Adaptive Algorithms for Automatic Link Selection in Multiple Access with Link Failures

Mevan Wijewardena
Dept. of Electrical and Computer Engineering
University of Southern California
Los Angeles, USA
mpathira@usc.edu

Michael J. Neely
Dept. of Electrical and Computer Engineering
University of Southern California
Los Angeles, USA
mjneely@usc.edu

***Abstract*—This paper focuses on the problem of automatic link selection in multi-channel multiple access control using bandit feedback. In particular, a controller assigns multiple users to multiple channels in a time slotted system, where in each time slot at most one user can be assigned to a given channel and at most one channel can be assigned to a given user. Given that user i is assigned to channel j, the transmission fails with a fixed probability $f_{i,j}$. The failure probabilities are not known to the controller. The assignments are made dynamically using success/failure feedback. The goal is to maximize the time average utility, where we consider an arbitrary (possibly nonsmooth) concave and entrywise nondecreasing utility function. The problem of merely maximizing the total throughput has a solution of always assigning the same user-channel pairs and can be unfair to certain users, particularly when the number of channels is less than the number of users. Instead, our scheme allows various types of fairness, such as proportional fairness, maximizing the minimum, or combinations of these by defining the appropriate utility function. We propose an algorithm for this task that is adaptive and gets within $\mathcal{O}(\log(T)/T^{1/3})$ of optimality over any interval of T consecutive slots over which the success probabilities do not change. This performance is improved to $\mathcal{O}(1/\sqrt{T})$ for single-channel problems with a minimum constraint on the rate of transmission attempts per user.**

***Index Terms*—Multi-armed bandit learning; Proportional fairness; Network utility maximization; Optimization; Stochastic control**

I. Introduction

We consider the Multiple Access Control (MAC) problem with n users and m channels in slotted time $t \in \mathbb{N}$. In each time slot, a controller has to assign the users to channels such that at most one user is assigned to a given channel and at most one channel is assigned to a given user. The channel assignments may fail. In particular, there exist $q_{i,j} \in [0,1]$ for each $(i,j) \in \{1,2,\dots,n\} \times \{1,2,\dots,m\}$, where in time slot t, given that the controller decided to assign user i to channel j, the assignment fails independently with probability $1-q_{i,j}$. The controller does not know the probabilities $q_{i,j}$. Instead, at the end of every slot, it receives feedback on whether the transmission for each assigned user-channel pair succeeded or failed.

This work was supported in part by one or more of: NSF CCF-1718477, NSF SpecEES 1824418.

Define the matrices $\boldsymbol{Y}(t), \boldsymbol{S}(t) \in \{0,1\}^{n\times m}$ and vector $\boldsymbol{X}(t) \in \{0,1\}^n$, where

$$S_{i,j}(t) = \begin{cases} 1 & \text{if link } i,j \text{ is successful in time slot } t \\ 0 & \text{otherwise,} \end{cases}$$

$$Y_{i,j}(t) = \begin{cases} 1 & \text{user } i \text{ is assigned to channel } j \text{ in time slot } t \\ 0 & \text{otherwise,} \end{cases}$$

and $X_i(t) = \sum_{j=1}^m Y_{i,j}(t)S_{i,j}(t)$ for all $i \in \{1,2,\dots,n\}$. Notice that $X_i(t) \in \{0,1\}$ denotes whether user i successfully transmitted during time slot t. The goal is to maximize $\lim_{T\to\infty} \phi(\mathbb{E}\{\overline{\boldsymbol{X}}(T)\})$ using feedback on the link failures, where $\phi : \mathbb{R}^n \to \mathbb{R}$ is a concave entrywise nondecreasing utility function known to the controller and $\overline{\boldsymbol{X}}(T) = \frac{1}{T}\sum_{t=1}^T \boldsymbol{X}(t)$. [1]

We also focus on establishing finite-time bounds. In particular, given a finite time horizon $T \in \mathbb{N}$, we require the algorithm to satisfy $\phi^{\text{opt}} - \phi\left(\mathbb{E}\{\overline{\boldsymbol{X}}(T)\}\right) \le g(T)$, where ϕ^{opt} is the optimal utility of the original problem and g is a nonnegative function such that $\lim_{T\to\infty} g(T) = 0$. In addition, we are looking for algorithms that are adaptive. Formally, consider a system in which the channel success probabilities may change. In such a system, given a $T \in \mathbb{N}$, we require

$$\phi^{\text{opt}} - \phi\left(\frac{1}{T}\mathbb{E}\left\{\sum_{t=T_0}^{T+T_0-1} \boldsymbol{X}(t)\right\}\right) \le g(T)$$

for any $T_0 \in \mathbb{N}$, irrespective of the success probabilities outside of the time frame $[T_0 : T_0+T-1]$, given that success probabilities remained constant in the frame. Here, ϕ^{opt} is the optimal utility of the original problem that uses the constant success probabilities in $[T_0 : T_0+T-1]$ of the above scenario. Note that g is the same function regardless of T_0.

This model is applicable in a multiple access scenario where n users are accessing m orthogonal channels [1]. Here, it is desirable for users to be scheduled to avoid collisions. Link failures occur when the receiver cannot decode packet transmissions. This can occur, for example, when a fixed

[1] The limit is assumed to exist for simplicity of this introduction; the precise goal is to maximize a $\liminf_{T\to\infty} \phi(\mathbb{E}\{\overline{\boldsymbol{X}}(T)\})$

transmission power is used, but channel conditions have random and unknown fluctuations so that the received signal strength is insufficient for decoding. Different links can have different properties (such as different geographic distances to the receiver), so they can also have different success probabilities. Our model can still be applied if the channels are non-orthogonal, such as in beyond 5G non-orthogonal multiple access (NOMA) schemes [2]. In particular, given that the interference on a channel due to other channels is independent of the user-channel assignment, the probability of a transmission failure due to unsuccessful interference cancellation can be captured in $q_{i,j}$ values.

The naive approach of assigning the users with links with the least failure probabilities leads to unfairness since, in such a scenario, users with high link failure probabilities will never be assigned. A common approach to solving this problem is to maximize a utility function of the time-averaged success. One utility function that can be used in our work is $\phi(\boldsymbol{x}) = \min\{x_1, \ldots, x_n\}$. This is a nonsmooth utility function that seeks to maximize the minimum time average success rate across all users. However, if there is one user with very low success probability, this utility function can cause almost all the resources to be devoted to that user, resulting in poor performance for all users. Another choice is $\phi(\boldsymbol{x}) = w_1 \min\{x_1, \ldots, x_n\} + w_2 \sum_{i=1}^{n} \log(1 + \beta x_i)$, where w_1, w_2, β are given nonnegative weights. The logarithmic term introduces a form of proportional fairness [3], [4]. See also discussion of different utility functions in [3]–[7]

A. Related work

Our problem has been well studied in the full information scenario when either the fluctuating channel conditions are known before transmission (called opportunistic scheduling) or when channel success probabilities are known in advance. Opportunistic scheduling has been considered using utility functions [8], Lyapunov drift [9], Frank-Wolfe [10]–[12], primal-dual [13], [14], and drift-plus-penalty [15]. The case when success probabilities are known in advance can be solved offline as a convex optimization problem using the mirror descent technique [16].

The problem becomes challenging when the success probabilities are not known, but we only receive bandit feedback on the successes. The problem has to be approached by combining ideas from optimizing functions of time averages with multi-armed bandit learning. The work on bandits with vector rewards and concave utility functions can be adapted for the single-channel case ($m = 1$) of our problem (See [17]–[19]). There are two main drawbacks to these approaches. First, the above works do not consider the matching constraints considered in our work. Next, they focus on upper confidence bound (UCB) techniques and hence are not adaptive. We propose an adaptive algorithm for the single-channel case combining ideas from the EXP3 algorithm [20] with Lyapunov optimization. The single-channel algorithm cannot be directly extended to the general multi-channel case. The main reason is the complexity of the inner problem arising in each iteration, a problem over the set of doubly stochastic matrices (Birkhoff polytope). This can be addressed using the computationally complex Sinkhorn's algorithm [21] in each iteration. The work of [22] uses follow the regularized leader approach to solve adversarial bandit problems over the set of doubly stochastic matrices. However, their algorithm relies on a computationally expensive Sinkhorn iteration. The work of [23] proposes an algorithm to solve online optimization problems over transport polytopes. All the inner iterations of their algorithm have explicit solutions thanks to the rounding trick introduced for transport polytopes by [24]. In our work, we adapt the rounding trick to the Birkhoff polytope and develop an algorithm for general m. Our paper also treats general (possibly nonsmooth) utility functions, so Frank-Wolfe methods that rely on smoothness cannot be used.

It should be noted that the concept of adaptiveness considered in this paper is different from the adversarial setting [20], although they have similarities. In adversarial settings, optimality is defined with respect to the rewards received throughout the time horizon. This is useful when no stochastic assumptions can be placed on the rewards. On the other hand, the adaptiveness considered in this paper is useful when we can place stochastic assumptions on the rewards, but the distributions may change from time to time. In this case, we can define an optimal strategy for each time frame in which the reward distribution remained constant. Hence, the goal in this frame is to learn the aforementioned strategy irrespective of the reward distributions of the past. An approach commonly used in problems of this flavor is minimizing dynamic regret [25], [26]. Here, the regret is modified to account for the changing environments and the regret bounds are in terms of some measure that captures the degree of change. Various algorithms are developed for the setting with linear utility functions using optimizing in phases/episodes [25], and sliding window-based algorithms [26]. We utilize a simpler notion of adaptiveness and develop an algorithm for the case with general utility functions and matching constraints. Our notion of adaptiveness is considered in [12] for the problem of utility optimal opportunistic scheduling.

B. Our Contributions

- We develop an algorithm to solve the problem of automatic link selection in multiple access, combining the ideas of multi-armed bandit learning and Lyapunov optimization. Although the classical MAB problem has been widely analyzed for maximizing linear utilities, they suffer from lack of fairness in assignments when applied to the considered problem. It is notable that our method allows either smooth or nonsmooth concave, entrywise nondecreasing utilities. We prove that the algorithm gets within $\mathcal{O}(T^{-1/3}\log(T))$ of optimality over any interval of T consecutive time slots during which the (unknown) success probabilities do not change. If these probabilities are different before T_0, but change to new probabilities during $\{T_0, T_0 + 1, \ldots, T_0 + T - 1\}$, our performance guarantees for the new interval are independent of behav-

iors before T_0, even though the algorithm does not know the exact time T_0 of the change. Hence, the algorithm is adaptive.
- We separately consider the special case $m = 1$ (single-channel case). This case does not require matching constraints and hence has a simpler adaptive algorithm with a faster analytical convergence guarantee.

C. Notation

For integers a, b, we use $[a : b]$ to denote the set of integers between a, b inclusive. We use $[a] = [1 : a]$. For $\boldsymbol{a} \in \mathbb{R}^k$, $\|\boldsymbol{a}\| = \sqrt{\sum_{i=1}^k a_i^2}$, $\|\boldsymbol{a}\|_1 = \sum_{i=1}^k |a_i|$, and $[\boldsymbol{a}]_+ \in \mathbb{R}^k$ is the vector with i-th entry $\max\{a_i, 0\}$. We use $\mathbf{1}_k$ to denote the k-dimensional vector of ones. When the dimension is clear from context, we use $\mathbf{1}$ instead of $\mathbf{1}_k$. For vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^k$, $\boldsymbol{c} = \boldsymbol{a} \odot \boldsymbol{b} \in \mathbb{R}^k$ is defined such that $c_i = a_i b_i$ for all $i \in [k]$. For matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{k\times l}$ we use $\|\boldsymbol{A}\| = \sqrt{\sum_{i=1}^k \sum_{j=1}^l A_{i,j}^2}$, $\|\boldsymbol{A}\|_1 = \sum_{i=1}^k \sum_{j=1}^l |A_{i,j}|$, and $\boldsymbol{C} = \boldsymbol{A} \odot \boldsymbol{B} \in \mathbb{R}^{k\times l}$ is defined such that $C_{i,j} = A_{i,j}B_{i,j}$ for all $i \in [k]$ and $j \in [l]$.

D. Definitions

In this subsection, we define some quantities that are useful throughout. Define $s = \max\{n, m\}$. Also, define the sets

$$\Delta_{l,\varepsilon} = \left\{\boldsymbol{p} \in \mathbb{R}^k : \sum_{i=1}^l p_i = 1, p_i \geq \varepsilon \ \forall i \in [l]\right\}, \text{ where } l \in \mathbb{N},$$

$$\mathcal{S}_\varepsilon^{\text{row}} = \left\{\boldsymbol{P} \in \mathbb{R}_+^{s\times s} : \sum_{k=1}^s P_{i,k} = 1, P_{i,j} \geq \varepsilon \ \forall i, j \in [s]\right\},$$

$$\mathcal{S}_\varepsilon^{\text{col}} = \left\{\boldsymbol{P} \in \mathbb{R}_+^{s\times s} : \sum_{k=1}^s P_{k,j} = 1, P_{i,j} \geq \varepsilon \ \forall i, j \in [s]\right\},$$

$$\mathcal{S}_\varepsilon^{\text{doub}} = \mathcal{S}_\varepsilon^{\text{col}} \cap \mathcal{S}_\varepsilon^{\text{row}}, \ \mathcal{S}_\varepsilon = \mathcal{S}_\varepsilon^{\text{col}} \cup \mathcal{S}_\varepsilon^{\text{row}}.$$

We also denote $\Delta_l = \Delta_{l,0}$, $\mathcal{S}^{\text{row}} = \mathcal{S}_0^{\text{row}}$, $\mathcal{S}^{\text{col}} = \mathcal{S}_0^{\text{col}}$, $\mathcal{S}^{\text{doub}} = \mathcal{S}_0^{\text{doub}}$, and $\mathcal{S} = \mathcal{S}_0$. We hide the dependence on s in the notation for sets for clarity.

E. Assumptions

Before moving on to the problem, we state our main assumptions.

A1 The function ϕ is concave, entrywise nondecreasing, and has bounded subgradients in $[0, 1]^n$, i.e, $|\phi_i'(\boldsymbol{x})| \leq B \ \forall i \in [n]$, $\boldsymbol{x} \in [0, 1]^n$. Hence, ϕ is $\sqrt{n}B$-Lipschitz continuous.

A2 We have access to the solution of the problem $\max_{\boldsymbol{x}\in[0,1]^n}[\phi(\boldsymbol{x}) + \sum_{i=1}^n c_i x_i]$, for all $\boldsymbol{c} \in \mathbb{R}_+^n$.
Note: This assumption is valid for most separable functions ϕ. For instance, when ϕ is a proportionally fair utility function type of the form $\phi(\boldsymbol{x}) = \sum_{i=1}^n \log(1+\beta x_i)$, where $\beta \in \mathbb{R}_+$, the problem has an explicit solution.

II. Problem Setup

Recall the definition of matrices $\boldsymbol{Y}(t), \boldsymbol{S}(t) \in \{0, 1\}^{n\times m}$ and vector $\boldsymbol{X}(t) \in \{0, 1\}^n$. In particular,

$$S_{i,j}(t) = \begin{cases} 1 & \text{if link } i, j \text{ is successful in time slot } t \\ 0 & \text{otherwise,} \end{cases}$$

$$Y_{i,j}(t) = \begin{cases} 1 & \text{user } i \text{ is assigned to channel } j \text{ in time slot } t \\ 0 & \text{otherwise,} \end{cases}$$

and $X_i(t) = \sum_{j=1}^m Y_{i,j}(t)S_{i,j}(t)$ for all $i \in \{1, 2, \ldots, n\}$. The problem of interest is

$$\text{(P1:)} \max_{\boldsymbol{Y}(1),\boldsymbol{Y}(2),\ldots} \liminf_{T\to\infty} \phi\left(\frac{1}{T}\sum_{t=1}^T \mathbb{E}\{\boldsymbol{X}(t)\}\right) \tag{1}$$

$$\text{s.t. } \boldsymbol{Y}(t) \text{ and } \boldsymbol{S}(\tau) \text{ are independent for all } t, \tau \in \mathbb{N} \text{ and } \tau \geq t \tag{2}$$

$$\boldsymbol{Y}(t) \text{ and } S_{i,j}(\tau) \text{ are independent for all } t, \tau \in \mathbb{N}, (i, j) \in [n] \times [m], \tau < t, Y_{i,j}(\tau) \neq 1 \tag{3}$$

$$\boldsymbol{Y}(t) \in \{0, 1\}^{n\times m} \ \forall t \in \mathbb{N} \tag{4}$$

$$\sum_{i=1}^n Y_{i,j}(t) \leq 1 \ \forall t \in \mathbb{N}, j \in [m] \tag{5}$$

$$\sum_{j=1}^m Y_{i,j}(t) \leq 1 \ \forall t \in \mathbb{N}, i \in [n] \tag{6}$$

$$X_i(t) = \sum_{j=1}^m Y_{i,j}(t)S_{i,j}(t) \forall t \in \mathbb{N}, \ i \in [n], \tag{7}$$

where constraint (2) ensures transmission decisions do not know success/failures before they happen; (3) ensures we cannot use information that is never observed. Define ϕ^{opt} as the optimal objective value of (P1).

III. Multi-Channel Algorithm

The algorithm uses three parameters $V > 0, \eta > 0$, and $\varepsilon \in (0, 1/s]$. We first introduce the ROUND function, a technique adapted from the one introduced in [24] to approximate a nonnegative matrix by a matrix in the transport polytope. This function takes a matrix $\boldsymbol{P} \in \mathbb{R}^{s\times s}$ as an input and outputs a doubly stochastic matrix. Then, we introduce our algorithm (Algorithm 1), which uses the ROUND function as a subroutine. Next, we provide explicit solutions to the algorithm's intermediate problems. In Theorem 2, we establish the performance bound of the algorithm. In section III-A, we provide intuitive explanations for the steps of the algorithm. In section III-B, we discuss the major steps in obtaining the error bound of Theorem 2. Finally, in section III-C, we discuss the adaptiveness of the algorithm.

ROUND($\boldsymbol{P}$) function for $\boldsymbol{P} \in \mathbb{R}_+^{s\times s}$:

1) Define the matrix $\boldsymbol{P}'$ (row normalization of $\boldsymbol{P}$)
$$P'_{i,j} = \begin{cases} \frac{P_{i,j}}{\sum_{l=1}^s P_{i,l}} & \text{if } \sum_{l=1}^s P_{i,l} > 1 \\ P_{i,j} & \text{otherwise.} \end{cases}$$
2) Define the matrix $\boldsymbol{P}''$ (column normalization of $\boldsymbol{P}'$)
$$P''_{i,j} = \begin{cases} \frac{P'_{k,j}}{\sum_{k=1}^s P'_{k,j}} & \text{if } \sum_{k=1}^s P'_{k,j} > 1 \\ P'_{i,j} & \text{otherwise.} \end{cases}$$
3) Define the output matrix $\boldsymbol{Q}$,
$$\boldsymbol{Q} = \begin{cases} \boldsymbol{P}'' + \frac{(\mathbf{1}-\boldsymbol{P}''\mathbf{1})(\mathbf{1}-(\boldsymbol{P}'')^\top\mathbf{1})^\top}{C} & \text{if } C \neq 0 \\ \boldsymbol{P}'' & \text{otherwise,} \end{cases}$$

where $C = \|\mathbf{1} - \boldsymbol{P}^{''}\mathbf{1}\|_1$.

It can be shown that ROUND$(\boldsymbol{P}) \in \mathcal{S}_{\varepsilon/s}^{\text{doub}}$ whenever $\boldsymbol{P} \in \mathcal{S}_\varepsilon$.

Algorithm 1: Multi-Channel Adaptive MAC

1 Initialize $\tilde{\boldsymbol{P}}(1) \in \mathcal{S}_\varepsilon^{\text{doub}}$ and the virtual queues $\boldsymbol{Q}(1) \in [0, BV+1]^n$ arbitrarily.
2 **for** *each time slot* $t \in \mathbb{N}$ **do**
3 Set $\boldsymbol{P}(t) = \text{ROUND}(\tilde{\boldsymbol{P}}(t))$ (This function yields $\boldsymbol{P}(t) \in \mathcal{S}_{\varepsilon/s}^{\text{doub}}$).
4 Using Birkhoff-von Neumann decomposition [27], find $r \in \mathbb{N}$ and permutation matrices $\boldsymbol{M}^1, \ldots, \boldsymbol{M}^r$ such that $\boldsymbol{P}(t) = \sum_{l=1}^r s_l(t)\boldsymbol{M}^l$, and $\boldsymbol{s}(t) \in \Delta_r$.
5 Sample $l_t \sim \boldsymbol{s}(t)$ and take action $\boldsymbol{Y}(t)$, where $Y_{i,j}(t) = M_{i,j}^{l_t}$ for all $i \in [n]$, $j \in [m]$, and receive $\boldsymbol{S}(t) \odot \boldsymbol{Y}(t)$ as feedback.
6 Compute the estimator $\hat{\boldsymbol{S}}(t)$ for $\boldsymbol{S}(t)$ using $\hat{S}_{i,j}(t) = S_{i,j}(t)Y_{i,j}(t)/P_{i,j}(t)$ for all $i \in [n]$, and $j \in [m]$.
7 Find $\boldsymbol{\gamma}(t+1) \in [0,1]^n$, and $\tilde{\boldsymbol{P}}(t+1) \in \mathcal{Q}$ using

$$\boldsymbol{\gamma}(t+1) = \arg\min_{\boldsymbol{\gamma}\in[0,1]^n} \Big[-V\phi(\boldsymbol{\gamma}) + \sum_{i=1}^n Q_i(t)\gamma_i \Big],$$

$$\tilde{\boldsymbol{P}}(t+1) = \arg\min_{\boldsymbol{P}\in\mathcal{Q}} \Big[-\sum_{i=1}^n\sum_{j=1}^m Q_i(t)\hat{S}_{i,j}(t)P_{i,j} + \frac{1}{\eta} D(\boldsymbol{P}\|\tilde{\boldsymbol{P}}(t)) \Big], \quad (8)$$

8 where $\mathcal{Q} = \mathcal{S}_\varepsilon^{\text{col}}$ if t is even, and $\mathcal{Q} = \mathcal{S}_\varepsilon^{\text{row}}$ if t is odd, and the divergence D is defined by

$$D(\boldsymbol{X}\|\boldsymbol{Y}) = \sum_{i=1}^s\sum_{j=1}^s X_{i,j}\log\left(\frac{X_{i,j}}{Y_{i,j}}\right) \quad (9)$$

for all $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{S}$.
9 Update the virtual queues

$$\boldsymbol{Q}(t+1) = [\boldsymbol{Q}(t) + \boldsymbol{\gamma}(t+1) - \boldsymbol{X}(t)]_+, \quad (10)$$

where $X_i(t) = \sum_{j=1}^m Y_{i,j}(t)S_{i,j}(t)$ for $i \in [n]$.

Below we focus on solving the intermediate problem (8).

Finding $\boldsymbol{\gamma}(t+1)$: Notice that this problem can be solved due to the Assumption **A2**.

Finding $\tilde{\boldsymbol{P}}(t+1)$: We only consider the case when t is even. The case when t is odd can be solved similarly. Notice that we can separately solve for each column of $\tilde{\boldsymbol{P}}(t+1)$. To solve for the j-th column of $\tilde{\boldsymbol{P}}(t+1)$, we define $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^s$, where

$$x_i = \begin{cases} \eta Q_i(t)\hat{S}_{i,j}(t) & \text{if } i \in [n], j \in [m] \\ 0 & \text{otherwise,} \end{cases}$$

and $\boldsymbol{y}$ is the j-th column of $\tilde{\boldsymbol{P}}(t)$. The problem to be solved is

$$\text{(P2:)} \min_{\boldsymbol{p}} -\sum_{i=1}^s x_i p_i + D_{\text{KL}}(\boldsymbol{p}\|\boldsymbol{y})$$
$$\text{s.t. } \boldsymbol{p} \in \Delta_{s,\varepsilon},$$

where $D_{\text{KL}}(\cdot\|\cdot)$ in this case is the KL-divergence. It should be noted that (P2) has a classic structure that is solved in [28]. In particular, we define $\boldsymbol{z}$, where $z_i = y_i\exp(x_i)$. First, assume that $\boldsymbol{z}$ is sorted in the nondecreasing order. It can be shown that there exists $l \in [0 : s-1]$ such that the vector $\boldsymbol{u}^l \in \mathbb{R}^s$ given by

$$u_j^l = \begin{cases} \varepsilon & \text{if } j \le l \\ \frac{z_j}{\sum_{i=l+1}^s z_i}(1-\varepsilon l) & \text{if } j > l \end{cases}$$

satisfies $u_j^l \ge \varepsilon$ for all $j \in [l+1 : s]$. Then it can be shown that $\boldsymbol{u}^l$ is the solution to (P2). Hence, to solve (P2) we calculate $\boldsymbol{u}^k$ for each $k \in [0 : s-1]$ and check the above condition.

Now, we state the error bound of Algorithm 1 as a theorem.

Theorem 1. *For any parameters $T, T_0 \in \mathbb{N}$, $\varepsilon \in (0, 1/s)$, $\eta, V > 0$, Algorithm 1 yields*

$$\begin{aligned} &\phi^{opt} - \phi\left(\mathbb{E}\left\{\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\boldsymbol{X}(t)\right\}\right) \\ &\le \frac{n}{V} + \frac{9\eta nms^2}{2\varepsilon V}(BV+1)^2 + \frac{\varepsilon nms(BV+1)}{V} \\ &\quad + \frac{s}{\eta VT}\log\left(\frac{1}{\varepsilon}\right) + \frac{n(BV+1)^2}{2VT} + \frac{nB(BV+1)}{T}. \end{aligned} \quad (11)$$

In particular

(1) *For any $\epsilon > 0$, choosing $V = \Theta(1/\epsilon)$, $\varepsilon = \Theta(\epsilon) < 1/s$, $\eta = \Theta(\epsilon^3)$, we have*

$$\phi^{opt} - \lim\inf_{T\to\infty}\phi\left(\mathbb{E}\left\{\frac{1}{T}\sum_{t=1}^T\boldsymbol{X}(t)\right\}\right) = O(\epsilon).$$

(2) *In the finite time horizon setting with $T, T_0 \in \mathbb{N}$, using $\eta = \Theta(1/T)$, $\varepsilon = \Theta(1/T^{1/3}) < 1/s$, and $V = \Theta(T^{1/3})$, we have*

$$\phi^{opt} - \phi\left(\mathbb{E}\left\{\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\boldsymbol{X}(t)\right\}\right) = O\left(\frac{\log(T)}{T^{1/3}}\right).$$

Proof. See the technical report [29]. □

A. Explanation of Algorithm 1

The idea of the algorithm is to find $\tilde{\boldsymbol{P}}(t) \in \mathcal{S}_\varepsilon$ as a column stochastic matrix and a row stochastic matrix alternatively in odd and even slots. Then we obtain $\boldsymbol{P}(t) \in \mathcal{S}^{\text{doub}}$ by approximating $\tilde{\boldsymbol{P}}(t)$ by a doubly stochastic matrix, using the rounding trick (Line 3). Next, we sample a permutation matrix from $\boldsymbol{P}(t)$ using the Birkhoff-von Neumann decomposition [27] (Lines 4-5). To perform the assignment in the time slot $t-1$, we discard the last $s-n$ rows or the last $s-m$ columns of the permutation matrix, depending on whether m or n is

larger. Next, using the feedback obtained from the assigned links, we compute the importance sampling-based estimator $\hat{\boldsymbol{S}}(t)$ of $\boldsymbol{S}(t)$ (Line 6). We use $\hat{\boldsymbol{S}}(t)$ to compute $\tilde{\boldsymbol{P}}(t+1)$ in (8).

The algorithm updates auxiliary variables $\boldsymbol{\gamma}(t) \in [0,1]^n$ and a virtual queue $\boldsymbol{Q}(t)$ in each iteration (Lines 7-9). These variables are used to deal with the nonlinearity of the function ϕ. In particular, instead of directly maximizing $\phi\left(\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\{\boldsymbol{X}(t)\}\right)$, we maximize $\mathbb{E}\left\{\sum_{t=T_0}^{T+T_0-1}\phi\left(\boldsymbol{\gamma}(t+1)\right)\right\}$ subject to the constraint that the time averages $\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\boldsymbol{\gamma}(t+1)$ and $\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\boldsymbol{X}(t)$ are close. We enforce the closeness of the time averages using queuing dynamics in (10) and by establishing a deterministic bound on $Q(t)$. Maximizing $\mathbb{E}\left\{\sum_{t=T_0}^{T+T_0-1}\phi\left(\boldsymbol{\gamma}(t+1)\right)\right\}$ is easier since unlike in $\phi\left(\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\{\boldsymbol{X}(t)\}\right)$, the sum over t is outside of the function ϕ, which allows us to utilize tools from classical bandit algorithms and Lyapunov optimization in our implementation.

B. Discussion of Theorem 1

In this section, we summarize the intuition of why Algorithm 1 gives the error bound of Theorem 1. The main idea of the algorithm is to make sure that the two differences

$$\phi^{\text{opt}} - \frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\}, \tag{12}$$

and

$$\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} - \phi\left(\sum_{t=T_0}^{T+T_0-1}\frac{\mathbb{E}\{\boldsymbol{X}(t)\}}{T}\right) \tag{13}$$

are small.

First, we discuss the importance of parameters V, ε and η. Like the standard EXP3 algorithm [20], η controls the amount of exploration required to learn $q_{i,j}$ values. The parameter V is used to trade-off between the degree to which the algorithm makes the two differences (12) and (13) small. Small V makes (13) small and (12) large. The parameter ε is required to ensure that the algorithm is adaptive.

Now, we explain the main intuition why the error bound of Theorem 1 holds. First, the ROUND function asserts the condition $\boldsymbol{P}(t) \in \mathcal{S}^{\text{doub}}_{\varepsilon/s}$ for all $t \in \mathbb{N}$. Next, notice that we find $\tilde{\boldsymbol{P}}(t)$ as the optimal value of the problem (8). However, the assignments (see lines 3-5 of Algorithm 1) are based on $\boldsymbol{P}(t)$. Hence, it is important to make sure that $\|\boldsymbol{P}(t) - \tilde{\boldsymbol{P}}(t)\|_1$ is small. By combining the properties of the ROUND function along with the idea of separately treating the even and the odd iterations of the algorithm, we can establish that $\|\boldsymbol{P}(t) - \tilde{\boldsymbol{P}}(t)\|_1 \leq \|\tilde{\boldsymbol{P}}(t+1) - \tilde{\boldsymbol{P}}(t)\|_1$ (see the technical report [29] for details). Notice that the decision in (8) penalizes the difference $\|\tilde{\boldsymbol{P}}(t+1) - \tilde{\boldsymbol{P}}(t)\|_1$ due to the final divergence term. Hence, $\|\boldsymbol{P}(t) - \tilde{\boldsymbol{P}}(t)\|_1$ is small.

Due to the properties of the queuing equation (10) and the auxiliary variables $\boldsymbol{\gamma}(t)$, we have the deterministic queue bound $Q_i(t) \leq BV + 1$ for all $i \in [n]$ and $t \in \mathbb{N}$. Combining the bound on $\|\boldsymbol{P}(t) - \tilde{\boldsymbol{P}}(t)\|_1$ with the deterministic queue bound, for any $T, T_0 \in \mathbb{N}$, we can bound the difference (12). From the queuing equation (10), the entrywise nondecreasing property of ϕ in Assumption **A1**, the deterministic queue bound and Jensen's inequality, we can bound the difference (13). Combining the two bounds, we get the result of Theorem 1. See [29] for the detailed proofs.

C. Adaptiveness

It turns out that the proof of Theorem 1 holds even if the success probabilities $q_{i,j}$ changed before time T_0, as long as they remained constant during $[T_0 : T_0 + T - 1]$. In this case, ϕ^{opt} in (11) is the optimal utility of (P1) that uses the constant success probabilities in $[T_0 : T_0 + T - 1]$ of the above scenario throughout the time horizon. Hence, the adaptiveness is satisfied.

IV. Single-channel algorithm

The absence of matching type constraints allows us to simplify Algorithm 1 in the single-channel scenario. We develop a separate adaptive algorithm for this case (Algorithm 2). Similar to Algorithm 1, this algorithm uses parameters $V > 0, \eta > 0$ and $\varepsilon \in (0, 1/n]$. In the algorithm, in the t-th iteration, we find $\boldsymbol{p}(t+1) \in \Delta_{n,\varepsilon}$, using the importance sampling based estimator of $\boldsymbol{S}(t)$. Then we use $\boldsymbol{p}(t+1)$ to sample the user in the $(t+1)$-th iteration. We also use auxiliary variables $\boldsymbol{\gamma}(t) \in [0,1]^n$ and a virtual queue $\boldsymbol{Q}(t)$, similar to Algorithm 1. We simplify the notations as $q_i = q_{i,1}$, $S_i(t) = S_{i,1}(t)$ and $Y_i(t) = Y_{i,1}(t)$ for all $i \in [n]$ and $t \in \mathbb{N}$.

Algorithm 2: Single-channel Adaptive MAC

1 Initialize $\boldsymbol{p}(1) \in \Delta_{n,\varepsilon}$, $\boldsymbol{\gamma}(1) \in [0,1]^n$, and the virtual queues $\boldsymbol{Q} \in [0, BV+1]^n$.
2 **for** *each iteration* $t \in [T]$ **do**
3 Sample $a_t \sim \boldsymbol{p}(t)$ and set $\boldsymbol{Y}(t) = \boldsymbol{e}_{a_t}$.
4 Receive feedback $\boldsymbol{S}(t) \odot \boldsymbol{Y}(t)$.
5 Compute the estimator $\hat{\boldsymbol{S}}(t)$ for $\boldsymbol{S}(t)$ using, $\hat{S}_i(t) = \frac{S_i(t)Y_i(t)}{p_i(t)}$ for all $i \in [n]$.
6 Find, $\boldsymbol{p}(t+1), \boldsymbol{\gamma}(t+1)$ using

$$\boldsymbol{\gamma}(t+1) = \arg\min_{\boldsymbol{\gamma}\in[0,1]^n}\Big[-V\phi(\boldsymbol{\gamma}) + \sum_{i=1}^{n} Q_i(t)\gamma_i\Big]$$

$$\boldsymbol{p}(t+1) = \arg\min_{\boldsymbol{p}\in\Delta_{n,\varepsilon}}\Big[-\sum_{i=1}^{n} Q_i(t)\hat{S}_i(t)p_i + \frac{1}{\eta}D_{\text{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t))\Big], \tag{14}$$

where D_{KL} is the KL-divergence.
7 Update the virtual queues,

$$\boldsymbol{Q}(t+1) = [\boldsymbol{Q}(t) + \boldsymbol{\gamma}(t+1) - \boldsymbol{X}(t)]_+, \tag{15}$$

where $\boldsymbol{X}(t) = \boldsymbol{Y}(t) \odot \boldsymbol{S}(t)$.

Notice that finding $\boldsymbol{\gamma}(t+1)$ in (14) is the same problem as in the multi-channel algorithm, whereas finding $\boldsymbol{p}(t+1)$ is the same as (P2).

Now, we introduce several lemmas that are useful in the solution without proof. See [29] for the proofs.

Lemma 1 (Pinsker's inequality)**.** *For $\boldsymbol{x}, \boldsymbol{y} \in \Delta_n$, we have that $D_{KL}(\boldsymbol{x}\|\boldsymbol{y}) \geq \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|_1^2 \geq \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2$.*

Lemma 2. *We have that $D_{KL}(\boldsymbol{x}\|\boldsymbol{y}) \leq \log\left(\frac{1}{\varepsilon}\right)$, for all $\boldsymbol{x} \in \Delta_n$ where $\boldsymbol{y} \in \Delta_{n,\varepsilon}$.*

Lemma 3. *We have for all $t \in \mathbb{N}$ and $i \in [n]$, $Q_i(t) \leq BV+1$.*

Define the drift $\Delta(t)$ as,

$$\Delta(t) = \frac{1}{2}\mathbb{E}\{\|\boldsymbol{Q}(t+1)\|^2\} - \frac{1}{2}\mathbb{E}\{\|\boldsymbol{Q}(t)\|^2\}. \tag{16}$$

We have the following bound on $\Delta(t)$.

Lemma 4. *We have that for all $t \in \{1, 2, \dots\}$,*

$$\Delta(t) \leq n + \sum_{i=1}^{n}\mathbb{E}\{\gamma_i(t+1)Q_i(t)\} - \sum_{i=1}^{n} q_i\mathbb{E}\{Q_i(t)p_i(t)\}.$$

Proof. The proof follows from standard Lyapunov drift analysis. See the technical report [29] for the full proof. □

Lemma 5. *We have that,*

$$\begin{aligned}
&-\sum_{i=1}^{n} q_i\mathbb{E}\left\{Q_i(t)p_i(t)\right\} \\
&\leq \frac{\eta n}{2\varepsilon}(BV+1)^2 + \frac{1}{\eta}\mathbb{E}\left\{D_{KL}(\boldsymbol{p}(t+1)\|\boldsymbol{p}(t))\right\} \\
&\quad - \sum_{i=1}^{n}\mathbb{E}\left\{Q_i(t)\hat{S}_i(t)p_i(t+1)\right\}.
\end{aligned}$$

Proof. See, [29] for details. □

Now, we introduce the following lemma.

Lemma 6. *We have that for any $T, T_0 \in \mathbb{N}$, $\boldsymbol{\gamma} \in [0,1]^n$, and $\boldsymbol{p} \in \Delta_{n,\varepsilon}$,*

$$\begin{aligned}
&VT\phi(\boldsymbol{\gamma}) - V\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} \\
&\leq nT + \frac{\eta nT}{2\varepsilon}(BV+1)^2 + \sum_{t=T_0}^{T+T_0-1}\sum_{i=1}^{n}(\gamma_i - q_ip_i)\mathbb{E}\{Q_i(t)\} \\
&\quad + \frac{1}{\eta}\log\left(\frac{1}{\varepsilon}\right) + \frac{n(BV+1)^2}{2}.
\end{aligned}$$

Proof. Adding $-V\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\}$ to the result of Lemma 4, we have that,

$$\begin{aligned}
&\Delta(t) - V\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} \\
&\leq n - V\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} + \sum_{i=1}^{n}\mathbb{E}\{\gamma_i(t+1)Q_i(t)\} \\
&\quad - \sum_{i=1}^{n} q_i\mathbb{E}\{Q_i(t)p_i(t)\} \\
&\leq_{(a)} n - V\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} + \sum_{i=1}^{n}\mathbb{E}\{\gamma_i(t+1)Q_i(t)\} \\
&\quad + \frac{\eta n}{2\varepsilon}(BV+1)^2 + \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}(t+1)\|\boldsymbol{p}(t))\right\} \\
&\quad - \sum_{i=1}^{n}\mathbb{E}\left\{Q_i(t)\hat{S}_i(t)p_i(t+1)\right\},
\end{aligned} \tag{17}$$

where (a) follows by Lemma 5. Combining the decision (14) with push-back lemma regarding minimizing strongly convex functions (see, for example [30]), we have that for any $\boldsymbol{\gamma} \in [0,1]^n$ and $\boldsymbol{p} \in \Delta_{n,\varepsilon}$,

$$\begin{aligned}
&- V\phi(\boldsymbol{\gamma}(t+1)) + \sum_{i=1}^{n}\gamma_i(t+1)Q_i(t) \\
&\quad + \frac{1}{\eta}D_{\mathrm{KL}}(\boldsymbol{p}(t+1)\|\boldsymbol{p}(t)) - \sum_{i=1}^{n}Q_i(t)\hat{S}_i(t)p_i(t+1) \\
&\leq -V\phi(\boldsymbol{\gamma}) + \sum_{i=1}^{n}Q_i(t)[\gamma_i - \hat{S}_i(t)p_i] + \frac{1}{\eta}D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t)) \\
&\quad - \frac{1}{\eta}D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t+1)).
\end{aligned}$$

Taking expectations of the above, we have that,

$$\begin{aligned}
&- V\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} + \sum_{i=1}^{n}\mathbb{E}\{\gamma_i(t+1)Q_i(t)\} \\
&\quad + \frac{\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}(t+1)\|\boldsymbol{p}(t))\right\}}{\eta} - \sum_{i=1}^{n}\mathbb{E}\{Q_i(t)\hat{S}_i(t)p_i(t+1)\} \\
&\leq -V\phi(\boldsymbol{\gamma}) + \sum_{i=1}^{n}\gamma_i\mathbb{E}\{Q_i(t)\} - \sum_{i=1}^{n}\mathbb{E}\{Q_i(t)\hat{S}_i(t)p_i\} \\
&\quad + \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t))\right\} - \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t+1))\right\} \\
&=_{(a)} -V\phi(\boldsymbol{\gamma}) + \sum_{i=1}^{n}\gamma_i\mathbb{E}\{Q_i(t)\} - \sum_{i=1}^{n}q_ip_i\mathbb{E}\{Q_i(t)\} \\
&\quad + \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t))\right\} - \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t+1))\right\}.
\end{aligned}$$

Here, (a) follows from $\mathbb{E}\{Q_i(t)\hat{S}_i(t)|\mathcal{H}(t)\} = q_iQ_i(t)$, where $\mathcal{H}(t) = \{\boldsymbol{Y}(1), \dots, \boldsymbol{Y}(t-1), \boldsymbol{Y}(1) \odot \boldsymbol{S}(1), \dots, \boldsymbol{Y}(t-1) \odot \boldsymbol{S}(t-1)\}$ is the history up to time t. This is because $Q_i(t)$ is $\mathcal{H}(t)$-measurable and $\mathbb{E}\{\hat{S}_i(t)|\mathcal{H}(t)\} = q_i$. Substituting the above in (17), we have

$$\begin{aligned}
\Delta(t) - V\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} &\leq n + \frac{\eta n}{2\varepsilon}(BV+1)^2 - V\phi(\boldsymbol{\gamma}) \\
&\quad + \sum_{i=1}^{n}(\gamma_i - q_ip_i)\mathbb{E}\{Q_i(t)\} + \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t))\right\} \\
&\quad - \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(t+1))\right\}.
\end{aligned}$$

Summing for $t \in \{T_0, T_0+1, \dots, T_0+T-1\}$ gives

$$\frac{1}{2}\mathbb{E}\{\|\boldsymbol{Q}(T+T_0)\|^2\} - \frac{1}{2}\mathbb{E}\{\|\boldsymbol{Q}(T_0)\|^2\}$$

$$
\begin{aligned}
&- V \sum_{t=T_0}^{T+T_0-1} \mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} \\
&\leq nT + \frac{\eta n T}{2\varepsilon}(BV+1)^2 - VT\phi(\boldsymbol{\gamma}) \\
&\quad + \sum_{t=T_0}^{T_0+T-1}\sum_{i=1}^{n}(\gamma_i - q_i p_i)\mathbb{E}\{Q_i(t)\} \\
&\quad + \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(T_0))\right\} - \frac{1}{\eta}\mathbb{E}\left\{D_{\mathrm{KL}}(\boldsymbol{p}\|\boldsymbol{p}(T+T_0))\right\} \\
&\leq nT + \frac{\eta n T}{2\varepsilon}(BV+1)^2 - VT\phi(\boldsymbol{\gamma}) \\
&\quad + \sum_{t=T_0}^{T_0+T-1}\sum_{i=1}^{n}(\gamma_i - q_i p_i)\mathbb{E}\{Q_i(t)\} + \frac{1}{\eta}\log\left(\frac{1}{\varepsilon}\right),
\end{aligned}
$$

where for the last inequality, we have used Lemma 2. Rearranging the above, and using $\|\boldsymbol{Q}(T_0)\|^2 \leq n(BV+1)^2$, and $\|\boldsymbol{Q}(T+T_0)\|^2 \geq 0$ we are done. □

Next, we have the following lemma.

Lemma 7. *We have for any $T, T_0 \in \mathbb{N}$*

$$
\begin{aligned}
&\phi\left(\frac{1}{T}\left(\sum_{t=T_0+1}^{T+T_0}\boldsymbol{\gamma}(t)\right)\right) \\
&\qquad \leq \phi\left(\frac{1}{T}\left(\sum_{t=T_0}^{T+T_0-1}\boldsymbol{X}(t)\right)\right) + \frac{nB(BV+1)}{T}.
\end{aligned}
$$

Proof. This follows combining the queuing equation (15) with Assumption **A1** and Lemma 3. See the technical report [29] for details. □

Now, we are ready to establish the performance bound of the algorithm.

Theorem 2. *For any $T, T_0 \in \mathbb{N}$, parameters $\varepsilon \in (0, 1/n)$, $\eta, V > 0$, we have that,*

$$
\begin{aligned}
\phi^{opt} - \phi\left(\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\{\boldsymbol{X}(t)\}\right) &\leq \frac{n}{V} + \frac{\eta n}{2\varepsilon V}(BV+1)^2 \\
&\quad + \frac{\varepsilon n^2}{V}(BV+1) + \frac{1}{\eta V T}\log\left(\frac{1}{\varepsilon}\right) + \frac{n(BV+1)^2}{2VT} \\
&\quad + \frac{nB(BV+1)}{T}.
\end{aligned}
$$

We also have the same two results as Theorem 1-(1), and Theorem 1-(2).

Proof. It can be shown that ϕ^{opt} is the optimal solution of the problem,

$$
\begin{aligned}
\text{(P3:)} \quad & \max_{\boldsymbol{p}\in\Delta_n, \boldsymbol{\gamma}\in[0,1]^n} \phi(\boldsymbol{\gamma}) && (18) \\
& \text{s.t. } p_i q_i \geq \gamma_i \ \forall i \in [n]. && (19)
\end{aligned}
$$

See [29] for a proof. Let $(\boldsymbol{p}^*, \boldsymbol{\gamma}^*)$ denote the optimal solution of (P3). Substituting $\boldsymbol{p}^*(1-\varepsilon n) + \varepsilon\mathbf{1}, \boldsymbol{\gamma}^*$ in Lemma 6, we have that

$$
\begin{aligned}
VT\phi^{\mathrm{opt}} - V\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\{\phi(\boldsymbol{\gamma}(t+1))\} &\leq nT + \frac{\eta n T}{2\varepsilon}(BV+1)^2 \\
&\quad + \sum_{t=T_0}^{T+T_0-1}\sum_{i=1}^{n}(\gamma_i^* - q_i p_i^*(1-\varepsilon n) - \varepsilon q_i)\mathbb{E}\{Q_i(t)\} \\
&\quad + \frac{1}{\eta}\log\left(\frac{1}{\varepsilon}\right) + \frac{n(BV+1)^2}{2} \\
&\leq nT + \frac{\eta n T}{2\varepsilon}(BV+1)^2 + \sum_{t=T_0}^{T+T_0-1}\sum_{i=1}^{n}(\gamma_i^* - q_i p_i^*)\mathbb{E}\{Q_i(t)\} \\
&\quad + \varepsilon\sum_{t=T_0}^{T+T_0-1}\sum_{i=1}^{n} n q_i p_i^* \mathbb{E}\{Q_i(t)\} + \frac{1}{\eta}\log\left(\frac{1}{\varepsilon}\right) + \frac{n(BV+1)^2}{2} \\
&\leq_{(a)} nT + \frac{\eta n T}{2\varepsilon}(BV+1)^2 + \varepsilon n^2 T(BV+1) + \frac{1}{\eta}\log\left(\frac{1}{\varepsilon}\right) \\
&\quad + \frac{n(BV+1)^2}{2}, \qquad (20)
\end{aligned}
$$

where (a) follows since $q_i p_i^* \geq \gamma_i^*$ (since $(\boldsymbol{p}^*, \boldsymbol{\gamma}^*)$ is feasible for (P3)), and Lemma 3. Now, we divide by VT and use the Jensen's inequality to obtain,

$$
\begin{aligned}
\phi^{\mathrm{opt}} - \mathbb{E}\left\{\phi\left(\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\boldsymbol{\gamma}(t+1)\right)\right\} &\leq \frac{n}{V} + \frac{\eta n}{2\varepsilon V}(BV+1)^2 \\
&\quad + \frac{\varepsilon n^2}{V}(BV+1) + \frac{1}{\eta V T}\log\left(\frac{1}{\varepsilon}\right) + \frac{n(BV+1)^2}{2VT}.
\end{aligned}
$$

Now, combining with Lemma 7, we have that,

$$
\begin{aligned}
&\phi^{\mathrm{opt}} - \mathbb{E}\left\{\phi\left(\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\boldsymbol{X}(t)\right)\right\} \\
&\leq \frac{n}{V} + \frac{\eta n}{2\varepsilon V}(BV+1)^2 + \frac{\varepsilon n^2}{V}(BV+1) + \frac{1}{\eta V T}\log\left(\frac{1}{\varepsilon}\right) \\
&\quad + \frac{n(BV+1)^2}{2VT} + \frac{nB(BV+1)}{T}.
\end{aligned}
$$

Using Jensen's inequality, we are done. □

A. Enforcing user fairness

Given $\theta \in (0, 1/n]$, consider the case where we require each user to transmit at least θ fraction of the time on average. Then, for (P1), we require the additional constraint of

$$
\liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\{Y_i(t)\} \geq \theta
$$

This simply transforms the constraint $\boldsymbol{p} \in \Delta_n$ of (P3) to $\boldsymbol{p} \in \Delta_{n,\theta}$. Using $\varepsilon = \theta$ in Algorithm 2, we can use $\boldsymbol{p}^*$ directly in (20) in Theorem 2 instead of $(1-\varepsilon n)\boldsymbol{p}^* + \varepsilon\mathbf{1}$, which gives

$$
\begin{aligned}
\phi^{\mathrm{opt}} - \phi\left(\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\left\{\boldsymbol{X}(t)\right\}\right) &\leq \frac{n}{V} + \frac{\eta n}{2\theta V}(BV+1)^2 \\
&\quad + \frac{1}{\eta V T}\log\left(\frac{1}{\theta}\right) + \frac{nB(BV+1)}{T} + \frac{n(BV+1)^2}{2VT}
\end{aligned}
$$

for all $T, T_0 \in \mathbb{N}$. Since θ is a constant, using $\eta = \Theta(1/T)$, and $V = \Theta(\sqrt{T})$, we have,

$$
\phi^{\mathrm{opt}} - \phi\left(\frac{1}{T}\sum_{t=T_0}^{T+T_0-1}\mathbb{E}\left\{\boldsymbol{X}(t)\right\}\right) = O\left(\frac{1}{\sqrt{T}}\right).
$$

B. Comparison with Algorithm 1

We compare the error bound of Theorem 2 with the error bound of Theorem 1 for $m = 1$. Notice that the dependence of the error on T is the same. But notice that the dependence of the error bound on n is worse for Algorithm 1. This is due to the $9ns^2$ term appearing in the bound of Theorem 1.

V. Simulations

We consider the function $\phi(\boldsymbol{x}) = \sum_{i=1}^{n} \log(1 + x_i)$ for the simulations. We consider two scenarios with $n = 5$, where in the first (Fig 1-Left) we use $m = 3$ (Algorithm 1), and in the second (Fig 1-Right) we use $m = 1$ (Algorithm 2). In each scenario, we run our algorithm for $T = 10^5$ time slots. The $q_{i,j}$ values are changed at $T/2$. It can be seen that in both cases, the algorithm converges to the corresponding optimal values. In addition, it can be seen that the algorithms adapt to the change of $q_{i,j}$ values.

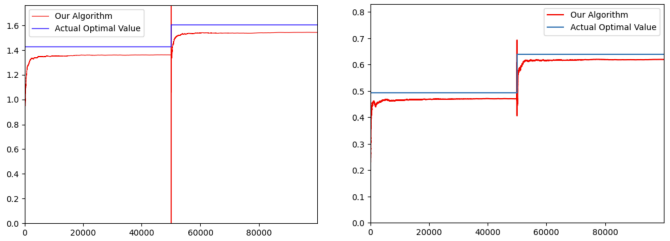

Fig. 1. ϕ^{opt} and objective value of algorithms vs t, **Left:** $m = 3$, **Right:** $m = 1$

VI. Conclusions

This paper focused on the problem of automatic link selection in multiple access with link failures. In particular, we solved a network utility maximization problem with bandit feedback on the link failures. Our algorithm was proven to provide near-optimal utility over any block of T slots where T is suitably large. Simulations depict the fast learning of efficient decisions and demonstrate quick adaptation when unknown probabilities change. Extending this work to consider time-correlated scenarios where the link successes are modulated by an unknown Markov chain is future work.

References

[1] T. Hwang, C. Yang, G. Wu, S. Li, and G. Ye Li, "OFDM and its wireless applications: A survey," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1673–1694, Aug. 2009.

[2] B. Makki, K. Chitti, A. Behravan, and M.-S. Alouini, "A survey of NOMA: Current status and open research challenges," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 179–189, Jan. 2020.

[3] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. vol. 8, no. 1 pp. 33-37, Jan.-Feb. 1997.

[4] F. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness, and stability," *Journ. of the Operational Res. Society*, vol. vol. 49, no. 3, pp. 237-252, Mar. 1998.

[5] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. vol. 8, no. 5, Oct. 2000.

[6] E. Altman, K. Avrachenkov, and A. Garnaev, "Generalized α-fair resource allocation in wireless networks," in *Proc. Conference on Decision and Control*, Dec. 2008.

[7] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *Proc. IEEE INFOCOM*, Mar. 2010.

[8] X. Liu, E. K. P. Chong, and N. B. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. vol. 41, no. 4, pp. 451-474, Mar. 2003.

[9] L. Tassiulas and A. Ephremides, "Dynamic server allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. vol. 39, no. 2, pp. 466-478, Mar. 1993.

[10] H. Kushner and P. Whiting, "Asymptotic properties of proportional-fair sharing algorithms," *Proc. 40th Annual Allerton Conf. on Communication, Control, and Computing, Monticello, IL*, Oct. 2002.

[11] R. Agrawal and V. Subramanian, "Optimality of certain channel aware scheduling policies," *Proc. 40th Annual Allerton Conf. on Communication, Control, and Computing, Monticello, IL*, Oct. 2002.

[12] M. J. Neely, "Convergence and adaptation for utility optimal opportunistic scheduling," *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 904–917, Jun. 2019.

[13] A. Eryilmaz and R. Srikant, "Joint congestion control, routing, and MAC for stability and fairness in wireless networks," *IEEE Journal on Selected Areas in Communications, Special Issue on Nonlinear Optimization of Communication Systems*, vol. vol. 14, pp. 1514-1524, Aug. 2006.

[14] A. Stolyar, "Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm," *Queueing Systems*, vol. vol. 50, no. 4, pp. 401-457, Aug. 2005.

[15] M. J. Neely, *Stochastic Network Optimization with Application to Communication and Queueing Systems*. Morgan & Claypool, 2010.

[16] J. C. Duchi, "Introductory lectures on stochastic optimization," *The mathematics of data*, vol. 25, pp. 99–186, Nov. 2018.

[17] S. Agrawal and N. R. Devanur, "Bandits with concave rewards and convex knapsacks," in *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, Jun. 2014, p. 989–1006.

[18] V. Do, E. Dohmatob, M. Pirotta, A. Lazaric, and N. Usunier, "Contextual bandits with concave rewards, and an application to fair ranking," in *The Eleventh International Conference on Learning Representations*, Feb. 2023.

[19] S. Agrawal and N. R. Devanur, "Bandits with global convex constraints and objective," *Operations Research*, vol. 67, no. 5, pp. 1486–1502, Aug. 2019.

[20] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "Gambling in a rigged casino: The adversarial multi-armed bandit problem," in *Proceedings of IEEE 36th annual foundations of computer science*. IEEE, Aug. 1995, pp. 322–331.

[21] R. Sinkhorn, "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 876 – 879, Jun. 1964.

[22] C. Chen, C. Zhao, and S. Li, "Simultaneously learning stochastic and adversarial bandits under the position-based model," in *AAAI Conference on Artificial Intelligence*, Jun. 2022.

[23] M. Ballu and Q. Berthet, "Mirror Sinkhorn: fast online optimization on transport polytopes," in *Proceedings of the 40th International Conference on Machine Learning*, ser. ICML'23, Jul. 2023.

[24] J. Altschuler, J. Niles-Weed, and P. Rigollet, "Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration," in *Advances in Neural Information Processing Systems*, Dec. 2017.

[25] P. Auer, P. Gajane, and R. Ortner, "Adaptively tracking the best bandit arm with an unknown number of distribution changes," in *Proceedings of the Thirty-Second Conference on Learning Theory*, vol. 99, Jun. 2019, pp. 138–158.

[26] W. C. Cheung, D. Simchi-Levi, and R. Zhu, "Learning to optimize under non-stationarity," in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, vol. 89. PMLR, Apr. 2019, pp. 1079–1087.

[27] G. Birkhoff, "Three observations on linear algebra," *Univ. Nac. Tacuman, Rev. Ser. A*, vol. 5, pp. 147–151, 1946.

[28] J. Huang, L. Golubchik, and L. Huang, "When Lyapunov drift based queue scheduling meets adversarial bandit learning," *IEEE/ACM Transactions on Networking*, vol. 32, no. 4, pp. 3034–3044, Aug. 2024.

[29] M. Wijewardena and M. J. Neely, "Automatic link selection in multi-channel multiple access with link failures," Jan. 2025. [Online]. Available: https://arxiv.org/abs/2501.14971

[30] X. Wei, H. Yu, and M. J. Neely, "Online primal-dual mirror descent under stochastic constraints," *Proc. ACM Meas. Anal. Comput. Syst.*, vol. 4, no. 2, Jun. 2020.