Time Complexity Analysis of Distributed Stochastic Optimization in a Non-Stationary Environment

B. N. Bharath and P. Vaishali
Dept. of ECE, PESIT Bangalore South Campus,
Bangalore 560100, INDIA
E-mail: bharathbn@pes.edu, vaishali.p.94@gmail.com

***Abstract*—In this paper, we consider a distributed stochastic optimization problem where the goal is to minimize the time average of a cost function subject to a set of constraints on the time averages of related stochastic processes called penalties. We assume that the state of the system is evolving in an independent and non-stationary fashion and the "common information" available at each node is distributed and delayed. Such stochastic optimization is an integral part of many important problems in wireless networks such as scheduling, routing, resource allocation and crowd sensing. We propose an approximate distributed Drift-Plus-Penalty (DPP) algorithm, and show that it achieves a time average cost (and penalties) that is within $\epsilon > 0$ of the optimal cost (and constraints) with high probability. Also, we provide a condition on the convergence time t for this result to hold. In particular, for any delay $D \geq 0$ in the common information, we use a coupling argument to prove that the proposed algorithm converges *almost surely* to the optimal solution. We use an application from wireless sensor network to corroborate our theoretical findings through simulation results.**

Index terms: Drift-plus-penalty, Lyapunov function, wireless networks, online learning, distributed stochastic optimization.

I. Introduction

Stochastic optimization is ubiquitous in various domains such as communications, signal processing, power grids, inventory control for product assembly systems and dynamic wireless networks [1]–[6]. A typical stochastic optimization problem involves designing control action for a given state of the system that minimizes the time average of a cost function subject to a set of constraints on the time average penalties [1], [2]. Both cost and penalties depend on the state of the system and the control actions taken by the users. For example, in a typical wireless application, the cost function refers to the instantaneous rate, and the penalty refers to the instantaneous power consumed. Further, the state here refers to the channel condition. An algorithm known as Drift-Plus-Penalty (DPP) (see [7]–[10]) is known to provide a solution for these problems with theoretical guarantees. At each time slot, the DPP method, an extension of the back-pressure algorithm [11], [12], finds a control action that minimizes a linear combination of the cost and the drift. In the problem that we consider, the drift is a measure of the deviation (of the penalties) from the constraints, and the penalty corresponds to the cost. The DPP algorithm is shown to achieve an approximately optimal solution even when the system evolves in a non-stationary fashion, and is robust to non-ergodic changes in the state [7].

The DPP algorithm mentioned above assumes that the control action is taken at a centralized unit where the complete state information is available. However, wireless network and crowd sensing applications require a distributed control action that uses only the delayed state information at each node [7], [13]. This calls for a *distributed* version of the DPP algorithm with theoretical guarantees. The author in [3] considers a relaxed version of the above problem. In particular, assuming i.i.d. states with correlated "common information," the author in [3] proposes a distributed DPP algorithm, and proves that the proposed algorithm is close to being optimal in the average sense. Several authors use the above results in various contexts such as crowd sensing [13], energy efficient scheduling in MIMO systems [14], to name a few. However, in many practical applications, the states evolve in a dependent and *non-stationary* fashion [10]. Thus, the following assumptions about the state made in [3] need to be relaxed: (i) independent and (ii) identically distributed. In addition, from a practical standpoint, it is important to investigate the rate of convergence of the distributed algorithm to the optimal. In this paper, we relax the assumption (ii) above, and unlike [3], we provide a Probably Approximately Correct (PAC) bound on the performance. Also, we prove an *almost sure* convergence of the proposed *distributed* algorithm to a constant within the optimal. We would like to emphasize that extending the analysis in [3] to non-stationary states is non-trivial. The only work that provides a "PAC type" result for the DPP algorithm is [15]. However, the authors consider i.i.d. states, and the decision is *centralized*. Moreover, the method used in [15] cannot be directly extended to a problem with non-stationary states since their proof requires the control action to be stationary, and this assumption in general is not true. Now, we highlight the contribution of our work.

A. Main Contribution of the Paper

In this paper, we consider a distributed stochastic optimization problem when the states evolve in an independent and *non-stationary* fashion. In particular, we assume that the state is asymptotically stationary, i.e., the probability measure π_t of the state $\omega(t) \in \Omega$ converges to a probability measure π as $t \to \infty$ in the $\mathcal{L}_1$-norm sense. This assumption makes the extension of the method in [3] non-trivial. When $\pi_t = \pi$ for all $t \in \mathbb{N}$, the author in [3] proves theoretical guarantees by making use of the equivalence between a Linear Program (**LP**) that is a function of π and the original stochastic optimization problem. However, when the probabilities are changing, this equivalence is difficult to establish. Instead, we show that the

978-3-901882-90-6/17 ©2017 IFIP

original problem is equivalent to a "perturbed" **LP**, which is a function of the limiting distribution π. Under mild conditions, we prove that the solution to the perturbed **LP** is approximately equal to that of the original problem. We use this result to prove theoretical guarantees for an approximate DPP algorithm that we propose in the paper. Moreover, unlike the previous works, we are more interested in providing sample complexity bounds rather than just dealing with the averages. The following are the main contributions of our work

1) For the above model, we show that with high probability, the average cost and penalties obtained by using the proposed approximate *distributed* DPP algorithm are within constants of the optimal solution and the constraints, respectively, provided the waiting time $t >$ a threshold (see Theorem 3). The threshold and the constants capture the degree of non-stationarity, and the number of samples used to compute an estimate of the state distribution.
2) Using the high probability result, we show that the cost corresponding to the proposed algorithm *almost surely* converges to a constant within $\epsilon_0 > 0$ of the optimal cost. We also show that the penalties induced by the proposed algorithm are within constants of the constraint values *almost surely*. It turns out that although the states are independent, the proposed algorithm induces dependencies across time in the cost and penalties. Thus, the method in [15] cannot be used as the proof there requires the control action to be stationary. To overcome this, we prove the PAC and the almost sure convergence results using a coupling argument, where the dependent sequence of the cost (and penalties) is replaced by an independent sequence which results in an error that is expressed in terms of the β_1-mixing coefficient; a term that captures the stochastic dependency across time (see Sec. II). Note that the β_1-mixing coefficient depends on the algorithm. In this paper, the β_1-mixing coefficient induced by the proposed approximate DPP algorithm is bounded using information theoretic technique. Further, in the centralized scenario with single user and i.i.d. states, we recover the results in [15] as a special case.
3) We show that due to non-stationarity of the states, the performance gap goes down slowly compared to i.i.d. states. This is captured through $\|\pi_t - \pi\|_1$ and a term that depends on the measure of the complexity of the probability space averaged with respect to π_t (see Theorem 3). Finally, we provide simulation results of a sensor network application, which is a particular use case scenario of the problem considered.

The paper is organized as follows. The motivation, problem statement, an approximate DPP Algorithm with related theoretical guarantees, a bound on the β_1-mixing coefficient, and simulation results are provided in Sec. II, Sec. III, Sec. IV, Sec. V and Sec. VI, respectively. Finally, Sec. VII concludes the paper.

Notation: We use $f(x) = \mathcal{O}(g(x))$, and $f(x) \preceq g(x)$ to mean $\lim_{x\to\infty} \frac{f(x)}{g(x)} = c$, and $\lim_{x\to\infty} \frac{f(x)}{g(x)} \leq c$, $c < \infty$, respectively.

II. Motivation

Towards motivating the system model studied in the paper, we consider a network of 3 sensors, where the sensor i observes the state $\omega_i(t) \in \{0, 1, 2, 3\}$, $i = 1, 2, 3$, and reports the observation to a central unit [3]. The meaning of various states depends on the application. For example, when the sensors are used to monitor vehicular traffic in a particular link, then $\omega_i(t) = 0, 1, 2, 3$ represent no, low, medium and high traffics, respectively. The reporting incurs a penalty in terms of the power consumed by the sensors to transmit the state information. The state $\omega(t) \triangleq \{\omega_1(t), \omega_2(t), \omega_3(t)\}$, $t \in \mathbb{N}$ in general is a stochastic process that evolves in a *non-stationary* fashion. Assume that the central unit trusts sensor 1 more than the others, which may be due to the fact that the sensing capability of this sensor is better than the others. The problem is to maximize the average of the following utility[1] function subject to the constraint that the average power consumed by each sensor is less than $\bar{P}$:

$$u_0(t) \triangleq \min\left\{\frac{\alpha_1(t)\omega_1(t)}{3} + \frac{\alpha_2(t)\omega_2(t) + \alpha_3(t)\omega_3(t)}{6}, 1\right\}, \tag{1}$$

where $\alpha_i(t) \in \{0, 1\}$, $i = 1, 2, 3$ are the decision variables. Note that if $\omega_i(t) = 3$ for $i = 1, 2, 3$, and $\alpha_i(t) = 1$ for $i = 2, 3$, then there is no increase in the utility if sensor 1 also decides to transmit, i.e., $\alpha_1(t) = 1$. However, none of the sensors know the entire state of the system. In this case, the sensor 1 may also choose to transmit, thus wasting its power leading to a suboptimal operation compared to a centralized scheme. In order to resolve this issue in a distributed setting, we assume that a delayed "common information" is available (see Sec. II of [3] for more details) using which each sensor picks one of the "pure strategies" (non-random function). For example, each sensor can acquire the information about the state $\omega(t)$ with a fixed delay $D > 0$. In this case, the "common information" can be some function of $\omega(t - D)$, using which each user can unanimously pick one of the different strategies. Here, each strategy maps to a unique set of states available at each user. Thus, the problem is to find the set of optimal decision variables in a distributed fashion with "common information" that maximizes the average of the above utility subject to the constraints on the average power. We are interested in finding a high probability result for the convergence of the utility to the optimal, which is one of the main difference compared to [3]. Next, we describe the system model that generalizes the above example.

III. System Model and Problem Statement

Consider a system comprising of N users making decisions in a distributed fashion at discrete time steps $t \in \{0, 1, 2, \ldots\}$ (see [3], [7], [10], [13]). Each user i observes a random state $\omega_i(t) \in \Omega_i$, and a "common information" $Y_c(t) \in \mathcal{Y}$ to make a control decision $\alpha_i(t) \in \mathcal{A}_i$, $i = 1, 2, \ldots, N$. Here, for each user i, Ω_i, $\mathcal{Y}$ and $\mathcal{A}_i$ denote the state space, common information space and action/control space, respectively. Let $\omega(t) \triangleq \{\omega_1(t), \omega_2(t), \ldots, \omega_N(t)\} \in \Omega$ and

[1] Maximizing the utility is equivalent to minimizing the negative cost.

$\alpha(t) \triangleq \{\alpha_1(t), \alpha_2(t), \ldots, \alpha_N(t)\} \in \mathcal{A}$, where $\Omega \triangleq \Omega_1 \times \Omega_2 \times \ldots \times \Omega_N$, and $\mathcal{A} \triangleq \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_N$. Also, let us assume that the number of possible values that $p_k(t)$ takes is finite and equal to $\mu_k \in \mathbb{N}$, $k = 1, \ldots, K$. The decision is said to be *distributed* if (see [3])

- There exists a function $f_i : \Omega_i \times \mathcal{Y} \to \mathcal{A}_i$, such that

$$\alpha_i(t) \triangleq f_i(\omega_i(t), Y_c(t)), \ i = 1, 2, \ldots, N, \qquad (2)$$

  where $Y_c(t)$ belongs to the common information set $\mathcal{Y}$.
- The common information $Y_c(t)$ is independent of $\omega(t)$ for every $t \in \mathbb{N}$.

At each time slot t, the decision $\alpha(t)$ and the state $\omega(t)$ result in a cost $p_0(t) \triangleq p_0(\alpha(t), \omega(t))$ and penalties $p_k(t) \triangleq p_k(\alpha(t), \omega(t))$, $k = 1, 2, \ldots, K$.

The central goal of the paper is to analyze an approximate distributed solution to the following problem when $\omega(t)$, $t \in \mathbb{N}$ is independent and *non-stationary*, $\mathbf{P_0}$:

$$\min_{\alpha(\tau)\in\mathcal{A}:\tau\in\mathbb{N}} \quad \limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E} p_0(\tau)$$

$$\text{subject to} \ \limsup_{t\to\infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E} p_k(\tau) \leq c_k, \ k = 1, 2, \ldots, K,$$

$$\alpha_i(\tau) \text{ satisfies (2)}, \ i = 1, 2, \ldots, N.$$

In the above, the expectation is jointly with respect to the distribution of the state $\omega(t)$ and a possible randomness in the decision $\alpha(t)$, $t \in \mathbb{N}$. Let $p^{(opt)}$ be the optimal cost corresponding to the problem $\mathbf{P_0}$. Note that the first equation in $\mathbf{P_0}$ represents the time average cost while the second and the third equations represent constraints on the penalties and the decisions, respectively. Informally, we are interested in proving a Probably Approximately Correct (PAC) type result and *almost sure* result of the following form [15]

- For every $\epsilon_k > 0$, with a probability of at least $1 - \delta_k$, $\frac{1}{t}\sum_{\tau=0}^{t-1} p_k^{(\approx)}(\tau) \leq c_k + \epsilon_k$ provided $t > \texttt{threshold}$, where $p_0^{(\approx)}(\tau)$ and $p_k^{(\approx)}(\tau)$, $k = 1, 2, \ldots, K$ are the cost and penalties, respectively, of the proposed approximate distributed scheme at $\tau \in \mathbb{N}$. Here $c_0 \triangleq p^{(opt)}$ is the optimal cost, and c_k, $k = 1, \ldots, K$ are as defined in $\mathbf{P_0}$ (see Theorem 3).
- The cost $\frac{1}{t}\sum_{\tau=0}^{t-1} p_k^{(\approx)}$ converges *almost surely* to c_k, where c_k, $k = 0, 1, \ldots, K$ are as defined above (see Theorem 6).

First, unlike the model in [3], we assume that the state $\omega(t)$ evolves in an independent and *non-stationary* fashion across time t. Note that the authors in [3] only provide simulation results for non-stationary $\omega(t)$ without providing a PAC type result. In particular, the distribution of $\omega(t)$ denoted $\pi_t(\omega)$, $\omega \in \Omega$ satisfies the following asymptotic stationarity property.

Assumption 1: Assume that there exists a probability measure $\pi(\omega)$ on Ω such that $\lim_{t\to\infty} \|\pi_t - \pi\|_1 = 0$.

For the sake of simplicity, we have made the above assumption. Note that this is the first step towards finding the PAC type result for the non-stationary states. The analysis of the algorithm for the non-stationary states is relegated to the future work. Note that the efficacy of the distributed algorithm depends on how accurately each node computes an estimate of π_t, $t \in \mathbb{N}$. Naturally, we expect the bounds that we derive to be a function of the complexity of the probability measure space from which the "nature" chooses $\pi_t(\omega)$. Let us assume that for each $t \in \mathbb{N}$, π_t is chosen from a set $\mathcal{P}$. Assuming that $\mathcal{P}$ is a closed set with respect to the $\mathcal{L}_1$*-norm*, we have $\pi \in \mathcal{P}$. One way of measuring the complexity is through the covering number, and the metric entropy of the set $\mathcal{P}$, which are defined as follows.

Definition 1: (see [16]) A δ-covering of $\mathcal{P}$ is a set $\mathcal{P}_c \triangleq \{\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_M\} \subseteq \mathcal{P}$ such that for all $\pi^{'} \in \mathcal{P}$, there exists a $\mathcal{P}_i \in \mathcal{P}_c$ for some $i = 1, 2, \ldots, M$ such that $\|\pi^{'} - \mathcal{P}_i\|_1 < \delta$. The smallest M denoted M_δ is called the covering number of $\mathcal{P}$. Further, $\mathcal{H}(\mathcal{P}, \delta) \triangleq \log M_\delta$ is called the *metric entropy*.

Note that in many practical scenarios, the available data at each time $t \in \mathbb{N}$ is delayed, and a data of size w_t, $t \in \mathbb{N}$ delayed by D slots will be used for estimation/inference purposes [3], [13]. The reason for making the sample size w_t depend on t becomes apparent later. Since $p_k(t)$, $k = 0, 1, 2, \ldots, K$ depend on $Y_c(t)$ for all t (see (2)), we have that the process $p_k(t)$ in general is a stochastically dependent sequence. The "degree" of correlation depends on the algorithm used. For $k = 0, 1, 2 \ldots, K$ and $s \in \mathbb{N}$, let $\mathbb{P}_{t,t+s}^{\texttt{ALG},k}(* \mid \mathcal{E})$ and $\mathbb{P}_{t}^{\texttt{ALG},k}(* \mid \mathcal{E})$ denote the joint and marginal distributions of $(p_k(t), p_k(t+s))$ and $p_k(t)$ conditioned on the event $\mathcal{E}$, respectively, induced by any algorithm ALG.[2] Note that if $p_k(t)$ and $p_k(t+s)$ are independent for each $t \in \mathbb{N}$ conditioned on some event $\mathcal{E}$, then $\left\| \mathbb{P}_{t,t+s}^{\texttt{ALG},k}(* \mid \mathcal{E}) - \mathbb{P}_{t}^{\texttt{ALG},k}(* \mid \mathcal{E}) \otimes \mathbb{P}_{t+s}^{\texttt{ALG},k}(* \mid \mathcal{E}) \right\|_{\text{TV}} = 0$. Thus, the difference above, maximized over all slots $t \in \mathbb{N}$ is a natural way of measuring the correlation between the sequences that are s time slots away. More precisely, we have the following definition (see [17] for a related definition).

Definition 2: The β_1 mixing coefficient of the process $p_k(t)$, $k = 0, 1, 2, \ldots, K$ conditioned on some event $\mathcal{E}$ is given by

$$\beta_{\texttt{ALG},k}(s, \alpha \mid \mathcal{E}) \triangleq \sup_{t\in\mathbb{N}, t\geq\alpha} \|\mathbb{M}_{t,s,k}(\mathcal{E})\|_{\text{TV}}, \qquad (3)$$

where $\mathbb{M}_{t,s,k}(\mathcal{E}) \triangleq \mathbb{P}_{t,t+s}^{\texttt{ALG},k}(* \mid \mathcal{E}) - \mathbb{P}_{t}^{\texttt{ALG},k}(* \mid \mathcal{E}) \otimes \mathbb{P}_{t+s}^{\texttt{ALG},k}(* \mid \mathcal{E})$, $s \geq 0$, $\alpha \geq 0$, $\mathbb{P}_{t}^{\texttt{ALG},k} \otimes \mathbb{P}_{t+s}^{\texttt{ALG},k}$ denotes the product distribution, and $\| * \|_{\text{TV}}$ is the total variational norm.

Note that in the definition of $\beta_{\texttt{ALG},k}(s, \alpha \mid \mathcal{E})$, we have used $t \geq \alpha$, which is required later in the proof of our main results. Further, if s is large, and the process is sufficiently mixing, then we expect that $\beta_{\texttt{ALG},k}(s, \alpha \mid \mathcal{E}) = 0$. This definition will be used to decouple a dependent stochastic process so that some of the large deviation bounds that are valid for independent sequences can be applied. The details of this approach will be clear in the proof of our main results. For notational convenience, let us denote the maximum and minimum values of $p_k(t)$, $k = 0, 1, 2, \ldots, K$ by $p_{\max,k}$ and $p_{\min,k}$, respectively. Further, let $(\Delta p)_{\max,k} \triangleq p_{\max,k} - p_{\min,k}$. In the following section, we propose an Approximate DPP (ADPP) algorithm with the associated theoretical guarantees. The β_1 coefficient for the ADPP algorithm will be $\beta_{\texttt{ADPP},k}(s, \alpha \mid \mathcal{E})$.

[2] In this paper, we propose a distributed Approximate DPP (ADPP) algorithm, and hence ALG will be ADPP.

IV. Algorithm and Main Results

In the following subsection, we prove that the optimal solution to $\mathbf{P_0}$ is close to a **LP**.

A. Approximately Optimal LP

Since the number of possible values that $p_k(t)$, $k = 0, 1, 2, \ldots, K$ take is finite, the number of possible strategies is also finite. Due to this, the question of whether $p_k(t)$ has restrictions such as convexity, linearity etc., does not matter. The approximate algorithm that we are going to propose chooses one of the *pure strategy* $\mathbf{S}(\omega) \triangleq \{\mathbf{s}_1(\omega_1), \mathbf{s}_2(\omega_2), \ldots, \mathbf{s}_N(\omega_N)\}$ based on the common information $Y_c(t)$, where $\mathbf{s}_i(\omega_i) \in \mathcal{A}_i$, and $\omega_i \in \Omega_i$, $i = 1, 2, \ldots, N$. The complexity of the solution that we propose depends on the number of possible strategies. For example, $\mathbf{s}_i(\omega_i)$ can be a simple threshold rule with the thresholds coming from a small finite set. The control action $\alpha_i(t)$ at the user i is chosen as a deterministic function of $\omega(t)$, i.e., $\alpha_i(t) \triangleq \mathbf{s}_i(\omega_i(t))$ for all $i \in \{1, 2, \ldots, N\}$ and for all $t \in \mathbb{N}$. Let the total number of such pure strategies be $F \triangleq \prod_{i=1}^{N} |\mathcal{A}_i|^{|\Omega_i|}$. Enumerating the F strategies, we get $\mathbf{S}^m(\omega)$, $m \in \{1, 2, \ldots, F\}$ and $\omega \in \Omega$. Each $\omega \in \Omega$ and the strategy $\mathbf{S}^m(\omega)$ result in a cost $p_k(\mathbf{S}^m(\omega), \omega)$, $k = 0, 1, 2, \ldots, K$. Note that it is possible to reduce F if the problem has a specific structure [3]. For each strategy $m \in \{1, 2, \ldots, F\}$, define the average cost/penalty as $r_{k,\pi'}^{(m)} \triangleq \sum_{\omega \in \Omega} \pi'(\omega) p_k(\mathbf{S}^m(\omega), \omega)$, where $k = 0, 1, 2, \ldots, K$ and the underlying distribution of $\omega \in \Omega$ is $\pi' \in \mathcal{P}_c$. As in [3], we consider a randomized algorithm where the strategy $m \in \{1, 2, \ldots, F\}$ is picked with probability $\theta_m(t)$ in an independent fashion across time t. Here, $\theta_m(t)$ is a function of the common information $Y_c(t)$. The corresponding average cost/penalty at time t becomes

$$\mathbb{E}p_k(t) = \sum_{m=1}^{F} \theta_m(t) \mathbb{E}_\lambda p_k(\mathbf{S}^m(\omega(t)), \omega(t)) = \sum_{m=1}^{F} \theta_m(t) r_{k,\lambda}^{(m)},$$

where $\lambda \in \{\pi_t, \pi, \mathcal{P}_i\}$, $i = 1, 2, \ldots, M_\delta$. In [3], it was shown that the problem $\mathbf{P_0}$ when $\pi_t = \pi$ for all $t \in \mathbb{N}$ ($\omega(t)$ is i.i.d.) is equivalent to the following **LP**:

$$\min_{\theta_1, \theta_2, \ldots, \theta_F} \sum_{m=1}^{F} \theta_m r_{0,\pi}^{(m)}$$
$$\text{s.t.} \sum_{m=1}^{F} \theta_m r_{k,\pi}^{(m)} \leq c_k, \ k = 1, 2, \ldots, K \text{ and } \sum_{m=1}^{F} \theta_m = 1. \quad (4)$$

In this paper, from **Assumption 1**, we have $\|\pi_t - \pi\|_1 \to 0$, as $t \to \infty$. With dense covering of the space $\mathcal{P}$, we expect that the limiting distribution is well approximated by $\mathcal{P}_i$ for some $i = 1, 2, \ldots, M_\delta$ in the covering set. More preciesely, $\mathcal{P}_{i^*} \triangleq \arg\min_{\mathcal{Q} \in \{\mathcal{P}_1, \ldots, \mathcal{P}_{M_\delta}\}} \|\pi - \mathcal{Q}\|_1$, and the corresponding distance be $d_{\pi, \mathcal{P}_{i^*}} \triangleq \|\pi - \mathcal{P}_{i^*}\|_1 < \delta$. Since the distribution of $\omega(t)$ is changing across time, directly applying Theorem 1 of [3] is not possible. However, from **Assumption** 1, we know that the distribution approaches a fixed measure $\pi \in \mathcal{P}_c$. Hence, we expect that the algorithm designed for $\pi \in \mathcal{P}_c$ or an approximation of π, i.e., $\mathcal{P}_{i^*}$ should eventually be close to the optimal algorithm. Therefore, we consider the following **LP** denoted $\mathbf{LP}_{\mathcal{P}_{i^*}}$:

$$\min_{\theta_1, \theta_2, \ldots, \theta_F} \sum_{m=1}^{F} \theta_m r_{0,\mathcal{P}_{i^*}}^{(m)}$$
$$\text{s.t.} \sum_{m=1}^{F} \theta_m r_{k,\mathcal{P}_{i^*}}^{(m)} \leq c_k, \ k = 1, 2, \ldots, K \text{ and } \sum_{m=1}^{F} \theta_m = 1. \quad (5)$$

Also, we assume that the solution to $\mathbf{LP}_{\mathcal{P}_{i^*}}$ exists and the optimal cost is absolutely bounded. Further, define

$$G(x) \triangleq \inf \left\{ \sum_{m=1}^{F} \theta_m r_{0,\mathcal{P}_{i^*}}^{(m)} : \Theta \in \mathcal{C}_{x,\Theta} \right\}, \quad (6)$$

where $\Theta \triangleq (\theta_1, \theta_2, \ldots, \theta_F)$, and for any $x \geq 0$, $\mathcal{C}_{x,\Theta} \triangleq \left\{ \Theta : \sum_{m=1}^{F} \theta_m r_{k,\mathcal{P}_{i^*}}^{(m)} \leq c_k + x, \ k = 1, 2, \ldots, K, \Theta \mathbf{1}^T = 1 \right\}$, where $\mathbf{1} \triangleq \{1, 1, \ldots, 1\} \in \mathbb{R}^F$. Note that $G(0)$ corresponds to $\mathbf{LP}_{\mathcal{P}_{i^*}}$. We make the following important smoothness assumption about the function $G(x)$.

Assumption 2: The function $G(x)$ is *c-Lipschitz continuous* around the origin, i.e., for some $c > 0$, we have

$$|G(x) - G(y)| \leq c\,|x - y|, \text{ for all } x, y \geq 0. \quad (7)$$

In the theorem to follow, given that **Assumption 2** is valid, we prove that the optimal cost of the linear optimization problem in (5) is "close" to the optimal cost of $\mathbf{P_0}$.

***Theorem** 1:* Let $p^{(\text{opt})}$ and $p_{\mathcal{P}_{i^*}}^{(\text{opt})}$ be the optimal solution to the problems $\mathbf{P_0}$ and $\mathbf{LP}_{\mathcal{P}_{i^*}}$, respectively. Then, under **Assumption 2**, we have $p_{\mathcal{P}_{i^*}}^{(\text{opt})} < p^{(\text{opt})} + (c+1)\Delta_{\pi, \mathcal{P}_{i^*}}$, where for any $\nu > 0$, $\Delta_{\pi, \mathcal{P}_{i^*}} \triangleq \max_{k=0,1,2,\ldots,K} b_{\max,k}(d_{\pi, \mathcal{P}_{i^*}} + \nu)$, and $b_{\max,k} \triangleq \max\{|p_{\max,k}|, |p_{\min,k}|\}$.

Proof: See Appendix A. ■

B. Approximate DPP (ADPP) Algorithm (ADPPA)

In this subsection, we present an online distributed algorithm that approximately solves the problem $\mathbf{P_0}$. We assume that at time $t \in \mathbb{N}$, all nodes receive feedback specifying the values of all the penalties and the states, namely, $p_1(t-D), p_2(t-D), \ldots, p_K(t-D)$ and $\omega(t-D)$. Recall that $D \geq 0$ is the delay in the feedback. Using this information, we construct the following set of queues

$$Q_k(t+1) = \max\{Q_k(t) + p_k(t-D) - c_k, 0\}, \quad (8)$$

$k = 1, 2, \ldots, K$, and $t \in \mathbb{N}$. These queues act as the common information, i.e., $Y_c(t) = \mathbf{Q}_t$, where $\mathbf{Q}_t \triangleq (Q_1(t), Q_2(t), \ldots, Q_K(t))$. Further, the past w_t samples of $\omega(t)$ given by $\{\omega(t-i), i = D, D+1, \ldots, D+w_t-1\}$ will be used to find an estimate of the state probabilities which is required for the algorithm that we propose. For all $k = 1, 2, \ldots, K$, we let $p_k(t) = 0$ when $t \in \{-1, -2, \ldots, -D\}$. The *Lyapunov* function is defined as

$$\mathcal{L}(t) \triangleq \frac{1}{2}\|\mathbf{Q}_t\|_2^2 = \frac{1}{2}\sum_{i=1}^{K} Q_i^2(t), \quad (9)$$

and the corresponding drift is given by $\Delta(t) \triangleq \mathcal{L}(t+1) - \mathcal{L}(t)$ for all $t \in \mathbb{N}$. A higher value of the drift indicates that the

constraints have been violated frequently in the past. Thus, the control action should be taken that simultaneously minimizes the drift and the penalty (cost). The DPP algorithm tries to find the optimal control action that minimizes an upper bound on the DPP term, which is the essence of the following lemma. The proof of the lemma follows directly from the proof of Lemma 5 of [3], and hence omitted.

***Lemma** 1:* For a fixed constant $V \geq 0$, we have $\mathbb{E}\left[\Delta(t+D) + Vp_0(t) \,|\, \mathbf{Q}_t\right] \leq B_t(1+2D) + V\sum_{m=1}^{F}\beta_m(t)r_{0,\pi_t}^{(m)} + \sum_{k=1}^{K}Q_k(t)\mathcal{C}_{i,k,t}$, where $\mathcal{C}_{i,k,t} \triangleq \sum_{m=1}^{F}\beta_m(t)r_{k,\pi_t}^{(m)} - c_k$, $r_{k,\pi_t}^{(m)} \triangleq \sum_{\omega\in\Omega}\pi_t(\omega)p_k(\mathbf{S}^m(\omega),\omega)$, $k = 0,1,2,\ldots,K$,

$$B_t \triangleq \max_{m\in\{1,2,\ldots,F\}} \frac{1}{2}\sum_{k=1}^{K}\sum_{\omega\in\Omega}\pi_t(\omega)\left|p_k(\mathbf{S}^m(\omega),\omega) - c_k\right|^2, \tag{10}$$

and, with a slight abuse of notation, $\beta_m(t)$ is the probability with which the strategy m is used at time t.

Note that as $t\to\infty$, $B_t \to B$. The expression for B can be obtained by replacing $\pi_t(\omega)$ by $\pi(\omega)$ in the expression for B_t. The algorithm to follow requires an estimate of $\pi_t(\omega)$, which can be computed using the past w_t samples by means of any estimate such as the sample average. However, when the space $\mathcal{P}$ is "simple", one can expect to compute an estimate of $\pi_t(\omega)$ more efficiently. For example, if the nature chooses $\omega(t)$ from a finite set of distributions ($M_\delta < \infty$ for all $\delta > 0$), then estimating the distribution corresponds to a hypothesis testing problem. Hence, by approximating the measure space $\mathcal{P}$ by a finite set of measures $\mathcal{P}_c$ gives us the flexibility to run a hypothesis testing to find an approximate distribution based on the available w_t samples through a likelihood ratio test. In the following, we provide the algorithm.

- **Algorithm:** Given the delayed feedback of size w_t at time slot $t\in\mathbb{N}$, i.e., $\omega(t-i-D)$, and $p_k(t-D)$, $i = 0,1,\ldots,w_t-1$ and for $k = 1,2,\ldots,K$, perform the following steps
  - **Step 1:** Find the probability measure in $\mathcal{P}_c$ that best fits the data, i.e., pick $\mathcal{P}_{j_t^*}\in\mathcal{P}_c$ such that

$$j_t^* \triangleq \arg\max_{j\in\{1,2,\ldots,M_\delta\}} \frac{1}{w_t}\sum_{\tau=t-D-w_t+1}^{t-D}\log\left(\mathcal{P}_j(\omega(\tau))\right). \tag{11}$$

  - **Step 2:** Choose $m_t\in\{1,2,\ldots,F\}$ (breaking ties arbitrarily) that minimizes the following:

$$Vr_{0,\mathcal{P}_{j_t^*}}^{(m_t)} + \sum_{k=1}^{K}Q_k(t)r_{k,\mathcal{P}_{j_t^*}}^{(m_t)}. \tag{12}$$

  - **Step 3:** Set $t\to t+1$, receive the delayed feedback, update the queues using (8), and go to **Step 1**.

We say that there is an error in the outcome of step 1 of the algorithm if $\mathcal{P}_{j_t^*}\neq\mathcal{P}_{i^*}$. Recall that i^* corresponds to the index of the probability measure in the covering set that is close to π in the $\mathcal{L}_1$ norm sense. The error event $\mathcal{E}_{\delta,t}$, $t\in\mathbb{N}$ is defined as those outcomes for which $j_t^*\neq i^*$. Further, let $\mathcal{E}_{[\tau:\tau+s]} \triangleq \bigcup_{t=\tau}^{\tau+s}\mathcal{E}_{\delta,t}$ to denote that there is an error in at least one of the time slot in the interval τ to $\tau+s$. In the following theorem, we state and prove our first result that will be used to prove the PAC type bound for the ADPP algorithm.

***Theorem** 2:* For the ADPP algorithm, for any $\epsilon_k > \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}p_k(\tau) - c_k + \frac{\alpha_t(p_{\max,k}-p_{\min,k})}{t-\alpha_t}$, and for constants $\alpha_t\in\mathbb{N}$, $u_t\in\mathbb{N}$ and $v_t\in\mathbb{N}$ such that $v_tu_t = t-\alpha_t$, we have

$$\Pr\left\{\frac{1}{t}\sum_{\tau=0}^{t-1}p_k(\tau) - c_k > \epsilon_k\right\} \leq u_t\exp\left\{\frac{-2\bar{\epsilon}_{t,k}^2v_t^2}{((\Delta p)_{\max,k})^2}\right\} + \sum_{\tau=\alpha_t}^{t}\Pr\left\{\mathcal{E}_{\delta,\tau}\right\} + (t-\alpha_t)\beta_{\texttt{ADPP},k}(u_t,\alpha_t\,|\,\mathcal{E}_{[\alpha_t:t]}^c), \tag{13}$$

where $\bar{\epsilon}_{t,k} \triangleq \frac{t\epsilon_{t,k}-\alpha_t(p_{\max,k}-p_{\min,k})}{t-\alpha_t}$, $\epsilon_{t,k} \triangleq \epsilon_k + c_k - \frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}p_k(\tau)$. Here, $c_0 = p^{(opt)}$, and c_k, $k=1,2,\ldots,K$ are the constraint variables in $\mathbf{P_0}$.

Proof: See Appendix C of [18] for more details. ■

The first term in the bound in Theorem 2 corresponds to the large deviation bound when $p_k(t)$'s are independent. The second term corresponds to an upper bound on the probability of error in the time slots α_t to t for decoding the correct index i^*; equivalently, this corresponds to an "incorrect" estimate of the distribution of the states in these slots. The last term captures the stochastic dependency of $p_k(t)$ across time $t\in\mathbb{N}$. In order to prove a high probability result, we need to find an expression for each of the terms in the bound. Next, we upper bound the error term $\Pr\{\mathcal{E}_{\delta,\tau}\}$ using the following assumption.

Assumption 3: Assume that for all $j = 1,2,\ldots,M_\delta$, $\mathcal{P}_j(\omega)\neq 0$, there exist constants $\alpha_\delta > \beta_\delta > 0$, such that $\alpha_\delta > \mathcal{P}_j(\omega) > \beta_\delta > 0$ for all $\omega\in\Omega$.

The above bound imposes the constraint that all the states occur with non-zero probability. In other words, the probability measures in the δ-covering set approximates the true measure by assigning non-zero probabilities to all possible $\omega\in\Omega$. We use the above assumption in the proof of the following lemma to bound the probability of error term in (13).

***Lemma** 2:* An upper bound on the probability of error is given by

$$\Pr\{\mathcal{E}_{\delta,\tau}\} \leq P_{e,\text{up}}^{(\tau)} \triangleq \begin{cases} q_{e,\text{up}}^{(\tau)} & \text{if } \tau > D + w_\tau - 1, \\ \frac{1}{M_\delta} & \text{otherwise,} \end{cases} \tag{14}$$

where $q_{e,\text{up}}^{(\tau)} \triangleq \exp\left\{-2\zeta_\delta\mathcal{D}_\tau^2w_\tau + \mathcal{H}(\mathcal{P},\delta)\right\}$, $\mathcal{D}_{\tau,j} \triangleq \frac{1}{w_\tau}\sum_{s=\tau-D-w_\tau+1}^{\tau-D}\mathbb{E}_{\pi_\tau}\log\left(\frac{\mathcal{P}_j(\omega(s))}{\mathcal{P}_{i^*}(\omega(s))}\right)$, $\zeta_\delta \triangleq \left[\log\left(\frac{\alpha_\delta}{\beta_\delta}\right)\right]^2$, $\mathcal{D}_\tau \triangleq \min_{j\neq i^*}\mathcal{D}_{\tau,j}$, and $\mathcal{H}(\mathcal{P},\delta) = \log M_\delta$ is the *metric entropy*. Further, when $u_t = \mathcal{O}(\sqrt{t})$, $v_t = \mathcal{O}(\sqrt{t})$, and $\alpha_t = \mathcal{O}(\sqrt{t})$, we have $\sum_{\tau=\alpha_t}^{t-1}\Pr\{\mathcal{E}_{\delta,\tau}\} \preceq (t-\alpha_t)S_{t,\delta}$, where $S_{t,\delta} \triangleq \exp\left\{-\phi_{\tau,t,\delta} + \mathcal{H}(\mathcal{P},\delta)\right\}$. In the above, $\phi_{\tau,t,\delta} \triangleq 2\zeta_\delta\left[\min_{\alpha_t\leq\tau\leq t}\mathcal{D}_\tau\right]^2N_{[\alpha_t:t]}$, $N_{[\alpha_t:t]} \triangleq \min_{\alpha_t\leq\tau\leq t}w_\tau$.

Proof: See Appendix D of [18]. ■

From the above lemma, we have that the error goes to zero exponentially fast as $\tau\to\infty$. The fact that $\sum_{\tau=\alpha_t}^{t-1}\Pr\{\mathcal{E}_{\delta,\tau}\} \preceq (t-\alpha_t)S_{t,\delta} \to 0$ exponentially fast as $t\to\infty$ will be used later in the paper to prove the almost sure convergence of the algorithm to the optimal. Now, it remains to find an upper bound on the first and the last term in (13). The following theorem uses the assumption above, (13) and (14) to provide a PAC result for the above algorithm.

***Theorem** 3:* Under **Assumptions 1-3**, for the proposed **Algorithm** with $\epsilon_0 = (c+1)\Delta_{\pi,\mathcal{P}_{i^*}} + \psi_t(\delta) + \frac{\alpha_t(p_{\max,k}-p_{\min,k})}{t-\alpha_t} + \epsilon$, $\epsilon_k = Q_{\rm up}(t) + \epsilon$, $k = 1, 2, \ldots, K$, and some finite positive constants V, C and c, the following holds.

1) For every $\epsilon > 0$, with a probability of at least $1 - \gamma_0$,

$$\frac{1}{t}\sum_{\tau=0}^{t-1} p_0(\tau) \leq p^{(opt)} + (c+1)\Delta_{\pi,\mathcal{P}_{i^*}} + \psi_t(\delta) + \zeta_{t,k} + \epsilon \quad (15)$$

provided $t \in \mathcal{T}_{t,0}$. Here, $\gamma_0 > \beta_0^*$, $\beta_0^* \triangleq (t-\alpha_t)\left[\beta_{\texttt{ADPP},0}(u_t, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]}) + S_{t,\delta}\right]$, where $S_{t,\delta}$ is as defined in Lemma 2, and $\zeta_{t,k} \triangleq \frac{\alpha_t(p_{\max,k}-p_{\min,k})}{t-\alpha_t}$.

2) For every $\epsilon > 0$, with a probability of at least $1 - \gamma_1$,

$$\frac{1}{t}\sum_{\tau=0}^{t-1} p_k(\tau) \leq c_k + Q_{\rm up}(t) + \frac{\alpha_t(p_{\max,k} - p_{\min,k})}{t-\alpha_t} + \epsilon, \quad (16)$$

$k = 1, 2, \ldots, K$, provided $t \in \mathcal{T}_{t,1}$. Here $\gamma_1 > \beta_1^*$, where $\beta_1^* \triangleq (t-\alpha_t)\left[\max_{k\neq 0}\beta_{\texttt{ADPP},k}(u_t, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]}) + S_{t,\delta}\right]$. In the above, $\mathcal{T}_{t,i} \triangleq \left\{t : (t-\alpha_t) > \frac{(\Delta p)_{\max,0}\, u_t}{\sqrt{2}\epsilon}\sqrt{\log\left(\frac{u_t}{\gamma_i - \beta_i^*}\right)}\right\}$, $i \in \{0,1\}$, $\Delta_{\pi,\mathcal{P}_{i^*}} = \max_{k=0,1,2,\ldots,K} b_{\max,k}(d_{\pi,\mathcal{P}_{i^*}} + \nu)$, and $\psi_t(\delta) \triangleq \frac{V(c+1)\bar{J}_t + \bar{H}_t + C/t}{V} + \frac{1+2D}{tV}\sum_{\tau=0}^{t-1} B_\tau P^{(\tau)}_{e,\rm up} + \frac{p_{\max,0}}{t}\sum_{\tau=0}^{t-1} P^{(\tau)}_{e,\rm up} + \frac{\rho}{Vt}\sum_{\tau=0}^{t-1}\tau P^{(\tau)}_{e,\rm up}$, where $\rho \triangleq \sum_{k=1}^{K}(p_{\max,k} - c_k)^2$, $\bar{J}_t \triangleq \max_{0\leq k\leq K} p_{\max,k}\left(\frac{1}{t}\sum_{\tau=0}^{t-1}\|\pi_\tau - \pi\|_1 + \delta\right)$, $\bar{H}_t \triangleq \frac{1+2D}{t}\sum_{\tau=0}^{t-1} B_\tau$. Further, $\mathcal{D}_{\tau,j}$, $\mathcal{D}_\tau$, ζ_δ, and $P^{(\tau)}_{e,\rm up}$ are as defined in Lemma 2. Also, $Q_{\rm up}(t) \triangleq \sqrt{\frac{VF}{t} + \frac{\Gamma_t}{t^2}}$ and $\Gamma_t \triangleq V(c+1)(\Delta_{\pi,\mathcal{P}_{i^*}} + \bar{J}_t) + \bar{H}_t + C + (1+2D)\sum_{\tau=0}^{t-1} B_\tau P^{(\tau)}_{e,\rm up} + p_{\max,0}\sum_{\tau=0}^{t-1} P^{(\tau)}_{e,\rm up} + \rho\sum_{\tau=0}^{t-1}\tau P^{(\tau)}_{e,\rm up}$ and $p_{\max,k}$, $k = 1, 2, \ldots, K$ is as defined earlier.

Proof: See the Appendix E of [18]. ■

Interpretation of Theorem 3: Note that γ_0 and γ_1 are lower bounded by $t - \alpha_t$ times the β_1-mixing coefficient and $S_{t,\delta}$. Thus, a high probability guarantees can be obtained provided the algorithm induces sufficient mixing and $S_{t,\delta}$ is small, i.e., the number of samples (w_τ) used to compute an estimate of $\pi(\omega)$ scales with τ. When $\omega(t)$ is i.i.d., both $\psi_\delta(t)$ and $Q_{\rm up}(t)$ reduces, leading to a smaller objective value and a better constraints satisfaction capability. This is due to the fact that $\|\pi_\tau - \pi\| = 0$ for all τ which reduces the value of $\bar{J}_t$. Note that unlike [19], the dependency on w_t is exponential instead of $\frac{1}{\sqrt{w_t}}$. Also, higher metric entropy, $\mathcal{H}(\delta, \mathcal{P})$ requires larger values of w_t for better performance. Equivalently, when the complexity of the model, $\mathcal{P}$ is low, then the learnability improves. Thus, as $t \to \infty$, we have a better result compared to [3], [19]. As $t \to \infty$, $\bar{H}_t$ goes to $B(1+2D)$ and $\bar{J}_t$ goes to zero. Further, both terms $\frac{1+2D}{tV}\sum_{\tau=0}^{t-1} B_\tau P^{(\tau)}_{e,\rm up}$ and $\frac{p_{\max,0}}{t}\sum_{\tau=0}^{t-1} P^{(\tau)}_{e,\rm up}$ go to zero since $\mathcal{D}_\tau$ goes to a constant for a large values of τ. To summarize, the average cost and penalties are close to the optimal with high probability provided the process $\omega(t)$ is sufficiently mixing, and the number of samples w_t monotonically increases with t. Note that w_t is a design parameter, and hence can be made to scale with t. Next, we prove a bound on $\beta_{\texttt{ADPP},k}$ when $D = 0$. Then, we state the result for any $D \geq 0$.

V. Bound On the Mixing Coefficient

By using the Pinsker's inequality that relates the total variational norm and the mutual information, we have the following bound [20]

$$\beta_{\texttt{ADPP},k}(s, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]}) \leq \sup_{t\geq\alpha_t}\sqrt{\frac{I(X_{k,t}; X_{k,t-s} \mid \mathcal{E}^c_{[\alpha_t:t]})}{2}}, \quad (17)$$

where $X_{k,t} \triangleq p_k(t)$, $I(X_{k,t}; X_{k,t-s} \mid \mathcal{E}^c_{[\alpha_t:t]})$ is the mutual information between random variables $p_k(t)$ and $p_k(t-s)$, $k = 0, 1, 2, \ldots, K$ conditioned on $\mathcal{E}^c_{[\alpha_t:t]}$, and any $s \in \mathbb{N}$. Later, we use $s = u_t$, as required. Thus, proving an upper bound on $\beta_{\texttt{ADPP},k}(s, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]})$ amounts to finding an upper bound on the conditional mutual information. To present our results, we use the following notations. Let $\mathbf{X}_t \triangleq (X_{0,t}, X_{1,t}, \ldots, X_{K,t})$, $\mathbf{X}_{\neq k,t} \triangleq (X_{1,t}, X_{2,t} \ldots, X_{k-1,t}, X_{k+1,t}, \ldots, X_{K,t})$, and as before, $\mathbf{Q}_t \triangleq (Q_1(t), Q_2(t), \ldots, Q_K(t))$. We first note that $I(\mathbf{X}_t; \mathbf{X}_{t-s} \mid \mathcal{E}^c_{[\alpha_t:t]}) = I(X_{k,t}; \mathbf{X}_{t-s} \mid \mathcal{E}^c_{[\alpha_t:t]}) + I(X_{\neq k,t}; \mathbf{X}_{t-s} \mid X_{k,t}, \mathcal{E}^c_{[\alpha_t:t]}) = I(X_{k,t}; X_{k,t-s} \mid \mathcal{E}^c_{[\alpha_t:t]}) + I(X_{k,t}; \mathbf{X}_{\neq k,t-s} \mid X_{k,t-s}, \mathcal{E}^c_{[\alpha_t:t]}) + I(\mathbf{X}_{\neq k,t}; \mathbf{X}_{t-s} \mid X_{k,t}, \mathcal{E}^c_{[\alpha_t:t]}) \geq I(X_{k,t}; X_{k,t-s} \mid \mathcal{E}^c_{[\alpha_t:t]})$, where the last inequality follows from the fact that the mutual information is non-negative. Thus, we have

$$\beta_{\texttt{ADPP},k}(s, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]}) \leq \sup_{t\geq\alpha_t}\sqrt{\frac{I(\mathbf{X}_t; \mathbf{X}_{t-s} \mid \mathcal{E}^c_{[\alpha_t:t]})}{2}}. \quad (18)$$

Let $\mathcal{Q}_t$ be the set of all vectors that $\mathbf{Q}_t$ takes at time t. Also, let $\mathcal{M}_t : \mathcal{Q}_t \to \{1, 2, \ldots, F\}$ be the rule induced by the ADPP algorithm that determines the strategy given the queue at time t. In order to obtain an upper bound on the mutual information, we state the following assumption about the conditional distribution of the process $\omega(t)$.

Assumption 4: For some $\kappa > 0$, $\mathbf{Q}_t \in \mathcal{Q}_t$, and for all $t \in \mathbb{N}$, we assume that the following bound is satisfied

$$\sup_{x,m,m'}\frac{\Pr\{\mathbf{X}_t = x \mid \mathcal{M}_t(\mathbf{Q}_t) = m, \mathcal{E}^c_{[\alpha_t:t]}\}}{\Pr\{\mathbf{X}_t = x \mid \mathcal{M}_t(\mathbf{Q}_t) = m', \mathcal{E}^c_{[\alpha_t:t]}\}} \leq e^\kappa \quad (19)$$

Note that a lower value of κ signifies the fact that the channel is noisy. For example, when $\kappa = 0$, we have uniform conditional distribution for all m and x leading to a completely noisy channel from Q_τ to X_τ. Next, we present an upper bound on $\beta_{\texttt{ADPP},k}(s, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]})$ for the $D \geq 0$ case.

A. Bound on $\beta_{\texttt{ADPP},k}(s, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]})$ when $D = 0$

In order to get insights on the proof of bounding the β_1 coefficient, we consider the centralized scheme, i.e., $D = 0$, and later we provide results for the $D > 0$ case. For $D = 0$, the queue update in the vector form becomes

$$\mathbf{Q}_{t+1} = \max\{\mathbf{Q}_t + \mathbf{X}_{\neq 0,t} - \mathbf{C}, 0\} \quad (20)$$

where $\mathbf{C} \triangleq (c_1, c_2, \ldots, c_K)$. Recall that **Step 2** of the **Algorithm** uses $\mathbf{Q}_t$ and the output from **Step 1** to find a pure strategy in a deterministic fashion that maximizes an upper bound on the drift-plus-penalty expression. Thus, the strategy is a deterministic function of the queue. Note that conditioned

on the event $\mathcal{E}^c_{[\alpha_t:t]}$, the output of **Step 1** is i^* for all time slots $\tau \in \{\alpha_t, \ldots, t\}$. Conditioned on $\mathcal{E}^c_{[\alpha_t:t]}$, this leads to the following Markov chain model

$$(\mathbf{Q}_{\alpha_t}, \mathbf{X}_{\alpha_t}) \longrightarrow (\mathbf{Q}_{\alpha_t+1}, \mathbf{X}_{\alpha_t+1}) \longrightarrow \ldots \longrightarrow (\mathbf{Q}_t, \mathbf{X}_t).$$

Fig. 1 depicts the graphical model representation of the above.

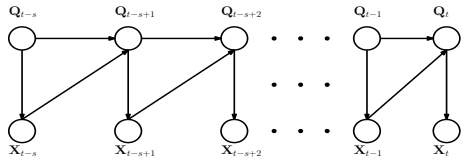


Fig. 1: Figure shows the graphical model corresponding to the ADPPA with $D = 0$ and time slots from $t - s \geq \alpha_t$ to t.

In order to prove an upper bound on the mutual information, we use the Strong Data Processing Inequality (SDPI) for the graphical model shown in Fig. 1. Note that the **Assumption 4** facilitates the proof of an upper bound on the β-mixing coefficient, as shown next.

***Theorem** 4:* Given **Assumption 4**, for $D = 0$, $\kappa < \log 3$, and for any $t \geq s \geq \alpha_t$, an upper bound on the β_1 mixing coefficient is given by

$$\beta_{\texttt{ADPP},k}(s, \alpha_t \mid \mathcal{E}^c_{[\alpha_t:t]}) \leq \frac{\theta^{(s-1)/2}}{\sqrt{2}} [\log \mu], k = 0, 1, 2, \ldots, K \tag{21}$$

where $\mu \triangleq F\,|\Omega|\,(K+1)$ is the number of possible values that $\mathbf{X}_t$ can take, $t \in \mathbb{N}$, and $\theta \triangleq \max\left\{\frac{(e^\kappa - 1)}{2}, \frac{1}{2}\right\} < 1$.

Proof outline: First, we use the graphical model shown in Fig. 1 along with the Strong Data Processing Inequality (SDPI) to bound $I(\mathbf{X}_t; \mathbf{X}_{t-s} \mid \mathcal{E}^c_{[\alpha_t:t]}) \leq \eta_{\texttt{ch}_1} I(\mathbf{Q}_t; \mathbf{X}_{t-s} \mid \mathcal{E}^c_{[\alpha_t:t]})$, where $\eta_{\texttt{ch}_1}$ is the Dobrushin coefficient [21]. Continuing this, we get $I(\mathbf{X}_t; \mathbf{X}_{t-s} \mid \mathcal{E}^c_{[\alpha_t:t]}) \leq \prod_{j=1}^{s-1} \eta_{\texttt{ch}_j} I(\mathbf{Q}_t; \mathbf{X}_{t-s} \mid \mathcal{E}^c_{[\alpha_t:t]})$. Next step is to use **Assumption 4** to show that $\eta_{\texttt{ch}_j} \leq \theta < 1$ for all j. See Appendix F of [18] for the details. ■

Note that $s = u_t$, and suppose u_t grows with t, then the Theorem says that the mixing coefficient goes down to zero exponentially fast with t. In order to get more insights, we will look at the asymptotic in the following subsection.

1) Almost sure convergence: Note that when $D = 0$, the authors in [15] prove the convergence of the algorithm to the optimal in probability. Here, we show an *almost sure* as well as a high probability convergence of the proposed ADPP algorithm to the optimal when $D = 0$, and $D > 0$.

***Lemma** 3:* Under **Assumptions 1-4**, for the proposed **Algorithm** with $D = 0$, $\alpha_t = \mathcal{O}(\sqrt{t})$, $w_t = \mathcal{O}(\sqrt{t})$, $V = \mathcal{O}(\sqrt{t})$, $\kappa < \log 3$, and some finite positive constant c, the following holds. For every $\epsilon > 0$, we have $\lim_{t\to\infty} \Pr\left\{\frac{1}{t}\sum_{\tau=0}^{t-1} p_0(\tau) \leq p^{(opt)} + \texttt{error}\right\} = 1$ and $\lim_{t\to\infty} \Pr\left\{\frac{1}{t}\sum_{\tau=0}^{t-1} p_k(\tau) \leq c_k + \epsilon\right\} = 1$, k = 1,2,…,K, where $\texttt{error} \triangleq (c+1)\Delta_{\pi,\mathcal{P}_{i^*}} + (c+1)\max_{0\leq k\leq K} p_{\max,k}\delta + \epsilon$. In the above, $\Delta_{\pi,\mathcal{P}_{i^*}} = \max_{k=0,1,2,\ldots,K} b_{\max,k}(d_{\pi,\mathcal{P}_{i^*}} + \nu)$, and $p_{\max,k}$, $k = 1, 2, \ldots, K$ is as defined earlier.

Proof: See Appendix G of [18]. ■

***Theorem** 5:* Under **Assumptions 1-4** with $D = 0$, $\alpha_t = \mathcal{O}(\sqrt{t})$, $w_t = \mathcal{O}(\sqrt{t})$, $V = \mathcal{O}(\sqrt{t})$, $\kappa < \log 3$, and $\infty > c > 0$, the following holds. For every $\epsilon > 0$, *almost surely*, $\lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} p_0(\tau) \leq p^{(opt)} + (c+1)\Delta_{\pi,\mathcal{P}_{i^*}} + (c+1)\max_{0\leq k\leq K} p_{\max,k}\delta + \epsilon$ and $\lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} p_k(\tau) \leq c_k + \epsilon$, $k = 1, 2, \ldots, K$. Here, $\Delta_{\pi,\mathcal{P}_{i^*}}$ and $p_{\max,k}$, $k = 1, \ldots, K$ are as defined earlier.

Proof: See Appendix H of [18]. ■

Next, we state the result for the $D > 0$ case.

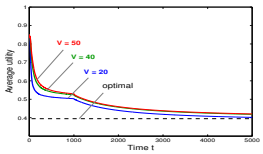
***Theorem** 6:* Under **Assumptions 1-4** with $D > 0$, $\alpha_t = \mathcal{O}(\sqrt{t})$, $w_t = \mathcal{O}(\sqrt{t})$, $V = \mathcal{O}(\sqrt{t})$, $\kappa < \frac{\log 3}{D}$, and some finite positive constant c, the following holds. For every $\epsilon > 0$, *almost surely*, $\lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} p_0(\tau) \leq p^{(opt)} + (c+1)\Delta_{\pi,\mathcal{P}_{i^*}} + (c+1)\bar{J} + \epsilon$ and $\lim_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1} p_k(\tau) \leq c_k + \epsilon$, $k = 1, 2, \ldots, K$. In the above, $\Delta_{\pi,\mathcal{P}_{i^*}}$, $\bar{J}$, and $p_{\max,k}$, $k = 1, 2, \ldots, K$ are as defined earlier.

Proof: See Sec. IV-B of [18]. ■

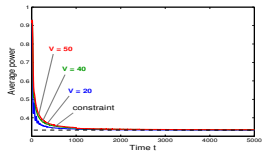
From Lemma 3 and Theorems 5 and 6, it is easy to see that the error can be reduced by reducing $\Delta_{\pi,\mathcal{P}_{i^*}}$, which amounts to reducing $d_{\pi,\mathcal{P}_{i^*}}$ and ν. Note that $d_{\pi,\mathcal{P}_{i^*}} < \delta$ can be reduced by reducing the error in the covering of the probability space $\mathcal{P}_c$. This comes at a cost of increased metric entropy since δ needs to be reduced. However, as $t \to \infty$, increased metric entropy does not effect the overall result. Further, for the centralized scheme with one sensor and $D = 0$, we get the almost sure result in [15] as a special case when $\kappa < \log 3$.

VI. Simulation Results

For the simulation setup, we consider the 3 sensors example of Sec. II. The problem is to maximize the average of the utility in (1) subject to an average power constraint of $1/3$. Here, the utility is the negative of the cost. The probability measure π_t is chosen from a set of 8 distributions, and converges to $\{0.1, 0.7, 0.1, 0.1\}$. Due to lack of space, we skip the details of the distribution that is used in the transient time. The optimal value of this is $p^{(opt)} = 0.394$. When $\alpha_i(t) = 1$, $i = 1, 2, 3$, a power of 1 watt each is consumed. Figures 2a and 2b show the plots of utility and penalty averaged over 1000 instantiations, versus time t for different values of V, $D = 10$ and $w_t = 40$ for all t, demonstrating the tradeoff in terms of V. For large values of t, the utility achieved by the algorithm with $V = 20$ is close to optimum while satisfying the constraints thereby confirming the optimality of the algorithm.



(a) Figure shows the plot of the average utility versus time.



(b) Figure shows the plot of the average power versus time.

VII. Concluding Remarks

In this paper, we considered a distributed stochastic optimization problem with independent and asymptotically stationary states. We showed that this stochastic optimization problem is approximately equal to a **LP** that is a function of the limiting distribution of the state. For the proposed

approximate DPP algorithm, we showed that with certain probabilities γ_0 and γ_1, the average cost and penalties are within constants of the optimal solution and the constraints, respectively, provided the waiting time $t >$ a threshold. The threshold is in terms of the mixing coefficient that indicates the non-stationarity of the cost/penalties. The approximation errors capture the degree of non-stationarity (i.e., $\|\pi_t - \pi\|_1$), the number of samples used to compute an estimate of the state distribution. Also, we have proved an almost sure convergence of the proposed algorithm to a constant close to the optimal. Finally, we presented simulations results to corroborate our theoretical findings.

APPENDIX A
PROOF OF THEOREM 1

In this proof, we use the fact that by decreasing the objective function and increasing the constraints c_k, $k = 1, 2, \ldots, K$ in $\mathbf{P_0}$ will result in a decreased optimal value. Consider the cost/penalties of the problem $\mathbf{P_0}$

$$\limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{\omega\in\Omega}\Gamma_{\tau,\omega,k} \overset{(a)}{=} \limsup_{t\to\infty,} \frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{\omega\in\Omega}\lambda_{k,\tau,\omega} \quad (22)$$

for $k = 0, 1, 2, \ldots, K$, any $t^{'} > 0$, $\Phi_{\pi_\tau,\pi} = \pi_\tau(\omega) - \pi(\omega)$ and $\Phi_{\pi,\mathcal{P}_{i^*}} = \pi(\omega) - \mathcal{P}_{i^*}(\omega)$. In the above, $\lambda_{k,\tau,\omega} \triangleq p_k(\tau)\left[\mathcal{P}_{i^*}(\omega) + \Phi_{\pi_\tau,\pi} + \Phi_{\pi,\mathcal{P}_{i^*}}\right]$, $\Gamma_{\tau,\omega,k} \triangleq \pi_\tau(\omega)p_k(\tau)$, $p_k(\tau) \triangleq p_k(\alpha(\tau), \omega(\tau))$ and (a) follows by adding and subtracting $\mathcal{P}_{i^*}(\omega)$ as mentioned earlier and $\pi(\omega)$. Since $\lim_{t\to\infty}\|\pi_t - \pi\|_1 = 0$, for every $\nu > 0$, there exists a $t^{'} \in \mathbb{N}$ such that for all $t > t^{'}$, $\|\pi_t - \pi\|_1 < \nu$. Using this $t^{'}$, and

$$\left|\sum_{\omega\in\Omega}\left(\pi_t(\omega) - \pi(\omega)\right)p_k(t)\right| \le \sum_{\omega\in\Omega}\left|\left(\pi_t(\omega) - \pi(\omega)\right)\right|\left|p_k(t)\right| \le \max\{|p_{\max,k}|, |p_{\min,k}|\}\nu \quad (23)$$

for every k and $t > t^{'}$, we have $-b_{\max,k}\nu \le \frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{\omega\in\Omega}\left(\pi_\tau(\omega) - \pi(\omega)\right)p_k(\tau) \le b_{\max,k}\nu$, where $b_{\max,k} \triangleq \max\{|p_{\max,k}|, |p_{\min,k}|\}$. Similarly, we have $-b_{\max,k}d_{\pi,\mathcal{P}_{i^*}} \le \mathcal{C}_t \le b_{\max,k}d_{\pi,\mathcal{P}_{i^*}}$, where $\mathcal{C}_t \triangleq \frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{\omega\in\Omega}\left(\pi(\omega) - \mathcal{P}_{i^*}(\omega)\right)p_k(\tau)$, and $d_{\pi,\mathcal{P}_{i^*}}$ is as defined earlier. Using the above two inequalities in (22), we get the following lower bound for all $k = 1, 2, \ldots, K$.

$$\limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}p_k(\tau) \ge \limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\sum_{\omega\in\Omega}\mathcal{P}_{i^*}(\omega)p_k(\tau) - \Delta_{\pi,\mathcal{P}_{i^*}}$$

where $\Delta_{\pi,\mathcal{P}_{i^*}} = \max_{k=0,1,2,\ldots,K} b_{\max,k}(d_{\pi,\mathcal{P}_{i^*}} + \nu)$. By using the above lower bound in $\mathbf{P_0}$, we get the following

$$\mathbf{P_1}: \quad \min_{\alpha(\tau)\in\mathcal{A}:\tau\in\mathbb{N}} \limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}p_0(t) - \Delta_{\pi,\mathcal{P}_{i^*}}$$

$$\text{s. t. } \limsup_{t\to\infty}\frac{1}{t}\sum_{\tau=0}^{t-1}\mathbb{E}p_k(t) \le c_k + \Delta_{\pi,\mathcal{P}_{i^*}}, \ k = 1, 2, \ldots, K,$$

$$\alpha_i(t) \text{ satisfies (2)}, \ i = 1, 2, \ldots, N,$$

where the expectation is taken with respect to $\mathcal{P}_{i^*}$. Note that the optimal cost obtained by solving $\mathbf{P_1}$ is smaller than p^{opt}. Further, the term $\Delta_{\pi,\mathcal{P}_{i^*}}$ is independent of the control action. It is evident from $\mathbf{P_1}$ that it is equivalent to $\mathbf{P_0}$ where the states $\omega(t)$ is i.i.d. whose distribution is $\mathcal{P}_{i^*}$. Using Theorem 1 of [3], it is easy to see that the solution to $\mathbf{P_1}$ is equal to $G(\Delta_{\pi,\mathcal{P}_{i^*}}) - \Delta_{\pi,\mathcal{P}_{i^*}}$,where $G(x)$ is as defined in (6). Thus, from **Assumption 2**, we have that $\left|p^{(\text{pert})}_{\mathcal{P}_{i^*}} - p^{(\text{opt})}_{\mathcal{P}_{i^*}}\right| < c\Delta_{\pi,\mathcal{P}_{i^*}} + \Delta_{\pi,\mathcal{P}_{i^*}} = (c+1)\Delta_{\pi,\mathcal{P}_{i^*}}$, where $p^{(\text{pert})}_{\mathcal{P}_{i^*}}$ denotes the optimal cost of $\mathbf{P}_1$. This leads to $p^{(\text{pert})}_{\mathcal{P}_{i^*}} > p^{(\text{opt})}_{\mathcal{P}_{i^*}} - (c+1)\Delta_{\pi,\mathcal{P}_{i^*}}$. But, we know that $p^{(\text{pert})}_{\mathcal{P}_{i^*}} \le p^{(\text{opt})}$, which implies that $p^{(\text{opt})}_{\mathcal{P}_{i^*}} < p^{(\text{opt})} + (c+1)\Delta_{\pi,\mathcal{P}_{i^*}}$. ■

REFERENCES

[1] E. N. Ciftcioglu, A. Yener, and M. J. Neely, "Maximizing quality of information from multiple sensor devices: The exploration vs exploitation tradeoff," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 883–894, 2013.

[2] M. J. Neely, "Dynamic optimization and learning for renewal systems," *IEEE Trans. on Automatic Control*, vol. 58, no. 1, pp. 32–46, 2013.

[3] ——, "Distributed stochastic optimization via correlated scheduling," *IEEE/ACM Trans. on Net.*, vol. 24, no. 2, pp. 759–772, 2016.

[4] M. Baghaie, S. Moeller, and B. Krishnamachari, "Energy routing on the future grid: A stochastic network optimization approach," in *IEEE Int. Conf. on Power System Technology (POWERCON)*, 2010, pp. 1–8.

[5] M. J. Neely and L. Huang, "Dynamic product assembly and inventory control for maximum profit," in *49th IEEE Conference on Decision and Control (CDC)*, 2010, pp. 2805–2812.

[6] S. Xu, G. Zhu, C. Shen, and B. Ai, "A qos-aware scheduling algorithm for high-speed railway communication system," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 2855–2860.

[7] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[8] M. J. Neely, E. Modiano, and C. P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. On Net.*, vol. 16, no. 2, pp. 396–409, 2008.

[9] L. Georgiadis, M. J. Neely, and L. Tassiulas, *Resource allocation and cross-layer control in wireless networks*. Now Publishers Inc, 2006.

[10] M. J. Neely, A. S. Tehrani, and A. G. Dimakis, "Efficient algorithms for renewable energy allocation to delay tolerant consumers," in *First IEEE Int. Conf. on Smart Grid Commun.,*, 2010, pp. 549–554.

[11] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. on Automatic Control*, vol. 37, no. 12, pp. 1936–1948, 1992.

[12] ——, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Trans. on Inf. Theory*, vol. 39, no. 2, pp. 466–478, 1993.

[13] Y. Han, Y. Zhu, and J. Yu, "A distributed utility-maximizing algorithm for data collection in mobile crowd sensing," *Proc. IEEE GLOBECOM, Austin, USA*, 2014.

[14] X. Zhang, S. Zhou, Z. Niu, and X. Lin, "An energy-efficient user scheduling scheme for multiuser mimo systems with rf chain sleeping," in *IEEE Wireless Commun. and Net. Conf. (WCNC)*, April 2013, pp. 169–174.

[15] X. Wei, H. Yu, and M. J. Neely, "A sample path convergence time analysis of drift-plus-penalty for stochastic optimization," *arXiv preprint arXiv:1510.02973*, 2015.

[16] S. A. Van de Geer, *Applications of empirical process theory*. Cambridge University Press Cambridge, 2000, vol. 91.

[17] V. Kuznetsov and M. Mohri, "Generalization bounds for time series prediction with non-stationary processes," in *Int. Conf. on Algorithmic Learning Theory, Springer*, 2014, pp. 260–274.

[18] B. N. Bharath and P. Vaishali, "Time complexity analysis of a distributed stochastic optimization in a non-stationary environment," 2017. [Online]. Available: https://arxiv.org/pdf/1701.02560v1.pdf

[19] M. J. Neely, S. T. Rager, and T. F. L. Porta, "Max weight learning algorithms for scheduling in unknown environments," *IEEE Trans. on Automatic Control*, vol. 57, no. 5, pp. 1179–1191, May 2012.

[20] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.

[21] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and bayesian networks," *arXiv. org preprint*, vol. 1508, 2016.