

Throughput Enhancement of Full-Duplex CSMA Networks Using Multiplayer Bandits

Jingwen Tong, *Student Member, IEEE*, Liqun Fu^{id}, *Senior Member, IEEE*, and Zhu Han^{id}, *Fellow, IEEE*

Abstract—This article studies the network-level throughput of a full-duplex (FD)-enabled CSMA network, considering the link's transmit power (TP) control, carrier-sensing threshold (CST) adjustment, and logarithm access intensity (LAI) adaptation. With the FD technique, a transmitter–receiver pair can transmit and receive simultaneously in the same frequency band. We aim to find the best combination of TP, CST, and LAI for each link to maximize the FD-CSMA network throughput. However, adjusting each link's TP and CST will change the network's carrier-sensing relation or contention graph, consequently leading to a computationally intractable network optimization problem. On the other hand, it is difficult to jointly optimize these three parameters in a fully distributed network. To overcome these, we first decompose this network optimization problem into two subproblems: 1) a joint control and scheduling problem in the transport- and media access control (MAC)-layer and 2) a parameter selection problem in the PHY-layer. Then, the multiplayer multiarmed bandit (MPMAB) framework has been introduced to address this problem by solving the two subproblems alternately. We put forth a fully distributed algorithm, named the stochastic and adversarial optimal FD-CSMA (SAO-FD-CSMA) algorithm, to solve the MPMAB problem by taking advantage of the optimization tool and the bandit theory. The numerical results show that the proposed algorithm outperforms the state-of-the-art bandit algorithms and can improve the network throughput by 43% compared with the random selection method.

Index Terms—CSMA, full duplex (FD), multiplayer multiarmed bandit (MPMAB), stochastic and adversarial optimal FD-CSMA (SAO-FD-CSMA) algorithm.

Manuscript received October 12, 2020; revised December 24, 2020 and February 24, 2021; accepted March 17, 2021. Date of publication March 23, 2021; date of current version July 23, 2021. The work of Liqun Fu was supported by the National Natural Science Foundation of China under Grant 61771017. The work of Zhu Han was supported by the U.S. National Science Foundation under Grant EARS-1839818, Grant CNS-1717454, Grant CNS-1731424, and Grant CNS-1702850. This article was presented in part at the IEEE Global Communications Conference, Big Island, HI, USA, Dec. 2019. (*Corresponding author: Liqun Fu.*)

Jingwen Tong is with the School of Informatics, Xiamen University, Xiamen 361005, China, and also with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA (e-mail: tongjingwen@stu.xmu.edu.cn).

Liqun Fu is with the School of Informatics and the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology Ministry of Education, Xiamen University, Xiamen 361005, China (e-mail: liqun@xmu.edu.cn).

Zhu Han is with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Seoul 446-701, South Korea (e-mail: hanzhu22@gmail.com).

Digital Object Identifier 10.1109/JIOT.2021.3068182

I. INTRODUCTION

CARRIER-SENSING multiple access with collision avoidance (CSMA/CA), which coordinates multiple-users' transmissions on the same channel, is a widely used random media access control (MAC) protocol in distributed communication systems [1]–[4]. The conventional CSMA networks are normally built on half-duplex (HD) radios. Recently, full duplex (FD), which enables a node to receive and transmit simultaneously in the same frequency band, has been considered as a new technique to boost the network throughput in future wireless systems [5].

The study of FD-enabled CSMA networks has attracted much research attention in recent years [6]–[10]. Compared with the traditional HD-CSMA networks, there are two distinct features in FD-CSMA networks. The most worthwhile feature is imperfect self-interference cancelation, which is the major problem that causes an FD-enabled link to prevent doubling its transmission rate. Particularly, when a link employs a high transmit power (TP), the impact of the residual self-interference (RSI) can be more severe [9]. The other feature is that an FD-link's spatial interference footprint is much heavier than an HD-link [8]. The reason is that an FD-link with two concurrent active nodes will cause a large carrier-sensing and interference range. As a result, the spatial reuse (SR), which allows multiple links to communicate concurrently in a limited area, is decreased. Furthermore, this influence can be more serious when the network density is high [10].

In light of the above problems, this article focuses on improving the network-level throughput of an FD-CSMA network by considering TP control, carrier-sensing threshold (CST) adjustment, and logarithm access intensity (LAI) adaptation. The ideas are that:

- 1) changing TP can reduce the interlink interference or improve the transmission rate;
- 2) adjusting CST can increase the number of concurrent transmission links;
- 3) tuning LAI can influence the CSMA parameters, which changes the proportion of each link's transmission per-unit time.

However, adjusting each link's TP and CST will change the carrier-sensing relation or contention graph, leading to a computationally intractable network optimization problem. On the other hand, it is difficult to jointly optimize these three parameters in a fully distributed network. To overcome these, we decompose this network optimization problem into two subproblems: 1) joint control and scheduling problem in the transport- and MAC-layer (i.e., subproblem 1) and 2) a

parameter selection problem in the physical (PHY)-layer (i.e., subproblem 2).

For the joint control and scheduling problem, a fully distributed HD-CSMA algorithm has been introduced in [11] to achieve the optimal performance by adaptively tuning the LAI parameter. Furthermore, Xie and Zhang [7] extended this algorithm to the FD-CSMA network along with some modifications to the CSMA protocol. For the parameter selection problem, many works consider the TP and CST selection problem in the HD-CSMA network. Gurses and Boutaba [12] and Yang *et al.* [13] investigated the multi-hop HD-CSMA networks capacity by considering the TP, CST, and medium access probability of p -persistent CSMA adjustment. Kim *et al.* [14] and Zhou *et al.* [15] targeted on improving the HD-CSMA network SR by tuning TP, CST, and data rate. However, these papers suffer high computational complexity caused by determining the feasible states repeatedly in the contention graph [16] and require some information exchange among links, increasing the system overhead. Therefore, it is critical to jointly optimize the TP, CST, and LAI parameters by solving these two subproblems efficiently, while maintaining a low computational complexity and system overhead in such a distributed network.

Recently, the multiplayer multiarmed bandit (MPMAB) framework, which captures the exploration and exploitation features among multiple players and actions, attracts increasing research attention in the communication society, such as resource allocation [17]–[19], and dynamic spectrum access [20]. There are three types of MPMAB frameworks in the literature.

- 1) All players have the same set of arms, two or more players selecting the same arm will cause collision without reward and the mean of the rewards at each arm is different for different players.
- 2) All players have the same set of arms, two or more players who select the same arm will cause collision without reward, but the mean of the rewards at each arm is the same for different players.
- 3) Each player has a different set of arms and independently chooses an arm from its local backlog, and the rewards of different players will not collide but may be relevant.

Most of the existing MPMAB works fall into the type 1) and need some information passing among players to coordinate their selection strategies.

In this article, we model subproblem 2 as an MPMAB framework; while subproblem 1 is considered as a mechanism that to produce the received rewards. In this MPMAB framework, the players are the FD-enabled links, the arms are the combinations of TP and CST, and the rewards are the links' normalized throughput. It belongs to the type 3) since we assume that each player has a private set of arms. Therefore, the collision will not happen because each player independently selects an arm from its local cache, and no information exchange among links is required. A closely related work was found in [21], considering the merits and demerits of SR in dense HD Wireless Local Area Networks (WLANs) by applying the MPMAB model. However, it only provide some numerical results without any theoretical analysis. Unlike

in [21], we give an upper regret bound for the proposed algorithm and clarify that the environment of this MPMAB problem is between stochastic and adversary. Moreover, we focus on improving the network-level throughput of an FD CSMA network by jointly considering TP, CST, and LAI.

We first design an optimal FD-CSMA algorithm for the subproblem 1 at each link to produce the received rewards. It turns out that the optimal FD-CSMA algorithm can converge to the optimal LAI by using a simple subgradient method [11]. Then, a stochastic and adversarial optimal (SAO) algorithm has been proposed for the subproblem 2 to find the best TP and CST pair for each link. The SAO algorithm will converge when the time horizon T is sufficiently large. At last, we put forth a stochastic and adversarial optimal FD-CSMA (SAO-FD-CSMA) algorithm to approximately address the throughput maximum problem by solving the subproblem 1 and the subproblem 2 alternately. In other words, the SAO-FD-CSMA algorithm is a combination of the optimal FD-CSMA algorithm and the SAO algorithm. Therefore, it can converge to the optimal LAI and the best combination of TP and CST, while achieving a sublinear regret.

The main contributions of this work are summarized in the following.

- 1) We study the network-level throughput of an FD-CSMA network by considering the TP control, CST adjustment, and LAI adaptation. We first formulate it as a utility maximum problem by using the *piecewise constant rate* and the continued time-reversible Markov network (CTMN) models.
- 2) Thereafter, we decompose it into two subproblems, i.e., a joint control and scheduling problem in the transport- and MAC-layer, and a parameter selection problem in the PHY-layer. We propose an optimal FD-CSMA algorithm for subproblem 1 and an SAO algorithm for subproblem 2, respectively. By merging the above two algorithms, we put forth an SAO-FD-CSMA algorithm to address the network throughput optimization problem.
- 3) We give a convergence analysis for the optimal FD-CSMA algorithm and derive an upper regret bound for the SAO-FD-CSMA algorithm.
- 4) To evaluate the proposed algorithm, we design a discrete event simulator (DESIM) to realize the FD-CSMA protocol.

The remainder of this article is organized as follows. Section II summarizes the related work and Section III introduces the system model. The problem formulation is given in Section IV. In Section V, we present an SAO-FD-CSMA algorithm for the throughput maximum problem. Section VI analyzes the performance of the SAO-FD-CSMA algorithm. Simulation results are presented in Section VII, and this article is concluded in Section VIII.

II. RELATED WORK

There is a wide arrange of works that study achieving the optimal throughput for the CSMA networks, which can be roughly classified into three groups, i.e., the optimization-based optimal CSMA, the game theory-based

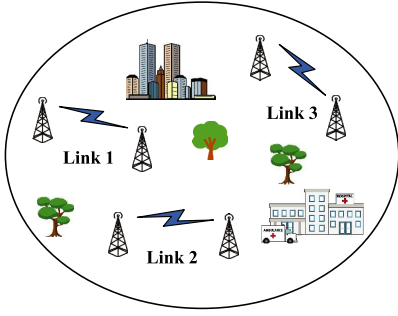


Fig. 1. Scenario of an FD-enabled CSMA network.

optimal CSMA, and the machine learning-based optimal CSMA. The optimization-based optimal CSMA was first considered by Tassiulas and Ephremides [22], where a maximal weight scheduling (MWS) has been introduced to achieve throughput optimum. But this centralized scheme requires to solve a maximum-weight independent set, which is an NP-hard problem [11]. To overcome this complexity, some low complexity algorithms for certain interference models [23] and low complexity algorithms that guarantee a portion of the capacity region [24] or a worst case performance [25] have been proposed. However, these approaches require certain information exchange among users, increasing system overhead and implementation complexity. In 2008, a simple, throughput optimally algorithm, called adaptive CSMA, was introduced in [11] that no information passing among links is required. Thereafter, increasing researches are interested in this so-called optimal CSMA [2], [26]. A suboptimal CSMA scheme has already been adopted for practical applications [27]. The game theory-based optimal CSMA was considered in [4], [16], and [28]. For a survey of game theory-based optimal CSMA, we refer authors to [28].

The machine learning-based optimal CSMA usually studies the objective of SR. Wilhelmi *et al.* [21] considered the channel, TP, and CST selection problem in IEEE 802.11 WLANs by using the bandit theory. Before this article, [29], with the same authors of [21], has investigated the SR in wireless networks by using a variety of bandit algorithms, including the ϵ -greed algorithm, the upper confidence bound (UCB1) [30], the exponential-weight algorithm for exploration and exploitation (Exp3) [31], and the Thompson sampling (TS) algorithm [32]. It states that the TS algorithm has the best performance in the wireless networks (without CSMA protocol), thereby it was directly applied in [21]. However, we note that the reward in the CSMA network is neither stochastic nor adversary because of the complicated interactions among links. As a result, the TS algorithm and the Exp3 algorithm can not directly apply here, especially when the number of links is large.

The problem of how to design an optimal bandit algorithm in both stochastic and adversarial regimes was first introduced in [33]. Then, a practical algorithm for both stochastic and adversarial regimes, called Exp3++ algorithm, has been proposed in [34] to achieve almost optimal performance. Based on this article, [31] also presents an algorithm that can achieve near optimal performance in both stochastic and

adversarial regimes. They derive a logarithmic pseudo-regret bound for the stochastic case and a polynomial pseudo-regret bound for the adversarial case. These works are all based on the Exp3 algorithm. However, [35] points out that the Exp3 algorithm is a special case of the implicitly normalized forecaster (INF) algorithm (or online mirror descent algorithm) by using negative entropy. Recently, based on the Tsallis entropy and the INF algorithm, [36] proposes an optimal algorithm (called the Tsallis-INF algorithm) for both stochastic and adversarial regimes. The numerical results show that the Tsallis-INF algorithm is better than the Exp3++ algorithm. In this article, we propose an SAO-FD-CSMA algorithm to solve this stochastic and adversarial bandit problem. The SAO-FD-CSMA algorithm is based on the Exp3 algorithm, which can significantly improve the network performance and outperforms the state-of-the-art bandit algorithms, such as the INF algorithm, TS algorithm, and UCB1 algorithm.

III. SYSTEM MODEL

We consider an FD-enabled CSMA network with K links distributively located in an area, as shown in Fig. 1. These links share the same frequency band and operate with the CSMA protocol. Each link consists of an FD-enabled transmitter-receiver pair (i.e., the Tx-Rx pair). With FD capability, both the transmitter and receiver can transmit and receive on the same band simultaneously. When the CSMA protocol is adopted, the carrier-sensing and random back-off mechanism are employed to coordinate the transmissions on different links.

A. Signal Model

Let P_k be the TP at the Tx-Rx pair of FD link k . Different FD links may use different TPs. We assume that the channels between $\text{Tx} \rightarrow \text{Rx}$ and $\text{Tx} \leftarrow \text{Rx}$ are symmetric. Thus, the received signals at Rx_k and Tx_k of link k are

$$\begin{cases} Y_{\text{Rx}_k} = X_{\text{Tx}_k} + \text{SI}_{\text{Rx}_k} + \text{AI}_{\text{Rx}_k} + n, & \text{Tx}_k \rightarrow \text{Rx}_k \\ Y_{\text{Tx}_k} = X_{\text{Rx}_k} + \text{SI}_{\text{Tx}_k} + \text{AI}_{\text{Tx}_k} + n, & \text{Rx}_k \rightarrow \text{Tx}_k \end{cases} \quad (1)$$

where X_{Rx_k} and X_{Tx_k} represent the transmit signals of link k . Without loss of generality, assume that both X_{Rx_k} and X_{Tx_k} follow the Gaussian distribution with zero mean and the same variance $P_k G_{\text{Tx}_k, \text{Rx}_k}$, where $G_{\text{Tx}_k, \text{Rx}_k}$ is the channel gain between Rx_k and Tx_k . In addition, n is the background noise with a nominal power of σ_n^2 .

As for FD communications, the RSI on each link that caused by the RF leakage and the imperfect self-interference cancellation, needs to be considered. Terms SI_{Tx_k} and SI_{Rx_k} in (1) represent the RSI attribute to nodes Tx_k and Rx_k , respectively. We assume that the Tx-Rx pair has the same RSI cancelation capability. According to [37], it is reasonable to assume that RSI follows a Rayleigh distribution with zero mean and variance $\chi_k P_k$, where χ_k is the self-interference suppression ratio. In addition, AI_{Tx_k} and AI_{Rx_k} are the accumulative interfering signals imposed on Tx_k and Rx_k , coming from other concurrent active in-band links. Then, the interference powers of AI_{Tx_k} and AI_{Rx_k} can be calculated as

$$\sigma_{I, \text{Tx}_k}^2 = \sum_{j \in \Lambda, j \neq k} P_j (G_{\text{Tx}_k, \text{Rx}_j} + G_{\text{Tx}_k, \text{Tx}_j}) \quad (2)$$

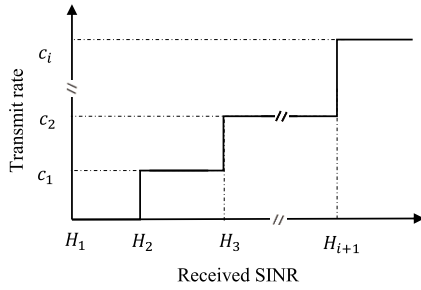


Fig. 2. Transmit rate versus received SINR.

and

$$\sigma_{I, \text{Rx}_k}^2 = \sum_{j \in \Lambda, j \neq k} P_j (G_{\text{Rx}_k, \text{Rx}_j} + G_{\text{Rx}_k, \text{Tx}_j}) \quad (3)$$

where P_j is the TP of link j . Note that set Λ can be determined by the carrier-sensing graph, which will be introduced in the forthcoming section. The received signal-to-interference-plus-noise ratios (SINRs) at Tx_k and Rx_k of link k are, respectively

$$\gamma_k^{\text{tx}} = \frac{P_k G_{\text{Tx}_k, \text{Rx}_k}}{\chi_k P_k + \sigma_{I, \text{Tx}_k}^2 + \sigma_n^2} \quad (4)$$

and

$$\gamma_k^{\text{rx}} = \frac{P_k G_{\text{Tx}_k, \text{Rx}_k}}{\chi_k P_k + \sigma_{I, \text{Rx}_k}^2 + \sigma_n^2}. \quad (5)$$

The achievable transmission rate depends on the received SINR and corresponds to the selected modulation and coding scheme [38]. In general, a link can only support a limited number of transmission rates. In this article, we adopt the *piecewise constant rate model* in [38] to identify each link's transmission rate according to their received SINR. Specifically, assume that link k has an available set of transmission rates, \mathbf{C}_k in ascending order, i.e., $c_1 < c_2 < \dots < c_{|\mathbf{C}_k|}$. Each c_i corresponds to an SINR interval $[H_i, H_{i+1}]$, as shown in Fig. 2. Thus, each SINR can map to a corresponding transmission rate. Note that, for an FD link, the transmission rates between $\text{Tx} \rightarrow \text{Rx}$ and $\text{Tx} \leftarrow \text{Rx}$ is different since the interference experienced by Rx and Tx may not be the same. For example, if $\gamma_k^{\text{rx}} \in [H_i, H_{i+1}]$ and $\gamma_k^{\text{tx}} \in [H_j, H_{j+1}]$, the achievable rates of $\text{Tx}_k \rightarrow \text{Rx}_k$ and $\text{Tx}_k \leftarrow \text{Rx}_k$ are c_i and c_j , respectively. Therefore, the total transmission rate of link k is

$$v_k = c_i + c_j. \quad (6)$$

B. Carrier Sensing Graph and Feasible Transmission State

In the 802.11 CSMA protocol, a node needs to sense the channel states before initiating its transmission. If the sensing result is idle, it waits a distributed interframe spacing time. It then uniformly selects a back-off time counter value from the range of $[0, \text{CW}]$, where CW refers to the contention window which is initially set to CW_{\min} . The back-off counter value is decremented by one for each slot if the channel remains idle. When the back-off counter reduces to zero, this station starts its transmission. If the channel is sensed to be busy before

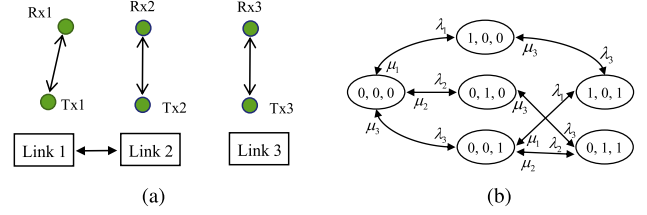


Fig. 3. (a) Three FD-links' network topology and the corresponding carrier-sensing graph. (b) State transition diagram of the time-reversible Markov process of the left network.

the back-off counter value reduces to zero, it is frozen and resumes when the channel is detected to be idle again.

The carrier-sensing operation of the FD-CSMA networks is different from that of the HD-CSMA networks. Here, we introduce the *pairwise carrier-sensing model* to capture the carrier-sensing relation. Specifically, consider two FD links, i.e., links i and j . We say link i can sense link j , if the following inequality is satisfied:

$$P_j (G_{\text{Tx}_i, \text{Rx}_j} + G_{\text{Tx}_i, \text{Tx}_j}) \geq S_i \quad (7)$$

where S_i is the CST^1 of link i . In a general network with multiple links, we define a carrier-sensing graph as an undirected graph, $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, to capture the carrier-sensing relation between links. In this graph, links are viewed as a set of vertices denoted by \mathcal{V} , and \mathcal{E} is the set of edges, which can be further written as a K -by- K matrix. That is, $\mathcal{E} = [\mathbf{e}_1; \mathbf{e}_2; \dots; \mathbf{e}_K]$, where the k th row vector is $\mathbf{e}_k = [e_{k,1}, e_{k,2}, \dots, e_{k,K}]$. If $e_{i,j} = 1$, this means that link i can sense link j , and $e_{i,j} = 0$ otherwise. We further assume that the sensing is symmetric, i.e., $e_{i,j} = e_{j,i}$.

Based on the carrier-sensing graph, we define the feasible transmission state as a subset of vertices such that no edge connects any two of them. To avoid some ambiguity with the notion Λ , we assume that there are N feasible transmission states in \mathcal{G} , denoted by $\mathbf{F} = \{\mathbf{f}^i, i = 0, 1, \dots, N-1\}$, where $\mathbf{f}^i = \{f_k^i, k = 1, 2, \dots, K\}$ is the i th feasible state. If $f_k^i = 1$, it means that link k can be active in the i th feasible state, and $f_k^i = 0$ otherwise. Note that a network with K links has total 2^K states, but the number of feasible states can be much less than this value because of the contentions among links. For example, there are three links in Fig. 3(a). The total number of states is $2^3 = 8$, but the number of feasible transmission states is only 6 because link 1 and link 2 cannot transmit simultaneously. The six feasible states are

$$\mathbf{F} = \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 0, 1), (0, 1, 1)\}. \quad (8)$$

Here, an empty state $(0, 0, 0)$ means that the original state of a network is $(0, 0, 0)$, i.e., no link is active.

C. Time-Reversible Markov Network and Throughput Calculation

Our work can fit into the "ideal CSMA" network ICN model [11]. As shown in Fig. 3(b), the transition among

¹We obscure the notions of energy detection threshold and CST by using CST for clear channel assessment in the CSMA protocol.

feasible states in ICN forms a CTMN. Therefore, the transmission durations of all feasible states can be obtained through computing its stationary distribution.

Let T_{cd} be the back-off counter value, which is a random variable with an arbitrary distribution. So the birth rate of a link (i.e., a link from idle state to transmission state) is $\lambda = 1/\mathbb{E}[T_{cd}]$, where $\mathbb{E}[\cdot]$ is the expectation operator. Similarly, the death rate of a link (i.e., a link from transmission state to an idle state) is $\mu = 1/\mathbb{E}[T_{tr}]$, where T_{tr} is the transmission time of a data packet with an arbitrary distribution [11]. We define a LAI ρ on a link as the ratio of its mean transmission time to its mean back-off time, i.e.,

$$\rho = \log\left(\frac{\mathbb{E}[T_{tr}]}{\mathbb{E}[T_{cd}]}\right) = \log\left(\frac{\lambda}{\mu}\right). \quad (9)$$

Specifically, let λ_k and μ_k denote the birth rate and death rate of link k , respectively. Given a set of LAIs $\boldsymbol{\rho} = \{\rho_1, \rho_2, \dots, \rho_K\}$, the stationary distribution of the feasible state \mathbf{f}^i is

$$\pi_{f^i} = \frac{1}{B_r} \exp\left(\sum_{k=1}^K f_k^i \rho_k\right), \quad i = 0, 1, \dots, N-1 \quad (10)$$

where

$$B_r = \sum_{j=0}^{N-1} \exp\left(\sum_{k=1}^K f_k^j \rho_k\right). \quad (11)$$

Note that the external summation of (11) operates on the set of all feasible transmission states in \mathbf{F} . Recall the state transition diagram in Fig. 3(b) and leveraging (10), the probabilities of the six feasible states in (8) are

$$\begin{aligned} \pi_{\mathbf{F}} &= \{\pi_{000}, \pi_{100}, \pi_{010}, \pi_{001}, \pi_{101}, \pi_{011}\} \\ &= \pi_{000} \left\{ 1, \frac{\lambda_1}{\mu_1}, \frac{\lambda_2}{\mu_2}, \frac{\lambda_3}{\mu_3}, \frac{\lambda_1\lambda_3}{\mu_1\mu_3}, \frac{\lambda_2\lambda_3}{\mu_2\mu_3} \right\} \end{aligned} \quad (12)$$

where $\pi_{000} = 1/B_r$, i.e.,

$$\pi_{000} = \frac{1}{1 + \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} + \frac{\lambda_3}{\mu_3} + \frac{\lambda_1\lambda_3}{\mu_1\mu_3} + \frac{\lambda_2\lambda_3}{\mu_2\mu_3}}. \quad (13)$$

The throughput of link k in a feasible state \mathbf{f}^i is $v_{k,i}\pi_{f^i}$ when link k is active (i.e., $f_k^i = 1$), where $v_{k,i} \in \mathbf{C}_k$ is the transmission rate at feasible state i of link k [21]. Then through summing over all feasible states, the throughput of link k can be obtained as

$$\Gamma_k = \sum_{i=0}^{N-1} v_{k,i} f_k^i \pi_{f^i} = \frac{1}{B_r} \sum_{i=0}^{N-1} v_{k,i} \prod_{k=1}^K f_k^i e^{f_k^i \rho_k} \quad (14)$$

where $v_{k,i} \in \mathbf{C}_k$ and $f_k^i \in \mathbf{F}$ are determined by the carrier sensing graph \mathcal{G} . Therefore, Γ_k is a function of ρ_k , P_k , and S_k .

IV. PROBLEM FORMULATION

We aim to improve the network-level throughput of an FD-enabled CSMA network by adjusting each link's TP, CST, and LAI. This problem (refer to a master problem) can be

formulated as the following optimization form:

$$\begin{aligned} (\mathbf{M}) \quad & \max_{\rho_k, P_k, S_k} \sum_{k \in \mathcal{V}} \log \Gamma_k \\ \text{s.t.} \quad & \Gamma_k = \sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} \\ & \rho_k \in \mathbb{R}^+, P_k \in \mathbf{P}_k, \text{ and } S_k \in \mathbf{S}_k \end{aligned} \quad (15)$$

where \mathbf{P}_k and \mathbf{S}_k are the available set of TP and CST on link k , respectively. Here, we consider the proportional fairness by using the logarithmic basis function \log . The objective function in the problem formulation together with our method can be generalized to any network utility function, as long as it is concave, increasing, and differentiable.

It is difficult to get an analytical solution for (15) because of the implicit relationship among P_k , S_k , v_k , and ρ_k . Also, the finite discrete values of P_k and S_k will result in a nonconvex problem of (15). Furthermore, there is no message passing among links in such a fully distributed CSMA network. To overcome these challenges, we propose to design a fully distributed online algorithm to approximately solve this throughput maximum problem. In order to better understanding the mechanism of the proposed online algorithm, we decompose the master problem (i.e., problem **M**) into two subproblems, i.e., subproblem 1 and subproblem 2. The subproblem 1 is a joint congestion control problem at the transport-layer and scheduling problem at the MAC-layer, accounting for the optimality of link's LAI when the TP and CST pair is fixed; while the subproblem 2 is a parameter selection problem at the PHY-layer by considering the TP control and CST adjustment, when the link's LAI is given. In the following, we show how to construct and solve the subproblem 1 and the subproblem 2 by using the optimization tool and the bandit theory.

A. Subproblem 1: Joint Congestion Control and Scheduling Problem

We first to obtain the subproblem 1 from the master problem (15) by fixing the TP and CST parameters.

Lemma 1: Given the link's TP and CST, problem **M** can be reformulated as a joint congestion control and scheduling problem, i.e.,

$$\begin{aligned} (\mathbf{P1}) \quad & \max_{\rho_k} V \sum_{k \in \mathcal{V}} \log \Gamma_k - \sum_{i=0}^{N-1} \pi_{f^i} \log \pi_{f^i} \\ \text{s.t.} \quad & \Gamma_k \leq \sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} \quad \forall k \in \mathcal{V} \\ & \sum_{i=0}^{N-1} \pi_{f^i} = 1 \text{ and } \rho_k \in \mathbb{R}^+ \end{aligned} \quad (16)$$

where V is a positive constant weighting factor.

Proof: See Appendix A. ■

According to [39], it is easy to prove that problem (16) is convex. Thus, the Lagrange function of (16) is given by

$$\begin{aligned} L(\Gamma, \pi; \beta, \eta) = & V \sum_{k \in \mathcal{V}} \log \Gamma_k - \sum_{i=0}^{N-1} \pi_{f^i} \log \pi_{f^i} \\ & + \sum_{k \in \mathcal{V}} \beta_k \left(\sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} - \Gamma_k \right) \\ & - \eta \left(\sum_{i=0}^{N-1} \pi_{f^i} - 1 \right) \end{aligned} \quad (17)$$

where β_k and η are the dual variables. Then, the Karush–Kuhn–Tucker (KKT) conditions of (16) are

$$\sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} - \Gamma_k \geq 0 \quad \forall k \in \mathcal{V} \quad (18)$$

$$\sum_{i=0}^{N-1} \pi_{f^i} = 1 \quad (19)$$

$$\beta_k \geq 0, \quad \forall k \in \mathcal{V} \quad (20)$$

$$\beta_k \left(\sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} - \Gamma_k \right) = 0, \quad \forall k \in \mathcal{V} \quad (21)$$

$$\frac{V}{\Gamma_k} - \beta_k = 0 \quad \forall k \in \mathcal{V} \quad (22)$$

$$-1 - \log \pi_{f^i} + \sum_{k \in \mathcal{V}} \beta_k v_{k,i} - \eta = 0, \quad i = 0, 1, \dots, N-1. \quad (23)$$

By solving the above equations, we obtain that

$$\eta^* = \log \left(\sum_{i=0}^{N-1} \exp \left(\sum_{k \in \mathcal{V}} \beta_k v_{k,i} \right) \right) - 1 \quad (24)$$

and

$$\pi_{f^i}^* = \frac{\exp(\sum_{k \in \mathcal{V}} \beta_k v_{k,i})}{\sum_{i=0}^{N-1} \exp(\sum_{k \in \mathcal{V}} \beta_k v_{k,i})}, \quad i = 0, 1, \dots, N-1. \quad (25)$$

In addition, by solving (21) and (22), a subgradient of dual variable β_k is obtained by

$$\dot{\beta}_k = \left[\frac{V}{\beta_k} - \sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} \right]^{Q_r} \quad (26)$$

where Q_r is the projection operation that bounds β_k into the interval of $[\beta_{\min}, \beta_{\max}]$. Both the dual variable β_k and the prime variable π_{f^i} have specific physical meanings, where β_k can be regarded as a virtual queue of link k and π_{f^i} is viewed as the probability of feasible transmission state f^i . These observations can be benefit in designing a distributed algorithm for problem **P1** in Section V-A.

B. Subproblem 2: Parameter Selection Problem

The goal of the parameter selection problem is to determine the best TP and CST for each FD-link when the LAI is given,

which can be directly written as

$$\begin{aligned} (\mathbf{P2}) \quad & \arg \max_{P_k, S_k} \sum_{k \in \mathcal{V}} \log \Gamma_k \\ \text{s.t.} \quad & \Gamma_k = \sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} \quad \forall k \in \mathcal{V} \\ & P_k \in \mathcal{P}_k, \text{ and } S_k \in \mathcal{S}_k. \end{aligned} \quad (27)$$

This is a nonconvex problem. Meanwhile, adjusting each link's TP and CST will change the network's carrier-sensing relation and contention graph, leading to a computationally intractable optimization problem. Therefore, the online learning method can be adopted to tradeoff the exploration and exploitation dilemma, i.e., each link needs to explore all TP and CST pairs sufficiently and to exploit the current best TP and CST pair as much as possible so as to reduce the performance loss. In fact, we incorporate **P2** into the MPMAB framework, where the FD-links are the players and the combinations of TP and CST are the arms. The rewards are the link's normalized throughput, which are generated by solving the subproblem 1. So it is easy to bound it in the range of $[0, 1]$. Time t is slotted and each player aims to selfishly maximize its total reward (namely, the link's accumulated throughput) up to time horizon T . In the following, we give some notations and terminologies to the MPMAB framework.

Let $\mathbf{K} = \{1, \dots, K\}$ be the set of players, referring to all the FD-links. Each player has an available set of arms to choose from, i.e., the combinations of TP and CST. Let $\mathcal{A}_k = \{a_{k,1}, a_{k,2}, \dots, a_{k,|\mathcal{A}_k|}\}$ denote the arm collection of link k . Note that set \mathcal{A}_k is the Descartes product of sets \mathcal{P}_k and \mathcal{S}_k , i.e., $\mathcal{A}_k = \mathcal{P}_k \otimes \mathcal{S}_k$, where $\mathcal{P}_k = \{P_{k,1}, P_{k,2}, \dots, P_{k,|\mathcal{P}_k|}\}$ and $\mathcal{S}_k = \{S_{k,1}, S_{k,2}, \dots, S_{k,|\mathcal{S}_k|}\}$. At the beginning of each time slot t , each player chooses an arm and then observes a reward. We assume that the time slot is sufficiently large enough so that the Markov chain can reach its steady state. In the simulation part, each time slot consists of 40 epoches.

The reward of link k by selecting arm i is defined as its normalized throughput

$$r_{a_{k,i}} \doteq \frac{\Gamma_{a_{k,i}}}{\Gamma_k^*} \quad (28)$$

where $\Gamma_k^* = \max_i \Gamma_k(a_{k,i})$, $i = 1, \dots, |\mathcal{A}_k|$, is the maximum observed throughput that link k experienced. Define the total rewards of link k by

$$R_k \doteq \sum_{t=1}^T r_{k, \xi_t^k}(t) \quad (29)$$

where ξ_t^k is the index of the selected arm at time slot t of link k .

Define the regret of link k as its performance metric by

$$\mathcal{R}eg_k \doteq \max_i \sum_{t=1}^T (r_{k,i}(t) - r_{k, \xi_t^k}(t)) \quad \forall k \in \mathbf{K} \quad (30)$$

where $i = 1, \dots, |\mathcal{A}_k|$. Here, we adopt the definition of *weak regret* [40] for our bandit problem, which is the difference between the best arm's rewards and the currently received

Algorithm 1 Optimal FD-CSMA Algorithm Run by Link k

- 1: **Initialize:** CSMA parameters λ_k , μ_k , hyper-parameter V , v , Q_r
- 2: For each epoch b , generate a packet with mean $1/\lambda_k$ and exponentially transmission duration (in terms of the number of slots) with mean $1/\mu_k$
- 3: Continuously run the FD-CSMA protocol
- 4: Record the number of successful transmissions N_r
- 5: Update the virtual queue β_k according to:
- 6: $\beta_k(b+1) \leftarrow [\beta_k(b) + v(b)(V/\beta_k(b) - v_k N_r)]^{Q_r}$
- 7: Adjust λ_k and μ_k by using $\beta_k = (1/v_k) \log(\lambda_k/\mu_k)$
- 8: Return to step 2

rewards. Therefore, the total *weak regret* of the MPMAB problem is simply computed as

$$\text{Reg} \doteq \sum_{k=1}^K \text{Reg}_k. \quad (31)$$

Note that, although we adopt the regret definition of adversarial MAB, the type of the rewards in the FD-CSMA network is between stochastic and adversary.

V. FD-SAO-CSMA ALGORITHM

In this section, we give an optimal FD-CSMA algorithm for **P1** in Section V-A and an SAO algorithm for **P2** in Section V-B. Thereafter, we present an SAO-FD-CSMA algorithm in Section V-C to approximately address the master problem by solving **P1** and **P2** alternately. These algorithms are all distributed and easy to implement.

A. Optimal FD-CSMA Algorithm for Subproblem 1

We first design a subgradient method for **P1** by using the subgradient of dual variable β_k in (26), which can converge to the optimal LAI. To achieve this, link k just needs to record the number of successful transmissions (accounting for the term $\sum_{i=0}^{N-1} \pi_{f^i}$) at each epoch. Then, it updates β_k by using (26). Through calculating the stationary distribution of the feasible state f^i in (10) and the equation of prime variable π_{f^i} in (25), we obtain that $\beta_k v_k = \rho_k$, i.e., $\beta_k = (1/v_k) \log(\lambda_k/\mu_k)$. At last, link k adjusts its CSMA parameters μ and λ such that $\beta_k = (1/v_k) \log(\lambda_k/\mu_k)$.

The optimal FD-CSMA algorithm was given in Algorithm 1, operated by link k . At the beginning of each epoch b , link k runs the FD-CSMA protocol which generates a packet with mean $1/\lambda_k$ and exponentially transmission duration with mean $1/\mu_k$. Then, link k records the number of successful transmissions, denoted by N_r . At the end of each epoch, link k updates its virtual queue β_k and the CSMA parameters according to the subgradient method and expression $\beta_k = (1/v_k) \log(\lambda_k/\mu_k)$, respectively. For the subgradient method, v is the step size that controls the convergence rate of Algorithm 1. Note that v_k is the data rate that does not change during each epoch.

We design a CSMA DESim to evaluate Algorithm 1 with MAC/PHY parameters (e.g., time slot, CW range) in 802.11g [39] based on the MATLAB platform. Table I shows the normalized throughput of different contention graphs in

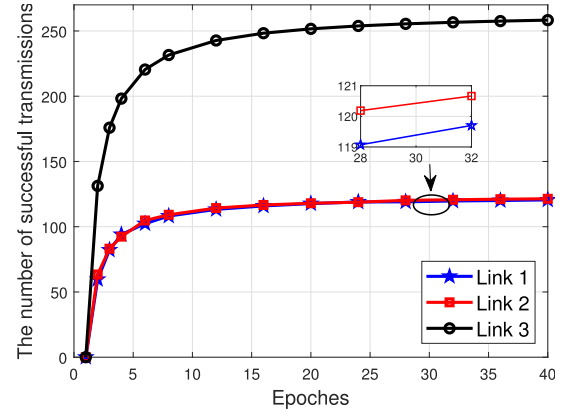


Fig. 4. Number of successful transmissions via epochs by running Algorithm 1 with contention graph of network topology shown in Fig. 3(a).

paper [41] (see Fig. 2) by using the DESim, back-of-the-envelope (BoE), and CTMN methods in HD-CSMA network with fixed LAI. All results are obtained running over 1000 Monte Carlo simulations. It can be seen that the normalized throughput of DESim is very close to that of BoE method in all the cases.² This indicates that the designed DESim method can essentially capture the features of the CSMA protocol.

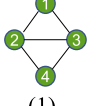
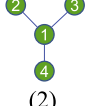
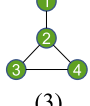
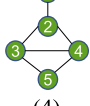
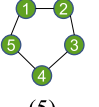
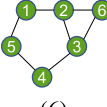
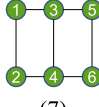
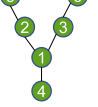
Based on the designed DESim, we then run the optimal CSMA algorithm under the FD-CSMA network with the contention graph of network topology in Fig. 3(a). Note that, by using the DESim, we can easily adapt the HD-CSMA network to the FD-CSMA network. Fig. 4 shows that the number of successful transmissions received in an epoch changes with the epoches when the step size $v = 0.01$ and the constant $V = 10^6$. The total epoches are $B_s = 40$ and the length of each epoch is $40\,000\,\mu s$. The projection range of dual variable β_k is set that $\beta_{\min} = \max\{0, \log(\mu_k^0 \log(10/9)/CW)\}$ and $\beta_{\max} = \log(\mu_k^0 \log 10)$, where μ_k^0 is the initial value of the expected transmission time of link k . From Fig. 4, we can see that the number of successful transmissions at three links increases with epoches, and finally reach their maximum value. This states that the optimal FD-CSMA algorithm can maximize the network throughput by adaptively adjusting each link's LAI parameter. Also, it can be seen that the performance of links 1 and 2 are about half of the link 3, since there is a contention relationship between links 1 and 2 as in Fig. 3(a).

B. Stochastic and Adversarial Optimality Algorithm for Subproblem 2

In bandit theory, the player's goal is to sequentially pick up the best action from a known set so as to maximize its accumulated reward. At each time slot, the player draws an arm and observes a reward from the environment. According to the received rewards, MAB problem can be typically classified into two groups, i.e., the stochastic MAB and the adversary MAB. However, we notice that, in the FD-CSMA network, the rewards (i.e., the normalized throughput) are neither stochastic nor adversary. This is because that the interactions among links

²For convenience, we just present the main results in here. More details please see [41] Fig. 2.

TABLE I
NORMALIZED THROUGHPUT VERSUS DIFFERENT CONTENTION GRAPHS OF NETWORK TOPOLOGIES
(AS IN [41] FIG. 2) BY USING THE DESIM, BoE, AND CTMN METHODS

# Contention graph	 (1)	 (2)	 (3)	 (4)
BoE	(1, 0, 0, 1)	(0, 1, 1, 1)	(1, 0, 0.5, 0.5)	(0.75, 0.25, 0.25, 0.25, 0.5)
DESIM	(0.96, 0.01, 0.01, 0.96)	(0.00, 0.98, 0.98, 0.98)	(0.98, 0.00, 0.60, 0.56)	(0.81, 0.26, 0.32, 0.31, 0.55)
CTMN	(0.74, 0.49, 0.49, 0.75)	(0.18, 0.86, 0.85, 0.85)	(0.95, 0.47, 0.71, 0.72)	(0.67, 0.50, 0.49, 0.50, 0.66)
# Contention graph	 (5)	 (6)	 (7)	 (8)
BoE	(0.4, 0.4, 0.4, 0.4, 0.4)	(1, 0, 0, 1, 0, 1)	(0.5, 0.5, 0.5, 0.5, 0.5, 0.5)	(0.2, 0.4, 0.4, 0.8, 0.6, 0.6)
DESIM	(0.47, 0.46, 0.47, 0.46, 0.45)	(0.94, 0.01, 0.02, 0.94, 0.02, 0.94)	(0.50, 0.48, 0.47, 0.51, 0.51, 0.48)	(0.22, 0.49, 0.48, 0.84, 0.66, 0.67)
CTMN	(0.61, 0.61, 0.61, 0.60, 0.61)	(0.85, 0.11, 0.11, 0.86, 0.14, 0.86)	(0.85, 0.84, 0.62, 0.62, 0.84, 0.87)	(0.48, 0.60, 0.60, 0.84, 0.75, 0.74)

form a game environment. In other words, each link only aims to selfishly maximize its own accumulated rewards, regardless of others.

Therefore, we propose an SAO algorithm for **P2**, as shown in Algorithm 2. The SAO algorithm can achieve the optimal regret bounds of $\mathcal{O}(\sqrt{KT \log T})$ for the adversary MAB and $\mathcal{O}(\sqrt{\sum_i \log T / \Delta_i})$ for the stochastic MAB, where Δ_i is the performance difference between the i th arm and the optimal arm. From Algorithm 2, we can see that the SAO algorithm is proceeding in phases. The length of each phase is controlled by the estimated total rewards R_k and the exponentially increasing bound θ_z . At each time slot, the player selects an arm (i.e., the combination of TP and CST) according to probability distribution q_k . Then, it receives the normalized throughput as a reward $r_{k,\xi_t^k}(t)$ bounded in $[0, 1]$. By observing this reward, it can update its selection strategy as in Exp3 algorithm. However, although our SAO algorithm is based on the Exp3 algorithm, but has the following features.

First, the SAO algorithm starts with a prior knowledge of the weight at each arm, which can help accelerate the convergence rate. The empirical probability mass function (PMF) of each arm of link k is given by

$$q_{k,i}(t) = (1 - \chi_z) \frac{\omega_{k,i}(t)}{\sum_{j \in \mathcal{A}} \omega_{k,j}(t)} + \frac{\chi_z}{|\mathcal{A}_k|} \quad (32)$$

where

$$\omega_{k,i}(t) = \hat{\omega}_{k,i}(1) \exp\left(\frac{\chi_z}{|\mathcal{A}_k|} \hat{R}_{k,i}(t)\right). \quad (33)$$

Thus, $q_{k,i}(t)$ is a function of the updated weight $\omega_{k,i}$ and the learning rate χ_z ; while the $\omega_{k,i}(t)$ is a function of the initial weight $\omega_{k,i}(1)$, the learning rate χ_z , and the accumulated estimated reward $\hat{R}_{k,i}(t)$. Therefore, if the initial weight

Algorithm 2 SAO Algorithm Run by Link k

```

1: Initialize:  $\hat{R}_{k,i}(1) = 0 \quad \forall i \in \mathcal{A}_k$ ,
2: Input:  $\omega_{k,i}(1)$  with prior information  $\hat{\omega}_{k,i}(1)$ 
3: for each phase  $z = 0, 1, 2, \dots$  do
4:   Set  $\chi_z$  and  $\theta_z$  according to (34) and (35)
5:   while  $\max_i \hat{R}_{k,i}(t) \leq \theta_z - |\mathcal{A}_k|/\chi_z$  do
6:     Compute all arms' PMF  $q_{k,i}(t)$  using (32)
7:     Draw an arm  $\xi_t^k$  according to  $q_k(t)$ 
8:     Run FD-CSMA protocol with DESIM
9:     Observe a transmission outcome  $r_{k,\xi_t^k}(t)$ 
10:    Compute the estimated reward of each arm:

$$\hat{r}_{k,i}(t) = \frac{r_{k,i}(t)}{q_{k,i}(t)} \mathbb{1}_{\xi_t^k=i} \quad \forall i \in \mathcal{A}_k$$

11:    Calculate the estimated total reward of each arm:

$$\hat{R}_{k,i}(t+1) = \hat{R}_{k,i}(t) + \hat{r}_{k,i}(t) \quad \forall i \in \mathcal{A}_k$$

12:    Update each arm's weight using (33)
13:    Update the time slot:  $t = t + 1$ 
14:   end while
15: end for

```

is replaced by the *prior* knowledge [i.e., $\hat{\omega}_{k,i}(1)$], Algorithm 2 can converge fast with a proper χ_z .

Second, we provide an upper bound for the maximum estimated accumulated rewards to reduce its variance by adjusting χ_z . According to (32), each player should start with a big χ_z to explore all the arms often enough; then it decrements by time slot to exploit the good estimation arm more greedily. Thus, χ_z can be selected as

$$\chi_z = \min \left\{ 1, \sqrt{\frac{|\mathcal{A}_k| \log(|\mathcal{A}_k|)}{(e-1)\theta_z}} \right\} \quad (34)$$

Algorithm 3 SAO-FD-CSMA Algorithm Run by Link k

- 1: **Initialize:** Parameters in Algorithm 1 and Algorithm 2
- 2: Choose a pair of TP and CST according to q_k
- 3: Determine the transmission data rate v_k
- 4: Run the optimal FD-CSMA algorithm in Algorithm 1
- 5: Compute link's normalized throughput r_k
- 6: Run the SAO algorithm for one slot (i.e., Algorithm 2)
- 7: Update arms' probability distribution q_k
- 8: Return to step 2

where

$$\theta_z = \frac{|\mathcal{A}_k| \log(|\mathcal{A}_k|)}{(e-1)} 4^z. \quad (35)$$

By substituting θ_z into (34), we have $\chi_z = 2^{-z}$ which is decreasing exponentially. When z is sufficiently large, the second term of (32) will tend to 0. As a result, the empirical PMF can converge to the actual PMF. That is, Algorithm 2 can choose the optimal arm with a high probability. Specifically, let $\hat{R}_{k,i}(t+1) = \sum_{s=1}^t \hat{r}_{k,i}(s)$ and according to $\hat{r}_{k,i}(t) = [r_{k,i}(t)/q_{k,i}(t)]\mathbb{1}_{\xi_t^k=i}$, it is easy to obtain that $\mathbb{E}[\hat{r}_{k,i}(t)] = r_{k,i}(t)$ and

$$\text{Var}[\hat{R}_{k,i}(t+1)] = \sum_{s=1}^t \left(\frac{1 - q_{k,i}(s)}{q_{k,i}(s)} r_{k,i}^2(s) \right) \quad (36)$$

where $\text{Var}[\cdot]$ is the variance operation. So we conclude that increasing the selection probability will reduce the estimation variance of $\hat{R}_{k,i}$, and hence a tighter upper regret bound can be achieved.

C. SAO-FD-CSMA Algorithm

We put forth an SAO-FD-CSMA algorithm, as shown in Algorithm 3, to address the master problem of (15) by running Algorithms 1 and 2 alternately. In the following, we show the feasibility of the proposed algorithm in the practical implementation and introduce its time structure.

All FD-links adopt the CSMA protocol to coordinate their transmissions as in [7]. In this distributed network, each FD-link runs the SAO-FD-CSMA algorithm asynchronously. First, each FD-link picks up a pair of TP and CST from its local cache. Note that the sets of TP and CST can be different for different FD-links. Then, according to the modulation scheme and the received SINR, the PHY-layer can determine a proper data rate for each FD-link. Meanwhile, the MAC-layer and the transport-layer need to activate the links with a high data rate and less contention so as to maximize the network throughput. In other words, through observing the number of successful transmissions, each FD-link can find the optimal LAI by running Algorithm 1 to solve subproblem 1. At last, by using the SAO bandit algorithm, each link can find the best TP and CST pair when the time horizon is sufficiently large. The detailed operation steps are given in Algorithm 3. Therefore, the proposed SAO-FD-CSMA algorithm is fully distributed and easy to implement in the practical network.

Fig. 5 is the time structure of Algorithm 3. We can see that each time slot includes B_s epoches, used to transmit data and to find the optimal LAI by running Algorithm 1. At the beginning of each time slot, the SAO bandit algorithm is used

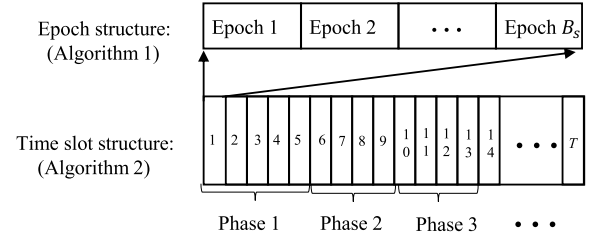


Fig. 5. Time structure of the SAO-FD-CSMA algorithm.

to choose a proper TP and CST pair for each link. During each epoch, the link runs the optimal FD-CSMA algorithm to find the CSMA protocol's optimal LAI. Then, the normalized throughput generated by solving subproblem 1 is used to calculate arms' probability distribution, i.e., q_k , which in turn influences the selection of TP and CST pair. To sum up, the proposed SAO-FD-CSMA algorithm is fully distributed and easy to implement in the practical network.

VI. PERFORMANCE ANALYSIS

In this section, we give a convergence analysis for the optimal FD-CSMA algorithm and an upper regret bound for the SAO-FD-CSMA algorithm. Since the SAO-FD-CSMA algorithm is a combination of Algorithms 1 and 2, the analysis of the upper regret bound of the SAO-FD-CSMA algorithm can be much different from the results in [40]. We need to provide a convergence analysis for Algorithm 1 before constructing this upper regret bound.

According to [42] and [43], the optimal CSMA algorithm can be regarded as a stochastic approximation algorithm with controlled Markov noise. The main difficulty in analyzing the convergence of Algorithm 1 is the fact that the updates in the virtual queues, and hence in the CSMA parameters, depending on the random service processes (N_r) at the receiver and transmitter. To proof the convergence of Algorithm 1, we need the following assumption.

Assumption 1: If $\beta_k^0 \in \mathbb{R}_+$ $\forall k \in \mathcal{V}$ solves, $\beta_k^0 v_{k,i} = V \exp(\sum_{i=0}^{N-1} f_k^i \pi_{f_i})$, then $\beta^{\min} < \beta_k^0 < \beta^{\max} \forall k \in \mathcal{V}$.

As a result, we can give the convergence result of the optimal FD-CSMA algorithm as following.

Theorem 1: Assume $\sum_0^\infty v(b) = \infty$ and $\sum_0^\infty v(b)^2 < \infty$. Under Assumption 1, for any initial condition $\beta_k \forall k \in \mathcal{V}$, optimal FD-CSMA converges in the following sense:

$$\lim_{b \rightarrow \infty} \beta_k(b) = \beta_k^* \text{ and } \lim_{b \rightarrow \infty} \Gamma_k(b) = \Gamma_k^*, \text{ almost surely}$$

where β_k^* and Γ_k^* are the solution of subproblem 1 in (16).

Proof: See the proof of [39, Th. 1]. ■

Remark 1: By choosing diminishing step-sizes, Algorithm 1 will converge to the optimal throughput. This means that there is an upper bound for each link's average throughput. In other words, the rewards of the SAO-FD-CSMA algorithm are bounded and can be normalized into the range $[0, 1]$. This observation is critical to derive an upper regret bound for the SAO-FD-CSMA algorithm.

Remark 2: We can choose a proper hyper-parameter V to control Algorithm 1's convergence rate. Specifically, a large

TABLE II
FD-CSMA NETWORK SIMULATION PARAMETERS

Symbol	Description	Value
\mathbf{P}_k	set of TP	(10, 15, 20) dBm
\mathbf{S}_k	set of CST	(-70, -80, -90) dBm
\mathbf{C}_k	set of data rate	(20, 50, 100, 150) Mbps
f_c	center frequency	5 GHz
B	bandwidth	40 MHz
t	length of time slot	9 μ s
DIFS/SIFS	DIFS and SIFS duration	34 μ s / 16 μ s
CW	contention window	32
L_{Data}	length of a data packet	12000 bits
L_{ACK}	length of ACK packet	304 bits
RTS/CTS	length of RTS and CTS	160 bits / 112 bits
Epoch	length of a data frame	4×10^4 μ s

value V tends to improve the efficiency of the algorithm; while the downside is that a large V decreases the convergence rate. Therefore, a proper V can be selected by carefully tradeoff the efficiency and the convergence rate of Algorithm 1.

Next, we give an upper regret bound for the SAO-FD-CSMA algorithm. Unlike the traditional MAB problems, rewards are well-defined in the case of the stochastic or adversarial bandit. Here, the rewards are between the stochastic bandit and adversarial bandit. On the other hand, the only constraint on the rewards is that they are bounded in the range of $[0, 1]$, which holds based on Theorem 1. For convenience, we mark $R_{k,\max} = \max_{j \in \mathcal{A}_k} \sum_{t=1}^T r_{k,j}(t)$ and $M_k = |\mathcal{A}_k|$. Then, the upper regret bound of the SAO-FD-CSMA algorithm can be constructed in Theorem 2.

Theorem 2: For any $M_k > 0$ where $k \in \mathbf{K}$

$$\begin{aligned} \text{Reg} \leq & \sum_{k \in \mathbf{K}} \left(8\sqrt{e-1} \sqrt{R_{k,\max} M_k \ln M_k} + 8(e-1)M_k \right. \\ & \left. + 2M_k \ln M_k \right) \end{aligned} \quad (37)$$

holds for any assignment of rewards and $T > 0$.

Proof: See Appendix B. ■

Remark 3: Theorem 2 shows that the regret of link k up to T will not exceed $\mathcal{O}(\sqrt{R_{k,\max} M_k \ln M_k})$. Compared with $\mathcal{O}(\sqrt{TM_k \ln M_k})$, which is the upper regret bound of the original Exp3, it is substantial tighter since $R_{k,\max} \leq T$ always holds. More importantly, $R_{k,\max} \leq T$ also indicates that when $T \rightarrow \infty$ the per-round network-level regret will tend to 0. In other words, the proposed algorithm will converge when the total time slot T is sufficiently large.

VII. NUMERICAL ANALYSIS

In this section, we conduct several simulations to evaluate the performance of the SAO-FD-CSMA algorithm and to compare the performances of FD-CSMA networks with HD-CSMA networks. The CSMA network parameters are chosen according to IEEE 802.11ax standard [44], as shown in Table II. All numerical results come from at least 1000 Monte Carlo trials.

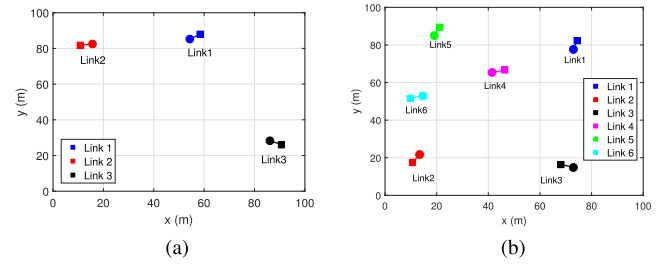


Fig. 6. (a) Network with three FD-links uniformly distributed in a (100×100) m² area. (b) Network with six FD-links uniformly distributed in a (100×100) m² area.

We first consider two network scenarios, where there are $K = 3$ and $K = 6$ FD-links uniformly distributed³ in a (100×100) m² area, as shown in Fig. 6(a) and (b), respectively. The transmitter (Tx) of each link was first generated uniformly in this square area. Then, we place the receivers (Rx) such that the angle and distance parameters of each link are generated from uniform distributions with $[0, 2\pi]$ and $[d_{\min}, d_{\max}]$, respectively. We adopt the exponent path-loss channel model in the simulation. Thus, the channel gain $G_{\text{Tx}_k, \text{Rx}_k}$ is calculated by $P_k D_0 d(\text{Tx}_k, \text{Rx}_k)^{-\alpha}$, where α is the path-loss exponent and $d(\text{Tx}_k, \text{Rx}_k)$ stands for the Euclidean distance between Tx_k and Rx_k of link k . The reference path gain D_0 is computed by $(G_t G_r l^2) / ((4\pi d_0)^2 L)$, where G_t and G_r are the transmit and receive gains, respectively. In addition, L is the loss of system hardware, and d_0 is the reference distance with unit m. Here, we set $G_t = 1$, $G_r = 1$, $L = 1$, and $d_0 = 1$ m. The term l is the wavelength of the central carrier frequency of 5 GHz. We assume that all links have the same available set of transmission rates, TPs, and CSTs, as given in Table II. Thus, the number of the combinations (or arms) of TPs and CSTs at each link is 3×3 . The background noise level is set to -95 dBm. Then, the received SINR at the TX and the RX node of an isolated FD link is at least 105 dB. As a result, it is reasonable to set the RSI cancellation parameter χ_k as 100 dB, for all $k \in \mathbf{K}$.

For the SAO-FD-CSMA algorithm, the parameters of step size ν , constant V , and the projection range of dual variable β_k are the same as in that of Fig. 4. We start with an optimistic initial value for the SAO-FD-CSMA algorithm to accelerate the convergence rate. The optimistic initial value can be some *prior* knowledge which usually comes from the device's self-configuration phase and proceeds during the network initialization. Thanks to MPMAB framework's online learning feature, the *prior* knowledge can be easily integrated into the SAO-FD-CSMA algorithm. In simulation, we give higher weights for the arms that performed better in the historical simulations as the *prior* knowledge. Each time slot consists of 40 time frames.⁴ In addition, Algorithm 2 precedes in phases and updates to next phase depending on the maximum estimated accumulated rewards and according to (34) and (35). Moreover, we adopt the round-Robin method to form

³Notice that the links' geographic position can be arbitrary distribution.

⁴To reduce the simulation time, we use total 40 epoches to approximately approach the optimal network throughput when the TP and CST are given. This approximation error is small when the number of epoches exceeds 40.

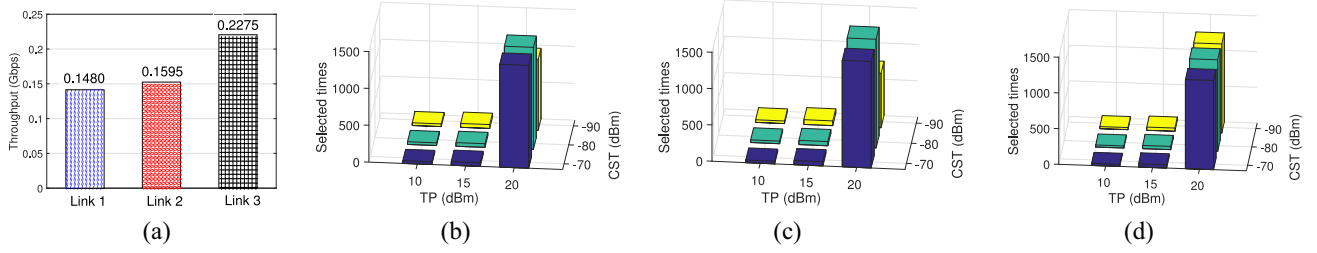


Fig. 7. (a) Average throughput of three FD links, (b-d) the number of selected times at each arm for each link, by running the SAO-FD-CSMA algorithm with an optimal LAI in the network scenario of Fig. 6(a). (a) Average throughput. (b) FD link 1 (Player 1). (c) FD link 2 (Player 2). (d) FD link 3 (Player 3).

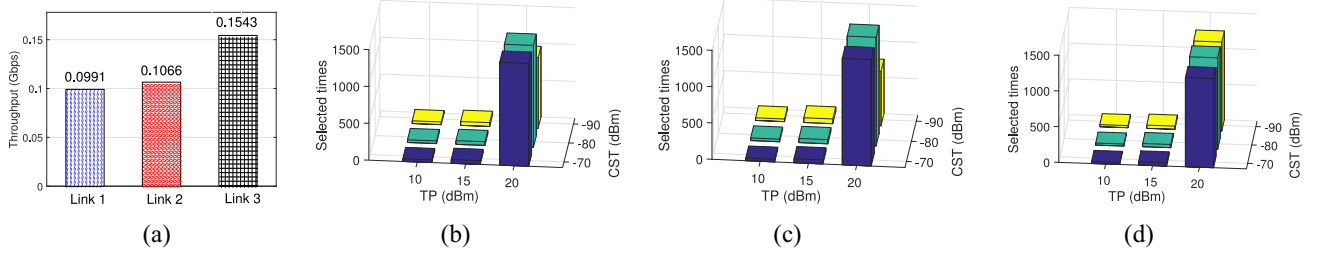


Fig. 8. (a) Average throughput of three FD-links, (b-d) The number of selected times at each arm for each link, by running the SAO-FD-CSMA algorithm with a constant LAI in the network scenario of Fig. 6(a). (a) Average throughput. (b) FD link 1 (Player 1). (c) FD link 2 (Player 2). (d) FD link 3 (Player 3).

the interaction mechanism among links. Thus, each link makes up one-third of the total time slots T .

Figs. 7 and 8 show the performance of the SAO-FD-CSMA algorithm with an optimal LAI and a constant LAI in the network scenario of Fig. 6(a), respectively, where the total time slot T is 12000. The Optimal LAI was obtained by the Optimal FD-CSMA algorithm. The path-loss exponent α is set to 4 and the distance between TX and RX nodes in an FD Link is between interval $[d_{\min}, d_{\max}] = [2, 5]$. From Fig. 7(a), we can see that the average throughput of link 1, link 2, and link 3 are 0.1480 Gb/s, 0.1595 Gb/s, and 0.2275 Gb/s, respectively. The total average throughput of three FD links is about 0.5350 Gb/s, which is much higher than the result in Fig. 8(a) that is just 0.3601. This indicates that the optimal FD-CSMA algorithm can significantly improve the network's performance. Also, it can be seen that the average throughput of link 3 is about twice as much as that of links 1 and 2. This is because links 1 and 2 are too close to each other which cannot transmit simultaneously, as shown in Fig. 6(a). In addition, Fig. 7(b)–(d) and Fig. 8(b)–(d) show the selected times of each arm at each link by running the SAO-FD-CSMA algorithm with an optimal LAI and a constant LAI. We can observe that all links prefer to choose a big TP and CST, since a big TP can boost the link's data rate, and a high CST can increase its transmission probability. However, because of the limited number of TP and CST at each link, the results in Figs. 7 and 8 may not exactly reveal the features of the SAO-FD-CSMA algorithm.

Next, we compare the proposed algorithm with various bandit algorithms in different network settings. The exploration and exploitation parameter of the Exp3 algorithm is set to 0.04 [40]. The upper bound of the estimated mean at each arm for the UCB1 algorithm is set to $[\hat{R}_{k,i}(t)/N_{k,i}(t)] + \sqrt{[2 \log t / N_{k,i}(t)]}$, where $N_{k,i}(t)$ is the number of times that

arm i of link k has been selected up to time slot t [30]. According to [35], the original Exp3 is a special case of the INF algorithm with the negative entropy. Here, we adopt the INF algorithm proposed in [36], which is based on the Tsallis entropy regularization. For the TS algorithm, we use the Gaussian distribution as its *prior* knowledge for the estimated average rewards [32]. Then each player chooses an arm according to the posterior probability that being the current best arm. In addition, the random selection method is simply that each player chooses a pair of TP and CST randomly at each time slot. Since these algorithms are all performed under the framework of Algorithm 1, we refer them as Exp3-FD-CSMA algorithm, UCB1-FD-CSMA algorithm, Tsallis-INF-FD-CSMA algorithm, and TS-FD-CSMA algorithm.

Before comparing the performance of the above algorithms, we give a time complexity analysis for the optimal value and the above algorithms. The optimal value of problem (15) is obtained by the exhaustive search method. Thus, the time complexity of the optimal value method is about $\mathcal{O}(\prod_{k=1}^K M_k B_s)$, where $M_k = |\mathcal{A}_k|$ and B_s is the total number of time epochs in a time slot. For the Tsallis-INF-FD-CSMA algorithm, it needs to find the normalization factor for the Tsallis entropy regularization by using the Newton method. So the time complexity of the Tsallis-INF-FD-CSMA algorithm is about $\mathcal{O}(TB_s \bar{W})$, where \bar{W} is the average number of iterations that the Newton method can reach a sufficient precision value. In addition, the other algorithms have the same time complexity which is linear in terms of the time horizon T , i.e., $\mathcal{O}(TB_s)$.

Fig. 9 compares the performances of the above algorithms, and the proposed SAO-FD-CSMA algorithm with *prior* knowledge and without *prior* knowledge in the network scenario of Fig. 6(a) where $\alpha = 4$ and $[d_{\min}, d_{\max}] = [2, 5]$. It can be seen that, except for the random selection method,

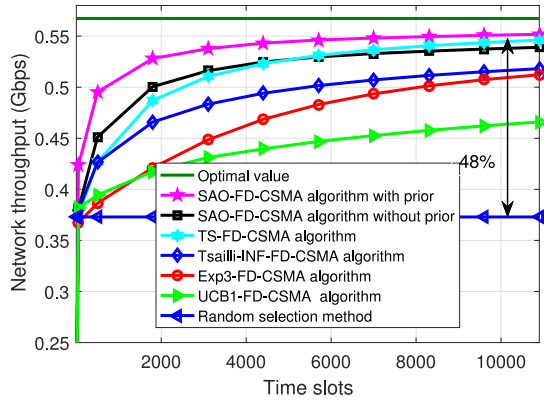


Fig. 9. Total network throughput of several algorithms in the network scenario of Fig. 6(a), where the total time slots is 12000.

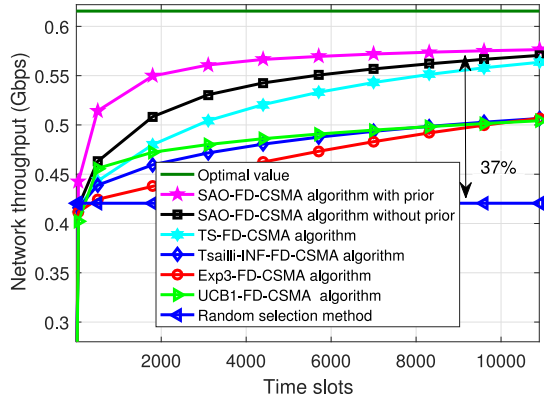


Fig. 10. Total network throughput of several algorithms in the network scenario of Fig. 6(b), where the total time slots is 12000.

all algorithms can converge. The SAO-FD-CSMA algorithm with *prior* knowledge has the fastest convergence rate and the best performance; while the TS-FD-CSMA algorithm is slightly lower than the proposed algorithm. Although the TS-FD-CSMA algorithm has the same time complexity as the proposed algorithm, the TS-FD-CSMA algorithm requires the prior distribution of the received rewards, which is difficult to obtain in practice. By contrast, the proposed algorithm does not require any assumptions on the received rewards. Furthermore, from Fig. 9, we can see that the performance of the Tsallis-INF-FD-CSMA algorithm is close to the proposed algorithm. However, the Tsallis-INF-FD-CSMA algorithm suffers a high time complexity due to repeated recall of the Newton method. It is well known that the UCB1 algorithm is designed for stochastic bandits and the Exp3-FD-CSMA algorithm is suitable for the adversarial bandits. Both UCB1-FD-CSMA algorithm and Exp3-FD-CSMA algorithm perform negatively here, indicating that the interaction among links forms an environment neither stochastic nor adversary. Furthermore, compared with the random selection method, the performance of the proposed algorithm is improved by about 48%.

Fig. 10 compares the performances of the above algorithms, and the proposed SAO-FD-CSMA algorithm with *prior* knowledge and without *prior* knowledge in the network scenario of Fig. 6(b). We can see that the total network throughput

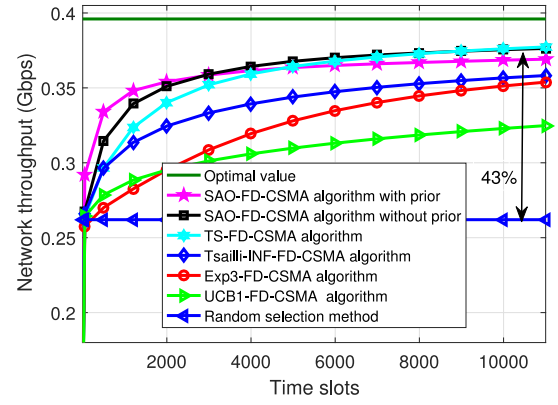


Fig. 11. Total network throughput of several algorithms for 10000 random network scenarios, where the total time slots is 12000.

is improved and these algorithms has the same increase trends as in Fig. 10. However, the performance gap between the optimal value and the proposed algorithm becomes large when the number of links increases from 3 to 6. The reasons are that 1) The more links in a fixed area the more contentions between links will occur, resulting in performance loss and 2) The more links in a fixed area the more time slots are needed for the proposed algorithm to traverse all the network states so as to convergence to the best arm.

Fig. 11 shows the performances of the above algorithms, and the proposed SAO-FD-CSMA algorithm with *prior* knowledge and without *prior* knowledge in 10000 random network scenarios where $\alpha = 4$ and $[d_{\min}, d_{\max}] = [2, 5]$. It can be observed that the total network throughput of all algorithms is lower than that in Fig. 9, but has the same increase trend. Here, the SAO-FD-CSMA algorithm with *prior* knowledge is slightly worse than the SAO-FD-CSMA algorithm without *prior* knowledge. This is reasonable because one cannot have *prior* information about all 10000 network scenarios. It can also be seen that the TS-FD-CSMA algorithm has a similar performance as the SAO-FD-CSMA algorithm without *prior* knowledge. However, the TS-FD-CSMA algorithm needs some *prior* distribution for the estimated average rewards. In addition, the performance of the proposed algorithm is improved by about 43% compared with the random selection method.

Fig. 12 depicts the average network throughput of the optimal value, SAO-FD-CSMA algorithm without *prior* knowledge, TS-FD-CSMA algorithm, Exp3-FD-CSMA algorithm, and random selection method for different numbers of links in 1000 random network scenarios when $T = 12000$, $\alpha = 4$, and $[d_{\min}, d_{\max}] = [2, 5]$. The average network throughput is calculated by $\sum_{k=1}^K \sum_{s=1}^T \Gamma_k(s) / T$. It can be seen that, except the Exp3-FD-CSMA algorithm, other algorithms first increase when the number of links increases from 3 to 5, and then drop fast when the number of links exceeds 5. The TS-FD-CSMA algorithm has the best performance when the number of links is less than 5. But it decreases faster and lower than the proposed algorithm when the number of links is greater than 6. This is because the more links in the network the more time is needed for The TS-FD-CSMA algorithm to traverse all the network states to better estimate each arm's

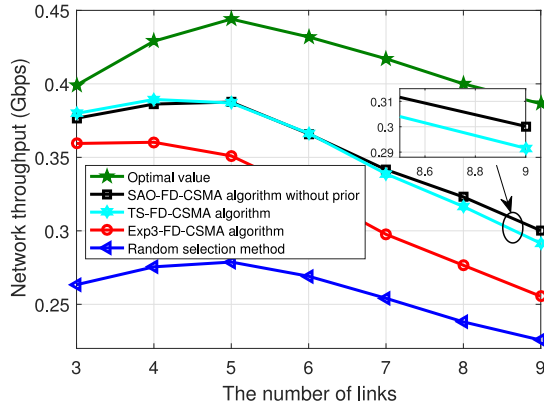


Fig. 12. Average network throughput of several algorithms versus the number of links for 1000 random network scenarios, where $T = 12000$.

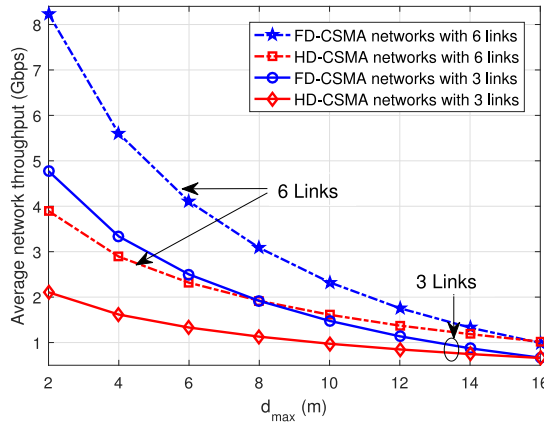


Fig. 13. Average network throughput versus the maximum link length d_{\max} of each Tx-Rx pair by using the proposed algorithm without *prior* knowledge for 1000 random network scenarios, where $T = 12000$.

average reward. The performance gaps between the optimal value and other algorithms increase with the number of links. This indicates that, in such a fully distributed network, the more contention among links the more performance loss will be suffered by the proposed algorithm, as well as other bandit algorithms when the time horizon is fixed. Therefore, it is better to choose a long time horizon T in high-density networks and a short time horizon T in low-density networks.

Finally, we compare the performances of the FD- and HD-enabled CSMA networks for different maximum link lengths and the path-loss exponents. The minimum link length is still $d_{\min} = 2$. In order to better capture the differences between the FD- and HD-CSMA networks, we remove the piecewise rate model in our simulation. Fig. 13 shows the average network throughput for different maximum link lengths by using the proposed algorithm without *prior* knowledge in 1000 random network scenarios, where $T = 12000$. We can observe that the network throughput of both the FD- and HD-CSMA networks decrease with the maximum link length. The reason is that increasing d_{\max} will decrease the received signal power and magnify the self-interference in the SINR. As a result, compared with the HD-CSMA networks, the FD-enabled CSMA networks suffer from more severe throughput degradation in both three links and six links cases. More interesting, the performance of FD-CSMA networks is smaller than that of

the HD-CSMA networks when $d_{\max} > 16$ (m) in both two cases.

VIII. CONCLUSION

In this article, we have investigated the network-level throughput of an FD-enabled CSMA network by considering TP control, CST adjustment, and LAI adaptation. We first formulated it as a network optimization problem by using the *piecewise constant rate* model and the CTMN model. Then, this problem was decomposed into two subproblems, i.e., a joint control and scheduling problem in the transport- and MAC-layer, and a parameter selection problem in the PHY-layer. For subproblem 1, we presented an optimal FD-CSMA algorithm to find the optimal LAI by using a subgradient method; while for subproblem 2, we proposed an SAO algorithm, which is both optimal in the stochastic and the adversarial environment, to select the best combination of TP and CST. By emerging the above two algorithms, we put forth an SAO-FD-CSMA algorithm to approximately address the throughput maximum problem by solving subproblem 1 and subproblem 2 alternately. To evaluate the proposed algorithms, we have designed a DESim method to simulate the FD-CSMA protocol. Theoretical results state that the proposed algorithm can converge to the optimal value when the time horizon $T \rightarrow \infty$. Finally, the numerical results verify the theoretical results and show its superior performance among the state-of-the-art bandit algorithms. In addition, we reveal that the FD-CSMA network cannot double its throughput compared to that of the HD-CSMA network in most cases, and even has worse performance than that of the HD-CSMA network in some special cases.

APPENDIX A PROOF OF LEMMA 1

According to [45], the master problem of (15) is equivalent to the following optimization problem:

$$\begin{aligned}
 (\mathbf{M}) \quad & \max_{\rho_k, P_k, S_k} V \sum_{k \in \mathcal{V}} \log \Gamma_k \\
 \text{s.t.} \quad & \Gamma_k \leq \sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} \\
 & \sum_{i=0}^{N-1} \pi_{f^i} = 1 \\
 & \rho_k \in \mathbb{R}^+, P_k \in \mathbf{P}_k, \text{ and } S_k \in \mathbf{S}_k \quad (38)
 \end{aligned}$$

where V is a positive constant weighting factor and the second constraint is the feasible capacity. Thus, when P_k and S_k is given, problem (38) can be transformed to the subproblem 1, i.e.,

$$\begin{aligned}
 (\mathbf{P1}) \quad & \max_{\rho_k} V \sum_{k \in \mathcal{V}} \log \Gamma_k - \sum_{i=0}^{N-1} \pi_{f^i} \log \pi_{f^i} \\
 \text{s.t.} \quad & \Gamma_k \leq \sum_{i=0}^{N-1} v_{k,i} \pi_{f^i} \quad \forall k \in \mathcal{V} \\
 & \sum_{i=0}^{N-1} \pi_{f^i} = 1 \text{ and } \rho_k \in \mathbb{R}^+ \quad (39)
 \end{aligned}$$

where $-\sum_{i=0}^{N-1} \pi_{f^i} \log \pi_{f^i}$ is the information entropy that gets the maximum value when all feasible states' probabilities are the same. So **P1** is equal to solving the objective of $\max \sum_{k \in \mathcal{V}} \log \Gamma_k$ by introducing a deviation of $\log(|\pi_{f^i}|)/V$ or $\log(N)/V$. By taking a large value of V (i.e., $V \gg N$) in the piratical application, the approximation error can be ignored.

APPENDIX B PROOF OF THEOREM 2

Let Z_k be the total number of phases of player K , i.e., $z = 1, 2, \dots, Z_k$. Define $B_{k,z}$ and $T_{k,z}$ be the first and last time slots of phase z , respectively. So phase z includes a sequence of trials, i.e., $\{B_{k,z}, B_{k,z} + 1, \dots, T_{k,z}\}$. To proof Theorem 2, we borrow the results of [40, Lemmas 4.2 and 4.3], where Lemma 4.2 was used to bound the regret produced at each phase and Lemma 4.3 limits the total number of phase at each player. For convenience, we reproduce them as following.

Lemma 2: For any arm j and every phase z at the k th player, the following inequality holds:

$$\sum_{t=B_{k,z}}^{T_{k,z}} r_{k,\xi_t^k}(t) \geq \sum_{t=B_{k,z}}^{T_{k,z}} \hat{r}_{k,j}(t) - 2\sqrt{e-1}\sqrt{\theta_z M_k \ln M_k}$$

where θ_z was given in Algorithm 2.

Lemma 3: The number of phases of the k th player Z_k satisfies

$$2^{Z_k-1} \leq \frac{e-1}{\ln M_k} + \sqrt{\frac{(e-1)\hat{R}_{k,\max}}{M_k \ln M_k}} + \frac{1}{2}.$$

First, according to Lemma 2, we have

$$\begin{aligned} R_k &= \sum_{t=1}^T r_{k,\xi_t^k}(t) = \sum_{z=0}^{Z_k} \sum_{t=B_{k,z}}^{T_{k,z}} r_{k,\xi_t^k}(t) \\ &\geq \max_j \sum_{z=0}^{Z_k} \left(\sum_{t=B_{k,z}}^{T_{k,z}} \hat{r}_{k,j}(t) - 2\sqrt{e-1}\sqrt{\theta_z M_k \ln M_k} \right). \end{aligned}$$

Note that Lemma 2 holds for any $j \in \mathcal{A}_k$. Recall that $\theta_z = 4^z M_k \ln(M_k)/(e-1)$, so

$$\begin{aligned} R_k &\geq \max_{j \in \mathcal{A}_k} \hat{R}_{k,j}(T+1) - 2M_k \ln M_k \sum_{z=0}^{Z_k} 2^z \\ &= \hat{R}_{k,\max} - 2M_k \ln M_k (2^{Z_k+1} - 1). \end{aligned}$$

Then, by applying Lemma 3, the above inequality can be further written as

$$\begin{aligned} R_k &\geq \hat{R}_{k,\max} + 2M_k \ln M_k \\ &\quad - 8M_k \ln M_k \left(\frac{e-1}{\ln M_k} + \sqrt{\frac{(e-1)\hat{R}_{k,\max}}{M_k \ln M_k}} + \frac{1}{2} \right) \\ &= \hat{R}_{k,\max} - 2M_k \ln M_k - 8(e-1)M_k \\ &\quad - 8\sqrt{e-1}\sqrt{\hat{R}_{k,\max} M_k \ln M_k}. \end{aligned} \quad (40)$$

For simplicity, let $f(x) = x - a\sqrt{x} - b$ ($x \geq 0$), where $a = 8\sqrt{e-1}\sqrt{M_k \ln M_k}$ and the constant term $b = 2M_k \ln M_k +$

$8(e-1)M_k$. Taking expectations on both left- and right-hand sides of (40) yields

$$\mathbb{E}[R_k] \geq \mathbb{E}\left[f\left(\hat{R}_{k,\max}\right)\right] \quad (41)$$

since f is differentiable and its second order derivation is positive for $x > 0$, and so function f is convex, by using Jensen's inequality, we have

$$\mathbb{E}\left[f\left(\hat{R}_{k,\max}\right)\right] \geq f\left(\mathbb{E}\left[\hat{R}_{k,\max}\right]\right) \quad (42)$$

where

$$\begin{aligned} \mathbb{E}\left[\hat{R}_{k,\max}\right] &= \mathbb{E}\left[\max_{j \in \mathcal{A}_k} \hat{R}_{k,j}(T+1)\right] \\ &\geq \max_{j \in \mathcal{A}_k} \mathbb{E}\left[\hat{R}_{k,j}(T+1)\right] = \max_{j \in \mathcal{A}_k} \sum_{t=1}^T r_{k,j}(t) = R_{k,\max}. \end{aligned}$$

Therefore, it is sufficient to consider the following two cases.

- 1) $R_{k,\max} > a^2/4$: It is easy to obtain that function f is increasing in the interval $(a^2/4, +\infty)$, and then $f(\mathbb{E}[\hat{R}_{k,\max}]) \geq f(R_{k,\max})$. Together with (41) and (42), we get $\mathbb{E}[R_k] \geq f(R_{k,\max})$, and this is equal to the k th player's upper regret bound.
- 2) $R_{k,\max} \leq a^2/4$: In the case, f is nonincreasing in $[0, a^2/4]$, so the maximum value was attained at point 0, which $f(0) = -b < 0 \leq \mathbb{E}[R_k]$.

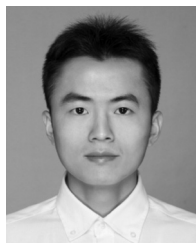
Thus, it is easy to consider this case.

Finally, summing over the link's index k , we get Theorem 2.

REFERENCES

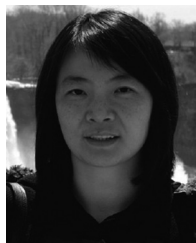
- [1] J. Tong, L. Fu, and Z. Han, "Throughput enhancement of full-duplex CSMA networks via adversarial multi-player multi-armed bandit," in *Proc. IEEE Global Telecom. Conf.*, Dec. 2019, pp. 1–6.
- [2] J. Ni, B. Tan, and R. Srikant, "Q-CSMA: Queue-length-based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 3, pp. 825–836, Jun. 2012.
- [3] A. Sabharwal, A. Khoshnevis, and E. Knightly, "Opportunistic spectral usage: Bounds and a multi-band CSMA/CA protocol," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 533–545, Jun. 2007.
- [4] J. Konorski, "A game-theoretic study of CSMA/CA under a back-off attack," *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1167–1178, Dec. 2006.
- [5] N. Rajatheva *et al.*, "White paper on broadband connectivity in 6G," Apr. 2020. [Online]. Available: arXiv:2004.14247.
- [6] S. Liu, L. Fu, and W. Xie, "Hidden-node problem in full-duplex enabled CSMA networks," *IEEE Trans. Mobile Comput.*, vol. 19, no. 2, pp. 347–361, Feb. 2020.
- [7] X. Xie and X. Zhang, "Does full-duplex double the capacity of wireless networks?" in *Proc. IEEE INFOCOM*, Apr. 2014, pp. 253–261.
- [8] S. Wang, V. Venkateswaran, and X. Zhang, "Fundamental analysis of full-duplex gains in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 3, pp. 1401–1416, Jun. 2017.
- [9] Y. Liao, T. Wang, L. Song, and Z. Han, "Listen-and-talk: Protocol design and analysis for full-duplex cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 1, pp. 656–667, Jan. 2017.
- [10] R. Doost-Mohammady, M. Y. Naderi, and K. R. Chowdhury, "Performance analysis of CSMA/CA based medium access in full duplex wireless communications," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1457–1470, Jun. 2016.
- [11] L. Jiang and J. Walrand, "A distributed CSMA algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 3, pp. 960–972, Jun. 2010.
- [12] E. Gurses and R. Boutaba, "Capacity of wireless multi-hop networks using physical carrier sense and transmit power control," in *Proc. IEEE Global Telecommun. Conf.*, Honolulu, HI, USA, Nov. 2009, pp. 1–6.
- [13] Y. Yang, J. C. Hou, and L.-C. Kung, "Modeling the effect of transmit power and physical carrier sense in multi-hop wireless networks," in *Proc. IEEE INFOCOM*, May 2007, pp. 2331–2335.

- [14] T.-S. Kim, H. Lim, and J. C. Hou, "Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multi-hop wireless networks," in *Proc. ACM MOBICOM*, Sep. 2006, pp. 366–377.
- [15] Z. Zhou, Y. Zhu, Z. Niu, and J. Zhu, "Joint tuning of physical carrier sensing, power and rate in high-density WLAN," in *Proc. Asia-Pac. Conf. Commun.*, Bangkok, Thailand, Oct. 2007, pp. 131–134.
- [16] H. Jang, S.-Y. Yun, J. Shin, and Y. Yi, "Game theoretic perspective of optimal CSMA," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 194–209, Jan. 2018.
- [17] H. Tibrewal, S. Patchala, M. K. Hanawal, and S. J. Darak, "Distributed learning and optimal assignment in multiplayer heterogeneous networks," Jan. 2019. [Online]. Available: arXiv:1901.03868.
- [18] A. Neogi, P. Chaporkar, and A. Karandikar, "Multi-player multi-armed bandit based resource allocation for D2D communications," Dec. 2018. [Online]. Available: arXiv:1812.11837.
- [19] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5667–5681, Nov. 2010.
- [20] H. Joshi, R. Kumar, A. Yadav, and S. J. Darak, "Distributed algorithm for dynamic spectrum access in infrastructure-less cognitive radio network," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Barcelona, Spain, Apr. 2018, pp. 1–6.
- [21] F. Wilhelmi, S. Barrachina-Muñoz, B. Bellalta, C. Cano, A. Jonsson, and G. Neu, "Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in WLANs," *J. Netw. Comput. Appl.*, vol. 127, pp. 26–42, Feb. 2019.
- [22] L. Tassioulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multi-hop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.
- [23] A. Eryilmaz, A. Ozdaglar, and E. Modiano, "Polynomial complexity algorithms for full utilization of multi-hop wireless networks," in *Proc. IEEE INFOCOM*, May 2007, pp. 499–507.
- [24] X. Wu, R. Srikant, and J. R. Perkins, "Scheduling efficiency of distributed greedy scheduling algorithms in wireless networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 6, pp. 595–605, Jun. 2007.
- [25] S.-Y. Yun, Y. Yi, J. Shin, and D. Y. Eun, "Optimal CSMA: A survey," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Singapore, Nov. 2012, pp. 199–204.
- [26] J. Lyu, Y. H. Chew, and W.-C. Wong, "A Stackelberg game model for overlay D2D transmission with heterogeneous rate requirements," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8461–8475, Oct. 2016.
- [27] A. Maatouk, M. Assaad, and A. Ephremides, "Energy efficient and throughput optimal CSMA scheme," *IEEE/ACM Trans. Netw.*, vol. 27, no. 1, pp. 316–329, Feb. 2019.
- [28] M. Ghazvini, N. Movahedinia, K. Jamshidi, and N. Moghim, "Game theory applications in CSMA methods," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1062–1087, 3rd Quart., 2013.
- [29] F. Wilhelmi, C. Cano, G. Neu, B. Bellalta, A. Jonsson, and S. Barrachina-Muñoz, "Collaborative spatial reuse in wireless networks via selfish multi-armed bandits," *Ad Hoc Netw.*, vol. 88, pp. 129–141, May 2019.
- [30] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, nos. 2–3, pp. 235–256, May. 2002.
- [31] P. Auer and C.-K. Chiang, "An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits," in *Proc. Conf. Learn. Theory*, Feb. 2016, pp. 116–120.
- [32] S. Agrawal and N. Goyal, "Further optimal regret bounds for Thompson sampling," *Artif. Intell. Stat.*, vol. 31, pp. 99–107, Jun. 2013.
- [33] S. Bubeck and A. Slivkins, "The best of both worlds: Stochastic and adversarial bandits," in *Proc. Conf. Learn. Theory*, Feb. 2012, pp. 1–23.
- [34] Y. Seldin and A. Slivkins, "One practical algorithm for both stochastic and adversarial bandits," in *Proc. ICML*, Feb. 2014, pp. 1287–1295.
- [35] J.-Y. Audibert and S. Bubeck, "Minimax policies for adversarial and stochastic bandits," in *Proc. 22nd Annu. Conf. Comput. Learn. Theory*, Jun. 2009, pp. 217–226.
- [36] J. Zimmert and Y. Seldin, "An optimal algorithm for stochastic and adversarial bandits," in *Proc. 22nd Int. Conf. Artif. Intell. Stat.*, Apr. 2019, pp. 467–475.
- [37] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Feb. 2014.
- [38] J.-G. Choi, C. Joo, J. Zhang, and N. B. Shroff, "Distributed link scheduling under SINR model in multi-hop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 22, no. 4, pp. 1204–1217, Aug. 2014.
- [39] J. Liu, Y. Yi, A. Proutiere, M. Chiang, and H. V. Poor, "Towards utility-optimal random access without message passing," *Wireless Commun. Mobile Comput.*, vol. 10, no. 1, pp. 115–128, 2010.
- [40] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The non-stochastic multiarmed bandit problem," *SIAM J. Comput.*, vol. 32, no. 1, pp. 48–77, Jan. 2003.
- [41] S. C. Liew, C. H. Kai, H. C. Leung, and P. Wong, "Back-of-the-envelope computation of throughput distributions in CSMA wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 9, pp. 1319–1331, Sep. 2010.
- [42] H. J. Kushner and G. G. Yin, *Stochastic Approximation and recursive algorithms and Applications*. New York, NY, USA: Springer, 2003.
- [43] V. S. Borkar, "Stochastic approximation with 'controlled' markov-noise," *Syst. Control Lett.*, vol. 55, no. 2, pp. 139–145, 2006.
- [44] E. Khorov, A. Kiryanov, A. Lyakhov, and G. Bianchi, "A tutorial on IEEE 802.11ax high efficiency WLANs," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 197–216, 1st Quart. 2019.
- [45] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2003.



Jingwen Tong (Student Member, IEEE) received the B.E. degree in electrical engineering from China Jiliang University, Hangzhou, China, in 2015, and the M.S. degree in electrical engineering from Ningbo University, Ningbo, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Communication Engineering, Xiamen University, Xiamen, China.

From 2019 to 2020 he was a visiting Ph.D. student with the University of Houston, TX, USA. His current research interests include multiarmed bandit and game theory.



Liqun Fu (Senior Member, IEEE) received the Ph.D. degree in information engineering from The Chinese University of Hong Kong, Hong Kong, in 2010.

She is a Full Professor with the School of Informatics, Xiamen University, Xiamen, China. She was a Postdoctoral Research Fellow with the Institute of Network Coding, The Chinese University of Hong Kong, from 2011 to 2013 and the ACCESS Linnaeus Center, KTH Royal Institute of Technology, Stockholm, Sweden, from 2013 to

2015. She was with ShanghaiTech University, Shanghai, China, as an Assistant Professor from 2015 to 2016. Her research interests are mainly in communication theory, optimization theory, and learning theory, with applications in wireless networks.

Prof. Fu served as the Technical Program Co-Chair of the GCCCN Workshop of the IEEE INFOCOM 2014, the Publicity Co-Chair of the GSN Workshop of the IEEE INFOCOM 2016, and the Web Chair of the IEEE WiOpt 2018. She also serves as a TPC Member for many leading conferences in communications and networking, such as the IEEE INFOCOM, ICC, and GLOBECOM. She is on the editorial board of IEEE ACCESS and the *Journal of Communications and Information Networks*.



Zhu Han (Fellow, IEEE) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 1997, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Maryland, College Park, MD, USA, in 1999 and 2003, respectively.

From 2000 to 2002, he was a Research and Development Engineer with JDSU, Germantown. From 2003 to 2006, he was a Research Associate with the University of Maryland. From 2006 to 2008, he was an Assistant Professor with Boise State

University, Boise, ID, USA. He is currently a John and Rebecca Moores Professor with the Electrical and Computer Engineering Department as well as with the Computer Science Department, University of Houston, Houston, TX, USA. His research interests include wireless resource allocation and management, wireless communications and networking, game theory, big data analysis, security, and smart grid.

Dr. Han received an NSF Career Award in 2010, the Fred W. Ellersick Prize of the IEEE Communication Society in 2011, the EURASIP Best Paper Award for the *EURASIP Journal on Advances in Signal Processing* in 2015, the IEEE Leonard G. Abraham Prize in the field of Communications Systems (Best Paper Award in IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS) in 2016, and several best paper awards in IEEE conferences. He is also the Winner of the 2021 IEEE Kyo Tomiyasu Award, for outstanding early to mid-career contributions to technologies holding the promise of innovative applications, with the following citation: "for contributions to game theory and distributed management of autonomous communication networks". He was an IEEE Communications Society Distinguished Lecturer from 2015 to 2018, and has been an AAAS Fellow since 2019, and an ACM Distinguished Member since 2019. He is 1% highly cited researcher since 2017 according to Web of Science.