Federated Best Arm Identification With Heterogeneous Clients

Zhirui Chen, P. N. Karthik, *Member, IEEE*, Vincent Y. F. Tan, *Senior Member, IEEE*, and Yeow Meng Chee, *Senior Member, IEEE*

***Abstract*— We study best arm identification in a federated multi-armed bandit setting with a central server and multiple clients, when each client has access to a *subset* of arms and each arm yields independent Gaussian observations. The goal is to identify the best arm of each client subject to an upper bound on the error probability; here, the best arm is one that has the largest *average* value of the means averaged across all clients having access to the arm. Our interest is in the asymptotics as the error probability vanishes. We provide an asymptotic lower bound on the growth rate of the expected stopping time of any algorithm. Furthermore, we show that for any algorithm whose upper bound on the expected stopping time matches with the lower bound up to a multiplicative constant (*almost-optimal* algorithm), the ratio of any two consecutive communication time instants must be *bounded*, a result that is of independent interest. We thereby infer that an algorithm can communicate no more sparsely than at exponential time instants in order to be almost-optimal. For the class of almost-optimal algorithms, we present the first-of-its-kind asymptotic lower bound on the expected number of *communication rounds* until stoppage. We propose a novel algorithm that communicates at exponential time instants, and demonstrate that it is asymptotically almost-optimal.**

***Index Terms*— Multi-armed bandits, best arm identification, federated learning.**

I. Introduction

THE problem of best arm identification [1], [2] deals with finding the best arm in a multi-armed bandit as quickly as possible, and falls under the class of optimal stopping problems in decision theory. This problem has been studied under two complementary regimes: (a) the *fixed-confidence* regime in which the goal is to minimise the expected time (number of samples) required to find the best arm subject to an upper bound on the error probability [1], [3], and (b) the *fixed-budget* regime in which the goal is to minimise the error probability subject to an upper bound on the number of samples [4], [5]. In this paper, we study best arm identification with fixed confidence.

A. Problem Setup and Objective

We consider a federated learning setup [6], [7] with a central *server* and M *clients* in which each client has access to a *subset* of arms from a K-armed bandit (*heterogeneous clients*). For $i \in [K] := \{1, \ldots, K\}$ and $m \in [M] := \{1, \ldots, M\}$, arm i of client m generates independent Gaussian *observations* with mean $\mu_{i,m}$ and unit variance. We assume that the clients do not communicate directly with each other, but instead communicate via the server. We let S_m denote the subset of arms accessible by client m. For each $i \in S_m$, we let μ_i denote the average of the values in $\{\mu_{i,m} : i \in S_m\}$. We define the *mean reward* of arm i at client m to be μ_i, which is consistent with other similar works in federated multi-armed bandits [8], [9], [10] in which $S_m = [K]$ for every m; in our work, we allow for the case when $S_m \subseteq [K]$ for every m. Defining the *best arm* of client m as $\arg\max_{i \in S_m} \mu_i$, the goal is to find the best arms of the clients with minimal expected stopping time, subject to an upper bound on the error probability. Figure 1 depicts the problem setup pictorially.

Because the mean reward of an arm is the average of the means from all clients having access to the arm, it is necessary for the clients to *communicate* with the server in order to determine their individual best arms. Intuitively, more frequent communication between the clients and the server implies smaller expected stopping time. Thus, there is a close interplay between (a) frequency of communication, and (b) expected stopping time. Also, intuitively, the smaller the error probability, the larger the expected stopping time. Our objectives in this paper are two-fold: (i) to provide a rigorous theoretical characterisation of the trade-off between (a) and (b), and (ii) to capture in precise mathematical terms the limiting growth rate of the expected stopping time as the error probability vanishes.

B. Motivating Examples

Systemic biases [11, Chapter 6] in data are common in federated multi-armed bandit problems, where the performance of an arm can vary significantly across clients due to variations in user behavior and contextual factors [8]. This poses a significant challenge for selecting the best arm, as the local estimates of the arm's performance may not reflect its true value across all clients. To address this challenge, we propose

Manuscript received 8 May 2023; revised 3 September 2023; accepted 24 November 2023. Date of publication 30 November 2023; date of current version 21 May 2024. This work was supported in part by the National Research Foundation Singapore and DSO National Laboratories through the AI Singapore Programme under AISG Award AISG2-RP-2020-018 and in part by the Singapore Ministry of Education Academic Research Fund Tier 2 under Grant A-8000423-00-00. *(Corresponding author: Zhirui Chen.)*

Zhirui Chen and Yeow Meng Chee are with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore 117576 (e-mail: zhiruichen@u.nus.edu; ymchee@nus.edu.sg).

P. N. Karthik is with the Department of Artificial Intelligence, Indian Institute of Technology Hyderabad, Hyderabad 502284, India (e-mail: pnkarthik1992@gmail.com).

Vincent Y. F. Tan is with the Department of Mathematics and the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119077 (e-mail: vtan@nus.edu.sg).

Communicated by A. Krishnamurthy, Associate Editor for Machine Learning and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2023.3338027.

Digital Object Identifier 10.1109/TIT.2023.3338027

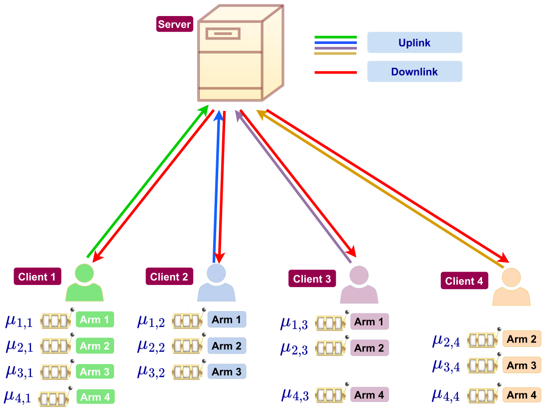

Fig. 1. A depiction of the federated learning setup with a central server and $M = 4$ clients, each having access to a subset of $K = 4$ arms. Client 1 has access to the subset $S_1 = \{1,2,3,4\}$, client 2 to the subset $S_2 = \{1,2,3\}$, client 3 to the subset $S_3 = \{1,2,4\}$, and client 4 to the subset $S_4 = \{2,3,4\}$. We assume that arm i of client m generates independent Gaussian rewards with mean $\mu_{i,m}$ and variance 1.

a definition of the best arm that considers both the local and global performance of each arm. Our proposed definition aggregates the local estimates of each arm's performance, which effectively reduces the overall bias and improves the estimation accuracy. To illustrate the importance of considering the global performance, we provide examples from market survey and democratic elections, where the best arm should be selected based on the average performance across all clients.

1) Market Survey: Consider M retailers (clients), each of whom sells products from a subset of K popular brands (arms). To determine the best product among their suite of products, suppose that each retailer conducts a market survey and collects consumers' ratings for different products. It is possible that different retailers accrue different expected rating scores (representative of $\mu_{i,m}$ in our problem setup) for the same brand (this corresponds to an arm having different means at different clients). For instance, skilled advertising by some retailers may influence customers' liking for certain brands over others, thereby forcing a certain degree of *bias* in the customers' ratings for products from such brands. Alternatively, in the event when the customers are asked to rate a product as "good", "satisfactory", "very good", etc. in the survey, there is possibility that these ratings are miscalibrated due to subjective differences in the perception of ratings by humans [12]. Thus, for instance, a "very good" rating may translate to a numerical score of 4 for one retailer and to 5 for another (say, on a scale of 0 to 5), depending on how the surveyors of various retailers perceive the customers' ratings. The above scenarios are only a handful examples of *systemic* biases in the collected data which make reliable estimation of the true value of a brand based on customer ratings a challenging task. In such scenarios, for each brand, it is logical to *average* the ratings across the retailers, and decide the best brand based on the average ratings, given that averaging is a definitive means of reducing the overall systemic bias. Our definition of the best arm (as $\arg\max_i \mu_i$, where μ_i is the *average* of $\mu_{i,m}$ values for $i \in S_m$) precisely accomplishes this.

2) Paper Reviewing Process: Another instance of miscalibration can be observed in the peer review process for academic papers. Consider a scenario where a set of K papers is distributed to M reviewers via an automated system. Each reviewer m is presented with a subset S_m of these papers. However, it is essential to acknowledge that the reviewers vary in terms of their expertise and seniority. Junior reviewers may evaluate the same paper quite differently from their senior counterparts. This disparity in evaluations could arise because junior reviewers tend to focus on different aspects of the papers, such as scrutinizing detailed proofs, while senior reviewers may prioritize assessing the broader impact and significance of the research. Consequently, it is highly probable that the same paper will receive a wide range of scores due to these differing evaluation criteria, and it is this phenomenon that we allude to as "miscalibration". Our proposal for mitigating this miscalibration is to take the *average* of the reviewers' scores; this is also commonly done in real-life.

3) Democratic Elections: In democratic elections involving multiple political parties represented across one or more states, one among a subset of parties (arms) is voted to power within each state (client). Favouritism in election—voting in favour of the party that has a demonstrated record of winning many past elections—is not uncommon; this is akin to voting in favor of party $\arg\max_{i\in S_m} \mu_{i,m}$, where $\mu_{i,m}$ is a collective measure of party i's performance (revenue generated, infrastructural improvements, etc.) as perceived by the people of state m. However, favouritism in voting is antithetical to the spirit of democracy, and calls for a more careful evaluation of a party's performance by the voters—one that gives a fair chance to non-favourite or new parties to come to power. Indeed, the voters of state m may want to exercise votes in favour of a non-favorite party that is seemingly performing well in *other states*, with the hope that it will deliver a similar (or better) performance in state m. Our proposal of evaluating party i's performance via μ_i (by averaging the performances across states) and voting in favour of party $\arg\max_i \mu_i$ precludes favouritism and supports voting in favour of the party that shows the greatest potential for performance overall.

C. Contributions

We now bring out the main contributions of this paper and highlight the challenges in the analysis. leftmargin=*

- We derive a problem instance-specific asymptotic lower bound on the expected stopping time (i.e., the time required to find the best arms of the clients). As in the prior works on best arm identification [13], [14], we show that given an error probability threshold $\delta \in (0,1)$, the lower bound scales as $\Omega(\log(1/\delta))$ (all logarithms are natural logarithms). We characterise the instance-dependent constant multiplying $\log(1/\delta)$. This constant, we show, is the solution to a max-inf optimisation problem in which the outer 'max' is over all probability distributions on the arms and the inner 'inf' is over the set of alternative problem instances, and is a measure of the "hardness" of the instance.

- The max-inf optimisation in the instance-dependent constant is seemingly hard to solve analytically. The hardness stems from set of alternative problem instances in the inner inf not admitting a closed-form expression, unlike in the prior works where simple closed-form expressions for the set of alternative instances exist. Notwithstanding this, we recast the inf over the (uncountably infinite) set of alternative problem instances as a min over the (finite) set of arms, and demonstrate that the max-min optimisation resulting from the latter can be solved analytically and differs from the true max-inf by at most a factor of 2.
- For any algorithm whose upper bound on the expected time to find the best arms of the clients matches the lower bound up to a multiplicative constant (an *almost-optimal algorithm*), we show that the ratio of any two consecutive communication instants *must be bounded*, a result that is of independent interest. That is, in order to achieve order-wise optimality in the expected time to find the best arms, an algorithm may communicate at most exponentially sparsely, e.g., at communication time instants of the form $t = \lceil (1+\lambda)^r \rceil$, $r \in \mathbb{N}$, for some $\lambda > 0$. In this sense, the class of all algorithms communicating exponentially sparsely (with different exponents) forms the *boundary* of the class of almost-optimal algorithms. Using this result, we show that given any error probability δ, there exists a sequence of problem instances with increasing hardness levels on which the expected number of communication rounds until stoppage grows with δ as $\Omega(\log\log(1/\delta))$ for any algorithm with a bounded ratio between consecutive communication time instants. This is the first-of-its kind result in the literature.
- We design a *Track-and-Stop*-based algorithm, called *Het*erogeneous *T*rack-and-*S*top (or HET-TS(λ) in short), that communicates only at exponential time instants of the form $t = \lceil (1+\lambda)^r \rceil$, $r \in \mathbb{N}$, for an input parameter $\lambda > 0$. We show that given any $\delta \in (0,1)$, the HET-TS(λ) algorithm (a) identifies the best arms correctly with probability greater than $1-\delta$, (b) is asymptotically almost-optimal up to the constant $2(1+\lambda)$, and (c) takes $O(\log\log(1/\delta))$ many communication rounds on the average. Here, λ serves as a tuning parameter to trade-off between the expected number of communication rounds and the expected stopping time.

D. Related Works

1) Federated Bandits: Best arm identification in the fixed-confidence regime for independent and identically distributed (i.i.d.) observations has been studied in [13] and [15]. The recent works [14], [16] extend the results of [13] to the setting of Markov observations from the arms. The problem of multi-armed bandits in federated learning has been studied in several recent works, including those with similar setups to our own [8], [9], [10], [17], [18], [19]. These works are generally classified as belonging to the class of "federated bandit" problems, first proposed by [9] in which each client has access to *all* the arms. The notion of *global mean* defined in these works coincides with our definition of *mean reward* (i.e., average of the arm means across the clients), and the goal is to design an algorithm that minimises the cumulative *regret* over a finite time horizon of T time units. The paper [20] studies best arm identification in a federated learning setting in which each client has access to a subset of arms that is disjoint from the arms subsets of the other clients, and the clients coordinate with each other to find the overall best arm (the arm with the largest mean); notice that in this setting, the best arm is necessarily the best arm of one of the clients. In our work, we allow for non-disjoint subsets of arms across the clients, and the best arm of one client may not necessarily be the best arm of another. The paper [21] studies an optimal stopping variant of the problem in [9] in which the uplink from each client to the server entails a fixed cost of $C \geq 0$ units, each client has access to *all* the arms, and the goal is to determine the arm with the largest mean at each client and also the arm with the largest *global mean* with minimal total cost, defined as the sum of the number of arm selections and the total communication cost. When each client has access to all the arms, our problem distils down to finding the arm with the largest global mean.

2) Collaborative Bandits: Another line of related works goes collectively by the name of *collaborative bandits* [22], [23], [24]. Here, each agent within a set of agents is capable of identifying the best arm in a *single* bandit environment without communication; the analytical task then is to quantify how much communication aids in reducing the overall sample complexity. In contrast, in our work, communication is clearly necessary to estimate the mean of the global best arm. This is the essential difference between our setting and that of collaborative bandits. Hillel et al. [24] initially carry out a study on pure exploration within the collaborative bandit framework. They demonstrate that a single communication round among agents is sufficient to identify the best arm efficiently. Tao et al. [22] quantifies the power of collaboration under limited interaction (or, communication steps), as interaction is expensive in many settings. They measure the running time of a distributed algorithm as the speedup over the best centralized algorithm where there is only one agent. Karpov et al. [23] study the problem of top-m arms identification in both fixed budget and fixed confidence cases under the setting of collaborative bandits. More recently, Karpov and Zhang [25] studied the problem of *fixed-budget* best arm identification in collaborative bandits on non-i.i.d. data. This bears similarities to the study of federated bandits. However, we tackle the problem of fixed-confidence best arm identification on a novel problem setting in which each client only has access to a possibly strict subset of arms (cf. Fig. 1).

3) Bandits With Communication Constraints: Additionally, some other studies take into account the effect of *communication constraints*, which is related to our work as communication or privacy constraints are often times incorporated into the federated learning setting. Hanna et al. [26] consider a bandit model in which the learner receives the reward value via a communication channel. Their findings indicate that a communication rate of 1-bit per time step is sufficient to achieve near-optimal regret when rewards are bounded. Mitra et al. [27] generalize this framework to the linear bandit setting. They show that a bit rate (i.e., number of bits per time step) linear in the dimension of the unknown parameter vector suffices to achieve near-optimal regret. Pase et al. [28] consider the Bayesian regret and demonstrate that to achieve sublinear regret, the bit rate needs to exceed $H(A^*)$, the entropy of the marginal distribution of the arm pulls under the optimal strategy. In addition, the model presented in Wang et al. [29] considers the *total number of bits* instead of the bit rate. They demonstrate that to achieve near-optimal regret, the total number of bits has to depend logarithmically on the horizon T. Mayekar et al. [30] recently analyzed communication-constrained bandits under additive Gaussian noise and showed that the regret depends on the capacity of the channel. Although these studies [26], [27], [28], [29], and [30] contribute towards understanding bandits with communication constraints, their focus is primarily on the number of bits of transmission in the communication channel. In contrast, our research focuses on the *frequency of transmission* in the communication channels from the clients to the server.

II. Notations and Preliminaries

For $n \in \mathbb{N} := \{1, 2, \ldots\}$, we let $[n] := \{1, \ldots, n\}$. We consider a federated multi-armed bandit with K arms, a central server, and M clients, in which each client has access to a subset of arms. For $m \in [M]$, let S_m denote the subset of arms accessible by client m. Without loss of generality, we assume that $|S_m| \geq 2$ for all m. Pulling arm $i \in S_m$ at time $t \in \mathbb{N}$ generates the observation $X_{i,m}(t)$ that is Gaussian distributed with mean $\mu_{i,m} \in \mathbb{R}$ and unit variance. A *problem instance* $v = (\{\mu_{i,1}\}_{i\in S_1}, \{\mu_{i,2}\}_{i\in S_2}, \ldots, \{\mu_{i,M}\}_{i\in S_M})$ is defined by the collection of the means of the arms in each client's set of accessible arms. For any $i \in [K]$, we define the *reward* $X_i(t)$ of arm i as the average of the observations obtained at time t from all clients m for which $i \in S_m$, i.e., $X_i(t) := \frac{1}{M_i}\sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}} X_{i,m}(t)$, where $M_i := \sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}}$ is the number of clients that have access to arm i. We let $\mu_i = \frac{1}{M_i}\sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}} \mu_{i,m}$ denote the mean reward of arm i. Arm i is said to be the *best arm* of client m if it has the largest mean reward among all the arms in S_m. We assume that each client has a *single* best arm, and we let $a_m^* := \arg\max_{i\in S_m} \mu_i$ denote the best arm of client m. We let $a^* := (a_m^*)_{m\in[M]} \in S_1 \times S_2 \times \ldots \times S_M$ denote the tuple of best arms. More explicitly, we write $a^*(v)$ to denote the tuple of best arms under the problem instance v, and let $\mathcal{P}$ be the set of the all problem instances with a single best arm at each client.

We assume that the clients and the central server are time-synchronised and that the clients communicate with the server at certain pre-defined time instants. Given $(M, K, \{S_m\}_{m=1}^{M})$, a problem instance v, and a confidence level $\delta \in (0, 1)$, we wish to design an algorithm for finding the best arm of each client with (a) the fewest number of time steps and communication rounds, and (b) error probability less than δ. By an algorithm, we mean a tuple $\Pi = (\Pi_{\mathrm{comm}}, \Pi_{\mathrm{cli}}, \Pi_{\mathrm{svr}})$ of (a) a strategy for *communication* between the clients and the server, (b) a strategy for selection of arms at each *client*, and (c) a combined stopping and recommendation rule at the *server*. The communication strategy Π_{comm} consists of the following components: (a) $\{b_r\}_{r\in\mathbb{N}}$: the time instants of communication, with $b_r \in \mathbb{N}$ and $b_r \leq b_{r+1}$ for all $r \in \mathbb{N}$, (b) Σ: the set of values transmitted from the server to each client, (c) Φ: the set of values transmitted from each client to the server; this is assumed to be identical for all the clients, (d) $\hbar_r : \Phi^{Mr} \to \Sigma$: a function deployed at the server, which aggregates the information transmitted from all the clients in the communication rounds $1, \ldots, r$, and generates an output value to be transmitted to each client, and (e) $\rho_r^m : (\mathbb{R} \times S_m)^{b_r} \to \Phi$: a function deployed at client $m \in [M]$, which aggregates the observations seen by client m in the time instants $1, \ldots, b_r$ from the arms in S_m, and generates an output value to be transmitted to the server.

The arms selection strategy Π_{cli} consists of component arm selection rules $\pi_t^m : (\mathbb{R} \times S_m)^t \times \Sigma \to S_m$, $m \in [M]$. Here, π_t^m takes as input the observations seen from the arms in S_m pulled by client m up to time t and the information received from the server to decide which arm in S_m to pull at time $t + 1$. Lastly, the stopping and recommendation strategy Π_{svr} at the server consists of the following components: (a) the stopping rule $\Upsilon_r : \Phi^{Mr} \to \{0, 1\}$ that decides whether the algorithm stops in the r-th communication round and (b) the recommendation rule $\Psi_r : \Phi^{Mr} \to S_1 \times S_2 \times \cdots \times S_M$ to output the empirical best arm of each client if the algorithm stops in the rth communication round. We let $\hat{a}_{\delta,m}$ denote the empirical best arm of client m output by the algorithm under confidence level δ, and define $\hat{a}_\delta := (\hat{a}_{\delta,m})_{m\in[M]}$.

We assume that all the functions defined above are Borel-measurable. Note that if an algorithm stops in the rth communication round, then its stopping time $\tau = b_r$. Given $\delta \in (0, 1)$, we say that an algorithm Π is *δ-probably approximately correct* (or δ-PAC) if $\mathbb{P}_v^{\Pi}(\tau_\delta < +\infty) = 1$ and $\mathbb{P}_v^{\Pi}(\hat{a}_\delta \neq a^*(v)) \leq \delta$ for any problem instance $v \in \mathcal{P}$; here, $\mathbb{P}_v^{\Pi}(\cdot)$ the probability measure induced by the algorithm Π and the problem instance v. Writing $\tau_\delta(\Pi)$ and $\mathfrak{r}_\delta(\Pi)$ to denote respectively the stopping time and the associated number of communication rounds corresponding to the confidence level δ under the algorithm Π, our interest is in the following optimisation problems:

$$\inf_{\Pi \text{ is } \delta\text{-PAC}} \mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)], \qquad \inf_{\Pi \text{ is } \delta\text{-PAC}} \mathbb{E}_v^{\Pi}[\mathfrak{r}_\delta(\Pi)]. \tag{1}$$

In (1), $\mathbb{E}_v^{\Pi}$ denotes expectation with respect to the measure $\mathbb{P}_v^{\Pi}$. Prior works [14], [15] show that the first term in (1) grows as $\Theta(\log(1/\delta))$ as $\delta \to 0$. We anticipate that a similar growth rate holds for our problem setting. Our objective is to precisely characterise

$$\liminf_{\delta\to 0} \inf_{\Pi \text{ is } \delta\text{-PAC}} \frac{\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)]}{\log(1/\delta)}. \tag{2}$$

In the following section, we present a lower bound for (2). Furthermore, we demonstrate that on any sequence of problem instances $\{v^{(l)}\}_{l=1}^{\infty}$ with increasing levels of "hardness" (to be made precise soon), the second term in (1) grows as $\Theta(\log\log(1/\delta))$ and obtain a precise characterisation of this growth rate, the first-of-its-kind result in the literature to the best of our knowledge.

III. Lower Bound: Converse

Below, we first derive a problem-instance specific asymptotic lower bound on the expected stopping time. Then, we present a simplification to the constant appearing in the lower bound and provide the explicit structure of its optimal solution. Next, we show that for any algorithm to be almost-optimal (in the sense to be made precise later in this section), the ratio of any two consecutive communication time instants must be bounded, a result that may be of independent interest. Using this result, we obtain an $O(\log\log(1/\delta))$ lower bound on the expected number of communication rounds for a sub-class of δ-PAC algorithms.

A. Lower Bound on the Expected Stopping Time

Let $\mathrm{Alt}(v) := \{v' \in \mathcal{P} : a^*(v) \neq a^*(v')\}$ denote the set of alternative problem instances corresponding to the problem instance v. Let Λ denote the simplex of probability distributions on K variables, and let $\Lambda_m := \{\alpha \in \Lambda : \alpha_i = 0 \ \forall\, i \notin S_m\}$ denote the subset of Λ corresponding to client $m \in [M]$. We write $\Gamma := \Lambda_1 \times \cdots \times \Lambda_M$ to denote the Cartesian product of $\{\Lambda_m\}_{m=1}^M$. The following proposition presents the first main result of this paper.

Proposition 1: For any $v \in \mathcal{P}$ and $\delta \in (0,1)$,

$$\inf_{\Pi \text{ is } \delta\text{-PAC}} \mathbb{E}_v^\Pi[\tau_\delta(\Pi)] \geq c^*(v) \log\left(\frac{1}{4\delta}\right), \tag{3}$$

where the constant $c^*(v)$ is given by

$$c^*(v)^{-1} = \max_{\omega\in\Gamma} \inf_{v'\in\mathrm{Alt}(v)} \sum_{m=1}^M \sum_{i\in S_m} \omega_{i,m} \frac{\big(\mu_{i,m}(v) - \mu_{i,m}(v')\big)^2}{2}. \tag{4}$$

Dividing both sides by $\log(1/\delta)$ and letting $\delta \to 0$ in (3),

$$\liminf_{\delta\to 0} \inf_{\Pi \text{ is } \delta\text{-PAC}} \frac{\mathbb{E}_v^\Pi[\tau_\delta(\Pi)]}{\log(1/\delta)} \geq c^*(v).$$

The term $c^*(v)$ defined in (4) is a measure of the "hardness" of the instance v and is the solution to a max-inf optimisation problem where the outer 'max' is over all M-ary probability distributions $\omega \in \Gamma$ such that $\sum_{i\in S_m} \omega_{i,m} = 1$ for all $m \in [M]$ (here, $\omega_{i,m}$ is the probability of pulling arm i of client m), and the inner 'inf' is over the set of alternative problem instances corresponding to the instance v. The proof of Proposition 1 is similar to the proof of [13, Theorem 1] and is omitted for brevity. The key ideas in the proof to note are (a) the transportation lemma of [15, Lemma 1] relating the error probability to the expected number of arm pulls and the Kullback–Leibler divergence between two problem instances v and $v' \in \mathrm{Alt}(v)$ with distinct best arm locations, and (b) Wald's identity for i.i.d. observations.

B. A Simplification

A close examination of the proof of the lower bound in [13] reveals that an important step in the proof therein is a further simplification of the max-inf optimisation in the instance-dependent constant; see [13, Theorem 5]. However, an analogous simplification of (4) is not possible as $\mathrm{Alt}(v)$ does not admit a closed-form expression.

Nevertheless, we propose the following simplification. For any $\omega \in \Gamma$ and instance v, let

$$g_v(\omega) := \inf_{v'\in\mathrm{Alt}(v)} \sum_{m=1}^M \sum_{i\in S_m} \omega_{i,m} \frac{(\mu_{i,m}(v) - \mu_{i,m}(v'))^2}{2} \tag{5}$$

denote the inner minimum in (4). Our simplification of (5) is given by

$$\widetilde{g}_v(\omega) := \min_{i\in[K]} \frac{\Delta_i^2(v)/2}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}} \frac{1}{\omega_{i,m}}}, \tag{6}$$

where for each $i \in [K]$,

$$\Delta_i(v) := \min_{m\in[M]:\, i\in S_m} \Big|\mu_i(v) - \max_{j\in S_m\setminus\{i\}} \mu_j(v)\Big|.$$

In particular, if $\omega_{i,m} = 0$ for some $m \in [K]$ and $i \in S_m$, then $\widetilde{g}_v(\omega) = 0$.

Notice that the infimum in (5) is over the *uncountably infinite* set $\mathrm{Alt}(v)$, whereas the simplified minimum in (6) is over the *finite* set $[K]$. Our next result shows that these two terms differ at most by a factor of 2.

Lemma 2: For $v \in \mathcal{P}$ and $\omega \in \Gamma$, let $g_v(\omega)$ and $\widetilde{g}_v(\omega)$ be as defined in (5) and (6) respectively. Then, $\frac{1}{2}\widetilde{g}_v(\omega) \leq g_v(\omega) \leq \widetilde{g}_v(\omega)$.

As a consequence of Lemma 2, it follows that $\widetilde{c}(v) := \max_{\omega\in\Gamma} \widetilde{g}_v(\omega)$ and $c^*(v) = \max_{\omega\in\Gamma} g_v(\omega)$ differ only by a multiplicative factor of 2. It is not clear if the optimiser of $c^*(v)$, if any, can be computed analytically. On the other hand, as we shall soon see, the optimiser of $\widetilde{c}(v)$ can be computed in closed-form and plays an important role in the design of an asymptotically almost-optimal algorithm.

Definition 3 (Balanced Condition): An $\omega \in \Gamma$ satisfies *balanced condition* if

$$\frac{\omega_{i_1,m_1}}{\omega_{i_2,m_1}} = \frac{\omega_{i_1,m_2}}{\omega_{i_2,m_2}}$$

for all $i_1, i_2 \in S_{m_1} \cap S_{m_2}$ and $m_1, m_2 \in [M]$.

That is, ω satisfies *balanced condition* if the ratios of the arm selection probabilities are consistent (or *balanced*) across the clients. The next result shows that $\max_{\omega\in\Gamma} \widetilde{g}_v(\omega)$ admits a solution that satisfies *balanced condition*.

Proposition 4: Given $v \in \mathcal{P}$, there exists $\omega \in \Gamma$ that attains the maximum in the expression for $\widetilde{c}(v)$ and satisfies the balanced condition.

Proposition 4 follows in a straightforward manner from a more general result, namely Theorem 14, which we state later in the paper.

Corollary 5: Let $\widetilde{\omega}(v) \in \Gamma$ be any M-ary probability distribution that attains the maximum in the expression for $\widetilde{c}(v)$ and satisfies balanced condition. Then, there exists a K-dimensional vector $G(v) = [G(v)_i]_{i\in[K]}$ such that

$$\widetilde{\omega}(v)_{i,m} = \frac{G(v)_i}{\sum_{i'\in S_m} G(v)_{i'}}, \quad i \in S_m,\ m \in [M].$$

Corollary 5 elucidates the rather simple form of the optimiser of $\widetilde{c}(v)$, one that is characterised by a K-dimensional vector $G(v)$ which, in the sequel, shall be referred to as the *global vector* corresponding to the instance v. We shall soon see that it plays an important role in the design of an almost-optimal algorithm for finding the best arms of the clients. In fact, we show that in order to inform each client of its arm selection probabilities, the server needs to broadcast only the global vector instead of sending a separate probability vector to each client, thereby leading to significantly less downlink network traffic, especially when M is large. For example, using a broadcast-type protocol instead of a unicast-type protocol such as user datagram protocol (UDP) for transmitting data from the server to clients over the internet is known to reduce the network traffic significantly [31, Chapter 20].

C. Lower Bound on the Expected Number of Communication Rounds

In this section, we present a lower bound on the expected number of communication rounds required by any "good" algorithm to find the best arms of the clients. By "good" algorithms, we mean the class of all *almost-optimal* δ-PAC algorithms as defined below.

Definition 6 (Almost-Optimal Algorithm): Given $\delta \in (0,1)$, and $\alpha \geq 1$, a δ-PAC algorithm Π is said to be *almost-optimal* up to a constant α if

$$\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)] \leq \alpha\, c^*(v) \log\left(\frac{1}{4\delta}\right) \qquad \forall v \in \mathcal{P}. \tag{7}$$

In addition, Π is said to be almost-optimal if it is almost-optimal up to a constant α for some $\alpha \geq 1$.

Definition 6 implies that the expected stopping time of an almost-optimal algorithm matches the lower bound in (3) up to the multiplicative constant α. Notice that the sparser (more infrequent) the communication between the clients and the server, the larger the time required to find the best arms of the clients. Because (7) implies that the expected stopping time of an almost-optimal algorithm cannot be infinitely large, it is natural to ask what is the sparsest level of communication achievable in the class of almost-optimal algorithms. The next result provides a concrete answer to this question.

Theorem 7: Fix $\delta \in (0, \frac{1}{4})$ and a δ-PAC algorithm Π with communication time instants $\{b_r\}_{r\in\mathbb{N}}$. If Π is almost-optimal, then $\sup_{r\in\mathbb{N}} \frac{b_{r+1}}{b_r} < +\infty$.

Theorem 7, one of the key results of this paper and of independent interest, asserts that the ratio of any two consecutive communication time instants of an almost-optimal algorithm *must be bounded.* An important implication of Theorem 7 is that an almost-optimal algorithm can communicate at most exponentially sparsely, i.e., at exponential time instants of the form $t = \lceil (1+\lambda)^r \rceil$, $r \in \mathbb{N}$, for some $\lambda > 0$. For instance, an algorithm that communicates at time instants that grow *super-exponentially* (i.e., $t = 2^{\kappa(r)}$ for any super-linear function $\kappa(r)$), does not satisfy the requirement in Theorem 7, and hence cannot be almost-optimal. In this sense, the class of all exponentially sparsely communicating algorithms (with different exponents) forms the *boundary* of the class of all almost-optimal algorithms.

The proof of Theorem 7 suggests that when an almost-optimal algorithm Π stops at time step $\tau_\delta(\Pi)$ and $\sup_{r\in\mathbb{N}} \frac{b_{r+1}}{b_r} \leq \eta$, at least $\Omega(\log_\eta(\tau_\delta(\Pi)))$ communication *rounds* must have occurred, i.e., $\mathfrak{r}_\delta(\Pi) = \Omega(\log_\eta(\tau_\delta(\Pi)))$ almost surely (a.s.). The next result relates $\log_\eta(\tau_\delta(\Pi))$ with $\log_\eta(\mathbb{E}[\tau_\delta(\Pi)])$.

Lemma 8: Let $\{v^{(l)}\}_{l=1}^{\infty} \subset \mathcal{P}$ be any sequence of problem instances with $\lim_{l\to\infty} c^*(v^{(l)}) = +\infty$. Given $\delta \in \left(0, \frac{1}{4}\right)$, for any almost-optimal algorithm Π and $\beta \in (0,1)$,

$$\liminf_{l\to\infty} \mathbb{P}_{v^{(l)}}^{\Pi} \left(\log\left(\tau_\delta(\Pi)\right) > \beta \log\left(\mathbb{E}_{v^{(l)}}^{\Pi}[\tau_\delta(\Pi)]\right)\right) \geq \frac{1}{4} - \delta.$$

Lemma 8 shows that $\log(\tau_\delta(\Pi)) = \Omega(\log(\mathbb{E}_{v^{(l)}}^{\Pi}[\tau_\delta(\Pi)])$ with a non-vanishing probability on a sequence of problem instances $v^{(l)}$ with increasing hardness levels. Proposition 1 implies that $\mathbb{E}_{v^{(l)}}^{\Pi}[\tau_\delta(\Pi)] = \Omega(\log(1/\delta))$, which in conjunction with Lemma 8 and the relation $\mathfrak{r}_\delta(\Pi) = \Omega(\log_\eta(\tau_\delta(\Pi)))$ a.s., yields $\mathfrak{r}_\delta(\Pi) = \Omega(\log_\eta \log(1/\delta))$ a.s., and consequently $\mathbb{E}[\mathfrak{r}_\delta(\Pi)] = \Omega(\log_\eta \log(1/\delta))$. The next result makes this heuristic precise.

Theorem 9: Fix $\{v^{(l)}\}_{l=1}^{\infty} \subset \mathcal{P}$ with $\lim_{l\to\infty} c^*(v^{(l)}) = +\infty$. Fix $\delta \in (0, \frac{1}{4})$. For any almost-optimal algorithm Π with communication time instants $\{b_r\}_{r\in\mathbb{N}}$ satisfying $\frac{b_{r+1}}{b_r} \leq \eta$ for all $r \in \mathbb{N}$,

$$\liminf_{l\to\infty} \frac{\mathbb{E}_{v^{(l)}}^{\Pi}[\mathfrak{r}_\delta(\Pi)]}{\log_\eta\left(c^*(v^{(l)}) \log\left(\frac{1}{4\delta}\right)\right)} \geq \frac{1}{4} - \delta.$$

Theorem 9 is the analogue of Proposition 1 for the number of communication rounds, and is the first-of-its-kind result to the best of our knowledge.

Remark 1: A natural desideratum in the lower bound on the expected number of communication rounds would be that it depends explicitly on α for δ-PAC algorithms Π that are almost optimal up to a constant α (the multiplicative gap from the lower bound $c^*(v)\ \log\left(\frac{1}{\delta}\right)$ as per Definition 6). However, we see that Theorem 9 is expressed in terms of η, a bound on the ratio between successive communication rounds $\frac{b_{r+1}}{b_r}$. Intuitively, it should hold that η is monotonically increasing in α. However, our proof strategies to establish the lower bounds in Theorems 7 and 9 are not amenable to elucidate the explicit dependence of $\mathbb{E}_v^{\Pi}[\mathfrak{r}_\delta(\Pi)]$ on α for algorithms Π that are asymptotically optimal up to constant α. We will see, however, that our algorithm HET-TS(λ) to be introduced in the next section makes this dependence explicit; see Theorem 11 where α is roughly 2η.

IV. THE HETEROGENEOUS TRACK-AND-STOP (HET-TS(λ)) ALGORITHM

In this section, we propose an algorithm for finding the best arms of the clients based on the well-known *Track-and-Stop* strategy [13], [15] that communicates exponentially sparsely. Known as Heterogeneous Track-and-Stop and abbreviated as HET-TS(λ) for an input parameter $\lambda > 0$, the individual components of our algorithm are described in detail below.

A. Communication Strategy

We set $b_r = \lceil (1+\lambda)^r \rceil$, $r \in \mathbb{N}$. In the rth communication round, each client sends to the server the empirical means of the observations seen from its arms up to time b_r. Note that

$$\hat{\mu}_{i,m}(t) := \frac{1}{N_{i,m}(t)} \sum_{s=1}^{t} \mathbf{1}_{\{A_m(s)=i\}} X_{i,m}(s) \tag{8}$$

is the empirical mean of $i \in S_m$ after t time instants, where $\hat{\mu}_{i,m}(t) = 0$ if $N_{i,m}(t) = 0$. In (8), $A_m(t)$ is the arm pulled by client m at time t, and $N_{i,m}(t) := \sum_{s=1}^{t} \mathbf{1}_{\{A_m(s)=i\}}$ is the number of times arm i of client m was pulled up to time t. On the downlink, for each $t \in \{b_r\}_{r\in\mathbb{N}}$, the server first computes the global vector $G(\hat{v}(t))$ according to the procedure outlined in Section VI and broadcasts this vector to each client. Here, $\hat{v}(t)$ is the empirical problem instance at time t, defined by the empirical arm means $\{\hat{\mu}_{i,m}(t) : i \in S_m, m \in [M]\}$ received from the clients. In particular, $G(\hat{v}(t)) = \mathbf{1}^K$ if $\hat{v}(t) \notin \mathcal{P}$, where $\mathbf{1}^K$ denotes the all-ones vector of length K.

B. Sampling Strategy at Each Client

We use a variant of the so-called *D-tracking* rule of [13] for pulling the arms at each of the clients. Accordingly, at any time t, client $m \in [M]$ first computes

$$\hat{\omega}_{i,m}(t) := \frac{G(\hat{v}(b_{r(t)}))_i}{\sum_{i' \in S_m} G(\hat{v}(b_{r(t)}))_{i'}}, \quad i \in S_m, \tag{9}$$

based on the global vector received from the server in the most recent communication round $r(t) := \min\{r \in \mathbb{N} : b_r \geq t\} - 1$ (with $b_0 := 0$), and subsequently pulls arm

$$A_m(t) \in \begin{cases} \arg\min\limits_{i\in S_m} N_{i,m}(t-1), & \min\limits_{i\in S_m} N_{i,m}(t-1) < \sqrt{\dfrac{t-1}{|S_m|}}, \\ \arg\min\limits_{i\in S_m} N_{i,m}(t-1) - t\,\hat{\omega}_{i,m}(t), & \text{otherwise}. \end{cases} \tag{10}$$

Ties, if any, are resolved uniformly at random. Notice that the rule in (10) ensures that in the long run, each arm is pulled at least $O(\sqrt{t})$ many times after t time instants.

C. Stopping and Recommendation Rules at the Server

We use a version of *Chernoff's stopping rule* at the server, as outlined below. Let

$$Z(t) := \inf_{v' \in \mathrm{Alt}(\hat{v}(t))} \sum_{m=1}^{M} \sum_{i\in S_m} N_{i,m}(t)\, \frac{(\mu'_{i,m} - \hat{\mu}_{i,m}(t))^2}{2},$$

where $\hat{v}(t)$ is the empirical problem instance at time t, defined by the empirical means $\{\hat{\mu}_{i,m}(t) : i \in S_m, m \in [M]\}$ received from the clients, and v' is defined by the means $\{\mu'_{i,m} : i \in S_m, m \in [M]\}$. Then, the (random) stopping time of the algorithm is defined as

$$\tau_\delta(\Pi_{\text{Het-TS}}) = \min\{t \in \{b_r\}_{r\in\mathbb{N}} : Z(t) > \beta(t,\delta), t \geq K\}, \tag{11}$$

where $\beta(t,\delta) = K' \log(t^2 + t) + f^{-1}(\delta)$, with $K' = \sum_{m=1}^{M} |S_m|$ and $f : (0, +\infty) \to (0,1)$ defined as

$$f(x) := \sum_{i=1}^{K'} \frac{x^{i-1} e^{-x}}{(i-1)!}, \quad x \in (0, +\infty). \tag{12}$$

Algorithm 1 HET-TS(λ): At Client $m \in [M]$

```
Require:
    $\delta \in (0,1)$: confidence level.
    $\lambda > 0$: communication frequency parameter.
    $\{b_r = \lceil (1+\lambda)^r \rceil : r \in \mathbb{N}\}$: communication instants.
    $S_m$: arms subset of the client.
Ensure: $\hat{a}_{\delta,m}$: the best arm in $S_m$.
Initialise $G \leftarrow \mathbf{1}^K$
for $t \in \{1, 2, \ldots\}$ do
    Compute $\{\hat{\omega}_{i,m}(t) : i \in S_m\}$ via (9).
    if $\min_{i\in S_m} N_{i,m}(t-1) < \sqrt{(t-1)/|S_m|}$ then
        Pull arm $A_m(t) \in \arg\min_{i\in S_m} N_{i,m}(t-1)$; resolve ties uniformly.
    else
        Pull arm $A_m(t) \in \arg\min_{i\in S_m} N_{i,m}(t-1) - t\,\hat{\omega}_{i,m}(t)$; resolve ties uniformly.
    end if
    Update the empirical means $\{\hat{\mu}_{i,m}(t) : i \in S_m\}$.
    if $t \in \{b_r : r \in \mathbb{N}\}$ then
        Send the empirical means $\{\hat{\mu}_{i,m}(t) : i \in S_m\}$ from client $m$ to the server.
        $G \leftarrow$ latest global vector received from the server.
    end if
    if Server signals to stop further arm pulls then
        Receive best arm $\hat{a}_{\delta,m}$ from the server.
        Break.
    end if
end for
return Best arm $\hat{a}_{\delta,m}$.
```

Our algorithm outputs $\hat{a}_{\delta,m} = \arg\max_{i\in S_m} \hat{\mu}_i(\tau_\delta)$ as the best arm of client $m \in [M]$, where $\hat{\mu}_i(\tau_\delta) = \frac{1}{M_i} \sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}}\, \hat{\mu}_{i,m}(\tau_\delta)$.

Remark 2: Notice that $\Pi_{\text{Het-TS}}$ only stops at the communication time instants $\{b_r\}_{r\in\mathbb{N}}$. This is evident from (11).

Remark 3: Our choice of f in (12) ensures that the map $x \mapsto f(x)$ is strictly monotone and continuous, and therefore admits an inverse, say $f^{-1}(\cdot)$. Furthermore, an important property of $f^{-1}(\cdot)$ that results from the careful construction of f as in (12) is that $\lim_{\delta\to 0} \frac{\log(1/\delta)}{f^{-1}(\delta)} = 1$. See Lemma 22 in the appendices for the proof.

The pseudo-code of HET-TS(λ) is presented in Algorithm 1 (for each client $m \in [M]$) and Algorithm 2 (for the server).

V. RESULTS ON THE PERFORMANCE OF HET-TS(λ)

In this section, we state the results on the performance of HET-TS(λ) which we denote alternatively by $\Pi_{\text{HET-TS}}$ (the input parameter λ being implicit). The first result below asserts that $\Pi_{\text{HET-TS}}$ is δ-PAC for any $\delta \in (0,1)$.

Theorem 10: HET-TS(λ) is δ-PAC for each $\delta \in (0,1)$.

The next result provides an asymptotic upper bound on the expected stopping time of HET-TS(λ) (or $\Pi_{\text{HET-TS}}$).

Theorem 11: Fix $\lambda > 0$, and let $b_r = \lceil (1+\lambda)^r \rceil$, $r \in \mathbb{N}$. Given any $v \in \mathcal{P}$ and $\delta \in (0,1)$, $\tau_\delta(\Pi_{\text{HET-TS}})$ satisfies

$$\mathbb{P}_v^{\Pi_{\text{HET-TS}}}\left(\limsup_{\delta\to 0} \frac{\tau_\delta(\Pi_{\text{HET-TS}})}{\log\left(\frac{1}{\delta}\right)} \leq 2\,(1+\lambda)\, c^*(v) \right) = 1.$$

Algorithm 2 HET-TS: At Central Server

```
Require:
    $\delta \in (0,1)$: confidence level.
    $\lambda > 0$: communication frequency parameter.
    $\{b_r = \lceil (1+\lambda)^r \rceil : r \in \mathbb{N}\}$: communication instants.
    $S_1, \ldots, S_M$: sets of clients' accessible arms.
Initialize $G \leftarrow \mathbf{1}^K$
for $t \in \{b_r : r \in \mathbb{N}\}$ do
    for $m \in [M]$ do
        Receive $\{\hat{\mu}_{i,m}(t) : i \in S_m\}$ from client $m$.
    end for
    if $t \geq K$ and $Z(t) > \beta(t,\delta)$ then
        for $m \in [M]$ do
            Compute the empirical best arm $\hat{a}_{\delta,m}$.
            Send $\hat{a}_{\delta,m}$ and signal of stop to client $m$.
        end for
        Break.
    end if
    Compute the global vector $G$ via the empirical means
        $\{\hat{\mu}_{i,m}(t) : i \in S_m,\, m \in [M]\}$.
    Broadcast vector $G$ to all the clients.
end for
```

Furthermore, $\mathbb{E}_v^{\Pi_{\text{HET-TS}}}[\tau_\delta(\Pi_{\text{HET-TS}})]$ satisfies

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_v^{\Pi_{\text{HET-TS}}}[\tau_\delta(\Pi_{\text{HET-TS}})]}{\log\left(\frac{1}{\delta}\right)} \leq 2\,(1+\lambda)\, c^*(v).$$

Thus, in the limit as $\delta \to 0$, $\Pi_{\text{HET-TS}}$ is asymptotically almost-optimal up to the constant $\alpha = 2\,(1+\lambda)$.

Theorem 11 lucidly demonstrates the trade-off between the frequency of communication, which is parameterized by λ, and the expected stopping time. Because HET-TS(λ) communicates at time instances $b_r = \lceil (1+\lambda)^r \rceil$, as λ increases, communication occurs with lesser frequency. This, however, leads to an increase in the multiplicative gap to asymptotic optimality, $2\,(1+\lambda)$. The factor $1+\lambda$ arises due to the necessity of communicating at time instances whose ratios $\frac{b_{r+1}}{b_r}$ are bounded; see Lemma 7. The other factor 2 (in $2\,(1+\lambda)$) arises from approximating $g_v(\omega)$ by $\widetilde{g}_v(\omega)$ in Lemma 2. This factor is required to ensure that the optimal solution to $\widetilde{c}(v)$ and the arm selection probabilities at each time instant can be evaluated in a tractable fashion.

Corollary 12: Fix $\lambda > 0$, and let $b_r = \lceil (1+\lambda)^r \rceil$, $r \in \mathbb{N}$. Given any $v \in \mathcal{P}$ and $\delta \in (0,1)$, $\mathfrak{r}_\delta(\Pi_{\text{HET-TS}})$ satisfies

$$\mathbb{P}_v^{\Pi_{\text{HET-TS}}}\left(\limsup_{\delta \to 0} \frac{\mathfrak{r}_\delta(\Pi_{\text{HET-TS}})}{\log_{1+\lambda}\left(\log\left(\frac{1}{\delta}\right) c^*(v)\right)} \leq 1\right) = 1.$$

Furthermore, $\mathbb{E}_v^{\Pi_{\text{HET-TS}}}[\mathfrak{r}_\delta(\Pi_{\text{HET-TS}})]$ satisfies

$$\limsup_{\delta \to 0} \frac{\mathbb{E}_v^{\Pi_{\text{HET-TS}}}[\mathfrak{r}_\delta(\Pi_{\text{HET-TS}})]}{\log_{1+\lambda}\left(\log\left(\frac{1}{\delta}\right) c^*(v)\right)} \leq 1.$$

In contrast to Theorem 11, Corollary 12 is a statement concerning the number of communication *rounds*. It says that the expectation of this quantity scales as $O(\log\log(1/\delta))$. This is perhaps unsurprising given that HET-TS(λ) communicates at time instants $b_r = \lceil (1+\lambda)^r \rceil$, $r \in \mathbb{N}$.

VI. Solving the Optimal Allocation

Recall from Section III-B that given any problem instance $v \in \mathcal{P}$, the optimal solution to $\max_{\omega \in \Gamma}\ \widetilde{g}_v(\omega)$ may be characterised by a K-dimensional global vector $G(v)$ (see Corollary 5 for more details). In this section, we provide the details on how to efficiently compute the global vector $G(v)$ corresponding to any problem instance $v \in \mathcal{P}$.

Consider the *relation* $R := \{(i_1, i_2) :\ \exists m \in [M], i_1, i_2 \in S_m\}$ on the arms. Let R_{e} be the equivalence relation generated by R, i.e., the smallest equivalence relation containing R. Clearly, the above equivalence relation R_{e} partitions $[K]$ into equivalence classes. Let $Q_1, \ldots, Q_L$ be the equivalence classes. For any $j \in [L]$, let

$$\widetilde{g}_v^{(j)}(\omega) := \min_{i \in Q_j} \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^{M} \mathbf{1}_{\{i \in S_m\}} \frac{1}{\omega_{i,m}}}.$$

We define $\widetilde{g}_v^{(j)}(\omega) = 0$ if there exists $m \in [M]$ and $i \in S_m \cap Q_j$ such that $\omega_{i,m} = 0$. In Eqn. (43) in Appendix F, we argue that the following optimisation problems admit a common solution:

$$\max_{\omega \in \Gamma} \widetilde{g}_v(\omega), \quad \left\{\max_{\omega \in \Gamma} \widetilde{g}_v^{(l)}(\omega), \quad l = 1, \ldots, L\right\}. \tag{13}$$

Definition 13 (Pseudo-Balanced Condition): Fix $v \in \mathcal{P}$. An $\omega \in \Gamma$ satisfies the *pseudo-balanced condition* if

$$\frac{\Delta_{i_1}^2(v)}{\frac{1}{M_{i_1}^2}\sum_{m=1}^{M} \frac{\mathbf{1}_{\{i_1 \in S_m\}}}{\omega_{i_1,m}}} = \frac{\Delta_{i_2}^2(v)}{\frac{1}{M_{i_2}^2}\sum_{m=1}^{M} \frac{\mathbf{1}_{\{i_2 \in S_m\}}}{\omega_{i_2,m}}}$$

for all $j \in [L]$ and $i_1, i_2 \in Q_j$.

The next result states that the common solution to (13) satisfies the *balanced* and *pseudo-balanced conditions*.

Theorem 14: For any $v \in \mathcal{P}$, the common solution to the optimization problems in (13) is unique and satisfies the balanced condition (Def. 3) and the pseudo-balanced condition. Let $\widetilde{w}(v)$ be the unique common solution to (13) corresponding to the instance v. Let $G(v)$ be the unique global vector characterising $\widetilde{w}(v)$ (see Corollary 5) with $G(v) > 0$ and

$$\|G^{(j)}(v)\|_2 = 1 \quad \forall j \in [L], \tag{14}$$

where $G^{(j)}(v) \in \mathbb{R}^{|Q_j|}$ denotes the sub-vector of $G(v)$ formed from the rows corresponding to the arms $i \in Q_j$. Let $H(v) \in \mathbb{R}^{K \times K}$ be the matrix defined by

$$H(v)_{i_1,i_2} := \frac{1}{\Delta_{i_1}^2(v) M_{i_1}^2} \sum_{m=1}^{M} \mathbf{1}_{\{i_1, i_2 \in S_m\}}, \ \ i_1, i_2 \in [K].$$

For $j \in [L]$, let $H^{(j)}(v) \in \mathbb{R}^{|Q_j| \times |Q_j|}$ be the sub-matrix of $H(v)$ formed from the rows and columns corresponding to the arms in Q_j. It is easy to verify that $\left(H^{(j)}(v)\right)^\top H^{(j)}(v) = H^{(j)}(v)\left(H^{(j)}(v)\right)^\top$. That is, $H^{(j)}(v)$ is a *normal matrix* [32, Chapter 2, Section 2.5] and therefore has $|Q_j|$ linearly independent eigenvectors. In Appendix H, we show that $G^{(j)}(v)$ is an eigenvector of the matrix $H^{(j)}(v)$ and that the eigenspace of $H^{(j)}(v)$ is one-dimensional. Building on these results, the main result of this section, a recipe for computing the global vector corresponding to an instance, is given below.

Proposition 15: Fix $j \in [L]$ and a problem instance $v \in \mathcal{P}$. Among any set of $|Q_j|$ linearly independent eigenvectors of $H^{(j)}(v)$, there exists only one vector $\mathbf{u}$ whose elements are all negative ($\mathbf{u} < \mathbf{0}$) or all positive ($\mathbf{u} > \mathbf{0}$). Furthermore,

$$G^{(j)}(v) = \begin{cases} -\frac{\mathbf{u}}{\|\mathbf{u}\|_2}, & \text{if } \mathbf{u} < \mathbf{0}, \\ \frac{\mathbf{u}}{\|\mathbf{u}\|_2}, & \text{if } \mathbf{u} > \mathbf{0}. \end{cases} \tag{15}$$

Proposition 15 provides an efficient recipe to compute the global vector $G(v)$: it is the unique eigenvector of $H^{(j)}(v)$ with either all-positive or all-negative entries and normalised to have unit norm. We use this recipe in our implementation of HET-TS(λ) on synthetic and real-world datasets (e.g., the MovieLens dataset [33]).

VII. Experimental Results

In this section, we corroborate our theoretical results by implementing HET-TS(λ) and performing a variety of experiments on a synthetic dataset and the MovieLens dataset.

A. Synthetic Dataset

In our synthetic dataset, the instance we used contains $M = 5$ clients and $K = 5$ arms. The expected mean reward $\mu_{i,m}$ is chosen uniformly at random from $[7-i, 7-i+1]$. As a consequence, μ_i is also uniformly random in $[7-i, 7-i+1]$.

To empirically evaluate the effect of various sets $\{S_m\}_{m\in[M]}$ on the expected stopping time, we consider different *overlap patterns* (or *multisets*) of the form $\mathcal{O}^{(p)} = \{S_1^{(p)}, S_2^{(p)}, \ldots, S_5^{(p)}\}$, where

$$\begin{aligned} \mathcal{O}^{(1)} &= \{\{1,2\},\{2,3\},\{3,4\},\{4,5\},\{5,1\}\}, \\ \mathcal{O}^{(2)} &= \{\{1,2,3\},\{2,3,4\},\{3,4,5\},\{4,5,1\},\{5,1,2\}\} \\ \mathcal{O}^{(3)} &= \{\{1,2,3,4\},\{2,3,4,5\},\{3,4,5,1\}, \\ &\qquad \{4,5,1,2\},\{5,1,2,3\}\}, \quad \text{and} \\ \mathcal{O}^{(4)} &= \{\{1,2,3,4,5\},\{1,2,3,4,5\},\{1,2,3,4,5\}, \\ &\qquad \{1,2,3,4,5\},\{1,2,3,4,5\}\}. \end{aligned} \tag{16}$$

Thus, the larger the index of the overlap pattern p, the larger the overlap among the sets $\{S_m^{(p)}\}_{m\in[M]}$, and therefore the larger the number of clients that have access to a fixed arm $i \in [K]$. The mean values $\{\mu_{i,m}\}_{i\in[K],m\in[M]}$, together with an overlap pattern $\mathcal{O}^{(p)}$, uniquely defines a problem instance $v = (\{\mu_{i,1}\}_{i\in S_1^{(p)}}, \{\mu_{i,2}\}_{i\in S_2^{(p)}}, \ldots, \{\mu_{i,M}\}_{i\in S_M^{(p)}})$.

1) Effect of Amount of Overlap: The empirical expected stopping times of HET-TS(λ) for $\lambda = 0.01$ are displayed in Fig. 2. It can be seen that as δ decreases, the empirical stopping time increases, as expected. More interestingly, note that for a fixed δ, the stopping time is *not monotone* in the amount of overlap. This is due to two factors that work in opposite directions as one increases the amount of overlap of S_m's among various clients. On the one hand, each client has access to more arms, yielding more information about the bandit instance for the client. On the other hand, with more arms, the set of arms that can potentially be the best arm for that particular client also increases. This observation is interesting and, at first glance, counter-intuitive.

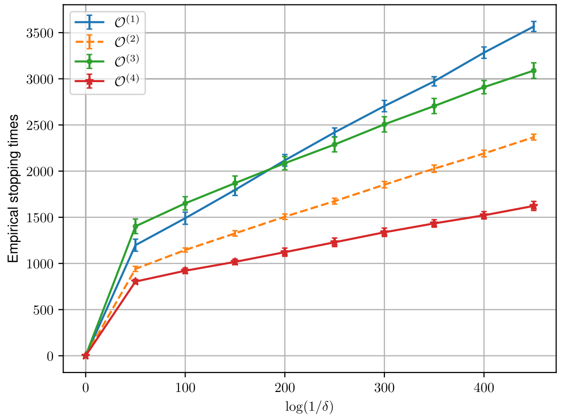

Fig. 2. Expected stopping times for various overlap patterns as described in (16).

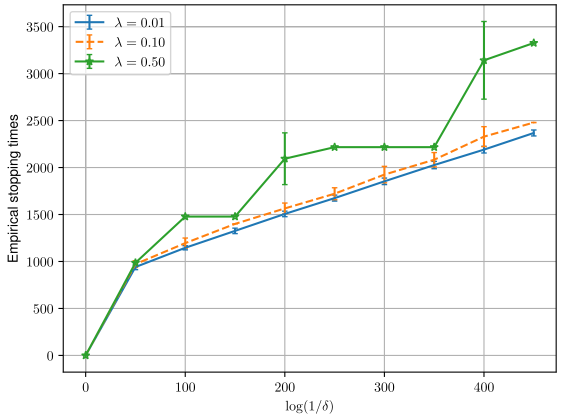

Fig. 3. Expected stopping times for various λ's and for overlap pattern $\mathcal{O}^{(2)}$.

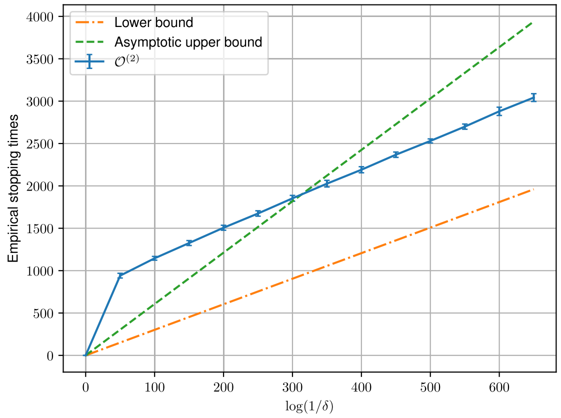

Fig. 4. Comparisons of the upper and lower bounds.

2) Effect of Communication Frequency: Recall that HET-TS(λ) communicates and stops at those time instants t of the form $b_r = \lceil (1+\lambda)^r \rceil$ for $r \in \mathbb{N}$. As λ increases, the communication frequency decreases. In other words, HET-TS(λ) is communicating at *sparser* time instants. Thus as λ grows, we should expect that the stopping times increase commensurately as the server receives less data per unit time. This is reflected in Fig. 3 where we use the instance with $\mathcal{O}^{(2)}$.

We note another interesting phenomenon, most evident from the curve indicated by $\lambda = 0.5$. The growth pattern of

TABLE I

A COMPARISON OF THE EMPIRICAL STOPPING TIMES OF HET-TS(λ) AND UNIFORM FOR $\lambda = 0.01$

$\log(1/\delta)$	10	50	100	200	500	1000
HET-TS(λ)	**32,473**	**32,798**	**33,457**	**34,129**	**35,870**	**38,458**
UNIFORM	252,184	254,706	259,826	265,048	278,568	292,778

the empirical stopping time has a piecewise linear shape. This is because HET-TS(λ) does not stop at any arbitrary integer time; it only does so at the times that correspond to *communication rounds* $b_r = \lceil (1+\lambda)^r \rceil$ for $r \in \mathbb{N}$. Hence, for δ and δ' sufficiently close, the empirical stopping times will be *exactly the same* with high probability. This explains the piecewise linear stopping pattern as $\log(1/\delta)$ grows.

3) Comparison to Theoretical Bounds: In the final experiment for synthetic data, we set $\lambda = 0.01$ and overlap pattern $\mathcal{O}^{(2)}$ as our instance v. In Fig. 4, we compare the empirical stopping time to the lower bound in Proposition 1 and the upper bound in Theorem 11. Recall that the asymptotic ratio of the expected stopping time $\mathbb{E}_v^{\Pi_{\text{HET-TS}}}[\tau_\delta(\Pi_{\text{HET-TS}})]$ to $\log(1/\delta)$ is $c^*(v)$ and $2(1+\lambda)c^*(v)$ in the lower and upper bounds respectively. We observe that as δ becomes sufficiently small, the slope of the empirical curve lies between the upper and lower bounds, as expected.

Furthermore, we see that $\mathbb{E}_v^{\Pi_{\text{HET-TS}}}[\tau_\delta(\Pi_{\text{HET-TS}})]/\log(1/\delta)$ is close to the lower bound, which strongly suggests our learned allocation $\widetilde{\omega}(\hat{v}(t))$ is very close to optimal allocation $\arg\max_{\omega\in\Gamma} g_{\hat{v}(t)}(\omega)$. We observe from Fig. 4 that the empirical performance or, more precisely, the slope of the expected stopping time as a function of $\log(1/\delta)$ is close to $(1+\lambda)\, c^*(v)$. This suggests that the factor $1+\lambda$ (in $2\,(1+\lambda)$) in Theorem 11 is unavoidable if we communicate at time instances that grow as $\Theta((1+\lambda)^r)$. The presence of the factor 2 (in $2\,(1+\lambda)$) is to enable the optimal allocation $\hat{\omega}_{i,m}(t)$ to be solved in a tractable fashion. For more details concerning this point, see the discussion following Theorem 11.

B. MovieLens Dataset

In the MovieLens dataset [33], there are about 2.2 million rating samples and 10,197 movies. Following the experimental settings in [21], we view each *country* and *genre* as a client and an arm, respectively. Besides, we normalize the rating score in the range of 0 to 100. We note that in the raw dataset that there are very few or even no samples for some combinations of country and genre. Thus, in our experiment we discard any country and genre pair with fewer than ten samples. As a result, we end up with 10,044 movies and $M = 48$ clients across $K = 19$ arms. It is natural that different clients have different arm sets in the dataset; this dovetails neatly with our problem setting in which S_m's need not be the same as one another and they need not be the full set $[K]$.

As in [17], we compare our algorithm to a baseline method which we call UNIFORM, having the same stopping rule as HET-TS(λ), but using a uniform sampling rule at each client (i.e., each client samples an arm uniformly at random at each time instant). We note that UNIFORM is δ-PAC for all $\delta \in (0,1)$. Our numerical results, which are obtained by averaging over four independent experiments and by setting $\lambda = 0.01$, are presented in Table I. We observed from our experiments that the statistical variations of the results are minimal (and virtually non-existent) as the algorithm necessarily stops at one of the time instants of the form $b_r = \lceil (1+\lambda)^r \rceil$ for $r \in \mathbb{N}$. Hence, "error bars" are not indicated. From Table I, we observe that the ratio of empirical stopping time between UNIFORM and HET-TS(λ) is *approximately eight*, showing that the sampling rule of HET-TS(λ) is highly effective in rapidly identifying the best arms in this real-world dataset.

VIII. CONCLUDING REMARKS AND FUTURE WORK

We studied best arm identification in a federated multi-armed bandit with heterogeneous clients in which each client can access a *subset* of the arms; this was mainly motivated from the unavailability of authorised vaccines in certain countries. We showed, among other results, that any *almost-optimal* algorithm must necessarily communicate such that the ratio of consecutive time instants is bounded, and that an algorithm may communicate at most exponentially sparsely while being almost-optimal. We proposed a track-and-stop-based algorithm that communicates exponentially sparsely and is almost-optimal up to an identifiable multiplicative constant in the regime of vanishing error probabilities. Future work includes carefully examining the effects of heterogeneity, possible corruptions, and the quantisation of various messages on the uplinks and downlinks. Additionally, an outstanding question concerns the derivation of a lower bound on the number of communication rounds as a function of α (the multiplicative gap to the lower bound per Definition 6) rather than η (which parameterizes the frequency of communication); see Remark 1.

APPENDIX

A. Proof of Lemma 2

Proof: Fix $v = \{\mu_{i,m} : i \in S_m,\, m \in [M]\} \in \mathcal{P}$ and $\omega \in \Gamma$. Let

$$\mathcal{C}(v) := \bigcup_{m\in[M]} \Big\{ (a_m^*(v), i) : i \in S_m \setminus \{a_m^*(v)\} \Big\}. \tag{17}$$

First, we note that by definition, $g_v(\omega) = 0$ if $\omega_{i,m} = 0$ for some $m \in [K]$ and $i \in S_m$. Therefore, it suffices to consider the case when $\omega_{i,m} > 0$ for all $m \in [K]$ and $i \in S_m$. In what follows, we abbreviate $\mu_{i,m}(v')$ and $\mu_i(v')$ to $\mu'_{i,m}$ and μ'_i respectively. We have

$$\begin{aligned} & g_v(\omega) \\ & = \inf_{v'\in\mathrm{Alt}(v)} \sum_{m=1}^{M} \sum_{i\in S_m} \omega_{i,m} \frac{(\mu_{i,m} - \mu'_{i,m})^2}{2} \\ & = \min_{(i_1,i_2)\in\mathcal{C}(v)} \inf_{\mu'_{i_1} < \mu'_{i_2}} \sum_{m=1}^{M} \sum_{i\in S_m} \omega_{i,m} \frac{(\mu_{i,m} - \mu'_{i,m})^2}{2} \end{aligned}$$

$$= \min_{(i_1,i_2)\in\mathcal{C}(v)} \inf_{\mu'_{i_1}\leq\mu'_{i_2}} \sum_{m=1}^{M} \Bigg[\mathbf{1}_{\{i_1\in S_m\}}\omega_{i_1,m}\frac{(\mu_{i_1,m}-\mu'_{i_1,m})^2}{2} + \mathbf{1}_{\{i_2\in S_m\}}\omega_{i_2,m}\frac{(\mu_{i_2,m}-\mu'_{i_2,m})^2}{2}\Bigg]$$

$$= \min_{(i_1,i_2)\in\mathcal{C}(v)} \frac{(\mu_{i_1}-\mu_{i_2})^2/2}{\frac{1}{M_{i_1}^2}\sum_{m=1}^{M}\frac{\mathbf{1}_{\{i_1\in S_m\}}}{\omega_{i_1,m}}+\frac{1}{M_{i_2}^2}\sum_{m=1}^{M}\frac{\mathbf{1}_{\{i_2\in S_m\}}}{\omega_{i_2,m}}}, \quad (18)$$

where (18) follows from the penultimate line by using the method of Lagrange multipliers and noting that the inner infimum in the penultimate line is attained at

$$\mu'_{i_1,m} = \mu_{i_1,m} - \frac{\mu_{i_1}-\mu_{i_2}}{M_{i_1}\omega_{i_1,m}\left(\sum_{i\in\{i_1,i_2\}}\sum_{m':i\in S_{m'}}\frac{1}{\omega_{i,m'}M_i^2}\right)} \quad \forall\, m : i_1 \in S_m,$$

$$\mu'_{i_2,m} = \mu_{i_2,m} + \frac{\mu_{i_1}-\mu_{i_2}}{M_{i_2}\omega_{i_2,m}\left(\sum_{i\in\{i_1,i_2\}}\sum_{m':i\in S_{m'}}\frac{1}{\omega_{i,m'}M_i^2}\right)} \quad \forall\, m : i_2 \in S_m.$$

From the definition of $\widetilde{g}_v(\omega)$ in (6), it is easy to verify that

$$\widetilde{g}_v(\omega) = \min_{(i_1,i_2)\in\mathcal{C}(v)} \min\left\{\frac{(\mu_{i_1}-\mu_{i_2})^2/2}{\frac{1}{M_{i_1}^2}\sum_{m=1}^{M}\mathbf{1}_{\{i_1\in S_m\}}\frac{1}{\omega_{i_1,m}}}, \frac{(\mu_{i_1}-\mu_{i_2})^2/2}{\frac{1}{M_{i_2}^2}\sum_{m=1}^{M}\mathbf{1}_{\{i_2\in S_m\}}\frac{1}{\omega_{i_2,m}}}\right\},$$

from which it follows that $\frac{\widetilde{g}_v(\omega)}{2} \leq g_v(\omega) \leq \widetilde{g}_v(\omega)$. □

B. Proof Proposition 1

We begin with a useful lemma that is used several times to establish various lower bounds.

Lemma 16: Let $T < \infty$ be any fixed time instant, and let $\mathcal{F}_T = \sigma(\{X_{A_m(t),m}(t), A_m(t) : t \in [T], m \in [M]\})$ be the history of all the arm pulls and rewards seen up to time T at all the clients under an algorithm Π. Let E be any event such that $\mathbf{1}_E$ is $\mathcal{F}_T$-measurable. Then, for any pair of problem instances v and v',

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\sum_{i\in S_m}\mathbb{E}_v^{\Pi}\left[\mathbf{1}_{\{A_m(t)=i\}}\ D_{\mathrm{KL}}(v_{i,m}\|v'_{i,m})\right] \geq d_{\mathrm{KL}}\left(\mathbb{P}_v^{\Pi}(E), \mathbb{P}_{v'}^{\Pi}(E)\right),$$

where $D_{\mathrm{KL}}(p\|q)$ denotes the Kullback–Leibler (KL) divergence between distributions p and q, and $d_{\mathrm{KL}}(x,y)$ denotes the KL divergence between two Bernoulli distributions with parameters x and y.

The proof of Lemma 16 follows along the exact same lines as the proof of [15, Lemma 19], and is hence omitted.

Proof: [Proof of Proposition 1] Fix $v \in \mathcal{P}$, and a δ-PAC policy Π, and $\delta \in (0,1)$. Let E denote the event that the empirical best arm is $a^*(v)$, i.e.,

$$E = \{\hat{a}_\delta = a^*(v)\}$$

By Lemma 16, for any $v' \in \mathrm{Alt}(v)$, we have,

$$\sum_{t=1}^{\infty}\sum_{m=1}^{M}\sum_{i\in S_m}\mathbb{E}_v^{\Pi}\left[\mathbf{1}_{\{A_m(t)=i\}}\ D_{\mathrm{KL}}(v_{i,m}\|v'_{i,m})\right] \geq d_{\mathrm{KL}}\left(\mathbb{P}_v^{\Pi}(E), \mathbb{P}_{v'}^{\Pi}(E)\right).$$

By using the fact that the arm reward distributions are Gaussian with unit variance, we have

$$\sum_{t=1}^{\infty}\sum_{m=1}^{M}\sum_{i\in S_m}\mathbb{E}_v^{\Pi}\left[\mathbf{1}_{\{A_m(t)=i\}}\frac{(\mu_{i,m}-\mu'_{i,m})^2}{2}\right] \geq d_{\mathrm{KL}}\left(\mathbb{P}_v^{\Pi}(E), \mathbb{P}_{v'}^{\Pi}(E)\right).$$

Now, by observing that $\mathbb{P}_v^{\Pi}(E) \geq 1-\delta$ and $\mathbb{P}_{v'}^{\Pi}(E) \leq \delta$ since all policies considered are δ-PAC, we obtain

$$\sum_{t=1}^{\infty}\sum_{m=1}^{M}\sum_{i\in S_m}\mathbb{E}_v^{\Pi}\left[\mathbf{1}_{\{A_m(t)=i\}}\frac{(\mu_{i,m}-\mu'_{i,m})^2}{2}\right] \geq \log\left(\frac{1}{4\delta}\right), \quad (19)$$

Then, denoting $\bar{\omega}_{i,m} := \frac{1}{\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)]}\sum_{t=1}^{\infty}\mathbb{E}_v^{\Pi}\left[\mathbf{1}_{\{A_m(t)=i\}}\right]$, (19) implies that

$$\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)]\sum_{m=1}^{M}\sum_{i\in S_m}\bar{\omega}_{i,m}\frac{(\mu_{i,m}-\mu'_{i,m})^2}{2} \geq \log\left(\frac{1}{4\delta}\right),$$

which leads to

$$\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)] \geq \frac{\log(\frac{1}{4\delta})}{\sum_{m=1}^{M}\ \sum_{i\in S_m}\bar{\omega}_{i,m}\frac{(\mu_{i,m}-\mu'_{i,m})^2}{2}}. \quad (20)$$

By using the fact that (20) holds for any $v' \in \mathrm{Alt}(v)$, we have

$$\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)] \geq \frac{\log(\frac{1}{4\delta})}{\inf_{v'\in\mathrm{Alt}(v)}\sum_{m=1}^{M}\ \sum_{i\in S_m}\bar{\omega}_{i,m}\frac{(\mu_{i,m}-\mu'_{i,m})^2}{2}}. \quad (21)$$

Next, by using the fact that $\bar{\omega} \in \Gamma$, (21) implies that

$$\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)] \geq \frac{\log(\frac{1}{4\delta})}{\sup_{\omega\in\Gamma}\inf_{v'\in\mathrm{Alt}(v)}\sum_{m=1}^{M}\sum_{i\in S_m}\omega_{i,m}\frac{(\mu_{i,m}-\mu'_{i,m})^2}{2}}. \quad (22)$$

By (18), $g_v(\omega) = \frac{1}{2}\inf_{v'\in\mathrm{Alt}(v)}\sum_{m=1}^{M}\sum_{i\in S_m}\omega_{i,m}(\mu_{i,m}(v)-\mu_{i,m}(v'))^2$ is continuous in ω. Furthermore, by noting that Γ is compact, (22) implies that

$$\mathbb{E}_v^{\Pi}[\tau_\delta(\Pi)] \geq \frac{\log(\frac{1}{4\delta})}{\max_{\omega\in\Gamma}\inf_{v'\in\mathrm{Alt}(v)}\sum_{m=1}^{M}\sum_{i\in S_m}\omega_{i,m}\frac{(\mu_{i,m}-\mu'_{i,m})^2}{2}},$$

as desired. This completes the proof. □

C. Proof of Theorem 7

Define $v_\dagger(\rho)$ to be a special problem instance in which the arm means are given by

$$\mu(v_\dagger(\rho))_{i,m} = \frac{i}{\sqrt{\rho}}, \quad m \in [M],\ i \in S_m. \tag{23}$$

Then, it follows that $\mu(v_\dagger(\rho))_i = \frac{i}{\sqrt{\rho}}$ for all $i \in [K]$. The following result will be used in the proof of Theorem 7.

Lemma 17: Given any $\rho > 0$, the problem instance $v_\dagger(\rho)$, defined in (23), satisfies

$$\frac{4\rho}{MK^2} \le c^*\big(v_\dagger(\rho)\big) \le 4K\rho.$$

Proof of Lemma 17: Recall the definition of $\mathcal{C}(\cdot)$ in (17). Let

$$\Delta_{\min}(\rho) := \min_{(i,j)\in\mathcal{C}(v_\dagger(\rho))} |\mu_i(v_\dagger(\rho)) - \mu_j(v_\dagger(\rho))|,$$

and let $\omega^{\text{trivial}} \in \Gamma$ be defined as

$$\omega^{\text{trivial}}_{i,m} = \frac{1}{|S_m|}, \quad m \in [M],\ i \in S_m.$$

Notice that

$$\begin{aligned}
&c^*(v_\dagger(\rho))^{-1} \\
&= \max_{\omega\in\Gamma} \inf_{v'\in\text{Alt}(v)} \sum_{m=1}^{M} \sum_{i\in S_m} \omega_{i,m} \frac{\big(\mu_{i,m}(v) - \mu_{i,m}(v')\big)^2}{2} \\
&\le \inf_{v'\in\text{Alt}(v)} \sum_{m=1}^{M} \sum_{i\in S_m} 1 \cdot \frac{\big(\mu_{i,m}(v) - \mu_{i,m}(v')\big)^2}{2} \\
&= g_{v_\dagger(\rho)}(\mathbf{1}^{K\times M}),
\end{aligned} \tag{24}$$

where $\mathbf{1}^{K\times M}$ denotes the all-ones matrix of dimension $K\times M$. Also notice that

$$\begin{aligned}
&c^*(v_\dagger(\rho))^{-1} \\
&= \max_{\omega\in\Gamma} \inf_{v'\in\text{Alt}(v)} \sum_{m=1}^{M} \sum_{i\in S_m} \omega_{i,m} \frac{\big(\mu_{i,m}(v) - \mu_{i,m}(v')\big)^2}{2} \\
&\ge \inf_{v'\in\text{Alt}(v)} \sum_{m=1}^{M} \sum_{i\in S_m} \omega^{\text{trivial}}_{i,m} \frac{\big(\mu_{i,m}(v) - \mu_{i,m}(v')\big)^2}{2} \\
&= g_{v_\dagger(\rho)}(\omega^{\text{trivial}}).
\end{aligned} \tag{25}$$

From (24) and (25), we have

$$\begin{aligned}
&g_{v_\dagger(\rho)}(\omega^{\text{trivial}}) \le c^*\big(v_\dagger(\rho)\big)^{-1} \le g_{v_\dagger(\rho)}(\mathbf{1}^{K\times M}) \\
&\overset{(a)}{\implies} \min_{(i,j)\in\mathcal{C}(v_\dagger(\rho))} \frac{(\mu_i(v_\dagger(\rho)) - \mu_j(v_\dagger(\rho)))^2/2}{\frac{1}{M_i^2}\sum_{m=1}^{M}\frac{\mathbf{1}_{\{i\in S_m\}}}{\omega^{\text{trivial}}_{i,m}} + \frac{1}{M_j^2}\sum_{m=1}^{M}\frac{\mathbf{1}_{\{j\in S_m\}}}{\omega^{\text{trivial}}_{j,m}}} \\
&\le c^*\big(v_\dagger(\rho)\big)^{-1} \\
&\le \min_{(i,j)\in\mathcal{C}(v_\dagger(\rho))} \frac{\big(\mu_i(v_\dagger(\rho)) - \mu_j(v_\dagger(\rho))\big)^2/2}{\frac{1}{M_i^2}\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}} + \frac{1}{M_j^2}\sum_{m=1}^{M}\mathbf{1}_{\{j\in S_m\}}} \\
&\overset{(b)}{\implies} \frac{\Delta^2_{\min}(\rho)}{4K} \le c^*\big(v_\dagger(\rho)\big)^{-1} \le \frac{M\Delta^2_{\min}(\rho)}{4} \\
&\overset{(c)}{\implies} \frac{1}{4K\rho} \le c^*\big(v_\dagger(\rho)\big)^{-1} \le \frac{MK^2}{4\rho} \\
&\implies \frac{4\rho}{MK^2} \le c^*\big(v_\dagger(\rho)\big) \le 4K\rho,
\end{aligned}$$

where (a) above follows from (18) of Lemma 2, in writing (b), we make use of the observation that for all $m \in [M]$ and $i \in S_m$,

$$\frac{1}{\omega^{\text{trivial}}_{i,m}} = |S_m| \le K,$$

and (c) makes use of the fact that $\Delta^2_{\min}(\rho) \in (\frac{1}{\rho}, \frac{K^2}{\rho})$. This completes the desired proof. □

Proof: [Proof of Theorem 7] Fix a confidence level $\delta \in (0, \frac{1}{4})$ arbitrarily, and let Π be δ-PAC and almost-optimal up to $\alpha \ge 1$. Suppose, on the contrary, that

$$\limsup_{r\to\infty} \frac{b_{r+1}}{b_r} = \infty. \tag{26}$$

Then, there exists an increasing sequence $\{z_l\}_{l=1}^\infty$ such that $\lim_{l\to\infty} \frac{b_{z_l}}{b_{z_l+1}} = 0$ and $b_{z_l} < b_{z_l+1}$ for all $l \in \mathbb{N}$. Let $T^*_\delta(v) := \log\left(\frac{1}{4\delta}\right) c^*(v)$.

Let

$$v^{(l)} := v_\dagger\left(\frac{\sqrt{b_{z_l+1} b_{z_l}}}{4\log(\frac{1}{4\delta})}\right), \quad \text{for all } l \in \mathbb{N}.$$

By Lemma 17, we then have

$$\frac{\sqrt{b_{z_l+1} b_{z_l}}}{MK^2 \log(\frac{1}{4\delta})} \le c^*(v^{(l)}) \le \frac{K\sqrt{b_{z_l+1} b_{z_l}}}{\log(\frac{1}{4\delta})}. \tag{27}$$

Also, we have

$$\frac{\sqrt{b_{z_l+1} b_{z_l}}}{MK^2} \le T^*_\delta(v^{(l)}) \le K\sqrt{b_{z_l+1} b_{z_l}}. \tag{28}$$

Let

$$\begin{aligned}
E_l := \{&\text{empirical best arms } \hat{a}_\delta = a^*(v^{(l)}) \\
&\text{and stopping time } \tau_\delta(\Pi) \le b_{z_l}\}, \quad l \in \mathbb{N},
\end{aligned}$$

be the event that (a) $\hat{a}_\delta = (\hat{a}_{\delta,m})_{m\in[M]}$, the vector of the empirical best arms of the clients at confidence level δ, equals the vector $a^*(v^{(l)})$, and (b) the stopping time $\tau_\delta(\Pi) \le b_{z_l}$. From Lemma 16, for any $l \in \mathbb{N}$, we have

$$\begin{aligned}
&\sum_{t=1}^{b_{z_l}} \sum_{m=1}^{M} \sum_{i\in S_m} \mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathbf{1}_{\{A_t(m)=i\}}\, D_{\text{KL}}(v^{(l)}_{i,m} \| v'_{i,m})\right] \\
&\quad \ge d_{\text{KL}}\left(\mathbb{P}^{\Pi}_{v^{(l)}}(E_l), \mathbb{P}^{\Pi}_{v'}(E_l)\right)
\end{aligned} \tag{29}$$

for all $v' \in \text{Alt}(v^{(l)})$. Note that

$$\begin{aligned}
&\mathbb{P}^{\Pi}_{v^{(l)}}(E_l) = 1 - \mathbb{P}^{\Pi}_{v^{(l)}}(E_l^c) \\
&\quad \overset{(a)}{\ge} 1 - \mathbb{P}^{\Pi}_{v^{(l)}}(\hat{a}_\delta \ne a^*(v^{(l)})) - \mathbb{P}^{\Pi}_{v^{(l)}}(\tau_\delta(\Pi) > b_{z_l}) \\
&\quad \overset{(b)}{=} 1 - \mathbb{P}^{\Pi}_{v^{(l)}}(\hat{a}_\delta \ne a^*(v^{(l)})) - \mathbb{P}^{\Pi}_{v^{(l)}}(\tau_\delta(\Pi) \ge b_{z_l+1}) \\
&\quad \overset{(c)}{\ge} 1 - \delta - \frac{\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]}{b_{z_l+1}} \\
&\quad \overset{(d)}{\ge} 1 - \delta - \frac{\alpha\ T^*(v^{(l)})}{b_{z_l+1}} \\
&\quad \overset{(e)}{\ge} 1 - \delta - \alpha K\sqrt{\frac{b_{z_l}}{b_{z_l+1}}},
\end{aligned}$$

where (a) above follows from the union bound, (b) follows by noting that $\mathbb{P}^{\Pi}_{v^{(l)}}(\tau_\delta(\Pi) \geq b_{z_l+1}) = \mathbb{P}^{\Pi}_{v^{(l)}}(\tau_\delta(\Pi) > b_{z_l})$ as $\tau_\delta(\Pi) \in \{b_r\}_{r\in\mathbb{N}}$ and $b_{z_l} < b_{z_l+1}$, (c) follows from Markov's inequality and the fact that $\mathbb{P}^{\Pi}_{v^{(l)}}(\hat{a}_\delta \neq a^*(v)) \leq \delta$ as Π is δ-PAC, (d) follows from the fact that $\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)] \leq \alpha\, T^*(v^{(l)})$ as Π is almost-optimal up to the constant α, and (e) follows from (28). Because the algorithm Π is δ-PAC, it can be shown that $\mathbb{P}^{\Pi}_{v'}(E_l) \leq \delta$ for all $v' \in \mathrm{Alt}(v^{(l)})$.

Continuing with (29) and using the fact that $d_{\mathrm{KL}}(x,y) \geq \log\left(\frac{1}{4\,\delta'}\right)$ whenever $x \geq 1-\delta'$ and $y \leq \delta'$ (see, for instance, [15]), setting $\delta' = \delta + \alpha K\sqrt{b_{z_l}/b_{z_l+1}}$, we have

$$\inf_{v'\in\mathrm{Alt}(v^{(l)})} \sum_{t=1}^{b_{z_l}} \sum_{m=1}^{M} \sum_{i\in S_m} \mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathbf{1}_{\{A_t(m)=i\}}\; D_{\mathrm{KL}}(v^{(l)}_{i,m} \| v'_{i,m})\right] \geq \log\left(\frac{1}{4\delta + 4\alpha\, K\sqrt{\frac{b_{z_l}}{b_{z_l+1}}}}\right)$$

which implies that

$$\inf_{v'\in\mathrm{Alt}(v^{(l)})} \sum_{t=1}^{b_{z_l}} \sum_{m=1}^{M} \sum_{i\in S_m} \frac{\mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathbf{1}_{\{A_t(m)=i\}}\right]}{b_{z_l}} \cdot \frac{(\mu^{(l)}_{i,m} - \mu'_{i,m})^2}{2} \geq \frac{1}{b_{z_l}} \log\left(\frac{1}{4\delta + 4\alpha\, K\sqrt{\frac{b_{z_l}}{b_{z_l+1}}}}\right). \tag{30}$$

The inequality in (30) implies that

$$\begin{aligned} (30) &\overset{(a)}{\Longrightarrow} \frac{b_{z_l}}{c^*(v^{(l)})} \geq \log\left(\frac{1}{4\delta + 4\alpha\, K\sqrt{\frac{b_{z_l}}{b_{z_l+1}}}}\right) \\ &\overset{(b)}{\Longrightarrow} MK^2 \log\left(\frac{1}{4\delta}\right) \sqrt{\frac{b_{z_l}}{b_{z_l+1}}} \geq \log \frac{1}{4\delta + 4\alpha\, K\sqrt{\frac{b_{z_l}}{b_{z_l+1}}}}. \end{aligned} \tag{31}$$

In the above set of inequalities, (a) follows from the definition of $c^*(v^{(l)})$, and (b) follows from (27). Letting $l \to \infty$ and using (26), we observe that the left-hand side of (31) converges to 0, whereas the right-hand side converges to $\log(\frac{1}{4\delta})$, thereby resulting in $0 \geq \log(\frac{1}{4\delta})$, a contradiction. This proves that $\limsup_{r\to\infty} \frac{b_{r+1}}{b_r} < \infty$. □

Proof: Fix $\delta \in (0, \frac{1}{4})$ and $\beta \in (0,1)$ arbitrarily. Let Π be almost-optimal up to a constant, say $\alpha \geq 1$. Let $\{v^{(l)}\}_{l=1}^{\infty}$ be any sequence of problem instances such that $\lim_{l\to\infty} c^*(v^{(l)}) = +\infty$, where this sequence must exist because of Lemma 17. Let $T_l := \left\lceil \left(\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]\right)^{\beta} \right\rceil$. Because $\lim_{l\to\infty} c^*(v^{(l)}) = +\infty$, we have

$$\lim_{l\to\infty} \frac{T_l}{\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]} = 0. \tag{32}$$

For $l \in \mathbb{N}$, let

$$\begin{aligned} F_l := \{&\text{empirical best arms } \hat{a}_\delta = a^*(v^{(l)}) \\ &\text{and stopping time } \tau_\delta(\Pi) \leq T_l\} \end{aligned}$$

be the event that (a) the vector of empirical best arms matches with the vector of best arms under $v^{(l)}$, and (b) the stopping time $\tau_\delta(\Pi) \leq T_l$. Also, let $p_l := \mathbb{P}^{\Pi}_{v^{(l)}}(\tau_\delta > T_l)$. From Lemma 16, we know that

$$\sum_{t=1}^{T_l} \sum_{m=1}^{M} \sum_{i\in S_m} \mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathbf{1}_{\{A_t(m)=i\}}\; D_{\mathrm{KL}}(v^{(l)}_{i,m} \| v'_{i,m})\right] \geq d_{\mathrm{KL}}\left(\mathbb{P}^{\Pi}_{v^{(l)}}(F_l), \mathbb{P}^{\Pi}_{v'}(F_l)\right)$$

for all problem instances v'. In particular, for $v' \in \mathrm{Alt}(v^{(l)})$, we note that for any $l \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}^{\Pi}_{v^{(l)}}(F_l) &\geq 1 - \mathbb{P}_{v^{(l)}}(\hat{a}_\delta \neq a^*(v^{(l)})) - \mathbb{P}_{v^{(l)}}(\tau_\delta > T_l) \\ &\geq 1 - \delta - p_l. \end{aligned}$$

Along similar lines, it can be shown that $\mathbb{P}^{\Pi}_{v'}(F_l) \leq \delta + p_l$ for any $v' \in \mathrm{Alt}(v^{(l)})$. Then, using the fact that $d(x,y) \geq \log\left(\frac{1}{4\delta'}\right)$ whenever $x \geq 1-\delta'$ and $y \leq \delta'$, setting $\delta' = \delta + p_l$, we have

$$\inf_{v'\in\mathrm{Alt}(v^{(l)})} \sum_{t=1}^{T_l} \sum_{m=1}^{M} \sum_{i\in S_m} \mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathbf{1}_{\{A_t(m)=i\}} D_{\mathrm{KL}}(v^{(l)}_{i,m} \| v'_{i,m})\right] \geq \log\left(\frac{1}{4\delta + 4p_l}\right),$$

which in turn implies that

$$\begin{aligned} \inf_{v'\in\mathrm{Alt}(v^{(l)})} \sum_{t=1}^{T_l} \sum_{m=1}^{M} \sum_{i\in S_m} \frac{\mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathbf{1}_{\{A_t(m)=i\}}\right]}{T_l} \cdot \frac{(\mu^{(l)}_{i,m} - \mu'_{i,m})^2}{2} &\geq \frac{1}{T_l} \log\left(\frac{1}{4\delta + 4p_l}\right) \\ \Longrightarrow \frac{T_l}{c^*(v^{(l)})} \geq \log\left(\frac{1}{4\delta + 4p_l}\right), \end{aligned} \tag{33}$$

for all $l \in \mathbb{N}$, where the last line above follows from the definition of $c^*(v^{(l)})$. Because Π is almost-optimal up to constant $\alpha \geq 1$, we have

$$c^*(v^{(l)}) \log\left(\frac{1}{4\delta}\right) \leq \mathbb{E}^{\Pi}_{v^{(l)}}(\tau_\delta(\Pi)) \leq \alpha\, c^*(v^{(l)}) \log\left(\frac{1}{4\delta}\right) \tag{34}$$

for all $l \in \mathbb{N}$. Combining (33) and (34), we get

$$\frac{T_l}{\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]} \geq \frac{\log\left(\frac{1}{4\delta+4p_l}\right)}{\alpha \log\left(\frac{1}{4\delta}\right)} \quad \text{for all } l \in \mathbb{N}. \tag{35}$$

Suppose now that there exists $\epsilon \in \left(0, \frac{1}{4} - \delta\right)$ such that

$$\liminf_{l\to\infty} \mathbb{P}^{\Pi}_{v^{(l)}}\left(\log(\tau_\delta(\Pi)) > \beta \log\left(\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]\right)\right) \leq \frac{1}{4} - \delta - \epsilon. \tag{36}$$

This implies from the definitions of T_l and p_l that there exists an increasing sequence $\{l_n : n \geq 1\}$ such that $p_{l_n} \leq \frac{1}{4} - \delta - \epsilon$ for all $n \geq 1$. Using this in (35), we get that

$$\begin{aligned} \limsup_{l\to\infty} \frac{T_l}{\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]} &\geq \limsup_{n\to\infty} \frac{T_{l_n}}{\mathbb{E}^{\Pi}_{v^{(l_n)}}[\tau_\delta(\Pi)]} \\ &\geq \frac{\log\left(\frac{1}{1-4\epsilon}\right)}{\alpha \log\left(\frac{1}{4\delta}\right)} > 0, \end{aligned}$$

which clearly contradicts (32). This proves that there is no $\epsilon \in \left(0, \frac{1}{4} - \delta\right)$ such that (36) holds, thereby establishing the desired result. □

D. Proof of Theorem 9

Proof: Fix a sequence of problem instances $\{v^{(l)}\}_{l=1}^{\infty}$ with $\lim_{l\to\infty} c^*(v^{(l)}) = +\infty$, a confidence level $\delta \in (0, \frac{1}{4})$, and an algorithm Π that is almost optimal up to a constant, say $\alpha \geq 1$. From Theorem 7, we know that there exists $\eta > 0$ such that

$$\mathfrak{r}_\delta(\Pi) \geq \log_\eta(\tau_\delta(\Pi)) \quad \text{almost surely.} \tag{37}$$

Also, from Lemma 8, we know that for any $\beta \in (0,1)$ and any sequence of problem instances $\{v^{(l)}\}_{l=1}^{\infty}$ with $\lim_{l\to\infty} c^*(v^{(l)}) = +\infty$,

$$\liminf_{l\to\infty}\ \mathbb{P}^{\Pi}_{v^{(l)}}\left(\log\left(\tau_\delta(\Pi)\right) > \beta \log\left(\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]\right)\right) \geq \frac{1}{4} - \delta. \tag{38}$$

Using (37) in (38), we have

$$\begin{aligned}
&\liminf_{l\to\infty}\ \mathbb{P}^{\Pi}_{v^{(l)}}\left(\mathfrak{r}_\delta(\Pi) > \beta\ \log_\eta\left(\mathbb{E}^{\Pi}_{v^{(l)}}[\tau_\delta(\Pi)]\right)\right) \\
&\geq \frac{1}{4} - \delta \quad \forall\ \beta \in (0,1) \\
\overset{(a)}{\Longrightarrow}\ &\liminf_{l\to\infty}\ \mathbb{P}^{\Pi}_{v^{(l)}}\left(\mathfrak{r}_\delta(\Pi) > \beta\ \log_\eta\left(\log\left(\frac{1}{4\delta}\right) c^*(v^{(l)})\right)\right) \\
&\geq \frac{1}{4} - \delta \quad \forall\ \beta \in (0,1) \\
\overset{(b)}{\Longrightarrow}\ &\liminf_{l\to\infty}\ \frac{\mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathfrak{r}_\delta(\Pi)\right]}{\log_\eta\left(\log\left(\frac{1}{4\delta}\right) c^*(v^{(l)})\right)} \\
&\geq \beta\left(\frac{1}{4} - \delta\right) \quad \forall\ \beta \in (0,1) \\
\overset{(c)}{\Longrightarrow}\ &\liminf_{l\to\infty}\ \frac{\mathbb{E}^{\Pi}_{v^{(l)}}\left[\mathfrak{r}_\delta(\Pi)\right]}{\log_\eta\left(\log\left(\frac{1}{4\delta}\right) c^*(v^{(l)})\right)} \geq \frac{1}{4} - \delta.
\end{aligned}$$

In the above set of inequalities, (a) follows from Proposition 1 and the hypothesis that Π is δ-PAC, (b) follows from Markov's inequality, and (c) follows from (b) by letting $\beta \to 1$. The desired result is thus established. □

E. Proof of Theorem 10

Below, we record some important results that will be useful for proving Theorem 10.

Lemma 18 [2, Lemma 33.8]: Let $Y_1, Y_2, \ldots$ be independent Gaussian random variables with mean μ and unit variance. Let $\hat{\mu}_n := \frac{1}{n}\sum_{i=1}^{n} Y_i$. Then,

$$\mathbb{P}\left(\exists\, n \in \mathbb{N} : \frac{n}{2}(\hat{\mu}_n - \mu)^2 \geq \log(1/\delta) + \log(n(n+1))\right) \leq \delta.$$

Lemma 19: Fix $n \in \mathbb{N}$. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $\mathbb{P}(Y_i \leq y) \leq y$ for all $y \in [0,1]$ and $i \in [n]$. Then, for any $\epsilon > 0$,

$$\mathbb{P}\left(\sum_{i=1}^{n} \log(1/Y_i) \geq \epsilon\right) \leq f_n(\epsilon)$$

where $f_n : (0,\infty) \to (0,1)$ is defined by

$$f_n(x) = \sum_{i=1}^{n} \frac{x^{i-1}e^{-x}}{(i-1)!}, \qquad x \in (0,\infty).$$

Proof of Lemma 19: First, for $i \in [n]$ we define the random variable $Z_i := F_i(Y_i)$, where F_i is the cumulative distribution function (CDF) of Y_i. Clearly, Z_i is a uniform random variable. Notice that $\mathbb{P}(Y_i \leq y) \leq y = \mathbb{P}(Z_i \leq y)$ for all $y \in (0,1)$, from which it follows that

$$\mathbb{P}\left(\sum_{i=1}^{n} \log(1/Y_i) \geq \epsilon\right) \leq \mathbb{P}\left(\sum_{i=1}^{n} \log(1/Z_i) \geq \epsilon\right).$$

Therefore, it suffices to prove Lemma 19 for the case when $Y_1, \ldots, Y_n$ are independent and uniformly distributed on $[0,1]$. Suppose that this is indeed the case. Then, we note that $\mathbb{P}\left(\sum_{i=1}^{n} \log(1/Y_i) \geq \epsilon\right) = \mathbb{P}\left(\prod_{i=1}^{n} Y_i \leq \exp(-\epsilon)\right)$. Let $h_s(x) := \mathbb{P}\left(\prod_{i=1}^{s} Y_i \leq x\right)$ for $s \in [n]$ and $x \in (0,1)$. We then have

$$\begin{aligned}
&h_1(x) = x, \\
\forall\ s > 1, \quad &h_s(x) = \int_0^1 h_{s-1}\left(\min\{x/y, 1\}\right)\, \mathrm{d}y \\
&\qquad\ \ = x + \int_x^1 h_{s-1}(x/y)\, \mathrm{d}y.
\end{aligned}$$

Using mathematical induction, we demonstrate below that

$$h_s(x) = \sum_{i=1}^{s} \frac{(\log\frac{1}{x})^{i-1}x}{(i-1)!} \tag{39}$$

for all $s \in [n]$ and $x \in (0,1)$.

Base Case: It is easy to verify that (39) holds for $s = 1$. For $s = 2$, we have

$$\begin{aligned}
h_2(x) &= x + \int_x^1 h_1\left(\frac{x}{y}\right)\, \mathrm{d}y = x + \int_x^1 \frac{x}{y}\, \mathrm{d}y \\
&= x + x\log\left(\frac{1}{x}\right),
\end{aligned}$$

thus verifying that (39) holds for $s = 2$.

Induction Step: Suppose now that (39) holds for $s = k$ for some $k > 2$. Then,

$$\begin{aligned}
h_{k+1}(x) &= x + \int_x^1 h_k(x/y)\, \mathrm{d}y \\
&\overset{(a)}{=} x + \int_x^1 \sum_{i=1}^{k} \frac{(\log\frac{y}{x})^{i-1}(\frac{x}{y})}{(i-1)!}\, \mathrm{d}y \\
&= x + \sum_{i=1}^{k} \int_x^1 \frac{(\log\frac{y}{x})^{i-1}(\frac{x}{y})}{(i-1)!}\, \mathrm{d}y \\
&= x + \sum_{i=1}^{k} \frac{x}{(i-1)!} \int_x^1 \frac{(\log\frac{y}{x})^{i-1}}{y}\, \mathrm{d}y \\
&\overset{(b)}{=} x + \sum_{i=1}^{k} \frac{x}{(i-1)!} \int_1^{1/x} \frac{(\log y')^{i-1}}{y'}\, \mathrm{d}y' \\
&\overset{(c)}{=} x + \sum_{i=1}^{k} \frac{x}{(i-1)!} \frac{(\log\frac{1}{x})^{i}}{i} \\
&= \sum_{i=1}^{k+1} \frac{(\log\frac{1}{x})^{i-1}x}{(i-1)!},
\end{aligned} \tag{40}$$

where (a) follows from the induction hypothesis, in writing (b) above, we set $y' = y/x$, and (c) follows by noting that

$$\int \frac{(\log y)^j}{y}\, \mathrm{d}y = \frac{1}{j+1}(\log y)^{j+1}.$$

This demonstrates that (40) holds for $s = k+1$.

Finally, we note that

$$\mathbb{P}\left(\sum_{i=1}^{n} \log\left(\frac{1}{Y_i}\right) \geq \epsilon\right) = h_n\left(e^{-\epsilon}\right) = \sum_{i=1}^{n} \frac{\epsilon^{i-1} e^{-\epsilon}}{(i-1)!} = f_n(\epsilon),$$

thus establishing the desired result. □

With the above ingredients in place, we are now ready to prove Theorem 10.

Proof of Theorem 10: Fix a confidence level $\delta \in (0,1)$ and a problem instance $v \in \mathcal{P}$ arbitrarily. We claim that $\tau_\delta(\Pi_{\text{Het-TS}}) < +\infty$ almost surely; a proof of this is deferred until the proof of Lemma 25. Assuming that the preceding fact is true, for $m \in [M]$ and $i \in S_m$, let

$$\begin{aligned}\xi_{i,m} := \sup_{t \geq K} \frac{N_{i,m}(t)}{2} \Big\{ &\left(\hat{\mu}_{i,m}(t) - \mu_{i,m}(v)\right)^2 \\ &- \log\left(N_{i,m}(t)(N_{i,m}(t)+1)\right)\Big\}.\end{aligned}$$

From Lemma 18, we know that for any confidence level $\delta' \in (0,1)$,

$$\mathbb{P}_v^{\Pi_{\text{Het-TS}}}(\xi_{i,m} \geq \log(1/\delta')) \leq \delta'.$$

Let $\xi'_{i,m} := \exp(-\xi_{i,m})$. Recall that $K' = \sum_{m=1}^{M} |S_m|$ From Lemma 19, we know that for any $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}_v^{\Pi_{\text{Het-TS}}}\left(\sum_{m\in[M]}\sum_{i\in S_m} \log(1/\xi'_{i,m}) \geq \epsilon\right) \leq f_{K'}(\epsilon) \\
&\overset{(a)}{\Longrightarrow} \mathbb{P}_v^{\Pi_{\text{Het-TS}}}\left(\sum_{m\in[M]}\sum_{i\in S_m} \xi_{i,m} \geq \epsilon\right) \leq f_{K'}(\epsilon) \\
&\overset{(b)}{\Longrightarrow} \mathbb{P}_v^{\Pi_{\text{Het-TS}}}\left(\sum_{m\in[M]}\sum_{i\in S_m} \xi_{i,m} \geq \epsilon\right) \leq f(\epsilon) \\
&\overset{(c)}{\Longrightarrow} \mathbb{P}_v^{\Pi_{\text{Het-TS}}}\left(\sum_{m\in[M]}\sum_{i\in S_m} \xi_{i,m} \geq f^{-1}(\delta)\right) \leq \delta, \qquad (41)
\end{aligned}$$

where (a) above follows from the definition of $\xi'_{i,m}$, (b) follows from the definition of f in (12), and in writing (c), we (i) make use of the fact that f is continuous and strictly decreasing and therefore admits an inverse, and (ii) set $\epsilon = f^{-1}(\delta)$. Eq. (41) then implies

$$\begin{aligned}\mathbb{P}_v^{\Pi_{\text{Het-TS}}}\Bigg(\forall\, t \geq K, &\sum_{m\in[M]}\sum_{i\in S_m} \frac{N_{i,m}(t)}{2}\left(\hat{\mu}_{i,m}(t) - \mu_{i,m}(v)\right)^2 \\ &\leq K' \log\left(t(t+1)\right) + f^{-1}(\delta)\Bigg) \geq 1-\delta,\end{aligned}$$

which in turn implies that

$$\begin{aligned}\mathbb{P}_v^{\Pi_{\text{Het-TS}}}\Bigg(\forall\, t \geq K &\sum_{m\in[M]}\sum_{i\in S_m} \frac{N_{i,m}(t)}{2}\left(\hat{\mu}_{i,m}(t) - \mu_{i,m}(v)\right)^2 \\ &\leq \beta(t,\delta)\Bigg) \geq 1-\delta. \qquad (42)\end{aligned}$$

Note that at the stopping time $\tau_\delta(\Pi_{\text{Het-TS}})$, we must have

$$\begin{aligned}\inf_{v' \in \text{Alt}(\hat{v}(\tau_\delta))} \sum_{m\in[M]}\sum_{i\in S_m} N_{i,m}(\tau_\delta) &\frac{(\mu_{i,m}(v') - \hat{\mu}_{i,m}(\tau_\delta))^2}{2} \\ &> \beta(\tau_\delta, \delta).\end{aligned}$$

Thus, we may write (42) equivalently as $\mathbb{P}_v^{\Pi_{\text{Het-TS}}}\left(v \notin \text{Alt}\left(\hat{v}(\tau_\delta(\Pi_{\text{Het-TS}}))\right)\right) \geq 1-\delta$, or equivalently, $\mathbb{P}_v^{\Pi_{\text{Het-TS}}}\left(a^*(v) = a^*(\hat{v}(\tau_\delta(\Pi_{\text{Het-TS}})))\right) \geq 1-\delta$. □

F. Proof of Theorem 11

We first state two results that will be used later in the proof of Theorem 11.

Lemma 20 [17], [34]: Let $f : S \times \Theta \to \mathbb{R}$ be a continuous function, and $\mathcal{D} : \Theta \rightrightarrows S$ be a compact-valued continuous correspondence. Let $f^* : \Theta \to \mathbb{R}$ and $D^* : \Theta \rightrightarrows S$ be defined by

$$f^*(\theta) = \max\{f(x,\theta) : x \in \mathcal{D}(\theta)\}$$

and

$$\begin{aligned}\mathcal{D}^*(\theta) &= \arg\max\{f(x,\theta) : x \in \mathcal{D}(\theta)\} \\ &= \{x \in \mathcal{D}(\theta) : f(x,\theta) = f^*(\theta)\}.\end{aligned}$$

Then f^* is a continuous function on Θ, and $\mathcal{D}^*$ is a compact-valued, upper hemicontinuous correspondence on Θ.

Lemma 21 [35, Lemma 17.6]: A singleton-valued correspondence is upper hemicontinuous if and only if it is lower hemicontinuous, in which case it is continuous as a function.

Lemma 22: Let f be as defined in (12). Then, $f^{-1}(\delta) = (1+o(1))\log(1/\delta)$ as $\delta \to 0$, i.e.,

$$\lim_{\delta\to 0} \frac{\log(1/\delta)}{f^{-1}(\delta)} = 1.$$

Proof: Let $x = f^{-1}(\delta)$. Then,

$$\begin{aligned}\lim_{\delta\to 0}\frac{\log(1/\delta)}{f^{-1}(\delta)} &\overset{(a)}{=} \lim_{x\to+\infty} \frac{\log(\frac{1}{f(x)})}{x} \overset{(b)}{=} \lim_{x\to+\infty} \frac{-f'(x)/f(x)}{1} \\ &\overset{(c)}{=} \lim_{x\to+\infty} \frac{\frac{x^{K'-1}e^{-x}}{(K'-1)!}}{\sum_{i=1}^{K'} \frac{x^{i-1}e^{-x}}{(i-1)!}} = \lim_{x\to+\infty} \frac{\frac{x^{K'-1}}{(K'-1)!}}{\sum_{i=1}^{K'} \frac{x^{i-1}}{(i-1)!}} = 1,\end{aligned}$$

where (a) above follows the fact that $x \to \infty$ as $\delta \to 0$, (b) follows from the L'Hospital's rule, and (c) makes use of the fact that $f'(x) = \frac{-x^{K'-1}e^{-x}}{(K'-1)!}$. This completes the proof. □

Before proceeding further, we introduce some additional notations. For any $j \in [L]$ and $m \in [M]$, let

$$\Lambda_m^{(j)} := \begin{cases} \Lambda_m, & \text{if } S_m \subseteq Q_j, \\ \{\mathbf{0}^K\}, & \text{otherwise,} \end{cases}$$

where $\mathbf{0}^K$ denotes the all-zeros vector of dimension K. For each $j \in [L]$, noting that $\prod_{i=1}^{M} \Lambda_i^{(j)} := \Lambda_1^{(j)} \times \ldots \Lambda_M^{(j)}$ is compact and that the mapping $\omega \mapsto \widetilde{g}_v^{(j)}(\omega)$ is continuous, there exists a solution to $\max_{\omega\in\prod_{i=1}^{M}\Lambda_i^{(j)}} \widetilde{g}_v^{(j)}(\omega)$. Let

$$\widetilde{\omega}^{(j)}(v) \in \operatorname*{arg\,max}_{\omega\in\prod_{i=1}^{M}\Lambda_i^{(j)}} \widetilde{g}_v^{(j)}(\omega),$$

Further, let $\widetilde{\omega}(v) := \sum_{j=1}^{L} \widetilde{\omega}^{(j)}(v)$. Then, it is easy to verify that $\widetilde{\omega}(v) \in \Gamma$ is a common solution to

$$\max_{\omega\in\Gamma} \widetilde{g}_v(\omega), \max_{\omega\in\Gamma} \widetilde{g}_v^{(1)}(\omega), \dots, \max_{\omega\in\Gamma} \widetilde{g}_v^{(L)}(\omega). \tag{43}$$

Note that such common solution above is unique (we defer the proof of this fact to Theorem 14), which then implies that the solution to $\arg\max_{\omega\in\prod_{i=1}^{M}\Lambda_i^{(j)}} \widetilde{g}_v^{(j)}(\omega)$ is unique. Hence, $\widetilde{\omega}^{(j)}(v)$ and $\widetilde{\omega}(v)$ are well-defined.

Lemma 23: Given any problem instance $v \in \mathcal{P}$, under $\Pi_{\text{Het-TS}}$,

$$\lim_{t\to\infty} \|\widetilde{\omega}(\hat{v}(t)) - \widetilde{\omega}(v)\|_\infty = 0 \quad \text{almost surely.} \tag{44}$$

Consequently, for any $m \in [M]$ and $i \in S_m$,

$$\lim_{t\to\infty} \left| \frac{N_{i,m}(t)}{t} - \widetilde{\omega}(v)_{i,m} \right| = 0 \quad \text{almost surely.} \tag{45}$$

Proof: Fix $j \in [L]$ and $v \in \mathcal{P}$ arbitrarily. By the strong law of large numbers, it follows that for any $i \in [K]$ and $m \in S_m$,

$$\begin{aligned} &\lim_{t\to\infty} \hat{\mu}_{i,m}(t) = \mu_{i,m}(v) \quad \text{almost surely} \\ &\implies \lim_{t\to\infty} \hat{\mu}_i(t) = \mu_i(v) \quad \text{almost surely} \qquad (46) \\ &\implies \lim_{t\to\infty} \Delta_i(\hat{v}(t)) = \Delta_i(v) \quad \text{almost surely.} \qquad (47) \end{aligned}$$

For any $v' \in \mathcal{P}$, note that $\widetilde{g}_{v'}^{(j)}(\omega)$ is a function of $(\Delta(v'), \omega)$ for $\Delta(v') \in (\mathbb{R}^+)^K$ and $\omega \in \prod_{i=1}^{M} \Lambda_i^{(j)}$. From Lemma 20 and Lemma 21 that for any $\epsilon_1 > 0$, there exists $\epsilon_2 > 0$ such that for all $v' \in \mathcal{P}$ with $\|\Delta(v) - \Delta(v')\|_\infty \le \epsilon_2$,

$$\|\widetilde{\omega}^{(j)}(v) - \widetilde{\omega}^{(j)}(v')\|_\infty \le \epsilon_1. \tag{48}$$

Combining (47) and (48), it follows that

$$\lim_{t\to\infty} \|\widetilde{\omega}^{(j)}(v) - \widetilde{\omega}^{(j)}(\hat{v}(t))\|_\infty = 0 \quad \text{almost surely,}$$

which in turn implies that

$$\lim_{t\to\infty} \|\widetilde{\omega}(v) - \widetilde{\omega}(\hat{v}(t))\|_\infty = 0 \tag{49}$$

almost surely. Recall the definition of $\hat{\omega}_{i,m}(t)$ in (9), which means $\hat{\omega}_{i,m}(t) = \widetilde{\omega}_{i,m}\big(\hat{v}(b_{r(t)})\big)$. Then, by (49) for any $m \in [M]$ and $i \in S_m$ we have

$$\lim_{t\to\infty} \|\widetilde{\omega}(v) - \hat{\omega}(t)\|_\infty = 0$$

almost surely.

Consequently, by [13, Lemma 17], for any $m \in [M]$ and $i \in S_m$,

$$\lim_{t\to\infty} \left| \frac{N_{i,m}(t)}{t} - \widetilde{\omega}(v)_{i,m} \right| = 0 \quad \text{almost surely.}$$

□

Lemma 24: Given any problem instance $v \in \mathcal{P}$, under $\Pi_{\text{Het-TS}}$,

$$\lim_{t\to\infty} \frac{Z(t)}{t} = g_v(\widetilde{\omega}(v)) \quad \text{almost surely.}$$

Proof: Fix $v \in \mathcal{P}$ arbitrarily. Define $\hat{N}(t) \in \Gamma$ as

$$\hat{N}_{i,m}(t) := \frac{N_{i,m}(t)}{t}, \quad i \in S_m,\ m \in [M]$$

Then,

$$\begin{aligned} &\frac{Z(t)}{t} \\ &= \inf_{v'\in\mathrm{Alt}(\hat{v}(t))} \sum_{m=1}^{M} \sum_{i\in S_m} \frac{N_{i,m}(t)}{t} \frac{(\mu_{i,m}(v') - \hat{\mu}_{i,m}(t))^2}{2} \\ &= \inf_{v'\in\mathrm{Alt}(\hat{v}(t))} \sum_{m=1}^{M} \sum_{i\in S_m} \hat{N}_{i,m}(t) \frac{(\mu_{i,m}(v') - \hat{\mu}_{i,m}(t))^2}{2} \\ &= g_{\hat{v}(t)}(\hat{N}(t)) \\ &\overset{(a)}{=} \min_{(i,j)\in\mathcal{C}(\hat{v}(t))} \frac{\big(\hat{\mu}_i(t) - \hat{\mu}_j(t)\big)^2/2}{\sum_{\iota\in\{i,j\}} \frac{1}{M_\iota^2} \sum_{m=1}^{M} \mathbf{1}_{\{\iota\in S_m\}} \frac{1}{\hat{N}_{\iota,m}(t)}}, \qquad (50) \end{aligned}$$

where (a) follows from (18) of Lemma 2. Because $v \in \mathcal{P}$ and $\lim_{t\to\infty} \hat{\mu}_i(t) = \mu_i$ almost surely for all $i \in [K]$ from (46), we get that

$$\lim_{t\to\infty} \mathcal{C}(\hat{v}(t)) = \mathcal{C}(v) \quad \text{almost surely,} \tag{51}$$

where $\mathcal{C}(\cdot)$ is as defined in (17). Combining (44), (45), (46), (51), and Lemma 23, we get that almost surely,

$$\begin{aligned} &\lim_{t\to\infty} \frac{Z(t)}{t} \\ &= \lim_{t\to\infty} \min_{(i,j)\in\mathcal{C}(\hat{v}(t))} \frac{\big(\hat{\mu}_i(t) - \hat{\mu}_j(t)\big)^2/2}{\sum_{\iota\in\{i,j\}} \frac{1}{M_\iota^2} \sum_{m=1}^{M} \mathbf{1}_{\{\iota\in S_m\}} \frac{1}{\hat{N}_{\iota,m}(t)}} \\ &= \min_{(i,j)\in\mathcal{C}(v)} \frac{\big(\mu_i(v) - \mu_j(v)\big)^2/2}{\sum_{\iota\in\{i,j\}} \frac{1}{M_\iota^2} \sum_{m=1}^{M} \mathbf{1}_{\{\iota\in S_m\}} \frac{1}{\widetilde{\omega}_{\iota,m}(v)}} \\ &= g_v(\widetilde{\omega}(v)). \end{aligned}$$

This completes the desired proof. □

Lemma 25: Given any confidence level $\delta \in (0,1)$,

$$\tau_\delta(\Pi_{\text{Het-TS}}) < +\infty \quad \text{almost surely.}$$

Proof: As a consequence of Lemma 24, we have

$$\begin{aligned} \lim_{t\to\infty} \frac{\beta(t,\delta)}{Z(t)} &= \lim_{t\to\infty} \frac{K' \log(t^2 + t) + f^{-1}(\delta)}{t\frac{Z(t)}{t}} \\ &= \lim_{t\to\infty} \frac{K' \log(t^2 + t) + f^{-1}(\delta)}{t g_v(\widetilde{\omega}(v))} \\ &= 0 \quad \text{almost surely.} \end{aligned}$$

Therefore, there almost surely exists $0 < T < +\infty$ such that $Z(t) > \beta(t,\delta)$ for all $t \ge T$, thus proving that $\tau_\delta(\Pi_{\text{HET-TS}})$ is finite almost surely. □

Lemma 26: Given any problem instance $v \in \mathcal{P}$ and $\epsilon \in (0, g_v(\widetilde{\omega}(v)))$, there exists $\delta_{\text{upper}}(v,\epsilon) > 0$ such that for any $\delta \in (0, \delta_{\text{upper}}(v,\epsilon))$,

$$t\, g_v(\widetilde{\omega}(v)) > \beta(t,\delta) + t\,\epsilon \tag{52}$$

for all $t \ge T_{\text{last}}(v,\delta,\epsilon)$, where

$$\begin{aligned} T_{\text{last}}(v,\delta,\epsilon) &:= 1 + \frac{f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon} \\ &+ \frac{K'}{g_v(\widetilde{\omega}(v)) - \epsilon} \log\left(\left(\frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon} \right)^2 + \frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon} \right). \end{aligned}$$

Proof: Fix $v \in \mathcal{P}$ and $\epsilon \in (0, g_v(\widetilde{\omega}(v)))$ arbitarily. Recall that $\beta(t,\delta) = K' \log(t^2+t) + f^{-1}(\delta)$.

To prove Lemma 26, it suffices to verify that

1) The derivative of the left-hand of (52) with respect to t is greater than that of the right-hand side of (52) for all $t \geq T_{\text{last}}(v,\delta,\epsilon)$, and
2) Eq. (52) holds for $t = T_{\text{last}}(v,\delta,\epsilon)$.

In order to verify that the condition in item 1 above holds, we note from Lemma 22 that $\lim_{\delta\to 0} f^{-1}(\delta) = +\infty$, as a consequence of which we get that there exists $\delta_{\text{upper}}(v,\epsilon) > 0$ such that for all $\delta \in (0, \delta_{\text{upper}}(v,\epsilon))$,

$$T_{\text{last}}(v,\delta,\epsilon) < \frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon}$$

and

$$\frac{f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon} \geq \frac{3K'}{g_v(\widetilde{\omega}(v)) - \epsilon}.$$

Notice that the derivative of the left-hand side of (52) with respect to t is equal to $g_v(\widetilde{\omega}(v))$, whereas that of the right-hand side of (52) is equal to $\frac{K'(2+\frac{1}{t})}{t+1} + \epsilon$. Hence, to verify the condition in item 1, we need to demonstrate that

$$g_v(\widetilde{\omega}(v)) - \epsilon > \frac{K'(2+\frac{1}{t})}{t+1} \quad \text{for all } t \geq T_{\text{last}}(v,\delta,\epsilon). \tag{53}$$

We note that for all $t \geq T_{\text{last}}(v,\delta,\epsilon)$,

$$\begin{aligned} t+1 > t \geq T_{\text{last}}(v,\delta,\epsilon) &\geq \frac{f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon} \\ &\geq \frac{3K'}{g_v(\widetilde{\omega}(v)) - \epsilon} \geq \frac{(2+\frac{1}{t})K'}{g_v(\widetilde{\omega}(v)) - \epsilon}, \end{aligned} \tag{54}$$

where in writing the last line above, we use the fact that $3 \geq 2 + \frac{1}{t}$ whenever $t \geq 1$. We then obtain (53) upon rearranging (54) and using the fact that $\epsilon > 0$. This verifies the condition in item 1. To verify the condition in item 2 above, we note that for all $T_{\text{last}}(v,\delta,\epsilon) \leq t \leq \frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}$, we have

$$\begin{aligned} t &\geq T_{\text{last}}(v,\delta,\epsilon) \\ &> \frac{f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon} \\ &\quad + \frac{K'}{g_v(\widetilde{\omega}(v)) - \epsilon} \log\left(\left(\frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon}\right)^2 + \frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon}\right) \\ &\geq \frac{f^{-1}(\delta)}{g_v(\widetilde{\omega}(v)) - \epsilon} + K' \log\left(t^2+t\right) \frac{1}{g_v(\widetilde{\omega}(v)) - \epsilon}. \end{aligned} \tag{55}$$

Equivalently, upon rearranging the terms in (55), we get

$$t\, g_v(\widetilde{\omega}(v)) > \beta(t,\delta) + t\,\epsilon \tag{56}$$

for all $T_{\text{last}}(v,\delta,\epsilon) \leq t \leq \frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}$. In particular, noting that (56) holds for $t = T_{\text{last}}(v,\delta,\epsilon)$ verifies the condition in item 2, and thereby completes the proof.

□

With the above ingredients in place, we are now ready to prove Theorem 11.

Proof of Theorem 11: Fix a problem instance $v \in \mathcal{P}$ arbitrarily. Given any $\epsilon > 0$, let $T_{\text{cvg}}(v,\epsilon)$ denote the smallest positive integer such that

$$\left|\frac{Z(t)}{t} - g_v(\widetilde{\omega}(v))\right| \leq \epsilon \quad \forall\, t \geq T_{\text{cvg}}(v,\epsilon).$$

From Lemma 24, we know that $T_{\text{cvg}}(v,\epsilon) < +\infty$ almost surely. Therefore, for any $\epsilon \in (0, g_v(\widetilde{\omega}(v)))$ and $\delta \in (0, \delta_{\text{upper}}(v,\epsilon))$, it follows from Lemma 26 that

$$Z(t) > \beta(t,\delta) \quad \forall\, t \geq \max\left\{T_{\text{cvg}}(v,\epsilon), T_{\text{last}}(v,\delta,\epsilon), K\right\} \tag{57}$$

almost surely, where $\delta_{\text{upper}}(v,\epsilon)$ and $T_{\text{last}}(v,\delta,\epsilon)$ are as defined in Lemma 26. Recall that $b_r = \lceil (1+\lambda)^r \rceil$ in the Het-TS algorithm. From (57), it follows that

$$\tau_\delta(\Pi_{\text{Het-TS}}) \leq (1+\lambda) \max\left\{T_{\text{cvg}}(v,\epsilon), T_{\text{last}}(v,\delta,\epsilon), K\right\} + 1$$

almost surely $\forall \epsilon \in (0, g_v(\widetilde{\omega}(v)))$ and $\forall \delta \in (0, \delta_{\text{upper}}(v,\epsilon))$, which implies that

$$\begin{aligned} \tau_\delta(\Pi_{\text{Het-TS}}) \leq (1+\lambda)\, T_{\text{cvg}}(v,\epsilon) &+ (1+\lambda)\, T_{\text{last}}(v,\delta,\epsilon) \\ &+ (1+\lambda)K + 1 \end{aligned} \tag{58}$$

almost surely. Then, $\forall \epsilon \in (0, g_v(\widetilde{\omega}(v)))$, the following set of relations hold almost surely:

$$\begin{aligned} &\limsup_{\delta\to 0} \frac{\tau_\delta(\Pi_{\text{Het-TS}})}{\log\left(\frac{1}{\delta}\right)} \\ &\leq \limsup_{\delta\to 0} \Bigg[\frac{(1+\lambda)\, T_{\text{cvg}}(v,\epsilon) + (1+\lambda)\, T_{\text{last}}(v,\delta,\epsilon)}{\log\left(\frac{1}{\delta}\right)} \\ &\qquad\qquad\qquad\qquad + \frac{(1+\lambda)K + 1}{\log\left(\frac{1}{\delta}\right)}\Bigg] \\ &\overset{(a)}{=} \limsup_{\delta\to 0} \frac{(1+\lambda)\, T_{\text{last}}(v,\delta,\epsilon)}{\log\left(\frac{1}{\delta}\right)} \\ &\overset{(b)}{=} \limsup_{\delta\to 0} \Bigg[\frac{(1+\lambda)\, \left(1 + \frac{f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}\right)}{\log\left(\frac{1}{\delta}\right)} \\ &+ \frac{(1+\lambda)\, \frac{K'}{g_v(\widetilde{\omega}(v))-\epsilon} \log\left(\left(\frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}\right)^2 + \frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}\right)}{\log\left(\frac{1}{\delta}\right)}\Bigg] \\ &\overset{(c)}{=} \frac{1+\lambda}{g_v(\widetilde{\omega}(v)) - \epsilon} \overset{(d)}{\leq} \frac{1+\lambda}{\frac{1}{2}c^*(v)^{-1} - \epsilon}, \end{aligned} \tag{59}$$

where (a) follows from the fact that $T_{\text{cvg}}(v,\epsilon)$ is not a function of δ and that $T_{\text{cvg}}(v,\epsilon) < +\infty$ almost surely, (b) follows from the definition of $T_{\text{last}}(v,\delta,\epsilon)$, (c) follows from Lemma 22, and (d) makes use of Lemma 2. Letting $\epsilon \to 0$ in (59), we get

$$\limsup_{\delta\to 0} \frac{\tau_\delta(\Pi_{\text{Het-TS}})}{\log\left(\frac{1}{\delta}\right)} \leq 2\,(1+\lambda)\, c^*(v) \quad \text{almost surely.}$$

Taking expectation on both sides of (58), we get

$$\begin{aligned} \mathbb{E}_v^{\Pi_{\text{Het-TS}}}[\tau_\delta(\Pi_{\text{Het-TS}})] \leq (1+\lambda)\, &\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[T_{\text{cvg}}(v,\epsilon)] \\ &+ (1+\lambda)\, T_{\text{last}}(v,\delta,\epsilon) + 1 \end{aligned}$$

for all $\epsilon \in (0, g_v(\widetilde{\omega}(v)))$ and $\delta \in (0, \delta_{\text{upper}}(v,\epsilon))$, from which it follows that

$$\limsup_{\delta\to 0} \frac{\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[\tau_\delta(\Pi_{\text{Het-TS}})]}{\log\left(\frac{1}{\delta}\right)}$$

$$\begin{aligned}
&\leq \limsup_{\delta\to 0}\Bigg[\frac{(1+\lambda)\,\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[T_{\text{cvg}}(v,\epsilon)]}{\log\left(\frac{1}{\delta}\right)} \\
&\qquad\qquad + \frac{(1+\lambda)\,T_{\text{last}}(v,\delta,\epsilon)+(1+\lambda)K+1}{\log\left(\frac{1}{\delta}\right)}\Bigg] \\
&\overset{(a)}{=} \limsup_{\delta\to 0}\frac{(1+\lambda)\,T_{\text{last}}(v,\delta,\epsilon)}{\log\left(\frac{1}{\delta}\right)} \\
&\overset{(b)}{=} \limsup_{\delta\to 0}\Bigg[\frac{(1+\lambda)\left(1+\frac{f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}\right)}{\log\left(\frac{1}{\delta}\right)} \\
&+ \frac{(1+\lambda)\frac{K'}{g_v(\widetilde{\omega}(v))-\epsilon}\,\log\left(\left(\frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}\right)^2+\frac{2f^{-1}(\delta)}{g_v(\widetilde{\omega}(v))-\epsilon}\right)}{\log\left(\frac{1}{\delta}\right)}\Bigg] \\
&\overset{(c)}{=} \frac{1+\lambda}{g_v(\widetilde{\omega}(v))-\epsilon} \\
&\overset{(d)}{\leq} \frac{1+\lambda}{\frac{1}{2}c^*(v)^{-1}-\epsilon},
\end{aligned} \tag{60}$$

where (a) follows from the fact that $\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[T_{\text{cvg}}(v,\epsilon)]$ does not depend on δ and that $\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[T_{\text{cvg}}(v,\epsilon)] < +\infty$ because of the following Lemma 27, (b) follows from the definition of $T_{\text{last}}(v,\delta,\epsilon)$, (c) follows from Lemma 22, and (d) makes use of Lemma 2. Letting $\epsilon \to 0$ in (60), we get

$$\limsup_{\delta\to 0}\frac{\mathbb{E}_v^{\Pi_{\text{Het-TS}}}(\tau_\delta(\Pi_{\text{Het-TS}}))}{\log\left(\frac{1}{\delta}\right)} \leq 2\,(1+\lambda)\,c^*(v).$$

This completes the desired proof. □

Lemma 27: Given any $\epsilon > 0$ and $v \in \mathcal{P}$,

$$\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[T_{\text{cvg}}(v,\epsilon)] < +\infty.$$

Proof of Lemma 27: Fix $v \in \mathcal{P}$ and $\epsilon > 0$. For any $\epsilon' > 0$, let $T'_N(v,\epsilon')$ denote the smallest positive integer such that for all $t \geq T'_N(v,\epsilon')$,

$$\left|\frac{N_{i,m}(t)}{t} - \widetilde{\omega}_{i,m}(v)\right| \leq \epsilon' \quad \forall m \in [M], i \in S_m.$$

Let $T'_\mu(v,\epsilon')$ denote the smallest positive integer such that for all $t \geq T'_\mu(v,\epsilon')$,

$$|\hat{\mu}_{i,m}(t) - \mu_{i,m}(v)| \leq \epsilon' \quad \forall m \in [M], i \in S_m.$$

From (50), we know that there exists $\epsilon'_1 \in (0,1)$ such that

$$T_{\text{cvg}}(v,\epsilon) \leq \max\{T'_N(v,\epsilon'_1), T'_\mu(v,\epsilon'_1)\}. \tag{61}$$

Let $T'_{\hat{\omega}}(v,\epsilon')$ denote the smallest positive integer such that for all $t \geq T'_{\hat{\omega}}(v,\epsilon')$,

$$|\hat{\omega}_{i,m}(t) - \widetilde{\omega}_{i,m}(v)| \leq \epsilon' \quad \forall m \in [M], i \in S_m.$$

Let $T'_{\widetilde{\omega}}(v,\epsilon')$ denote the smallest positive integer such that for all $t \geq T'_{\widetilde{\omega}}(v,\epsilon')$,

$$|\widetilde{\omega}_{i,m}(\hat{v}(t)) - \widetilde{\omega}_{i,m}(v)| \leq \epsilon' \quad \forall m \in [M], i \in S_m.$$

Recall the definition of $\hat{\omega}_{i,m}(t)$ in (9). We then have

$$T'_{\hat{\omega}}(v,\epsilon') \leq (1+\lambda)T'_{\widetilde{\omega}}(v,\epsilon') + 1 \quad \forall \epsilon' > 0. \tag{62}$$

In addition, by the D-tracking rule in (10), we have for all $t \geq \max\{\frac{T'_{\hat{\omega}}(v,\epsilon')}{\epsilon'}, \frac{1}{(\epsilon')^2}\}$ and all $\epsilon' \in (0,1)$

$$\begin{aligned}
&N_{i,m}(t) \\
&\overset{(a)}{\leq} \max\{N_{i,m}(T'_{\hat{\omega}}(v,\epsilon')), t(\widetilde{\omega}_{i,m}(v)+\epsilon')+1, \sqrt{t}+1\} \\
&\leq \max\{t\epsilon', t(\widetilde{\omega}_{i,m}(v)+\epsilon')+1, t\epsilon'+1\} \\
&\leq t(\widetilde{\omega}_{i,m}(v)+\epsilon')+1.
\end{aligned} \tag{63}$$

In (a), the first term inside $\max$ follows from the fact that $\epsilon' < 1$, while the second and third terms follow from (10). Consequently, for all $t \geq \max\{\frac{T'_{\hat{\omega}}(v,\epsilon')}{\epsilon'}, \frac{1}{\epsilon'^2}, \frac{K}{\epsilon'}\}$ and all $\epsilon' \in (0,1)$, we have

$$\begin{aligned}
N_{i,m}(t) &= t - \sum_{j\neq i: j\in S_m} N_{j,m}(t) \\
&\geq t - t\sum_{j\neq i: j\in S_m}(\widetilde{\omega}_{j,m}(v)+\epsilon') - K \\
&= t(\widetilde{\omega}_{i,m}(v) - K\epsilon') - K \\
&\overset{(a)}{\geq} t(\widetilde{\omega}_{i,m}(v) - (K+1)\epsilon'),
\end{aligned} \tag{64}$$

where (a) follows $t > \frac{K}{\epsilon'}$. Then, from (63) and (64), we have

$$\begin{aligned}
&T'_N(v,\epsilon'_1) \\
&\leq \max\left\{\frac{(K+1)T'_{\hat{\omega}}(v,\frac{\epsilon'_1}{K+1})}{\epsilon'_1}, \frac{(K+1)^2}{(\epsilon'_1)^2}, \frac{K(K+1)}{\epsilon'_1}\right\} \\
&\leq \frac{(K+1)T'_{\hat{\omega}}(v,\frac{\epsilon'_1}{K+1})}{\epsilon'_1} + \frac{(K+1)^2}{(\epsilon'_1)^2} + \frac{K(K+1)}{\epsilon'_1} \\
&\overset{(a)}{\leq} \frac{(K+1)((1+\lambda)T'_{\widetilde{\omega}}(v,\frac{\epsilon'_1}{K+1})+1)}{\epsilon'_1} + \frac{(K+1)^2}{(\epsilon'_1)^2} \\
&\quad + \frac{K(K+1)}{\epsilon'_1},
\end{aligned} \tag{65}$$

where (a) follows from (62). From (48), we know that there exists $\epsilon'_2 > 0$ such that

$$T'_{\widetilde{\omega}}\left(v, \frac{\epsilon'_1}{K+1}\right) \leq T'_\mu(v,\epsilon'_2). \tag{66}$$

Combining (61), (65), and (66), we have

$$\begin{aligned}
T_{\text{cvg}}(v,\epsilon) &\leq T'_\mu(v,\epsilon'_1) + \frac{(K+1)((1+\lambda)T'_\mu(v,\epsilon'_2)+1)}{\epsilon'_1} \\
&\quad + \frac{(K+1)^2}{(\epsilon'_1)^2} + \frac{K(K+1)}{\epsilon'_1}.
\end{aligned}$$

Note that by the force exploration in D-tracking rule, there exists a constant $T_{\text{start}} > 0$ (depending only on K) such that for all $t \geq T_{\text{start}}$,

$$N_{i,m}(t) \geq \sqrt{\frac{t-1}{K}} - 1 \quad \forall m \in [M], i \in S_m,$$

which in turn implies that for $\epsilon' \in \{\epsilon'_1, \epsilon'_2\}$,

$$\begin{aligned}
&\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[T'_\mu(v,\epsilon')] \\
&\leq T_{\text{start}} + \sum_{t=T_{\text{start}}}^{+\infty}\mathbb{P}_v^{\Pi_{\text{Het-TS}}}(T'_\mu(v,\epsilon') > t) \\
&\leq T_{\text{start}} + \sum_{t=T_{\text{start}}}^{+\infty} 2MK\exp\left(-\frac{1}{2}\left(\sqrt{\frac{t-1}{K}}-1\right)(\epsilon')^2\right) \\
&< +\infty.
\end{aligned}$$

Hence, it follows that

$$\mathbb{E}_v^{\Pi_{\text{Het-TS}}}[T_{\text{cvg}}(v,\epsilon)] < +\infty.$$

thus concluding the proof. □

G. Proof of Theorem 14

Fix a problem instance $v \in \mathcal{P}$ arbitrarily. Recall that there exists a common solution to (13) (see the discussion after (43)). The following results show that this solution satisfies *balanced condition* (Def. 3) and *pseudo-balanced condition* (Def. 13) and that it is unique.

Lemma 28: The common solution to (13) satisfies the pseudo-balanced condition.

Proof: Let $\widetilde{\omega}(v)$ be a common solution to (13). Let

$$Q_j^{\min} := \mathop{\arg\min}_{i \in Q_j} \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}} \frac{1}{\widetilde{\omega}_{i,m}(v)}}, \quad j \in [L].$$

Suppose $\widetilde{\omega}(v)$ does not meet the *pseudo-balanced condition*. Then, there exists $l \in [L]$ such that $Q_l \neq Q_l^{\min}$. We now recursively construct an $\omega \in \Gamma$ such that $\widetilde{g}_v^{(l)}(\omega) > \widetilde{g}_v^{(l)}(\widetilde{\omega}(v))$, thereby leading to a contradiction.

Step 1: Initialization: Set $\omega^{(0)} := \widetilde{\omega}(v)$ and $Q^{(0)} := Q_l^{\min}$.

Step 2: Iterations: For each $s \in \{0, 1, 2, \ldots |Q_l^{\min}| - 1\}$, note that there exists $i_1 \in Q^{(s)}$, $i_2 \in Q_l \setminus Q^{(s)}$, and $m' \in [M]$ such that $i_1, i_2 \in S_{m'}$. Let $\epsilon > 0$ be sufficiently small so that

$$\frac{\Delta_{i_2}^2(v)}{\frac{1}{M_{i_2}^2}\sum_{m=1}^{M} \mathbf{1}_{\{i_2\in S_m\}} \frac{1}{\omega_{i_2,m}^{(s)} - \epsilon}} > \min_{i \in Q_l} \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}} \frac{1}{\widetilde{\omega}_{i,m}(v)}} = \widetilde{g}_v^{(l)}(\widetilde{\omega}(v)).$$

Then, we construct $\omega^{(s+1)}$ as $\forall\, m \in [M], i \in S_m$,

$$\omega_{i,m}^{(s+1)} := \begin{cases} \omega_{i,m}^{(s)} - \epsilon, & \text{if } i = i_2, m = m' \\ \omega_{i,m}^{(s)} + \epsilon, & \text{if } i = i_1, m = m', \\ \omega_{i,m}^{(s)}, & \text{otherwise}, \end{cases}$$

and set $Q^{(s+1)} := Q^{(s)} \setminus \{i_1\}$. We then have

$$\frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}} \frac{1}{\omega_{i,m}^{(s+1)}}} > \widetilde{g}_v^{(l)}(\widetilde{\omega}(v)) \quad \forall\, i \in Q_l \setminus Q^{(s+1)},$$

and

$$\frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^{M} \mathbf{1}_{\{i\in S_m\}} \frac{1}{\omega_{i,m}^{(s+1)}}} = \widetilde{g}_v^{(l)}(\widetilde{\omega}(v)) \quad \forall\, i \in Q^{(s+1)}.$$

By following the above procedure for $s \in \{0, 1, 2, \ldots, |Q_l^{\min}|\}$, we arrive at $\omega^{(|Q_l^{\min}|)}$ such that $\widetilde{g}_v^{(l)}(\omega^{(|Q_l^{\min}|)}) > \widetilde{g}_v^{(l)}(\widetilde{\omega}(v))$, which is the clearly a contradiction. □

Lemma 29: The common solution to (13) satisfies *balanced condition*.

Proof: Let $\widetilde{\omega}(v)$ be a common solution to (13). From Lemma 28, we know that $\widetilde{\omega}(v)$ satisfies *pseudo-balanced condition*. Suppose now that $\widetilde{\omega}(v)$ does not satisfy *balanced condition*. Then, there exists $l \in [L]$, $m_1, m_2 \in [M]$, and $i_1, i_2 \in S_{m_1} \cap S_{m_2} \subseteq Q_l$ such that

$$\frac{\widetilde{\omega}_{i_1,m_1}(v)}{\widetilde{\omega}_{i_2,m_1}(v)} > \frac{\widetilde{\omega}_{i_1,m_2}(v)}{\widetilde{\omega}_{i_2,m_2}(v)}. \tag{67}$$

Because $\widetilde{\omega}(v)$ satisfies *pseudo-balanced condition*, we must have

$$\frac{\Delta_{i_2}^2(v)}{\frac{1}{M_{i_2}^2}\sum_{m=1}^{M} \mathbf{1}_{\{i_2\in S_m\}} \frac{1}{\widetilde{\omega}_{i_2,m}(v)}} = \frac{\Delta_{i_1}^2(v)}{\frac{1}{M_{i_1}^2}\sum_{m=1}^{M} \mathbf{1}_{\{i_1\in S_m\}} \frac{1}{\widetilde{\omega}_{i_1,m}(v)}} = \widetilde{g}_v^{(l)}(\widetilde{\omega}(v)).$$

Note that (67) implies that $\frac{\widetilde{\omega}_{i_2,m_1}(v)}{\widetilde{\omega}_{i_2,m_2}(v)} < \frac{\widetilde{\omega}_{i_1,m_1}(v)}{\widetilde{\omega}_{i_1,m_2}(v)}$. Let ρ be any value such that

$$\left(\frac{\widetilde{\omega}_{i_2,m_1}(v)}{\widetilde{\omega}_{i_2,m_2}(v)}\right)^2 < \rho < \left(\frac{\widetilde{\omega}_{i_1,m_1}(v)}{\widetilde{\omega}_{i_1,m_2}(v)}\right)^2. \tag{68}$$

Using the fact that the derivative of $x \mapsto \frac{1}{x}$ is $-\frac{1}{x^2}$, we have

$$\frac{1}{\widetilde{\omega}_{i_1,m_1}(v) - \rho\epsilon} - \frac{1}{\widetilde{\omega}_{i_1,m_1}(v)} = \frac{\rho\epsilon}{\widetilde{\omega}_{i_1,m_1}^2(v)} + o(\epsilon) \quad \text{as } \epsilon \to 0,$$

and

$$\frac{1}{\widetilde{\omega}_{i_1,m_2}(v)} - \frac{1}{\widetilde{\omega}_{i_1,m_2}(v) + \epsilon} = \frac{\epsilon}{\widetilde{\omega}_{i_1,m_2}^2(v)} + o(\epsilon) \quad \text{as } \epsilon \to 0.$$

Here $o(\epsilon)$ is a function in ϵ that satisfies $\lim_{\epsilon\to 0}\frac{o(\epsilon)}{\epsilon} = 0$. By combining these equations,

$$\begin{aligned} &\frac{1}{\widetilde{\omega}_{i_1,m_1}(v) - \rho\epsilon} - \frac{1}{\widetilde{\omega}_{i_1,m_1}(v)} - \left(\frac{1}{\widetilde{\omega}_{i_1,m_2}(v)} - \frac{1}{\widetilde{\omega}_{i_1,m_2}(v) + \epsilon}\right) \\ &= \frac{\rho\epsilon}{\widetilde{\omega}_{i_1,m_1}^2(v)} + o(\epsilon) - \left(\frac{\epsilon}{\widetilde{\omega}_{i_1,m_2}^2(v)} + o(\epsilon)\right) \\ &= \epsilon\left[\frac{\rho}{\widetilde{\omega}_{i_1,m_1}^2(v)}\big(1 + o_\epsilon(1)\big) - \frac{1}{\widetilde{\omega}_{i_1,m_2}^2(v)}\big(1 + o_\epsilon(1)\big)\right] \end{aligned} \tag{69}$$

where $o_\epsilon(1)$ is a term that vanishes as $\epsilon \downarrow 0$. From (68), we have,

$$\rho < \frac{\widetilde{\omega}_{i_1,m_1}^2(v)}{\widetilde{\omega}_{i_1,m_2}^2(v)} \iff \frac{\rho}{\widetilde{\omega}_{i_1,m_1}^2(v)} < \frac{1}{\widetilde{\omega}_{i_1,m_2}^2(v)} \tag{70}$$

By (69) and (70), there exists $\epsilon_1 > 0$ such that for all $\epsilon \in (0, \epsilon_1]$,

$$\frac{1}{\widetilde{\omega}_{i_1,m_1}(v) - \rho\epsilon} - \frac{1}{\widetilde{\omega}_{i_1,m_1}(v)} - \left(\frac{1}{\widetilde{\omega}_{i_1,m_2}(v)} - \frac{1}{\widetilde{\omega}_{i_1,m_2}(v) + \epsilon}\right) < 0.$$

In other words, for all $\epsilon \in (0, \epsilon_1]$,

$$\frac{1}{\widetilde{\omega}_{i_1,m_1}(v)} + \frac{1}{\widetilde{\omega}_{i_1,m_2}(v)} > \frac{1}{\widetilde{\omega}_{i_1,m_1}(v) - \rho\epsilon} + \frac{1}{\widetilde{\omega}_{i_1,m_2}(v) + \epsilon}. \tag{71}$$

Similarly, there exists $\epsilon_2 > 0$ for any $\epsilon \in (0, \epsilon_2]$ such that

$$\frac{1}{\widetilde{\omega}_{i_2,m_1}(v)} + \frac{1}{\widetilde{\omega}_{i_2,m_2}(v)} > \frac{1}{\widetilde{\omega}_{i_2,m_1}(v) + \rho\epsilon} + \frac{1}{\widetilde{\omega}_{i_2,m_2}(v) - \epsilon}. \tag{72}$$

Set $\epsilon = \min\{\epsilon_1, \epsilon_2\}$. Let $\omega' \in \Gamma$ be defined for all $m \in [M]$ and $i \in S_m$ as

$$\omega'_{i,m} := \begin{cases} \widetilde{\omega}_{i,m}(v) - \rho\epsilon, & \text{if } i = i_1, m = m_1 \\ \widetilde{\omega}_{i,m}(v) + \epsilon, & \text{if } i = i_1, m = m_2 \\ \widetilde{\omega}_{i,m}(v) + \rho\epsilon, & \text{if } i = i_2, m = m_1 \\ \widetilde{\omega}_{i,m}(v) - \epsilon, & \text{if } i = i_2, m = m_2 \\ \widetilde{\omega}_{i,m}(v), & \text{otherwise.} \end{cases}$$

Then, from (71) and (72), we have $\forall\, i \in \{i_1, i_2\}$

$$\frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\omega'_{i,m}}} > \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)}}, \tag{73}$$

and $\forall\, i \in Q_l \setminus \{i_1, i_2\}$

$$\frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\omega'_{i,m}}} = \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)}}. \tag{74}$$

We then consider the following two cases.

Case 1: $Q_l = \{i_1, i_2\}$. In this case, it follows from (73) that $\widetilde{g}_v^{(l)}(\omega') > \widetilde{g}_v^{(l)}(\widetilde{\omega}(v))$, which contradicts with the fact that $\widetilde{\omega}(v)$ is an optimum solution to $\max_{\omega\in\Gamma}\widetilde{g}_v(\omega)$.

Case 2: $\{i_1, i_2\} \subsetneq Q_l$. In this case, it follows from (74) that $\widetilde{g}_v^{(j)}(\omega') = \widetilde{g}_v^{(j)}(\widetilde{\omega}(v))$ for all $j \in [L]$, which implies that ω' is a common solution to (13) just as $\widetilde{\omega}(v)$ is. However, note that the right-hand sides of (74) and (73) are equal because $\widetilde{\omega}(v)$ satisfies *pseudo-balanced condition*. As a result, it follows that $\forall\, i \in Q_l \setminus \{i_1, i_2\}$

$$\frac{\Delta_{i_1}^2(v)}{\frac{1}{M_{i_1}^2}\sum_{m=1}^M \mathbf{1}_{\{i_1\in S_m\}}\frac{1}{\omega'_{i_1,m}}} > \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\omega'_{i,m}}}.$$

This shows that ω' does not meet *pseudo-balanced condition*, thereby contradicting Lemma 28. □

Lemma 30: The common solution to (13) is unique.

Proof: Suppose that $\widetilde{\omega}(v)$ and $\widetilde{\omega}'(v)$ are two common solutions to (13). Suppose further that $\widetilde{\omega}(v) \neq \widetilde{\omega}'(v)$. In the following, we arrive at a contradiction. Let $\widetilde{\omega}^{\rm avg}(v) := (\widetilde{\omega}(v) + \widetilde{\omega}'(v))/2$. From Lemma 28, we know that $\widetilde{\omega}(v)$ and $\widetilde{\omega}'(v)$ meet *pseudo-balanced condition*. This implies that

$$\frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)}} = \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}'_{i,m}(v)}}, \tag{75}$$

for all $i \in [K]$, which in turn implies that

$$\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)} = \sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}'_{i,m}(v)}.$$

Using the relation $\frac{2}{(a+b)/2} < \frac{1}{a} + \frac{1}{b}$ whenever $a, b > 0$ and $a \neq b$, we get that

$$\frac{2}{\widetilde{\omega}^{\rm avg}_{i,m}(v)} < \frac{1}{\widetilde{\omega}_{i,m}(v)} + \frac{1}{\widetilde{\omega}'_{i,m}(v)} \tag{76}$$

for all $m \in [M]$, $i \in S_m$ such that $\widetilde{\omega}_{i,m}(v) \neq \widetilde{\omega}'_{i,m}(v)$. Let $Q_{\rm diff} := \{\iota \in [K] : \exists\, m, \widetilde{\omega}_{\iota,m}(v) \neq \widetilde{\omega}'_{\iota,m}(v)\}$. As a consequence of (76), for all $i \in Q_{\rm diff}$, we have

$$\begin{aligned} &\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{2}{\widetilde{\omega}^{\rm avg}_{i,m}(v)} \\ &\quad < \sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)} + \sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}'_{i,m}(v)} \\ &\quad \overset{(a)}{\Longrightarrow} \sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}^{\rm avg}_{i,m}(v)} < \sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)} \\ &\quad \Longrightarrow \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum\limits_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}^{\rm avg}_{i,m}(v)}} > \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum\limits_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)}}, \end{aligned} \tag{77}$$

where (a) follows from (75). In addition, it is clear to observe that for all $i \in [K] \setminus Q_{\rm diff}$,

$$\frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}^{\rm avg}_{i,m}(v)}} = \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}_{i,m}(v)}}. \tag{78}$$

By (77) and (78) we know that $\widetilde{g}_v^{(j)}(\widetilde{\omega}^{\rm avg}(v)) \geq \widetilde{g}_v^{(j)}(\widetilde{\omega}(v))$ for each $j \in [L]$, which implies that $\widetilde{\omega}^{\rm avg}(v)$ is a common solution to (13). Now, there must exist $l \in [L]$ such that $Q_{\rm diff} \cap Q_l \neq \emptyset$, and then we consider two cases.

Case 1: $Q_{\rm diff} \cap Q_l = Q_l$. Eq. (77) implies that $\widetilde{g}_v^{(l)}(\widetilde{\omega}^{\rm avg}(v)) > \widetilde{g}_v^{(l)}(\widetilde{\omega}(v))$, which contradicts the fact that $\widetilde{\omega}(v)$ and $\widetilde{\omega}^{\rm avg}(v)$ are the common solution to (13).

Case 2: $Q_{\rm diff} \cap Q_l \subsetneq Q_l$. Note that the right-hand side of (77) and (78) are equal because $\widetilde{\omega}(v)$ meets pseudo-balanced condition. Hence, there exists $i_1, i_2 \in Q_l$ such that

$$\frac{\Delta_{i_1}^2(v)}{\frac{1}{M_{i_1}^2}\sum_{m=1}^M \mathbf{1}_{\{i_1\in S_m\}}\frac{1}{\widetilde{\omega}^{\rm avg}_{i_1,m}(v)}} > \frac{\Delta_{i_2}^2(v)}{\frac{1}{M_{i_2}^2}\sum_{m=1}^M \mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}^{\rm avg}_{i_2,m}(v)}}$$

which means that $\widetilde{\omega}^{\rm avg}(v)$ violates *pseudo-balanced condition*, a contradiction to Lemma 28. □

Finally, Theorem 14 follows from Lemma 28, Lemma 29, and Lemma 30.

H. Proof of Proposition 15

Before proving Proposition 15, we first present two important results. The first result below asserts that $G^{(j)}(v)$ is an eigenvector of $H^{(j)}(v)$ with $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$ as its eigenvalue for every $j \in [L]$.

Lemma 31: Given $v \in \mathcal{P}$, $G^{(j)}(v)$ is a eigenvector of $H^{(j)}(v)$ with eigenvalue $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$ for all $j \in [L]$, i.e.,

$$H^{(j)}(v)\, G^{(j)}(v) = \frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}\, G^{(j)}(v) \quad \forall\, j \in [L]. \tag{79}$$

Proof: Fix $j \in [L]$ and a problem instance v arbitrarily. Recall that $\widetilde{\omega}(v)$ is the unique common solution (13) and $G(v)$ is the global vector characterising $\widetilde{\omega}(v)$ uniquely (via (14)).

From Lemma 28, we know that $\widetilde{\omega}(v)$ satisfies *pseudo-balanced condition*, which implies that for any $i \in Q_j$,

$$\begin{aligned}
&\frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}(v)_{i,m}}} = \widetilde{g}_v^{(j)}(\widetilde{\omega}(v))\\
&\implies \frac{\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}}\frac{1}{\widetilde{\omega}(v)_{i,m}}}{M_i^2\Delta_i^2(v)} = \frac{1}{\widetilde{g}_v^{(j)}(\widetilde{\omega}(v))}\\
&\overset{(a)}{\implies} \frac{\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}}\frac{\sum_{\iota\in S_m}G(v)_\iota}{G(v)_i}}{M_i^2\Delta_i^2(v)} = \frac{1}{\widetilde{g}_v^{(j)}(\widetilde{\omega}(v))}\\
&\implies \frac{\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}}\sum_{\iota\in S_m}G(v)_\iota}{M_i^2\Delta_i^2(v)} = \frac{G(v)_i}{\widetilde{g}_v^{(j)}(\widetilde{\omega}(v))}\\
&\implies \sum_{\iota\in Q_j}G(v)_\iota\left(\frac{\sum_{m=1}^{M}\mathbf{1}_{\{i,\iota\in S_m\}}}{M_i^2\Delta_i^2(v)}\right) = \frac{G(v)_i}{\widetilde{g}_v^{(j)}(\widetilde{\omega}(v))}. \qquad (80)
\end{aligned}$$

In the above set of implications, (a) follows from (14). Noting that (80) is akin to (79) completes the desired proof. □

The second result below asserts that the eigenspace of $H^{(j)}(v)$ is one-dimensional.

Lemma 32: Given $v \in \mathcal{P}$, the dimension of the eigenspace of $H^{(j)}(v)$ associated with the eigenvalue $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$ is equal to one for all $j \in [L]$.

Proof: Fix $j \in [L]$ and a problem instance $v \in \mathcal{P}$ arbitrarily. It is easy to verify that $G^{(j)}(v)$ has strictly positive entries (else, $\widetilde{g}_v(\widetilde{\omega}(v)) = 0$). Suppose that $\mathbf{u} \in \mathbb{R}^{|Q_j|}$ is another eigenvector of $H^{(j)}(v)$ corresponding to the eigenvalue $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$ and $\{\mathbf{u}, G^{(j)}(v)\}$ is linearly independent. Let

$$\mathbf{u}' := G^{(j)}(v) + \epsilon\mathbf{u},$$

where $\epsilon > 0$ is any number such that each entry of $\mathbf{u}'$ is strictly positive. Let $\omega' \in \Gamma$ be defined as

$$\forall\, m \in [M], i \in S_m, \quad \omega'_{i,m} = \begin{cases} \frac{\mathbf{u}'_{\mathrm{Idx}(i)}}{\sum_{\iota\in S_m}\mathbf{u}'_{\mathrm{Idx}(\iota)}} & \text{if } i \in Q_j\\ \widetilde{\omega}(v)_{i,m} & \text{otherwise,}\end{cases}$$

where for any $i \in Q_j$, $\mathrm{Idx}(i) \in [|Q_j|]$ represents the index of arm i within the arms set Q_j. Then, it follows from the definition of ω' that $\widetilde{g}_v^{(l)}(\widetilde{\omega}(v)) = \widetilde{g}_v^{(l)}(\omega')$ for all $l \neq j$ and $\omega' \neq \widetilde{\omega}(v)$. Note that $\mathbf{u}'$ is also an eigenvector of $H^{(j)}(v)$ corresponding to the eigenvalue $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$. This means that for all $i \in Q_j$,

$$\begin{aligned}
&\sum_{\iota\in Q_j}\mathbf{u}'_{\mathrm{Idx}(\iota)}\left(\frac{\sum_{m=1}^{M}\mathbf{1}_{\{i,\iota\in S_m\}}}{M_i^2\Delta_i^2(v)}\right) = \frac{\mathbf{u}'_{\mathrm{Idx}(i)}}{\widetilde{g}_v^{(j)}(\widetilde{\omega}(v))}\\
&\implies \frac{\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}}\sum_{\iota\in S_m}\mathbf{u}'_{\mathrm{Idx}(\iota)}}{M_i^2\Delta_i^2(v)} = \frac{\mathbf{u}'_{\mathrm{Idx}(i)}}{\widetilde{g}_v^{(j)}(\widetilde{\omega}(v))}\\
&\implies \frac{\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}}\frac{\sum_{\iota\in S_m}\mathbf{u}'_{\mathrm{Idx}(\iota)}}{\mathbf{u}'_{\mathrm{Idx}(i)}}}{M_i^2\Delta_i^2(v)} = \frac{1}{\widetilde{g}_v^{(j)}(\widetilde{\omega}(v))}\\
&\implies \frac{\Delta_i^2(v)}{\frac{1}{M_i^2}\sum_{m=1}^{M}\mathbf{1}_{\{i\in S_m\}}\frac{1}{\omega'_{i,m}}} = \widetilde{g}_v^{(j)}(\widetilde{\omega}(v)). \qquad (81)
\end{aligned}$$

From (81), it is clear that $\widetilde{g}_v^{(l)}(\widetilde{\omega}(v)) = \widetilde{g}_v^{(l)}(\omega')$ for all $l \in [L]$, which contradicts Lemma 30. Thus, there is no eigenvector $\mathbf{u} \in \mathbb{R}^{|Q_j|}$ of $H^{(j)}(v)$ corresponding to the eigenvalue $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$ such that $\mathbf{u}$ and $G^{(j)}(v) > 0$ are linearly independent. This completes the desired proof. □

1) Proof of Proposition 15: Let $\mathbf{v}$ be any eigenvector of $H^{(j)}(v)$ whose eigenvalue is not equal to $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$. Because $H^{(j)}(v)$ is a normal matrix, its eigenvectors corresponding to distinct eigenvalues are orthogonal [32, Chapter 2, Section 2.5]. This implies that $\langle\mathbf{v}, G^{(j)}(v)\rangle = 0$, where $\langle\cdot,\cdot\rangle$ denotes the vector inner product operator. Note that $G^{(j)}(v)$ has strictly positive entries. Therefore, the entries of $\mathbf{v}$ cannot be all positive or all negative.

From Lemma 32, we know that any eigenvector $\mathbf{v}'$ associated with the eigenvalue $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$ should satisfy

$$\mathbf{v}' = \alpha G^{(j)}(v), \quad \text{for some } \alpha \in \mathbb{R}\setminus\{0\},$$

which implies that the entries of $\mathbf{v}'$ are either all positive or all negative. Also, Lemma 32 implies that among any complete set of eigenvectors of $H^{(j)}(v)$, there is only one eigenvector $\mathbf{u}$ with eigenvalue $\frac{1}{g_v^{(j)}(\widetilde{\omega}(v))}$. From the exposition above, it then follows that the entries of $\mathbf{u}$ must be all positive or all negative. Noting that $G^{(j)}(v)$ has unit norm (see (14)), we arrive at the form in (15). This completes the proof.

ACKNOWLEDGMENT

The authors would like to sincerely thank the two anonymous reviewers for providing thoughtful and constructive feedback, which contributed to enhancing the overall quality of this paper.

REFERENCES

[1] E. Even-Dar, S. Mannor, Y. Mansour, and S. Mahadevan, "Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems," *J. Mach. Learn. Res.*, vol. 7, no. 6, pp. 1–27, 2006.

[2] T. Lattimore and C. Szepesvári, *Bandit Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2020.

[3] K. Jamieson, M. Malloy, R. Nowak, and S. Bubeck, "lil'ucb: An optimal exploration algorithm for multi-armed bandits," in *Proc. Conf. Learn. Theory*, 2014, pp. 423–439.

[4] J.-Y. Audibert, S. Bubeck, and R. Munos, "Best arm identification in multi-armed bandits," in *Proc. Conf. Learn. Theory*, 2010, pp. 41–53.

[5] S. Bubeck, R. Munos, and G. Stoltz, "Pure exploration in finitely-armed and continuous-armed bandits," *Theor. Comput. Sci.*, vol. 412, no. 19, pp. 1832–1852, Apr. 2011.

[6] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, 2017, pp. 1273–1282.

[7] P. Kairouz et al., "Advances and open problems in federated learning," *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, 2021.

[8] Z. Zhu, J. Zhu, J. Liu, and Y. Liu, "Federated bandit: A gossiping approach," in *Proc. ACM SIGMETRICS/Int. Conf. Meas. Model. Comput. Syst.*, May 2021, pp. 3–4.

[9] C. Shi and C. Shen, "Federated multi-armed bandits," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 9603–9611.

[10] Z. Yan, Q. Xiao, T. Chen, and A. Tajer, "Federated multi-armed bandit via uncoordinated exploration," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 5248–5252.

[11] L. Kirkup and R. B. Frenkel, *An Introduction to Uncertainty in Measurement: Using the GUM (Guide to the Expression of Uncertainty in Measurement)*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[12] J. Wang and N. B. Shah, "Your 2 is my 1, your 3 is my 9: Handling arbitrary miscalibrations in ratings," in *Proc. 18th Int. Conf. Auton. Agents MultiAgent Syst.*, 2019, pp. 864–872.

[13] A. Garivier and E. Kaufmann, "Optimal best arm identification with fixed confidence," in *Proc. Conf. Learn. Theory*, 2016, pp. 998–1027.

[14] V. Moulos, "Optimal best Markovian arm identification with fixed confidence," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 5606–5615.

[15] E. Kaufmann, O. Cappé, and A. Garivier, "On the complexity of best-arm identification in multi-armed bandit models," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 1–42, 2016.

[16] P. N. Karthik, K. S. Reddy, and V. Y. F. Tan, "Best arm identification in restless Markov multi-armed bandits," *IEEE Trans. Inf. Theory*, vol. 69, no. 5, pp. 3240–3262, May 2023.

[17] J. Yang, Z. Zhong, and V. Y. F. Tan, "Optimal clustering with bandit feedback," 2022, *arXiv:2202.04294*.

[18] C. Shi, C. Shen, and J. Yang, "Federated multi-armed bandits with personalization," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2021, pp. 2917–2925.

[19] A. Dubey and A. Pentland, "Differentially-private federated linear bandits," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6003–6014.

[20] A. Mitra, H. Hassani, and G. Pappas, "Exploiting heterogeneity in robust federated best-arm identification," 2021, *arXiv:2109.05700*.

[21] K. S. Reddy, P. N. Karthik, and V. Y. F. Tan, "Almost cost-free communication in federated best arm identification," in *Proc. 37th AAAI Conf. Artif. Intell.*, 2023, pp. 8378–8385.

[22] C. Tao, Q. Zhang, and Y. Zhou, "Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits," in *Proc. IEEE 60th Annu. Symp. Found. Comput. Sci. (FOCS)*, Nov. 2019, pp. 126–146.

[23] N. Karpov, Q. Zhang, and Y. Zhou, "Collaborative top distribution identifications with limited interaction (Extended Abstract)," in *Proc. IEEE 61st Annu. Symp. Found. Comput. Sci. (FOCS)*, Nov. 2020, pp. 160–171.

[24] E. Hillel, Z. S. Karnin, T. Koren, R. Lempel, and O. Somekh, "Distributed exploration in multi-armed bandits," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 26, 2013, pp. 854–862.

[25] N. Karpov and Q. Zhang, "Collaborative best arm identification with limited communication on non-IID data," 2022, *arXiv:2207.08015*.

[26] O. A. Hanna, L. Yang, and C. Fragouli, "Solving multi-arm bandit using a few bits of communication," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2022, pp. 11215–11236.

[27] A. Mitra, H. Hassani, and G. J. Pappas, "Linear stochastic bandits over a bit-constrained channel," in *Proc. Learn. Dyn. Control Conf.*, 2023, pp. 1387–1399.

[28] F. Pase, D. Gündüz, and M. Zorzi, "Rate-constrained remote contextual bandits," *IEEE J. Sel. Areas Inf. Theory*, vol. 3, no. 4, pp. 789–802, Dec. 2022.

[29] Y. Wang, J. Hu, X. Chen, and L. Wang, "Distributed bandit learning: Near-optimal regret with efficient communication," 2019, *arXiv:1904.06309*.

[30] P. Mayekar, J. Scarlett, and V. Y. F. Tan, "Communication-constrained bandits under additive Gaussian noise," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 24236–24250.

[31] W. R. Stevens, B. Fenner, and A. M. Rudoff, *UNIX Network Programming: The SocketsNetworking API*, vol. 1, 3rd ed. Reading, MA, USA: Addison-Wesley, 2003.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[33] I. Cantador, P. Brusilovsky, and T. Kuflik, "Second workshop on information heterogeneity and fusion in recommender systems (HetRec2011)," in *Proc. 5th ACM Conf. Recommender Syst.*, Oct. 2011, pp. 387–388.

[34] R. K. Sundaram, *A First Course in Optimization Theory*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[35] D. A. Charalambos and C. B. Kim, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, 3rd ed. Berlin, Germany: Springer, 2006.

Zhirui Chen received the B.Eng. degree from the School of Data and Computer Science, Sun Yat-sen University, China, in 2017. He is currently pursuing the Ph.D. degree with the Department of Industrial Systems Engineering and Management, National University of Singapore, Singapore. His research interests include reinforcement learning, and especially, multi-armed bandits.

P. N. Karthik (Member, IEEE) received the B.E. degree in electronics and communications from the R. V. College of Engineering, Bengaluru, in 2014, and the dual M.Sc. (Engg.) and Ph.D. degree from the Indian Institute of Science, Bengaluru, in 2021. He is currently an Assistant Professor with the Department of Artificial Intelligence, Indian Institute of Technology Hyderabad. Prior to assuming this role, he was a Research Fellow with the Institute of Data Science, National University of Singapore. His research interests include multi-armed bandits, federated learning, transfer learning, sequential analysis, Markov decision processes, and stochastic adaptive control.

Vincent Y. F. Tan (Senior Member, IEEE) was born in Singapore, in 1981. He received the B.A. and M.Eng. degrees in electrical and information science from Cambridge University in 2005 and the Ph.D. degree in electrical engineering and computer science (EECS) from the Massachusetts Institute of Technology (MIT) in 2011.

He is currently a Professor with the Department of Mathematics and the Department of Electrical and Computer Engineering (ECE), National University of Singapore (NUS). His research interests include information theory, machine learning, and statistical signal processing. He is a member of the IEEE Information Theory Society Board of Governors. He received the MIT EECS Jin-Au Kong Outstanding Doctoral Thesis Prize in 2011, the NUS Young Investigator Award in 2014, the Singapore National Research Foundation (NRF) Fellowship (Class of 2018), and the NUS Young Researcher Award in 2019. He regularly serves as the Area Chair of prominent machine learning conferences, such as the International Conference on Learning Representations (ICLR) and the Conference on Neural Information Processing Systems (NeurIPS). He was an IEEE Information Theory Society Distinguished Lecturer from 2018 to 2019. He is currently serving as a Senior Area Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING and an Associate Editor in Machine Learning and Statistics for IEEE TRANSACTIONS ON INFORMATION THEORY.

Yeow Meng Chee (Senior Member, IEEE) received the B.Math. degree in computer science, and combinatorics and optimization and the M.Math. and Ph.D. degrees in computer science from the University of Waterloo, Waterloo, ON, Canada, in 1988, 1989, and 1996, respectively. He is currently a Professor of design and engineering with the National University of Singapore. Prior to this, he was a Professor of mathematical sciences with Nanyang Technological University, the Program Director of interactive digital media research and development with the Media Development Authority of Singapore, a Post-Doctoral Fellow with the University of Waterloo and the IBM's Zurich Research Laboratory, the General Manager of the Singapore Computer Emergency Response Team, and the Deputy Director of Strategic Programs at the Infocomm Development Authority, Singapore. His research interests include the interplay between combinatorics and computer science/engineering, particularly in combinatorial design theory, coding theory, extremal set systems, and their applications. He is a fellow of the Institute of Combinatorics and its Applications. He is an Editor of the *Journal of Combinatorial Theory, Series A*.