}(t))$. Finally, we present the numerical results that show the convergence speeds and price-of-anarchy of proposed learning dynamics from our game oCSMA(β) for various β. Since network utility has negative value in our framework, to get more intuitive values, we use GAT (Geometric Average of user Throughput) instead, which is defined as $(\Pi_{i\in V} s_i(\boldsymbol{r}(t)))^{1/n}$. Note that under the proportional fairness, maiximizing GAT equals to maximizing the aggregate log utilities. The network topology under which our results are presented here is one that leads to the 5×5 grid interference graph. More results are provided in our technical report [45].

*(i) The game dynamics **SA-BRD, SA-JD,** and **SA-GD** converge to the unique non-trivial NE.* Fig. 2(a) demonstrates the Theorem 2, *i.e.*, convergence of intensity and GAT to the unique non-trivial NE under ***SA-BRD, SA-JD*** and ***SA-GD***, where we use $\beta = 1.0$ and $\alpha = 0.5$. We see that all dynamics converge to the same value after long iterations. The convergence speeds of all algorithms do not show much difference.

*(ii) The convergence speed of three proposed learning dynamics is much faster than that of **JW** and **EJW**.* Here, we run the simulation under the same setup as in Fig. 2(a). Fig. 2(b) shows the traces of transmission intensities and GATs of three proposed dynamics, ***JW***, and ***EJW***. Since the convergence patterns of three proposed dynamics are all similar, we plot only ***SA-BRD*** out of three game dynamics and compare with other conventional algorithms. Regarding the intensity, we observe that ***SA-BRD*** converges within 4×10^6 frames, while ***JW*** and ***EJW*** require more than 10^7 iterations. The GAT of ***SA-BRD*** also converges faster than that of other algorithms. Although the intensities of ***JW*** and ***EJW*** seem to converge in



(a) Convergence of SA-based learning dynamics: intensity and GAT. (b) Comparison of SA-BRD with JW and EJW. (c) Tradeoff: efficiency and convergence speed.

Fig. 2. Numerical results for 5×5 grid interference graph

early iterations, they are still growing down very slowly to the point to which ***SA-BRD*** converges.

(iii) Tradeoff between efficiency and convergence speed. As stated in Theorem 1, PoA of ***SA-BRD***, ***SA-JD*** and ***SA-GD*** is asymptotically $1/\beta$. To support it through numerical examples, we vary β and plot the GAT at the converged non-trivial NE. For the relation between convergence speed and β, we also measure the convergence time to reach the NE. Fig. 2(c) shows that, as β grows, SA-based dynamics require exponentially long time to reach the equilibrium point, and the corresponding point becomes closer to the socially optimal point. According to the numerical experiments, the GAT with $\beta = 3.0$ is 0.4986 and converges after almost 5×10^8 iterations, while that with $\beta = 1.0$ is 0.4342 and converges after 4×10^6 iterations.

VII. Conclusion

Despite a large array of game-theoretic studies on wireless MAC, to the best of knowledge, this is the first game-theoretic work that is utility optimal over the maximum throughput region in wireless multi-hop networks. We start by designing a CSMA game whose equilibrum properties such as uniqueness and price-of-anarchy are first analyzed, and we propose three game-learning dynamics based on the idea of stochastic approximation technique. Our theoretical findings exploit the unique features of CSMA, where the price function is smartly designed so that (non-trivial) NE is unique and is very close to the socially optimal point as well as fully-distributed dynamics are feasible.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (NRF-2013R1A2A2A01067633), and the Center for Integrated Smart Sensors funded by the MSIP as Global Frontier Project (CISS-2012M3A6A6054195).

Appendix: Proof of Theorem 2

Due to space limitation, we focus on summarizing the key proof procedures.

Proof of convergence of SA-JD and SA-GD. We first define following sequence $\{\boldsymbol{\nu}(t), t \in \mathbb{N}\}$ as:

$$\nu_i(t) \quad = \quad r_i(t) - (1-\alpha) r_i(t-1),$$

$$r_i(t) \quad = \quad \alpha \frac{\nu_i(t)}{\alpha} + (1-\alpha) r_i(t-1). \qquad (15)$$

It is noteworthy that $r_i(t)$ is an exponential moving average of $\{\frac{\nu_i(j)}{\alpha} | 1 \le j \le t\}$.

Then, ***SA-JD***'s update rule is represented as

$$\begin{aligned}
&\nu_i(t+1) = \alpha\beta U_i'(\bar{s}_i(t)) \\
&= \alpha\beta U_i'\left(\bar{s}_i(t-1) - \frac{1}{t}(\bar{s}_i(t-1) - \hat{s}_i(t))\right) \\
&\approx \alpha\beta\left(U_i'(\bar{s}_i(t-1)) - \frac{1}{t}(\bar{s}_i(t-1) - \hat{s}_i(t)) U_i''(\bar{s}_i(t-1))\right) \\
&= \alpha\beta U_i'(\bar{s}_i(t-1)) + \frac{1}{t} g(\nu_i(t))(\bar{s}_i(t-1) - \hat{s}_i(t)) \\
&= \nu_i(t) + \frac{1}{t} g(\nu_i(t)) \left(U_i'^{-1}\left(\frac{\nu_i(t)}{\alpha\beta}\right) - \hat{s}_i(t)\right),
\end{aligned}$$

where $g(x) = -\alpha\beta U_i''(U_i'^{-1}(\frac{x}{\alpha\beta}))$, and $g(x) > 0$ due to concavity of utility functions.

Since $r_i \in [r^{\min}, r^{\max}]$, from (15), there exist M and L such that

$$\left|\frac{g(\nu_i(t))}{\alpha}\left(U_i'^{-1}\left(\frac{\nu_i(t)}{\alpha\beta}\right) - \hat{s}_i(t)\right)\right| < M \text{ and } |\nu_i(t)| < L,$$

for all t. For $\varepsilon > 0$, let $T(\varepsilon) := \frac{4\log(\frac{\varepsilon\alpha}{4L})}{\varepsilon \log(1-\alpha)}$. Then, for all $t \ge T(\varepsilon)$, $\left|\frac{\nu_i(t)}{\alpha} - r_i(t)\right| \le \varepsilon \cdot M$, because [4]

$$\begin{aligned}
\left|\frac{\nu_i(t)}{\alpha} - r_i(t)\right| \quad &= \quad \left|\frac{\nu_i(t)}{\alpha} - \sum_{j=0}^{t-1} \frac{\nu_i(t-j)}{\alpha}\alpha(1-\alpha)^j\right| \\
&\le \quad \sum_{j=0}^{t-1}\left|\frac{\nu_i(t)}{\alpha} - \frac{\nu_i(t-j)}{\alpha}\right|\alpha(1-\alpha)^j \\
&\overset{(a)}{\le} \quad \frac{\varepsilon t/4}{t - \varepsilon t/4} M + \frac{2L}{\alpha}\sum_{j=\varepsilon t/4}^{t-1}\alpha(1-\alpha)^j \\
&\le \quad \frac{\varepsilon}{2} M + \frac{2L}{\alpha}(1-\alpha)^{\varepsilon t/4} \;\le\; \varepsilon, \qquad (16)
\end{aligned}$$

where (a) comes from the followings:

$$\sum_{j=0}^{\varepsilon t/4 - 1}\left|\frac{\nu_i(t)}{\alpha} - \frac{\nu_i(t-j)}{\alpha}\right|\alpha(1-\alpha)^j$$

[4] Here, we use just $\varepsilon t/4$ instead of $\lceil \varepsilon t/4 \rceil$ for notionaly simplicity.

$$\leq \sum_{j=1}^{\varepsilon t/4-1} \sum_{k=1}^{j} \left| \frac{\nu_i(t-k+1)}{\alpha} - \frac{\nu_i(t-k)}{\alpha} \right| \alpha(1-\alpha)^j$$

$$\leq \sum_{j=1}^{\varepsilon t/4-1} \frac{j \cdot M}{t-j} \alpha(1-\alpha)^j \;\leq\; \frac{\varepsilon t/4}{t-\varepsilon t/4} M.$$

Therefore, $\lim_{t\to\infty} \frac{\nu_i(t)}{\alpha} - r_i(t) = 0$.

We now, from the discrete-time sequence $\{\boldsymbol{\nu}(t), t \in \mathbb{N}\}$, define a continuous-time interpolation $\bar{\boldsymbol{\nu}}(\cdot)$ with $\tau(k) := \sum_{j=1}^{k} \frac{g(\nu_i(j))}{j}$. We make the continuous-time interpolation such that, for all $t \in [\tau(k), \tau(k+1))$,

$$\bar{\nu}_i(t) = \nu_i(k) + (\nu_i(k+1) - \nu_i(k))(t-k),$$

and also denote $\bar{s}_i(t) = \hat{s}_i(k) \cdot \mathbf{1}_{\tau(k)\leq t<\tau(k+1)}$, where $\mathbf{1}_E$ is the indicator function for the event E.

From the nice properties of the utility function (continuous and twice-differentiable) and by restricting our attention to the compact strategy set $[r^{\min}, r^{\max}]$, it is not hard to check that this algorithm satisfies sufficient conditions of a classical stochastic approximation algorithm with controlled Markov noise [19], [46] for convergence. In particular, we have that when t is large, $\bar{\boldsymbol{\nu}}$ is well approximated by the solution $\tilde{\boldsymbol{\nu}}$ of the following ODE system: for all i,

$$\dot{\tilde{\nu}}_i = g(\tilde{\nu}_i)[U_i'^{-1}\left(\frac{\tilde{\nu}_i}{\alpha\beta}\right) - \sum_{\boldsymbol{\sigma}} \pi_{\boldsymbol{\sigma}}^{\tilde{\boldsymbol{\nu}}/\alpha} \sigma_i]. \tag{17}$$

Furthermore, the above ODE converges to a unique fixed point $\boldsymbol{\nu}^{**}$ such that $\boldsymbol{r}^{**} = \frac{\boldsymbol{\nu}^{**}}{\alpha}$ and

$$\alpha\beta U_i'(s_i(\boldsymbol{r}^{**})) = \alpha\beta U_i'(s_i(\frac{\boldsymbol{\nu}^{**}}{\alpha})) = \nu_i^{**} = \alpha r_i^{**}.$$

Thus, we conclude that $\boldsymbol{r}(t)$ converges to $\boldsymbol{r}^{**}$ such that

$$r_i^{**} = \beta U_i'(s_i(\boldsymbol{r}^{**})),$$

and it is clear that $\boldsymbol{r}^{**} = \boldsymbol{r}^{\mathrm{NE}}$, which completes the proof.

In case of ***SA-GD***, the convergence can be proved similarly to ***SA-JD***, just by considering $\alpha \frac{\partial s_i(\boldsymbol{r})}{\partial r_i}$ instead of α in ***SA-JD***.

References

[1] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling for maximum throughput in multihop radio networks," *IEEE Transactions on Automatic Control*, vol. 37, no. 12, pp. 1936–1949, December 1992.

[2] C. Joo and N. B. Shroff, "Performance of random access scheduling schemes in multi-hop wireless networks," in *Proc. of Infocom*, 2007.

[3] X. Lin and S. Rasool, "Constant-time distributed scheduling policies for ad hoc wireless networks," in *Proc. of CDC*, 2006.

[4] L. Tassiulas, "Linear complexity algorithms for maximum throughput in radionetworks and input queued switches," in *Proc. of Infocom*, 1998.

[5] E. Modiano, D. Shah, and G. Zussman, "Maximizing throughput in wireless networks via gossiping," in *Proc. of Sigmetrics*, New York, NY, USA, 2006.

[6] A. Eryilmaz, A. Ozdaglar, and E. Modiano, "Polynomial complexity algorithms for full utilization of multi-hop wireless networks," in *Proc. of Infocom*, 2007.

[7] S. Sanghavi, L. Bui, and R. Srikant, "Distributed link scheduling with constant overhead," in *Proc. of Sigmetrics*, 2007.

[8] P. Chaporkar, K. Kar, and S. Sarkar, "Throughput guarantees through maximal scheduling in wireless networks," in *Proc. of Allerton*, 2005.

[9] X. Lin and N. B. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proc. of Infocom*, 2005.

[10] X. Wu and R. Srikant, "Bounds on the capacity region of multi-hop wireless networks under distributed greedy scheduling," in *Proc. of Infocom*, 2006.

[11] L. Jiang and J. Walrand, *Scheduling and Congestion Control for Wireless and Processing Networks*. Morgan-Claypool, 2010.

[12] Y. Yi and M. Chiang, "Stochastic network utility maximization: A tribute to kelly's paper published in this journal a decade ago," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 421–442, 2008.

[13] N. Li and J. R. Marden, "Designing games for distributed optimization," *J. Sel. Topics Signal Processing*, vol. 7, no. 2, pp. 230–242, 2013.

[14] S. Hart and A. M. Colell, "Uncoupled dynamics do not lead to nash equilibrium," *American Economic Review*, vol. 93, no. 5, pp. 1830–1836, 2003.

[15] P. H. Young, "Learning by trial and error," *Games and Economic Behavior*, vol. 65, no. 2, pp. 626–643, March 2009.

[16] Z. Ghahramani, "Graphical models: Parameter learning," in *Handbook of Brain Theory and Neural Networks*, 2003.

[17] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Computation.*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.

[18] L. Jiang and J. Walrand, "A distributed CSMA algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 3, pp. 960 –972, June 2010.

[19] J. Liu, Y. Yi, A. Proutiere, M. Chiang, and H. V. Poor, "Towards utility-optimal random access without message passing," *Wiley Journal of Wireless Communications and Mobile Computing*, vol. 10, no. 1, pp. 115–128, Jan. 2010.

[20] P. A, Y. Yi, T. Lan, and M. Chiang, "Resource allocation over network dynamics without timescale separation," in *Proc. of Infocom*, 2010.

[21] S. Rajagopalan, D. Shah, and J. Shin, "Network adiabatic theorem: An efficient randomized protocol for contention resolution," in *Proc. of Sigmetrics*, 2009.

[22] J. Ni, B. Tan, and R. Srikant, "Q-csma: Queue-length based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks," in *Proc. of Infocom*, 2010.

[23] A. MacKenzie and S. Wicker, "Stability of multipacket slotted aloha with selfish users and perfect information," in *Proc. of Infocom*, 2003.

[24] E. Altman, R. El-Azouzi, and T. Jimenezy, "Slotted aloha as a game with partial information," *Computer Networks*, vol. 45, no. 6, pp. 701–713, 2004.

[25] Y. Jin and G. Kesidis, "Equilibria of a noncooperative game for heterogeneous users of an aloha network," *IEEE Communications Letters*, vol. 6, no. 7, pp. 282–284, 2002.

[26] G. Kesidis, Y. Jin, A. P. Azad, and E. Altman, "Stable nash equilibria of aloha medium access games under symmetric, socially altruistic behavior," in *Proc. of CDC*, 2010.

[27] I. Menache and N. Shimkin, "Fixed-rate equilibrium in wireless collision channels," *Network Control and Optimization (NET-COOP)*, pp. 23–32, 2007.

[28] Y. Jin, G. Kesidis, and J. W. Jang, "A channel aware mac protocol in an aloha network with selfish users," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 128–137, 2012.

[29] M. Cagalj, S. Ganeriwal, I. Aad, and J. P. Hubaux, "On smart cheating in csma/ca networks," in *Proc. of Infocom*, 2005.

[30] Y. Jin and G. Kesidis, "Distributed contention window control for selfish users in ieee 802.11 wireless lans," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 6, pp. 1113–1123, 2007.

[31] J. W. Lee, A. Tang, J. Huang, M. Chiang, and A. Calderbank, "Reverse-engineering mac: a non-cooperative game model," *IEEE Journal on Selected Areas in Communications*, vol. 25, pp. 1135–1147, 2007.

[32] J. W. Lee, M. Chiang, and R. A. Calderbank, "Utility-optimal medium access control: reverse and forward engineering," in *Proc. of Infocom*, 2006.

[33] T. Cui, L. Chen, and S. Low, "A game-theoretic framework for medium access control," *IEEE Journal on Selected Areas in Communications*, vol. 26, pp. 1116–1127, 2008.

[34] L. Chen, S. Low, and J. Doyle, "Random access game and medium access control design," *IEEE/ACM Trans. Netw.*, vol. 18, no. 4, pp. 1303–1316, Aug. 2010.

[35] A. H. M. Rad, J. Huang, M. Chiang, and V. Wong, "Utility-optimal random access: Reduced complexity, fast convergence, and robust performance," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 898–911, 2009.

[36] A. H. M. Rad, J. Huang, M. Chiang, and V. W. S. Wong, "Utility-optimal random access without message passing," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1073–1079, 2009.

[37] N. Hegde and A. Proutiere, "Simulation-based optimization algorithms with applications to dynamic spectrum access," in *Proc. of CISS*, 2012.

[38] L. Jiang, D. Shah, J. Shin, and J. Walrand, "Distributed random access algorithm: Scheduling and congestion control," *IEEE Transactions on Information Theory*, vol. 56, no. 12, pp. 6182 –6207, dec. 2010.

[39] J. Lee, J. Lee, Y. Yi, S. Chong, and M. Chiang, "Implementing utility-optimal csma," in *Proceedings of Allerton Conference*, 2009.

[40] B. Nardelli, J. Lee, K. Lee, Y. Yi, S. Chong, E. W. Knightly, and M. Chiang, "Experimental evaluation of optimal csma," in *Proc. of Infocom*, 2011.

[41] J. Lee, H. Lee, Y. Yi, S. Chong, B. Nardelli, E. W. Knightly, and M. Chiang, "Making 802.11 dcf near-optimal: Design, implementation, and evaluation," in *Proc. of SECON*, 2013.

[42] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[43] L. Chen, T. Cui, S. H. Low, and J. C. Doyle, "A game-theoretic model for medium access control," in *Proc. of WICON*, 2007.

[44] S. D. Flam, "Equilibrium, evolutionary stability and gradient dynamics," *International game Theory Review*, vol. 4, no. 4, pp. 357–370, 2002.

[45] H. Jang, S. Yun, J. Shin, and Y. Yi, "Adaptive CSMA for utility maximization: A game-theoretic approach," Dept. of EE, KAIST, http://lanada.kaist.ac.kr/pub/ocsma_game.pdf, Tech. Rep., 2013.

[46] V. S. Borkar, *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, 2008.