Computer Networks 104 (2016) 43–54

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

Entropy-based active learning for wireless scheduling with incomplete channel feedback☆,☆☆

Mehmet Karaca[a,*], Ozgur Ercetin[b], Tansu Alpcan[c]

[a] Department of Electrical and Information Technology, Lund University, Lund, Sweden
[b] Faculty of Engineering and Natural Science, Sabanci University, Istanbul, Turkey
[c] Dept. of Electrical and Electronic Engineering, The University of Melbourne, Melbourne, Australia

ARTICLE INFO

Article history:
Received 27 June 2015
Revised 29 January 2016
Accepted 2 May 2016
Available online 3 May 2016

Keywords:
Opportunistic scheduling
Queue stability
Limited information
Machine learning

ABSTRACT

Most of the opportunistic scheduling algorithms in literature assume that full wireless channel state information (CSI) is available for the scheduler. However, in practice obtaining full CSI may introduce a significant overhead. In this paper, we present a learning-based scheduling algorithm which operates with partial CSI under general wireless channel conditions. The proposed algorithm predicts the instantaneous channel rates by employing a Bayesian approach and using Gaussian process regression. It quantifies the uncertainty in the predictions by adopting an entropy measure from information theory and integrates the uncertainty to the decision-making process. It is analytically proven that the proposed algorithm achieves an ϵ fraction of the full rate region that can be achieved only when full CSI is available. Numerical analysis conducted for a CDMA based cellular network operating with high data rate (HDR) protocol, demonstrate that the full rate region can be achieved our proposed algorithm by probing less than 50% of all user channels.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

A challenging open problem in wireless networks is the efficient allocation of limited and time-varying resources among multiple users to satisfy their requirements. The problem is exacerbated by the highly dynamic nature of wireless channels due to multiple superimposed random effects caused by mobility and multi-path fading. In many cases, acquiring extensive information on wireless channel characteristics is simply infeasible as a result of prohibitive overhead costs and hard constraints. In yet other cases, the wireless channel may be highly non-stationary that by the time the information is obtained, it becomes outdated due to channels' fast-changing nature. Hence, scheduling decisions should be made based on partial and outdated channel state information.

One of the main assumptions in prior works [2] is that the exact and complete channel state information (CSI) of all users is available at every time slot. Under this assumption, the seminal work by Tassiulas and Ephremides has shown that the opportunistic *Max-Weight* scheduling algorithm is throughput-optimal, i.e., it can stabilize the network whenever this is possible [2]. Max-Weight algorithm is a simple index policy which schedules the user with the largest queue length and rate product at each time slot.

In this paper, we investigate scheduling in a multi-user downlink wireless network where only partial channel state information can be acquired due to the band-limited feedback channel (Fig. 1). We present a joint CSI acquisition and scheduling algorithm which operates without any a priori knowledge on the distribution of channel states. The proposed algorithm tracks the states of the channels by using a learning algorithm and by judiciously probing a set of users whose channel states may have changed. At each slot, the algorithm schedules a user among the set of probed users, which has the highest queue backlog and transmission rate product.

Our work relies on a recent learning and optimization framework developed in [3], wherein the exploration and exploitation trade-off is explicitly quantified as a multi-objective meta optimization problem. In this paper, we investigate a trade-off between scheduling a user with the highest queue-rate product (exploitation), and probing of users with outdated channel observations (exploration). The solution of this trade-off problem requires the prediction of the instantaneous user channel states,

☆ A preliminary version of the paper was presented at the International Workshop on Smart Communication Protocols and Algorithms (SCPA) co-located with IEEE ICC in Ottawa, Canada on June 2012 [1].

☆☆ This work was in part done when Mehmet Karaca was with Sabanci University, Istanbul, Turkey.

* Corresponding author.
E-mail address: mehmet.karaca@eit.lth.se, mehmetkrc@sabanciuniv.edu (M. Karaca).

http://dx.doi.org/10.1016/j.comnet.2016.05.001
1389-1286/© 2016 Elsevier B.V. All rights reserved.

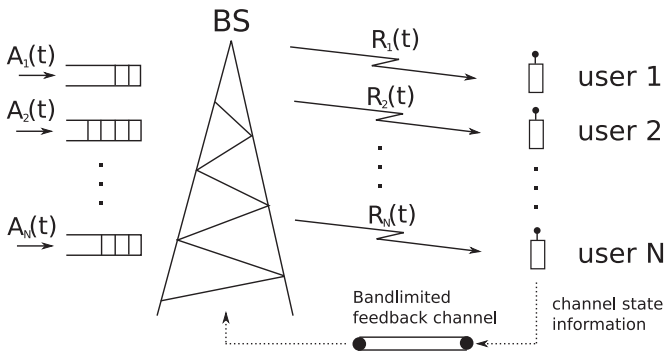

Fig. 1. Cellular downlink network.

and the measurement of the associated level of uncertainty in the prediction. We adopt a Bayesian approach, and use Gaussian processes as a state-of-the-art regression method to predict the instantaneous user channel states.

Gaussian process regression is a powerful nonlinear interpolation tool, where the inference of continuous values are made with respect to a Gaussian process prior [4]. Although the inference of instantaneous channel gains is with a Gaussian process prior, this does not assume that the underlying channel model is Gaussian. In fact, as demonstrated by our numerical experiments, our approach is applicable to a wide range of channel models including time-correlated and even non-stationary channels. Another unique feature of our algorithm is that the uncertainty in the predicted channel state is quantified explicitly by the *entropy measure* from the information theory. Our algorithm weighs the level of uncertainty eliminated by probing a channel against the aspiration to schedule the user with the maximum weight to determine a set of users probed at every slot.

Our contributions are summarized as follows: i-) we first define a general Max-Weight-like policy which makes scheduling decisions based on the predicted values of instantaneous channel rates rather than their exact values. Based on the channel prediction errors, we define the achievable rate region of this algorithm as compared to the full rate region achieved by the Max-Weight algorithm with complete CSI. Specifically, we analytically show that with this policy, ϵ fraction of the full rate region can be obtained. We also explicitly compute ϵ under certain conditions; ii-) Next, based on this general policy we investigate a multi-objective framework where the exploration and exploitation tradeoff of probing different users is identified. In this framework, the information obtained by probing a user channel is modeled with the help of Shannon's entropy formula according to the past observations of the channel; iii-) Then, we specify in detail our channel predictor used to predict instantaneous CSI, and suitable for both stationary and non-stationary channels based on Gaussian Process Regression; iv-) Lastly, we perform an extensive number simulations using High Data Rate (HDR) protocol [5] with a realistic channel model. We compare the performance of our algorithm with that of the state-of-the-art channel prediction method based on Autoregression (AR) [6].

The organization of our paper is given as follows: Section 2 summarizes the literature on opportunistic algorithms scheduling with a partial CSI, and learning methods previously used for the control of wireless networks. Section 3 presents the system model used in this paper. In Section 4, the general Max-Weight type policy and its performance in terms of achievable rate region are presented. In Section 5, GPR is explained in detail. The performance of the proposed algorithm is evaluated numerically in Section 6. Finally, we conclude the paper in Section 7.

2. Related work

It was shown that Max-Weight algorithm scheduling the user with the highest queue backlog and transmission rate product at every time slot is throughput optimal [2]. An important assumption of Max-Weight algorithm is that it requires complete knowledge of channel states at the beginning of each time slot. Investigating the performance of Max-Weight algorithm with incomplete CSI has been an active research area, and we classify the previous works on this area into two main categories: in the first category, it is assumed that the channel distributions of the users are known in advance whereas the main assumption of the works in the second category is that the channel distributions are not known a priori but users have only a specific channel distribution such as iid, Markovian or, in general, a stationary channel distribution.

In [7–9], the authors proposed joint scheduling and channel feedback algorithms by considering the problem of stabilizing the network of queues with incomplete CSI. These algorithms were shown to be throughput-optimal under the assumption that channel distributions are known a priori and they are independent and identically distributed (iid). These works are within the first category in our classification.

For the case when the channel distributions are not known a priori but can only have a specific distribution, which refers to the second category, several joint scheduling and probing algorithms were presented in [10–12], and [13]. In [10] and [11], the authors proposed algorithms that estimate the user channel statistics by assuming that the channels are iid. The problem of joint prediction of channel states and scheduling to optimize a long term metric under stability and other resource constraints was studied in [10]. The work in [11] proposed a probabilistic algorithm which at every slot decides to either explore a user channel state or exploit the slot to transmit data. In [12], the authors have presented a joint scheduling and channel estimation algorithm for correlated ON/OFF Markovian channels. In [13], we have developed an algorithm which probes only those users with sufficiently good channel quality and schedules the user with the maximum weight at each transmission opportunity. The underlying system model considered in [13] is different than the one used in this paper, since in [13] feedback from as many users as needed can be obtained by tolerating a reduction in data transmission rates. The authors in [14–19], studied the problem of scheduling with limited information under various aims and techniques such as utility maximization, thresholds based policies, distributed solution for cost reduction, etc. We refer the readers to [20] for a summary of different techniques used to reduce the overhead of obtaining CSI, e.g., quantization of CSI, beamforming or precoding. It is important to note that the common assumption of all these works is that the wireless channel has a well-defined stationary distribution.

In this work, without making the assumptions in the works in both of these categories, a learning based approached is utilized. Learning algorithms been applied to various problems in communication networks where there is limited information on network states such as routing [21], spectrum allocation [22], interference mitigation [23], multi-channel cognitive networks [24], combinatorial network optimization [25], multi-channel access [26]. These problems were solved by using reinforcement learning [21,22], Q-learning [23] techniques, or by modeling them as multi-armed bandit [24,25], [26], problems. Furthermore, studies have shown that such learning based future channel prediction techniques can provide more efficient spectrum utilization [27].

The studies which apply various learning techniques, [21–26], focused on finding a single solution assuming that the stochastic processes underlying the channel characteristics are stationary. Although these methods may provide provably optimal solutions under some special cases, none of them can adapt to the changes in

the underlying process, if it is non-stationary. The need for tracking nonstationary environments via learning, although widely acknowledged, has not been extensively pursued previously. In [28], the authors investigated the performance of Max-Weight algorithm over non-stationary channels by assuming that the base station acquires all CSI at every scheduling time. By employing Gaussian Process Regression (GPR) this paper focuses on continuous prediction of the channel process rather than finding a single stationary solution when only limited CSI is available at the base station.

Gaussian Process Regression (GPR) is a popular regression method for predicting and tracking of continuous processes, and it is widely used for practical problems. In [29], the authors addressed the problem of localization in a cellular network via GPR. In [30], GPR was employed to select a group of sensors in an environmental sensor network to accurately track the ambient conditions while minimizing the total energy consumption. Similarly, in [31] the authors employed GPR for efficient deployment of sensors. In [32], GPR was used to estimate a Rayleigh channel, and it was shown that GPR is a superior estimation technique.

In this paper, we propose a joint scheduling and channel feedback algorithm which does not aim to estimate the channel statistics, and thus, it is more suitable for realistic non-stationary channel models [33]. Unlike all aforementioned approaches, our proposed approach assumes neither stationarity nor a particular distribution for channels. Thus, it is practically more applicable. In our preliminary work [1], we have applied a similar algorithm assuming the same model in [13] without giving any analytical performance guarantees.

3. System model

We consider a multiuser downlink network with N users and a single base station (BS) as shown in Fig. 1. Time is slotted, and a non-interference model is adopted, where only one user transmits at any given time slot, and there is no interference from neighboring cells. Each user channel experiences independent quasi-static Rayleigh fading, in which the channel gain is constant over the duration of a time slot [1], and it is varying *continuously* from slot to slot [34]. [2] The gain of the channel between the BS and user n, $n \in \{1, 2, \ldots, \mathcal{N}\}$ at time t is denoted by $c_n(t)$, and its value is determined according to an arbitrary probability distribution. As described in the subsequent section, our algorithm which relies on Gaussian Process Regression (GPR), does not make any assumption on the channel distribution.

The instantaneous channel rate between the BS and user n, $R_n(t)$, is defined as the mutual information between the output symbols of the base station and the input symbols at user n over slot t. The maximum value of $R_n(t)$, is obtained when the input symbols are chosen from a Gaussian-distributed input alphabet, i.e.,

$$R_n(t) \leq BW \log_2 \left(1 + P|c_n(t)|^2\right) \text{ bits}, \tag{1}$$

where BW is the bandwidth of the channel, and P is the noise normalized transmit power. We assume that both BW and P are exogenous variables over which we have no control, i.e., we do not consider bandwidth or power control in this paper.

The base station does not have the knowledge of the channel states of the receivers at the beginning of the slot, but it has to acquire this information by probing the users. At the beginning of each time slot, t, the base station broadcasts a pilot signal with a fixed and known power. Each user n determines its CSI, $c_n(t)$, by measuring its received Signal-to-Noise-Ratio. The base station has a dedicated and band-limited feedback control channel that allows receiving channel state feedback from at most $L < N$ users and L is upper bounded by the bandwidth of the feedback channel. We also assume that L is fixed and remains constant throughout the system operation. Such a feedback channel model closely represents practical systems such as HDR[5] and LTE [35]. For instance, in a LTE system [35], CSI is quantized with 4 bits, and physical uplink control channel (PUCCH) with format 2 can be configured for transmission of CSI, which can carry up to 20 bits of feedback information. Let $\mathcal{S}(t)$ be the set of users for which the channel state information is acquired at time slot t. Then, the cardinality of set $\mathcal{S}(t)$ is L or $|\mathcal{S}(t)| = L$, for all t.

The base station maintains a separate queue for each user n. Packets arrive according to a stationary arrival process that is independent across users and time slots. Let $A_n(t)$ be the amount of data arriving into the queue of user n at time slot t with an average arrival rate $\lambda_n = \mathbb{E}[A_n(t)]$. There is a departure from the queue of user n, whenever that user is selected for transmission. Let $\mathcal{J}_n(t)$ represent the scheduler decision, where $\mathcal{J}_n(t) = 1$ if user $n \in \mathcal{S}(t)$ is scheduled for transmission in slot t, and $\mathcal{J}_n(t) = 0$ otherwise. By definition, at most one user can be served at a time slot, i.e., $\sum_{n=1}^{N} \mathcal{J}_n(t) = 1$, for all t.

The dynamics of the queue length process at user n is given as follows:

$$Q_n(t+1) = [Q_n(t) + A_n(t) - R_n(t)\mathcal{J}_n(t)]^+, \tag{2}$$

where $[x]^+ = \max(x, 0)$. Let $\mathbf{Q}(t) = [Q_1(t), Q_2(t), \ldots, Q_N(t)]$ denote the vector of user queue lengths. The objective is to schedule a user at every time slot, so that all of the user queues remain stable at the given arrival rates. A queue is stable if its mean length is finite.

At a given slot, Max-Weight algorithm [2] schedules the user n^* for which the transmission rate weighted by the queue length is the maximum, i.e.,

$$n^* = \underset{n}{\operatorname{argmax}}\, W_n(t) = \underset{n}{\operatorname{argmax}}\, Q_n(t)R_n(t), \tag{3}$$

According to (3), Max-Weight scheduler requires the transmission rates $R_n(t)$ for all users. Instead, we investigate the case where the scheduler only has this information for a subset $S(t)$ of users at a given slot. We denote by $\pi \in \mathcal{F}$, a joint scheduling and channel probing policy which selects the pair $(n, \mathcal{S}(t))$ at every slot t, where $\mathcal{F}$ is the set of all feasible policies. Given $\mathcal{S}(t)$, n is determined in a way similar to Max-Weight rule, i.e.,

$$n = \underset{i \in \mathcal{S}(t)}{\operatorname{argmax}}\; Q_i(t)R_i(t). \tag{4}$$

Note that in previous studies such as [7–9], $\mathcal{S}(t)$ was determined given the knowledge of the stationary distribution of the wireless channels. However, in practice, it is not possible to know the exact channel distributions a priori to system operation. Moreover, if the channel distributions are non-stationary, then the knowledge of channel distribution becomes obsolete.

4. Scheduling with incomplete CSI

4.1. A General result

In this section, we study the scheduling problem with incomplete CSI over general fading channel model without assuming a priori channel distribution. Let $\pi(\eta)$ be a joint policy which employs an arbitrary channel prediction algorithm η to predict the channel states at each slot. The quantity $\hat{c}_n^{(\eta)}(t)$ is the estimated

[1] In practice, the channel gain may change during a time slot especially over very fast fading channel. In this work, for analytical tractability we assume the gain remains constant during each time slot.

[2] We consider both stationary and non-stationary behavior of the fading channels. We call a stationary channel if the statistics such as mean and variance of the channel gain remain constant over time. Otherwise, it is considered as non-stationary channel.

CSI of user n at the beginning of time t under policy η. Let $\hat{R}_n^{(\eta)}(t)$ denote the predicted transmission rate of user n at time t which is defined according to (1) by replacing $c_n(t)$ by $\hat{c}_n^{(\eta)}(t)$. The prediction error is defined as $e_n^{(\eta)}(t) = |\hat{R}_n^{(\eta)}(t) - R_n(t)|$. We assume that $R_{min} < \hat{R}_n^{(\eta)}(t) < R_{max}$, and $e_n^{(\eta)}(t) < e_{max}$ for all n, and t.

Under policy $\pi(\eta)$, $\mathcal{S}(t)$ and n are determined as follows: Based on the predicted channel states by prediction algorithm η, L users with the highest *estimated* transmission rate and queue length product are added to $\mathcal{S}(t)$. After acquiring CSI from users in $\mathcal{S}(t)$, a user in $\mathcal{S}(t)$ is scheduled according to (4). Policy $\pi(\eta)$ is given in Algorithm 1 .

Algorithm 1: Policy $\pi(\eta)$

(1) probing decision:

- Step 1: Sort

  $W_n \triangleq Q_n(t)\hat{R}_n^{(\eta)}(t)$,

  in a descending order. Tie is broken randomly.
- Step 2: Construct $\mathcal{S}(t)$ by selecting the first$L$ users in this order.

(2) scheduling decision:
The base station acquires CSI of each user in $\mathcal{S}(t)$ and user $n^* \in \mathcal{S}(t)$ is scheduled to transmit,

$$n^* = \underset{n \in \mathcal{S}(t)}{\operatorname{argmax}}\, Q_n(t)R_n(t). \tag{5}$$

i.e., $\mathcal{J}_{n^*}(t) = 1$, and updates queue lengths according to(2).

For stationary and ergodic time-varying channels, we define *achievable rate region* as the convex hull of the set of arrival rate vectors $\Lambda = (\lambda_1, \ldots, \lambda_N)$ for which there exists an appropriate scheduling policy that stabilizes the network. When the exact channel information for all users is known, i.e., $L = N$, the achievable rate region is the largest. Let Λ_h denote this *hypothetical rate region*, the boundary of which can never be achieved in real systems [36]. It was shown that Max-Weight algorithm with full CSI stabilizes the network for all arrival rate vectors in Λ_h [2].

We next analyze the performance of policy $\pi(\eta)$. Note that depending on the quality of employed prediction method, and the choice of users probed at each slot, policy $\pi(\eta)$ may or may not schedule the user with the actual highest weight at each slot. Given the backlog process, $\mathbf{Q}(t)$, we need to determine how often a policy $\pi(\eta)$ chooses the user with the *actual* maximum queue backlog-rate product. Let $\rho^{\pi(\eta)}(\mathbf{Q}(t))$ be this probability which is defined as:

$$\rho^{\pi(\eta)}(\mathbf{Q}(t)) = \Pr\left[\underset{n}{\operatorname{argmax}}\, Q_n(t)R_n(t) = k \,\middle|\, \underset{n}{\operatorname{argmax}}\, Q_n(t)\hat{R}_n^{(\eta)}(t) = k, \mathbf{Q}(t)\right].$$

The following theorem characterizes the achievable rate region of policy $\pi(\eta) \in \mathcal{F}$, i.e., $\Lambda^{\pi(\eta)}$, as compared to that of Max-Weight algorithm, Λ_h.

Theorem 1. *For all* $\mathbf{Q}(t)$ *and some given* $0 < \epsilon < 1$, *if*

$$\rho^{\pi(\eta)}(\mathbf{Q}(t)) \geq \epsilon$$

then, an ϵ *fraction of the full rate region can be achieved, i.e.,* $\Lambda^{\pi(\eta)} \subseteq \epsilon \cdot \Lambda_h$.

Proof. The proof relies on a theorem given in [37], and the calculation of expected weighted rates obtained by full CSI Max-Weight algorithm and $\pi(\eta) \in \mathcal{F}$ for any given $\mathbf{Q}(t)$. The details of the proof are given in Appendix Appendix A. □

We note that the largest value of ϵ that can be supported by a prediction policy depends on the channel statistics, and it cannot be obtained for a general case. However, to demonstrate the typical structure of ϵ, we consider a simple example where there are two users with identical channel distributions receiving service from the same base station.

Example. The channel gain between the base station and a user is assumed to be iid Rayleigh fading channel with parameter μ. We assume that at most one user can be probed at each slot, i.e., $L = 1$.

Lemma 2. *Under high SNR assumption,* ϵ *is given by,*

$$\epsilon = \exp\left(-\frac{\mu}{P}\left[1 - e^{-e_{max}\left(1+\frac{R_{max}}{R_{min}}\right)}\right]e^{\frac{(R_{max})^2}{R_{min}}}\right)$$

Proof. The proof is provided in Appendix Appendix B. □

Note that ϵ increases, (which in turn increases achievable rate region according to Theorem 1), as the maximum prediction error e_{max} decreases.

4.2. Exploration-exploitation tradeoff

Under policy $\pi(\eta)$ given in Algorithm 1, the quality of prediction depends on both the prediction method and the set of users probed at each slot, $\mathcal{S}(t)$. It is possible that under policy $\pi(\eta)$ some channels may not be probed for a long time if users have small queue backlogs. The states of such channels may not be predicted accurately, especially if they exhibit fast-fading characteristics. Therefore, a good policy should acquire the channel states of not only users with high queue backlog-rate products but also users whose channel states may have changed significantly since the last probing.

We define the *information of an unexplored channel* as the reduction of uncertainty in the channel state given its past observations. Let $I_n^{\pi(\eta)}(t)$ denote the *information* of channel state of user n under policy $\pi(\eta)$ at the beginning of time slot t given past observations. The value of $I_n^{\pi(\eta)}(t)$ depends on the channel prediction algorithm, and it is defined using the Shannon's entropy formula [38] as explained in Section 5. The information obtained by probing a channel whose state was observed recently and many times before is less than the channel which has not been probed for a long time, since the uncertainty in the state of the latter is higher.

Hence, we have two inter-related objectives:

- objective 1: $\max \sum_{n=1}^{N} Q_n(t)\hat{R}_n^{\pi(\eta)}(t)$,
  making a scheduling decision that stabilizes a network with the largest possible achievable region
- objective 2: $\max \sum_{n=1}^{N} I_n^{\pi(\eta)}(t)$,
  probing the set of users that minimizes the channel prediction error.

We seek a *modified* version of policy $\pi(\eta)$, which determines a subset of users probed by considering both objectives, and schedules a user out of this subset according to Max-Weight algorithm. The most common approach to find the solution of a multi-objective optimization problems is the weighted sum method [39]. The problem is stated as follows:

$$\max_{\pi(\eta)\in\mathcal{F}} \sum_{n=1}^{N} \alpha_1 Q_n(t)\hat{R}_n^{\pi(\eta)}(t) + \alpha_2 I_n^{\pi(\eta)}(t), \tag{6}$$

where α_1 and α_2 are the weights assigned to each objective according to their relative importance.

The scheduling and probing decisions depend not only on the queue sizes and the estimated channel rates as was the case in the original Max-Weight algorithm, but also on the uncertainty in each channel state given its past observations. The problem

(6) exhibits the well-known "*Exploration vs. Exploitation*" trade-off by exploiting the users with high backlog-rate product and exploring the current state of the channels with outdated CSI. In the following sections, we deal with a modified version of this problem, where we divide the objective function in (6) by α_1, and define a single weight $\xi = \alpha_2/\alpha_1$. Note that when ξ is tuned to higher (lower) values, we track the channels more (less) closely.

4.3. Multi-objective scheduling and CSI feedback

Multi-Objective Scheduling and Feedback (MOSF) algorithm is given in Algorithm 2 . The algorithm takes into account both the estimated transmission rates of users and the information acquired from each user.

Algorithm 2: MOSF Algorithm

(1) probing decision:

- Step 1: Sort

  $W_n \triangleq Q_n(t)\hat{R}_n^{\pi(\eta)}(t) + \xi I_n(t)$,

  in a descending order. Tie is broken randomly.
- Step 2: Construct $\mathcal{S}(t)$ by selecting the first$L$ users in this order.

(2) scheduling decision:
The base station acquires CSI of each user in $\mathcal{S}(t)$ and user $n^* \in \mathcal{S}(t)$ is scheduled to transmit,

$$n^* = \underset{n \in \mathcal{S}(t)}{\operatorname{argmax}}\, Q_n(t)R_n(t). \tag{7}$$

i.e., $\mathcal{J}_{n^*}(t) = 1$, and updates queue lengths according to(2).

Lemma 3. *Given L, ξ, $\mathbf{Q}(t)$, $\hat{R}_n^{\pi(\eta)}(t)$ and $I_n(t)$ for each user at time slot t, MOSF algorithm solves (6).*

Proof. The proof is straightforward, and it is omitted for brevity. □

Next, we analyze the performance of MOSF algorithm. Let n^* be the user scheduled by Max-Weight algorithm with complete CSI at time t. Under the worst case scenario, i.e., $L = 1$, MOSF algorithm schedules user n^* if the maximum prediction error is below a certain threshold as given by the following Lemma.

Lemma 4. *For $L = 1$, MOSF algorithm schedules user n^* at time t, if the maximum prediction error e_{max} satisfies the following inequality for all $n \neq n^*$,*

$$e_{max} \leq \frac{Q_{n^*}(t)R_{n^*}(t) - Q_n(t)R_n(t) + \xi(I_{n^*}(t) - I_n(t))}{Q_{n^*}(t) + Q_n(t)}, \tag{8}$$

Proof. The proof is provided in Appendix Appendix C. □

Remark. Lemma 4 gives the sufficient condition for the same user to be scheduled by MOSF and Max-Weight algorithms. Note that this condition always holds when $I_{n^*}(t) \geq I_n(t)$ for all $n \neq n^*$, and $\xi \to \infty$ since in this case the user which has the maximum weight and the maximum information is scheduled. On the other hand, the condition does not hold if there is at least one user such that $I_{n^*}(t) \leq I_n(t)$ and if ξ is high. This is because, as ξ increases MOSF algorithm emphasizes the information rather than the backlog-rate product.

Note that as long as MOSF schedules the user with the actual maximum weight correctly at each slot (i.e., the condition in (8) is satisfied), then the rate region achieved by MOSF approaches that of the maximum achievable rate region, Λ_h. The achievable rate region of MOSF can be obtained based on Theorem 1, and Lemma 4. Let p_{min} be the probability that (8) holds for all t. Then, we have the following theorem.

Theorem 5. *MOSF algorithm can achieve a fraction ϵ of the full rate region, Λ_h, where $\epsilon = p_{min}$.*

Proof. The proof follows the same lines as the proof of Theorem 1 and relies on the calculation of expected weighted rates obtained by Max-Weight algorithm with complete CSI and MOSF algorithm for any given $\mathbf{Q}(t)$. The sketch of the proof is given in Appendix Appendix D. □

The exact value of p_{min} depends on the channel characteristics, and it can be calculated in a similar fashion for a given system as was done in Lemma 2. As the number of probed users increases, i.e., as L increases, the prediction error decreases. This is due to the fact that channels are more frequently probed, which in turn helps track the channel states more closely. On the other hand, since the channel states are tracked more closely, the uncertainty in the states of the channels decrease. As a result, the information acquired from an unexplored channel decreases as well, i.e., $I_n(t)$ decreases.

5. Predicting channel states using Gaussian process regression

The implementation of MOSF algorithm involves predicting $R_n(t)$ by employing a particular prediction algorithm η, and measuring $I_n(t)$ for each channel. The prediction of a variable such as $R_n(t)$ is known as the *regression problem* in pattern recognition literature. In this work, we employ Gaussian Process Regression (GPR) to track the variation of channel states.

Let $\mathcal{D}_n(t) = (\mathbf{c}_n, \boldsymbol{\tau}_n)$ denote the set of observations of channel n at the beginning of time slot t, where $\mathbf{c}_n = \{c_n^1, c_n^2, \ldots, c_n^w\}$ denotes the set of the latest w CSI values taken at times, $\boldsymbol{\tau}_n = \{\tau_n^1, \tau_n^2, \ldots, \tau_n^w\}$, and $\tau_n^i < t$, for all $\tau_n^i \in \boldsymbol{\tau}_n$, $i \in \{1, 2, \ldots, w\}$. Let $\hat{c}_n(t)$ be the *predicted* state of the channel n at the beginning of time slot t. The value of $\hat{c}_n(t)$ is predicted by GPR given $\mathcal{D}_n(t)$, as described next.

Let $p(c_n(t)|t, \mathcal{D}_n(t))$ be a posterior distribution of channel n. Note that the foundation of the approach adopted in GPR is Bayesian inference, where the idea is to choose an a priori model and update it with actual experimental data observed. According to GPR, a posteriori distribution is Gaussian with mean $\hat{c}_n(t)$ and variance $v_n(t)$. Specifically, Gaussian process is specified by the kernel function, $k_n(\tau_n^i, \tau_n^j)$, that describes the correlation of channel n between two of its measurements taken at times τ_n^i and τ_n^j. Defining a valid class of kernel functions play a key role for Gaussian processes, as it assures consistency of the model specification. To ensure the validity, kernel function has to be positive definite kernel function.

The squared exponential kernel function or the Gaussian kernel function,

$$k_n(\tau_n^i, \tau_n^j) = \sigma_{n,f}^2 \exp\left[-\frac{(\tau_n^i - \tau_n^j)^2}{2l_n^2}\right]. \tag{9}$$

is widely used when the underlying model function is smooth. However, it is possible that the smoothness of channel state may vary due to obstacles or moving objects etc. In this case, it is more suitable to use a non-stationary covariance function to adapt to this change [40]. We define $\theta_n = (\sigma_{n,f}^2, l_n)$ as the hyperparameter[3]

[3] The hyperparameters of the kernel function can be optimized to further increase the performance of GPR in terms of prediction quality.

set for user n. Given $\mathcal{D}_n(t)$, the estimated channel state $\hat{c}_n(t)$ and variance $v_n(t)$ are determined as follows:

$$\hat{c}_n(t) = \mathbf{k}_n^T(t)\mathbf{K}_n^{-1}\mathbf{c}_n, \tag{10}$$

$$v_n(t) = k_n(t,t) - \mathbf{k}_n^T(t)\mathbf{K}_n^{-1}\mathbf{k}_n(t), \tag{11}$$

where $\mathbf{K}_n$ is a $w \times w$ matrix composed of elements $k_n(\tau_n^i, \tau_n^j)$ for $1 \le i, j \le w$ and $\mathbf{k}_n(t)$ is a vector with elements $k_n(\tau_n^i, t)$ for $\forall \tau_n^i \in \boldsymbol{\tau}_n$. The network scheduler can easily predict the CSI of users at time t by using (10). The variance $v_n(t)$ is used as a representative of the level of uncertainty in the predictions, $I_n(t)$. The calculation of $\hat{c}_n(t)$ in (10) requires the inversion of a $w \times w$ matrix, which is the basic complexity of GPR. In general, the complexity of GPR is $\mathcal{O}(w^3)$, where w is the number of recent channel observations [4]. We note that using a fewer number of recent CSI may result in a higher prediction error. Therefore, one can conclude that there is a trade-off between the complexity and the prediction quality. In our simulations, we observe that the prediction error is sufficiently low even for a small value of w (e.g. $w = 3$) so the complexity can be easily handled.

Shannon's *entropy* of a discrete random variable A is defined as $H(A) = \sum_s p_s \log_s(\frac{1}{p_s})$, where $p(.)$ is the probability distribution function of A [38]. Shannon entropy is often used as a measure of uncertainty of a quantity. In our context, the current realization of CSI, $c_n(t)$, is a random variable. Accordingly, let $H_n^0(c_n(t)|t, \mathcal{D}_n(t))$ and $H_n^1(c_n(t)|t, \mathcal{D}_n(t))$ denote the entropy of the random variable $c_n(t)$ before and after time slot t, respectively when $\mathcal{D}_n(t)$ is given. If channel n is probed at time t, then $H_n^1(c_n(t)|t, \mathcal{D}_n(t))$ will be zero since the channel state is known exactly. Otherwise, the uncertainty increases, i.e., $H_n^1(c_n(t)|t, \mathcal{D}_n(t)) > H_n^0(c_n(t)|t, \mathcal{D}_n(t))$. Hence, the information acquired by probing channel of user n is the reduction in its uncertainty, which is simply the difference between its entropies before and after the probing:

$$I_n(t) = H_n^0(c_n(t)|t, \mathcal{D}_n(t)) - H_n^1(c_n(t)|t, \mathcal{D}_n(t)).$$

The following lemma is similar to the one given in [3], and establishes that the information obtained by probing a channel is equal to the variance of the estimate of the state of that channel.

Lemma 6. *Given $\mathcal{D}_n(t), \forall n = 1, \ldots, N$, finding the channel that has the highest information at time slot t is equal to finding the channel which has the highest variance at that time slot, i.e.,*

$$i^* = \underset{1 \le n \le N}{\operatorname{argmax}}\, I_n(t) = \underset{1 \le n \le N}{\operatorname{argmax}}\, v_n(t). \tag{12}$$

Proof. Since $H_n^1(c_n(t)|t, \mathcal{D}_n(t)) = 0$ after probing, $I_n(t)$ is simply

$$I_n(t) = H_n^0(c_n(t)|t, \mathcal{D}_n(t)). \tag{13}$$

Note that according to GPR a posterior distribution of state of channel given $\mathcal{D}_n$ is

$$p(c_n(t)|t, \mathcal{D}_n) \sim \mathcal{N}(\hat{c}_n(t); v_n(t)). \tag{14}$$

Then, the entropy of this Gaussian distribution is given by,

$$H_n^0(c_n(t)|t, \mathcal{D}_n) = \frac{1}{2}\log(2\pi e v_n(t)). \tag{15}$$

Hence, $i^* = \operatorname{argmax}_{1 \le n \le N} I_n(t) = \operatorname{argmax}_{n \in \mathcal{N}} v_n(t)$. □

6. Numerical analysis

In our simulations, we model a single cell CDMA downlink transmission utilizing high data rate (HDR) [5]. The base station serves 16 users and keeps a separate queue for each user. Time is slotted with length $T_s = 1.67$ ms as defined in HDR specifications. Packets arrive at each slot according to Bernoulli distribution: The size of a packet is 128 bytes which corresponds to the size of an HDR packet. The wireless channel is modeled as *correlated* Rayleigh fading channel according to Jakes' model [41]. The bandwidth of the system is $BW = 1.25$ MHz and transmission power of the base station is $P = 10$ dB. The transmission rate of the user channel is taken as the maximum value given in (1). Doppler frequency of each channel is randomly chosen in the range $f_d = [5, 20]$ Hz. We divide the users into two groups with eight users in each by considering the effect of both slow and fast fading. Note that the channel gain of a user with fast fading changes considerable from slot to slot. On the other hand, for a slow fading channel the variations in the channel gain changes slowly with time. For more rigorous approach, we characterize a channel based on its normalized Doppler frequency i,e. $f_d T_s$ by following the work in [42]. Specifically, we consider that if the normalized Doppler frequency is less than 0.02 then that channel is slow fading. Otherwise, the channel exhibits a fast fading. Hence, the users in the first group experience slow fading, i.e., $f_d/f_s \le 0.02$, and the users in the second group experience fast fading, i.e., $f_d/f_s \ge 0.02$, where $f_s = 600Hz$ is the sampling rate of the channel. We set $\sigma_{n,f} = 1$ and $l_n = 1$ for all users. The simulation is run for 10^5 slots, which is sufficiently long for the average queue sizes to reach their steady-state values

Auto-regression (AR) models were previously used in the literature to estimate the future values of the fading coefficients. We compare MOSF with an algorithm named LAR in [6], which also adopts an AR model. According to the AR model, the current CSI of a user can be predicted when p previous CSIs of that user are given, where p is the order of AR model. After predicting the current channel states of all users, LAR algorithm probes L users with the highest backlog-estimated rate product and schedules the user with the maximum weight in the set of probed users at every slot. Recall that MOSF algorithm employs GPR for channel state prediction based on w most recent observations. We empirically observe that the minimum prediction error by GPR is achieved when $p = 2$ with AR and $w = 3$, and thus, these values are taken throughout all experiments. The performance of the algorithms are measured in terms of their average queue sizes and average estimation errors. Note that the average queue size is an indicator to the average delay experienced by the users. Also note that by inspecting the average rate of change of queue sizes, we can approximately determine the maximum arrival rates that can be supported by different algorithms. The lower bound for average queue sizes is given by Max-Weight algorithm which has full CSI at every time slot. The estimation error in LAR and MOSF is measured as average absolute error (AAE):

$$AAE = \lim_{t \to \infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{n=1}^{N} \left| c_n(t) - \hat{c}_n^{\pi}(t) \right|.$$

6.1. Effect of information

We first conduct an experiment to show the effect of taking into account the uncertainty in the channel states while making probing decisions. In order to keep the plots simple we only depict results for $L = 4$. Fig. 2a shows the average total queue sizes with respect to arrival rates. When the information is not taken into account, i.e., $\xi = 0$, the network can be stabilized for small arrival rates. For instance, for $\xi = 0$ the stabilizable arrival rate is approximately 5 packets/slot, whereas for $\xi = 10^3$, the network can be stabilized for arrival rates up to 7.5 packets/slot. Fig. 2b depicts the average error in channel estimation. Clearly, as ξ increases, the estimation error decreases since the channels are tracked more accurately for higher values of ξ.

6.2. Achievable rate region

Fig. 3 depicts the average total queue sizes of MOSF and LAR with respect to arrival rates. The exploitation factor is set to be

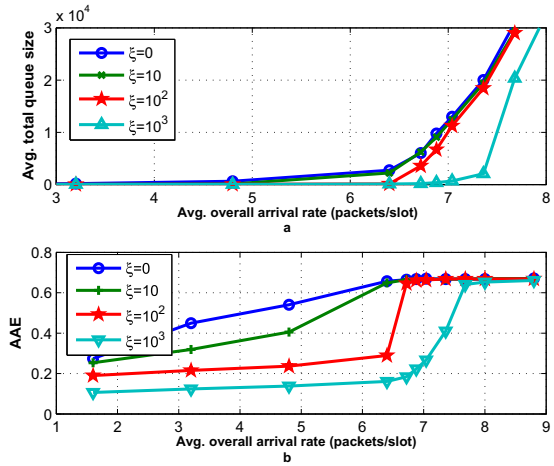

Fig. 2. Average total queue backlogs (Fig. 2a) and absolute channel estimation error (Fig. 2b) vs. arrival rate for varying ξ when L=4.

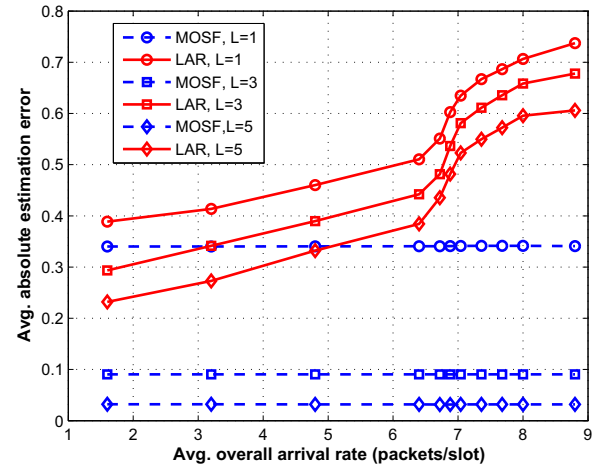

Fig. 4. Average absolute channel estimation error with respect to arrival rates.

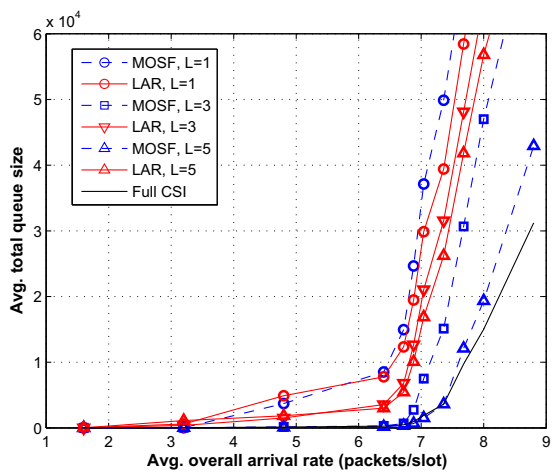

Fig. 3. Average total queue backlogs with MOSF and LAR.

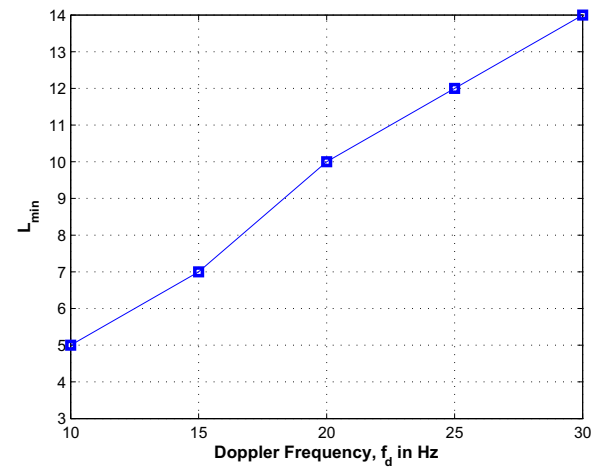

Fig. 5. The required size of the feedback channel for MOSF to achieve Λ_h.

$\xi = 10^5$. As shown in Fig. 3, for $L = 1$ MOSF algorithm is not throughput-optimal since it cannot to stabilize the network for all arrival rates. For instance, when the arrival rate is around 3 packets/slot, the average total queue size suddenly increases, which shows the instability of the network. This is because the learning algorithm does not have sufficiently frequent observations to accurately predict the channel state. However, as L is increased to $L = 3$ the network is stabilized for higher arrival rates (i.e., it stabilizes the network for 6.5 packets/slot), but MOSF still does not stabilize the network for all arrival rates. When we further increase the feedback to $L = 5$, then MOSF algorithm has a comparable performance as that of Max-Weight algorithm with full CSI. On the other hand, LAR algorithm also achieves its best performance when $L = 5$, however, even in that case it cannot stabilize the network for all arrival rates. Hence, we conjecture that LAR achieves approximately smaller rate region as compared to MOSF algorithm. Next, we investigate the performance of MOSF and LAR algorithms in terms of their average absolute error (AAE). As depicted in Fig. 4, as the queue sizes increase, the estimation error increases with LAR. The increase in the error with MOSF algorithm is small and it is more robust to changes in queue sizes than LAR. Moreover, the best error performance for both prediction algorithms is achieved with $L = 5$ and when the arrival rate is at its lowest value. In that case, the average absolute error with MOSF is 0.03 whereas it is equal to 0.23 with LAR. Then, we determine the minimum number of channels, L_{min}, required to stabilize the network for a given arrival rate and to achieve the similar delay performance as with Max-Weight algorithm with full CSI. We set the average total arrival to the network 5 packets/slot. As depicted in Fig. 5, when $f_d = 10$ Hz for all channels, the base station has to probe at least $L_{min} = 5$ channels to achieve the same rate region and the average delay performance. As f_d increases L_{min} increases as well. This is due to the fact that at a higher Doppler frequency the channel states vary faster, which in turn necessitates more observations for the learning algorithm to accurately predict the current channel state. Hence, the base station should collect CSI more frequently as f_d increases. Finally, we compare the performance of the users with slow and fast fading when MOSF is employed. The normalized Doppler frequency for slow fading users is set to 0.003 whereas the fast fading users have a normalized Doppler frequency of 0.03. Fig. 6 depicts the performance of those users in terms of average total queue sizes with respect to the average total arrival rate for different values of L. Interestingly, when $L = 1$ the average queue size of the users with slow fading is higher than that of the users with fast fading. This is due to the fact that MOSF algorithm does not only take into account the channel state information but also the estimation variance. Since the variance is higher for the users with fast fading MOSF algorithm gives higher priority

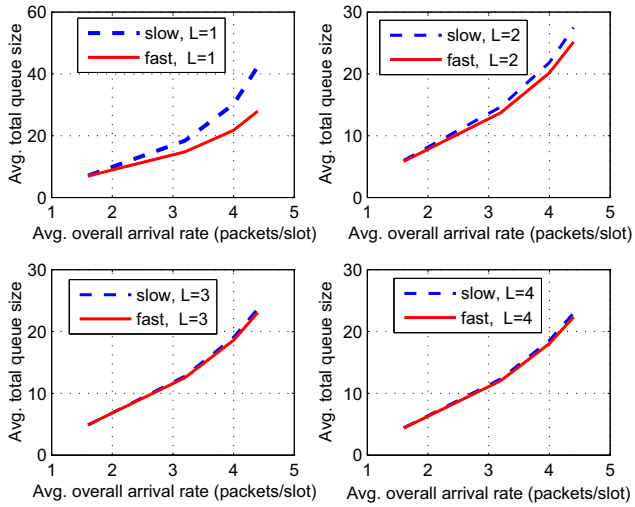

Fig. 6. Performance of slow and fast fading users.

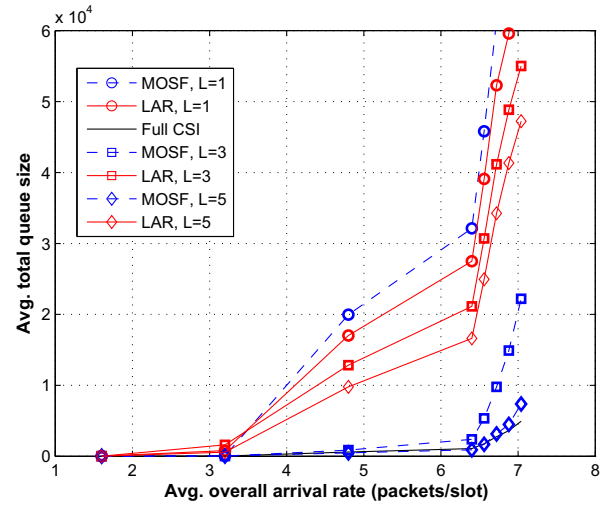

Fig. 7. Average total queue backlogs with MOSF and LAR over non-stationary channels.

to these users for channel access. Thus, their average queue sizes are less then that of slow fading users. From Fig. 6, similar results can be observed when $L = 2$. However, when $L = 3$ and $L = 4$, the estimation error is sufficiently low for both type of users, and hence the average queue sizes for all users become the same.

6.3. Performance over Non-stationary channels

As a non-stationary scenario, we assume that the velocity of users (i.e., in turn their Doppler frequencies) vary in time, and the base station does not have any knowledge about these changes. In our numerical experiment, the normalized Doppler frequency f_d/f_s is increased at a constant rate from 0.003 to 0.03 until $t = 5 \times 10^4$. After 5×10^4, it is decreased again at a constant rate. This model approximately represents the movement of a mobile user such that the user velocity is increasing from $t = 0$ to $t = 5 \times 10^4$ and it is slowing from $t = 5 \times 10^4$ to $t = 10^5$. Fig. 7 depicts that MOSF algorithm outperforms LAR in terms of rate region even in the dynamically changing channels. This is due to the fact that the prediction of GPR depends on the most recent channel observations and how fast the channels change rather than the distribution of channels.

7. Conclusion

In this paper, we considered scheduling in a downlink multi-user setting, where the base station can only probe a limited number of users due to the limited bandwidth on the uplink feedback channel. We have presented a joint scheduling and channel probing algorithm that can operate in a stationary and non-stationary network scenarios. The algorithm is based on an active learning framework that quantifies the reward of learning the current state of the system by using the entropy measure. Based on this measure, the scheduler makes an intelligent trade off between having a more up-to-date picture of the system and maximizing the overall system throughput. The proposed algorithm first decides the set of channels that should be probed at the beginning of each time slot. The set of channels is determined by considering not only the queue sizes and the estimated transmission rates but also the information to be obtained by probing a channel. We apply Gaussian Process Regression technique to predict CSI at each time slot based on the previously observed CSI. In numerical results, we show that the base station using MOSF can stabilize the network and achieve a similar delay performance as compared to full CSI Max-Weight algorithm by probing less than half of the users at every slot. Possible directions for future work include the investigation of the case where the channel gain changes within a time slot. Another possible future direction is to investigate the scheduling problem with incomplete CSI when interference from neighboring cells is present.

Appendix A. Proof of Theorem 1

We consider the worst case in which at most one user is probed at every slot, i.e., $L = 1$. In this case, the probed user is always the scheduled user. Note that when $L > 1$, we may achieve a larger rate region. We also drop η in $\pi(\eta)$ for notational simplicity. Let $\mathcal{J}^f(t) = 1$ if user n is scheduled when full CSI available, otherwise $\mathcal{J}^f(t) = 0$. Similarly, $\mathcal{J}_n^\pi(t) = 1$ if user n is scheduled with policy π. Otherwise, $\mathcal{J}_n^\pi(t) = 0$.

Consider also the following functions :

$$g_f(\mathbf{Q}(t)) = \mathbb{E}\left[\sum_{n=1}^{N} Q_n(t)R_n(t)\mathcal{J}_n^f(t)|\mathbf{Q}(t)\right], \tag{A.1}$$

$$g_\pi(\mathbf{Q}(t)) = \mathbb{E}\left[\sum_{n=1}^{N} Q_n(t)R_n(t)\mathcal{J}_n^\pi(t)|\mathbf{Q}(t)\right]. \tag{A.2}$$

The function (A.1) gives the expected weighted-sum rate according to Max-Weight algorithm with full CSI, whereas (A.2) is the expected-sum rate when at most $L = 1$ channel is probed with policy π. Our aim is to determine $\mathcal{J}_n^\pi(t)$ at every time slot so that $g_f(\mathbf{Q}(t))$ is close to $g_\pi(\mathbf{Q}(t))$. We define the event χ such that χ occurs if policy π and full CSI Max-Weight algorithm schedule the same user at a time slot, i.e.,

$$\operatorname*{argmax}_n Q_n(t)R_n(t) = \operatorname*{argmax}_n Q_n(t)\hat{R}_n^\pi(t).$$

We denote the probability of event χ as $\rho^\pi(\mathbf{Q}(t))$. We use the following theorem given in [37] to prove our main result Theorem 1.

Theorem 7. *[37] If for some $\epsilon > 0$ policy π guarantees*

$$g_\pi(\mathbf{Q}(t)) \geq \epsilon g_f(\mathbf{Q}(t)) \tag{A.3}$$

for all $\mathbf{Q}(t)$, then policy π can achieve a fraction ϵ of hypothetical rate region, Λ_h.

Note that $g_\pi(\mathbf{Q}(t))$ can be rewritten as follows,

$$g_\pi(\mathbf{Q}(t)) = \mathbb{E}\left[\sum_{n=1}^{N} Q_n(t)R_n(t)\mathcal{I}_n^\pi(t)|\mathbf{Q}(t),\chi\right]\rho^\pi(\mathbf{Q}(t)) + \mathbb{E}\left[\sum_{n=1}^{N} Q_n(t)R_n(t)\mathcal{I}_n^\pi(t)|\mathbf{Q}(t),\chi'\right](1-\rho^\pi(\mathbf{Q}(t))).$$

Note that when event χ occurs, the following equality is true,

$$g_f(\mathbf{Q}(t)) = \mathbb{E}\left[\sum_{n=1}^{N} Q_n(t)R_n(t)\mathcal{I}_n^\pi(t)|\mathbf{Q}(t),\chi\right].$$

Thus, we have

$$g_\pi(\mathbf{Q}(t)) = g_f(\mathbf{Q}(t))\rho^\pi(\mathbf{Q}(t)) + \mathbb{E}\left[\sum_{n=1}^{N} Q_n(t)R_n(t)\mathcal{I}_n^\pi(t)|\mathbf{Q}(t),\chi'\right](1-\rho^\pi(\mathbf{Q}(t))).$$

Note that,

$$\mathbb{E}\left[\sum_{n=1}^{N} Q_n(t)R_n(t)\mathcal{I}_n^\pi(t)|\mathbf{Q}(t),\chi'\right](1-\rho^\pi(\mathbf{Q}(t))) \geq 0$$

Hence,

$$g_\pi(\mathbf{Q}(t)) \geq g_f(\mathbf{Q}(t))\rho^\pi(\mathbf{Q}(t)) \tag{A.4}$$

By dividing both sides of (A.4) by $g_f(\mathbf{Q}(t))$, we obtain,

$$\frac{g_\pi(\mathbf{Q}(t))}{g_f(\mathbf{Q}(t))} \geq \rho^\pi(\mathbf{Q}(t)) \tag{A.5}$$

Thus, if $\rho^\pi(\mathbf{Q}(t)) \geq \epsilon$, then $g_\pi(\mathbf{Q}(t))/g_f(\mathbf{Q}(t)) \geq \epsilon$. Hence, according to theorem [37], the scheduling policy with estimated channel rates can achieve as least ϵ fraction of Λ_h.

Appendix B. Proof of Lemma 2

Let transmission rates of users be defined as $R_1(t)$ and $R_2(t)$, and their queue sizes be defined as $Q_1(t)$, $Q_2(t)$, respectively. For analytical simplicity, we assume high SNR approximation, i.e., $R_n(t) \approx \ln(P|c_n(t)|^2)$ and we drop η in $\pi(\eta)$ and time index for notational simplicity. Then, we determine the probability that the same user is scheduled by policy π and full CSI Max-Weight algorithm. Formulary:

$$p^\pi = \Pr\left(Q_1R_1 \geq Q_2R_2 | Q_1\hat{R}_1^\pi \geq Q_2\hat{R}_2^\pi, Q_1, Q_2\right)$$

We assume the worst case scenario in which the estimation error is e_{max} and R_1 is overestimated whereas R_2 is underestimated. In this scenario we have,

$$Q_1\hat{R}_1^\pi \geq Q_2\hat{R}_2^\pi$$
$$Q_1(R_1 + e_{max}) \geq Q_2(R_2 - e_{max})$$
$$Q_1R_1 - Q_2R_2 \geq -e_{max}(Q_1 + Q_2)$$

Let us define $K \triangleq e_{max}(Q_1 + Q_2)$. Now, p^π can be rewritten as follows:

$$\begin{aligned} p^\pi &= \Pr(Q_1R_1 \geq Q_2R_2 | Q_1R_1 - Q_2R_2 \geq -K, Q_1, Q_2) \\ &= \frac{\Pr(Q_1R_1 \geq Q_2R_2 \text{ and } Q_1R_1 - Q_2R_2 \geq -K | Q_1, Q_2)}{\Pr(Q_1R_1 - Q_2R_2 \geq -K | Q_1, Q_2)} \\ &= \frac{\Pr(Q_1R_1 - Q_2R_2 \geq 0 | Q_1, Q_2)}{\Pr(Q_1R_1 - Q_2R_2 \geq -K | Q_1, Q_2)} \end{aligned}$$

We first calculate the following probability,

$$p_{num}^\pi = \Pr(Q_1R_1 - Q_2R_2 \geq 0 | Q_1, Q_2).$$

Since $|c_1(t)|^2$ and $|c_2(t)|^2$ are identically distributed exponential random variables with parameter μ, p_{num}^π is determined as follows:

$$\begin{aligned} p_{num}^\pi &= \Pr\left(R_1 \geq \frac{Q_2R_2}{Q_1} \mid Q_1, Q_2\right) \\ &= \Pr\left(P|c_1|^2 \geq \frac{e^{aR_2}}{P} \mid Q_1, Q_2\right) \\ &= \int_0^\infty \Pr\left(|c_1|^2 \geq \frac{e^{ar_2}}{P} \mid Q_1, Q_2, R_2 = r_2\right) f_{R_2}(r_2) d_{r_2} \end{aligned}$$

where $a = \frac{Q_2}{Q_1}$ and $f_{R_2}(r_2)$ is the pdf of R_2 which is given by,

$$f_{R_2}(r_2) = \frac{\mu}{P} e^{r_2} e^{-\frac{\mu}{P}e^{r_2}}, \text{ for } r_2 \geq 0.$$

Hence,

$$\begin{aligned} &\Pr\left(|c_1|^2 \geq \frac{e^{ar_2}}{P} \mid Q_1, Q_2, R_2 = r_2\right) \\ &= \int_{\frac{e^{aR_2}}{P}}^\infty \mu e^{-\mu c_1} dc_1 \\ &= e^{-\frac{\mu}{P}e^{ar_2}} \end{aligned}$$

Then, p_{num}^π is given by,

$$\begin{aligned} p_{num}^\pi &= \int_0^\infty e^{-\mu e^{\frac{ar_2}{P}}} f_{R_2}(r_2) d_{r_2} \\ &= \int_0^\infty \frac{\mu}{P} e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{ar_2}} d_{r_2} \end{aligned}$$

Let $b = \frac{K}{Q_1}$. Next, we determine the following probability,

$$p_{denum}^\pi = \Pr(Q_1R_1 - Q_2R_2 \geq -K | Q_1, Q_2).$$

By following the same way, p_{denum}^π is given as,

$$p_{denum}^\pi = \int_0^\infty \frac{\mu}{P} e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{ar_2 - b}} d_{r_2}$$

Hence, we have,

$$\begin{aligned} p^\pi &= \frac{p_{num}^\pi}{p_{denum}^\pi} \\ &= \frac{\int_0^\infty \frac{\mu}{P} e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{ar_2}} d_{r_2}}{\int_0^\infty \frac{\mu}{P} e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{ar_2 - b}} d_{r_2}} \end{aligned} \tag{B.1}$$

Since $R_{min} < r_2 < R_{max}$, p^π can be approximated as follows,

$$p^\pi \approx \frac{\int_{R_{min}}^{R_{max}} e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{ar_2}} d_{r_2}}{\int_{R_{min}}^{R_{max}} e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{-b}e^{ar_2}} d_{r_2}} \tag{B.2}$$

Note that the following inequality holds,

$$e^{-\frac{\mu}{P}e^{-b}e^{ar_2}} \leq e^{-\frac{\mu}{P}e^{ar_2} + \mu B}$$

where $B = e^{aR_{max}}(1 - e^{-b})$. Hence,

$$\int_0^\infty e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{-b}e^{ar_2}} d_{r_2} \leq e^{\mu B} \int_0^\infty e^{r_2} e^{-\frac{\mu}{P}e^{r_2}} e^{-\frac{\mu}{P}e^{ar_2}} d_{r_2}$$

By applying this result to (B.1), we have

$$p^\pi \geq e^{-\mu B} \tag{B.3}$$

We know also that

$$Q_2 \leq Q_1 \frac{\hat{R}_1^\pi}{\hat{R}_2^\pi}$$

Since $R_{min} \le \hat{R}_i^{\pi} \le R_{max}$, $i = 1, 2$ we have,

$$Q_2 \le Q_1 \frac{R_{max}}{R_{min}}$$

By applying this result to (B.3), we have,

$$p^{\pi} \ge \exp\left(-\frac{\mu}{P}\left[1 - e^{-e_{max}\left(1+\frac{R_{max}}{R_{min}}\right)}\right] e^{\frac{(R_{max})^2}{R_{min}}}\right) \ \forall t.$$

Hence, $\epsilon = \exp\left(-\frac{\mu}{P}\left[1 - e^{-e_{max}\left(1+\frac{R_{max}}{R_{min}}\right)}\right] e^{\frac{(R_{max})^2}{R_{min}}}\right)$. Policy π can achieve ϵ fraction of rate region of Max-Weight algorithm with full CSI.

Appendix C. Proof of Lemma 4

Since n^* is the scheduled user with Max-Weight algorithm with full CSI, the following inequality is true,

$$Q_{n^*}(t)R_{n^*}(t) \ge Q_n(t)R_n(t) \ \forall n \ne n^*$$

When MOSF algorithm decides to schedule user n^* then the following inequality must be satisfied,

$$Q_{n^*}(t)\hat{R}_{n^*}(t) + \xi I_{n^*}(t) \ge Q_n(t)\hat{R}_n(t) + \xi I_n(t), \ \forall n \ne n^*$$

We consider the worst case scenario in which $e_n(t) = e_{max}$, for all n and GPR underestimates for channel n^* whereas it overestimates for all other channels, i.e.,

$$\hat{R}_{n^*}(t) = R_{n^*}(t) - e_{max}$$

$$\hat{R}_n(t) = R_n(t) + e_{max}, \forall n \ne n^*$$

Hence, the following inequality must be true so that MOSF algorithm schedules user n^* in the worst case,

$$Q_{n^*}(t)(R_{n^*}(t) - e_{max}) + \xi I_{n^*}(t) \ge Q_n(t)(R_n(t) + e_{max}) + \xi I_n(t), \forall n \ne n^*$$

Thus, e_{max} should satisfy the following condition,

$$e_{max} \le \frac{Q_{n^*}(t)R_{n^*}(t) - Q_n(t)R_n(t) + \xi(I_{n^*}(t) - I_n(t))}{Q_{n^*}(t) + Q_n(t)} \tag{C.1}$$

Appendix D. Proof of Theorem 5

Let $\mathcal{J}_n^m(t)$ represent the scheduling decision with MOSF algorithm. $\mathcal{J}_n^m(t) = 1$ if user n is scheduled with MOSF. Otherwise, $\mathcal{J}_n^m(t) = 0$. Let the condition in Lemma 4 hold with probability $\rho^m(\mathbf{Q}(t))$ at time slot t. We consider the following function:

$$g_m(\mathbf{Q}(t)) = \mathbb{E}\left[\sum_n Q_n(t)R_n(t)\mathcal{J}_n^m(t)|\mathbf{Q}(t)\right]$$

Using arguments similar to those in Theorem 1, we have

$$\frac{g_m(\mathbf{Q}(t))}{g_f(\mathbf{Q}(t))} \ge \rho^m(\mathbf{Q}(t))$$

If $\rho^m(\mathbf{Q}(t)) \ge p_{min}$ for all t, then MOSF can achieve a fraction $\epsilon = p_{min}$ of rate region Λ_h.

References

[1] M. Karaca, T. Alpcan, O. Ercetin, Smart scheduling and feedback allocation over non-stationary wireless channels, ICC'12WSSCPA, 2012.
[2] L. Tassiulas, A. Ephremides, Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks, IEEE Trans. Autom. Control 37 (12) (1992) 1936–1948.
[3] T. Alpcan, A framework for optimization under limited information, J. Global Optim. 55 (3) (2012) 681–706.
[4] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning), The MIT Press, 2005.
[5] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, S. Viterbi, CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users, IEEE Commun. Mag. 38 (7) (2000) 70–77.
[6] A. Duel-Hallen, Fading channel prediction for mobile radio adaptive transmission systems, Proc. IEEE 95 (12) (2007) 2299–2313.
[7] A. Gopalan, C. Caramanis, S. Shakkottai, On wireless scheduling with partial channel-state information, in: Proc. of the 45th Allerton Conference, 2007.
[8] M. Ouyang, L. Ying, On optimal feedback allocation in multichannel wireless downlinks, Mobihoc, 2011.
[9] P. Chaporkar, A. Proutiere, H. Asnani, A. Karandikar, Scheduling with limited information in wireless systems, Mobihoc, 2009.
[10] M.J. Neely, Max weight learning algorithms with application to scheduling in unknown environments, IEEE Trans. Autom. Control 57 (5) (2012) 1179–1191.
[11] W. Ouyang, S. Murugesan, A. Eryilmaz, N.B. Shroff, Scheduling with rate adaptation under incomplete knowledge of channel/estimator statistics, in: 48th Allerton Conference on Communication, Control, and Computing, 2010.
[12] W. Ouyang, S. Murugesan, A. Eryilmaz, N.B. Shroff, Exploiting channel memory for joint estimation and scheduling in downlink networks, INFOCOM, 2011.
[13] M. Karaca, Y. Sarikaya, O. Ercetin, T. Alpcan, H. Boche, Joint opportunistic scheduling and selective channel feedback, IEEE Trans. Wireless Commun. 12 (6) (2013) 3024–3034.
[14] X. Qin, R. Berry, Opportunistic splitting algorithms for wireless networks, in: INFOCOM, 6, 2004, pp. 1662–1672.
[15] J. Huang, V. Subramanian, R. Agrawal, R. Berry, Joint scheduling and resource allocation in uplink ofdm systems for broadband wireless access networks, IEEE J. Sel. Areas Commun. 27 (2) (2009) 226–234.
[16] V. Hassel, D. Gesbert, M.-S. Alouini, G. Oien, A threshold-based channel state feedback algorithm for modern cellular systems, IEEE Trans. Wireless Commun 6 (7) (2007) 2422–2426.
[17] T. Tang, R. Heath, Opportunistic feedback for downlink multiuser diversity, IEEE Commun. Lett 9 (10) (2005) 948–950.
[18] M. Deghel, M. Assaad, M. Debbah, Queueing stability and csi probing of a tdd wireless network with interference alignment, IEEE International Symposium on Information Theory (ISIT), 2015.
[19] B. Li, A. Eryilmaz, Distributed channel probing for efficient transmission scheduling in wireless networks, IEEE Trans. Mobile Comput. 14 (6) (2015) 1176–1188.
[20] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, M. Andrews, An overview of limited feedback in wireless communication systems, IEEE J. Sel. Areas Commun. 26 (8) (2008) 1341–1365.
[21] J.A. Boyan, M.L. Littman, Packet routing in dynamically changing networks: a reinforcement learning approach, Advances in Neural Information Processing Systems, 1994.
[22] F. Fu, M. van der Schaar, Learning to compete for resources in wireless stochastic games, IEEE Trans. Veh. Technol. 58 (4) (2009) 1904–1919.
[23] M. Bennis, D. Niyato, A Q-learning based approach to interference avoidance in self-organized femtocell networks, Proceeding. of GLOBECOM Workshops (GC Wkshps), 2010.
[24] K. Liu, Q. Zhao, Distributed learning in multi-armed bandit with multiple players, IEEE Trans. Signal Process 58 (11) (2010) 5667–5681.
[25] Y. Gai, B. Krishnamachari, R. Jain, Combinatorial network optimization with unknown variables: multi-armed banditswith linear rewards and individual observations, IEEE/ACM Trans. Netw. 20 (5) (2012) 1466–1478.
[26] K. Liu, Q. Zhao, B. Krishnamachari, Dynamic multichannel access with imperfect channel state detection, IEEE Trans. Signal Process 58 (5) (2010) 2795–2808.
[27] A. Duel-Hallen, H.H.S. Hu, Long-range prediction of fading signals: enabling adaptive transmission for mobile radio channels, IEEE Signal Process. Mag. 17 (3) (2000) 62–75.
[28] M. Andrews, L. Zhang, Scheduling over nonstationary wireless channels with finite rate sets, IEEE/ACM Trans. Netw. 14 (5) (2006) 1067–1077.
[29] A. Schwaighofer, M. Grigoras, V. Tresp, C. Hoffmann, Gpps: a gaussian process positioning system for cellular networks, Advances in Neural Information Processing Systems, 2004.
[30] D. Gu, H. Hu, Spatial gaussian process regression with mobile sensor networks, IEEE Trans. Neural Netw. Learn. Syst 23 (8) (2012) 1279–1290.
[31] A. Krause, A. Singh, C. Guestrin, Near-optimal sensor placements in gaussian processes: theory, efficient algorithms and empirical studies, J. Mach. Learn. Res. 9 (2) (2008) 235–284.
[32] F. Perez-Cruz, S.V. Vaerenbergh, J. Murillo-Fuentes, M. Lazaro-Gredilla, I. Santamaria, Gaussian processes for nonlinear signal processing, IEEE Signal Process. Mag. 30 (4) (2013) 40–50.
[33] B. Radunovic, A. Proutiere, D. Gunawardena, P. Key, Exploiting channel diversity in white spaces, Technical Report MSR-TR-2011-53, 2011.
[34] D.K. Aktas, M.P. Fitz, Distance spectrum analysis of spacetime trellis-coded modulations in quasi-static rayleigh-fading channels, IEEE Trans. Inf. Theory 49 (12) (2003) 3335–3344.
[35] 3GPP TS 36.212. "Multiplexing and channel coding." 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA). URL: http://www.3gpp.org.
[36] J.G. Andrews, S. Shakkottai, R. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar, A. Yener, Rethinking information theory for mobile ad hoc networks, IEEE Commun. Mag. 46 (12) (2008) 94–101.
[37] A. Eryilmaz, R. Srikant, J. Perkins, Stable scheduling policies for fading wireless channels, IEEE/ACM Trans. Netw. 13 (2005) 411–424.

[38] C.E. Shannon, Communication theory of secrecy systems, Bell Syst. Tech. 28 (3) (1949) 565–715.
[39] R.T. Marler, J.S. Arora, Survey of multi-objective optimization methods for engineering, Structural and Multidisciplinary Optimization (2004).
[40] C. Paciorek, M. Schervish, Nonstationary covariance functions for gaussian process regression, in: Advances in Neural Information Processing Systems, 2004, pp. 273–280.
[41] W.C. Jakes, Microwave Mobile Communications, Piscataway NJ IEEE Press, 1994.
[42] L. Zhong, F. Alajaji, G. Takahara, A model for correlated rician fading channels based on a finite queue, IEEE Trans. Veh. Technol. 57 (1) (2008) 79–89.

Mehmet Karaca received the BS degree in Telecommunication Engineering from Istanbul Technical University, Istanbul, Turkey in 2006 and the MS and PhD degrees in Electronics Engineering from Sabanci University, Istanbul, Turkey in 2008 and 2013, respectively. He worked as a wireless systems engineer at Airties Wireless Networks, Turkey for two years. He is currently postdoctoral researcher at Lund University, Sweden. His research interests include the design and analysis of scheduling and resource allocation algorithms for wireless networks, stochastic optimization and machine learning, the next generation IEEE WLANs.

Ozgur Ercetin received the BS degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey, in 1995 and the MS and PhD degrees in electrical engineering from the University of Maryland, College Park, in 1998 and 2002, respectively. Since 2002, he has been with the Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul. He was also a visiting researcher at HRL Labs, Malibu, CA, Docomo USA Labs, CA, and The Ohio State University, OH. His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models, architectures and protocols of wireless systems, and stochastic optimization.

Tansu alpcan received the B.S. degree in electrical engineering from Bogazici University, Istanbul, Turkey in 1998. He received the M.S. and Ph.D. degrees in electrical and computer engineering from University of Illinois at Urbana-Champaign in 2001 and 2006, respectively. His research involves applications of distributed decision making, game theory, optimisation, and control to various security and resource allocation problems in complex and networked systems. He is recipient of multiple research and best paper awards from UIUC and IEEE. He has played a role in organization of several workshops and conferences such as IEEE Infocom, ICC, GameComm, and GameSec as TPC member, associate editor, co-chair, chair, and steering board member. He is the (co-)author of more than 100 journal and conference articles, an edited volume, as well as the book "Network Security: A Decision and Game Theoretic Approach" published by Cambridge University Press in 2011. He has worked as a senior research scientist in Deutsche Telekom Laboratories, Berlin, Germany, between 2006-2009, and as Assistant Professor in Technical University of Berlin from 2009 until 2011. He has joined the Department of Electrical and Electronic Engineering in the University of Melbourne as Senior Lecturer in October 2011.