

# Exploiting Heterogeneity in Robust Federated Best-Arm Identification

Aritra Mitra, George J. Pappas, and Hamed Hassani \*

## Abstract

We study a federated variant of the best-arm identification problem in stochastic multi-armed bandits: a set of clients, each of whom can sample only a subset of the arms, collaborate via a server to identify the best arm (i.e., the arm with the highest mean reward) with prescribed confidence. For this problem, we propose **Fed-SEL**, a simple communication-efficient algorithm that builds on successive elimination techniques and involves local sampling steps at the clients. To study the performance of **Fed-SEL**, we introduce a notion of arm-heterogeneity that captures the level of dissimilarity between distributions of arms corresponding to different clients. Interestingly, our analysis reveals the benefits of arm-heterogeneity in reducing both the sample- and communication-complexity of **Fed-SEL**. As a special case of our analysis, we show that for certain heterogeneous problem instances, **Fed-SEL** outputs the best-arm after just one round of communication. Our findings have the following key implication: unlike federated supervised learning where recent work has shown that statistical heterogeneity can lead to poor performance, one can provably reap the benefits of both local computation and heterogeneity for federated best-arm identification. As our final contribution, we develop variants of **Fed-SEL**, both for federated and peer-to-peer settings, that are robust to the presence of Byzantine clients, and hence suitable for deployment in harsh, adversarial environments.

## 1 Introduction

Federated Learning (FL) has recently emerged as a popular framework for training statistical models using diverse data available at multiple clients (mobile phones, smart devices, health organizations, etc.) [1–5]. Two key challenges intrinsic to FL include statistical heterogeneity that arises due to differences in the data sets of the clients, and stringent communication constraints imposed by bandwidth considerations. Several recent works [6–11] have reported that under statistical heterogeneity, FL algorithms that perform local computations to save communication can exhibit poor performance. Notably, these observations concern the task of supervised learning, which has almost exclusively been the focus of FL thus far.

In a departure from the standard supervised learning setting, the authors in [12] recently showed that for the problem of federated clustering, one can in fact benefit from a notion of heterogeneity defined suitably for their setting. This is a rare finding that naturally begs the following question.

*Are there other classes of cooperative learning problems where, unlike federated supervised learning, local computations and specific notions of heterogeneity can provably help?*

---

\*The authors are with the Department of Electrical and Systems Engineering, University of Pennsylvania. Email: {amitra20, pappasg, hassani}@seas.upenn.edu. This work was supported by NSF Award 1837253, NSF CAREER award CIF 1943064, and the Air Force Office of Scientific Research Young Investigator Program (AFOSR-YIP) under award FA9550-20-1-0111.

In this paper, we answer the above question in the affirmative by introducing and studying a federated variant of the *best-arm identification* problem [13–19] in stochastic multi-armed bandits. In our model, a set of arms is partitioned into disjoint arm-sets and each arm-set is associated with a unique client. Specifically, each client can only sample arms from its respective arm-set. This modeling assumption is consistent with the standard FL setting where each client typically has access to only a portion of a large data set; we will provide further motivation for our model shortly. The task for the clients is to coordinate via a central server and identify the best arm with high probability, i.e., with a given fixed confidence.<sup>1</sup> To achieve this task, communication is clearly *necessary* since each client can only acquire information about its own arm-set in isolation. Our formulation naturally captures statistical heterogeneity: other than the requirement of sharing the same support, we allow distributions of arms across arm-sets to be arbitrarily different. In what follows, we briefly describe a couple of motivating applications.

**Applications in Robotics and Sensor Networks:** Our work is practically motivated by the diverse engineering applications (e.g., distributed sensing, estimation, localization, target-tracking, etc.) of cooperative learning [20–24]. For instance, consider a team of robots deployed over a large environment and tasked with detecting a chemical leakage based on readings from sensors spread out over the environment. Since it may be infeasible for a single robot to explore the entire environment, each robot is instead assigned the task of exploring a sub-region. The robots explore their respective sub-regions and coordinate via a central controller. We can abstract this application into our model by mapping robots to clients, controller to server, sensors to arms, and sub-regions to arm-sets. Based on this abstraction, detecting the sensor with the highest readings directly translates into the problem of identifying the best arm.

**Applications in Recommendation Systems:** As another concrete application, consider a web-recommendation system for restaurants in a city. People in the city upload ratings of restaurants to this system based on their personal experiences. Based on these ratings, the system constantly updates its scores, and provides recommendations regarding the best restaurant(s) in the city. To capture this scenario via our model, one can interpret people as clients, the web-recommendation system as the server, and restaurants as arms. Since it is somewhat unreasonable to expect that a person visits every restaurant in the city, the rationale behind a client sampling only a subset of the arms makes sense in this context as well.

We have thus justified the reason for considering heterogeneous action spaces (arm-sets) for the clients. In contrast, as we discuss in Section 1.2, prior related works make the arguably unrealistic assumption of a common action space for all clients, i.e., it is assumed that each client can sample every arm. As far as we are aware, this is the first work to study a federated/distributed variant of the heterogeneous multi-armed bandit (MAB) best-arm identification problem. The core question posed by this problem is the following: *Although a client can sample only a subset of the arms itself, how can it still learn the best arm via collaboration?*

In this context, we will focus on two important themes relevant to each of the applications we described above: *communication-efficiency* and *robustness*. To be more precise, we are interested in designing best-arm identification algorithms that (i) minimize communication between the clients and the server, and (ii) are robust to the presence of clients that act adversarially. Minimizing communication is motivated by low-bandwidth considerations in the context of sensor networks, and privacy concerns in the context of recommendation systems. Moreover, for the latter application, we would ideally like to avoid scenarios where a few deliberately biased reviews culminate in a bad recommendation. The need for robustness has a similar rationale in distributed robotics and sensor networks. Having motivated the setup, we now describe our main contributions.

---

<sup>1</sup>We will provide a precise problem formulation in Sec. 2 where we explain what is meant by a “best arm”.

## 1.1 Contributions

The specific contributions of this paper are as follows.

- **Problem Formulation:** Our first contribution is to introduce the *federated best-arm identification problem* in Section 2, and its robust variants in Sections 5 and 6. As we explained earlier in the introduction, the problems we study have diverse practical applications that span collaborative learning in sensor networks, distributed robotics, and recommendation systems.

In [12], the authors had conjectured that specific notions of heterogeneity - together with careful analyses - may provide benefits for a plethora of problems in federated learning. By identifying federated best-arm identification as one such problem, our work is one of the earliest efforts to formally support this conjecture.

- **Algorithm:** In Section 3, we propose **Fed-SEL** - a simple, communication-efficient algorithm that builds on successive elimination techniques [13, 14], and comprises of two distinct phases. During Phase I, each client aims to identify the locally best arm in its own arm-set; this requires no communication. Phase II involves periodic, encoded communication between the clients and the server (akin to standard FL algorithms) to determine the best arm among the locally best arms. As we discuss later in the paper, having two separate phases contributes towards (i) significantly reducing communication; and (ii) tolerating adversaries.

- **Benefits of Heterogeneity:** To analyze **Fed-SEL**, we introduce a notion of *arm-heterogeneity* in Section 4. Roughly speaking, our definition of heterogeneity captures the “closeness” between distributions of arms within an arm-set relative to that across different arm-sets. In Theorem 1, we rigorously characterize the performance of **Fed-SEL**. Our analysis reveals that more the level of heterogeneity in a given instance of the problem, fewer the number of communication rounds it takes for **Fed-SEL** to output the best arm at the prescribed confidence level, i.e., **Fed-SEL adapts** to the level of heterogeneity. We also identify problem instances where one round of communication suffices to learn the best arm. Since fewer rounds leads to fewer arm-pulls, we conclude that *heterogeneity helps reduce both the sample- and communication-complexity* of **Fed-SEL**.

As far as we are aware, Theorem 1 is the only result to demonstrate provable benefits of heterogeneity in federated/distributed MAB problems. In the process of arriving at this result, we derive new bounds on the Lambert  $W$  function [25] that may be of independent interest; see Lemma 5 in Appendix A.1.

- **Robustness to Byzantine adversaries:** By now, there are several works that explore the challenging problem of tackling adversaries in supervised learning [26–35]. However, there is no analogous result for the distributed MAB problem we consider here. We fill this gap by developing robust variants of **Fed-SEL** that output the best arm despite the presence of Byzantine adversarial clients, both in federated and peer-to-peer settings. Our robustness results, namely Theorems 2 and 3, highlight the role of information- and network structure-redundancy in combating adversaries.

Overall, for the problem of cooperative best-arm identification, our work develops simple, intuitive algorithms that are communication-efficient, benefit from heterogeneity, and practically appealing since they can be deployed in harsh, adversarial environments.

## 1.2 Related Work

In what follows, we position our work in the context of relevant literature.

- **Multi-Armed Bandits in Centralized and Distributed Settings:** Best-arm identification is a classical problem in the MAB literature and is typically studied with two distinct objectives: (i) maximizing the probability of identifying the best arm given a fixed sample budget [16–18]; and (ii) minimizing the number of arm-pulls (samples) required to identify the best arm at a given fixed

confidence level [13–15]. In this paper, we focus on the fixed confidence setting; for an excellent survey on this topic, see [19].

The best-arm identification problem has also been explored in distributed *homogeneous* environments [36, 37] where every agent (client) can sample all the arms, i.e., each agent can identify the best-arm *on its own*, and communication serves to reduce the sample-complexity (number of arm-pulls) at each agent. A fundamental difference of our work with [36, 37] is that communication is *necessary* for the heterogeneous setting we consider. Moreover, unlike the vast majority of distributed MAB formulations that focus on minimizing the cumulative regret metric [38–57], the goal of our study is best-arm identification within a PAC framework.

A couple of recent works demonstrate that in the context of multi-agent stochastic linear bandits, if the parameters of the agents share a common low-dimensional representation [53], or exhibit some other form of similarity [54], then such structure can be exploited to improve performance (as measured by regret). Thus, these works essentially show that *similarity helps*. Our findings complement these results by showing that depending upon the problem formulation and the performance measure, *dissimilarity can also help*.

• **Robust Multi-Armed Bandits:** A recent stream of works has focused on understanding the impacts of adversarial corruptions in MAB problems, both in the case of independent arms [58–61], and also for the linear bandit setting [62–64]. In these works, it is assumed that an attacker with a bounded attack budget can perturb the stochastic rewards associated with the arms, i.e., the attack is *structured* to a certain extent. The Byzantine attack model we consider in this work is different from the reward-corruption models in [58–64] in that it allows a malicious client to act *arbitrarily*. Thus, the techniques used in these prior works do not apply to our setting.

Like our work, worst-case adversarial behavior is also considered in [65], where a *blocking* approach is used to provide regret guarantees that preserve the benefits of collaboration when the number of adversaries is small. While the authors in [65] assume a complete graph and focus on regret minimization, our analysis in Section 6 accounts for general networks, and our goal is best-arm identification, i.e., we focus on the pure exploration setting.

• **Federated Optimization:** The standard FL setup involves periodic coordination between the server and clients to solve a distributed optimization problem [1–4]. The most popular algorithm for this problem, FedAvg [2], has been extensively analyzed both in the homogeneous setting [66–72], and also under statistical heterogeneity [6, 73–76]. Several more recent works have focused on improving the convergence guarantees of FedAvg by drawing on a variety of ideas: proximal methods in FedProx [77]; normalized aggregation in FedNova [78]; operator-splitting in FedSplit [10]; variance-reduction in Scaffold [11] and S-Local-SVRG [79]; and gradient-tracking in FedLin [80]. Despite these promising efforts, there is little to no theoretical evidence that justifies performing local computations - a key feature of FL algorithms - under statistical heterogeneity.

*Heterogeneity can hurt:* Indeed, even in simple, deterministic settings with full client participation, it is now known that FedAvg and FedProx fail to match the basic centralized guarantee of linear convergence to the global minimum for strongly-convex, smooth loss functions [6–10]. Moreover, for algorithms such as Scaffold [11], S-Local-SVRG [79], and FedLin [80] that do guarantee linear convergence to the true minimizer, their linear convergence rates demonstrate no apparent benefit of performing multiple local gradient steps under heterogeneity. To further exemplify this point, a lower bound is presented in [80] to complement the upper bounds in [11], [79], and [80]; see also Remark 2.2 in [79]. Thus, under heterogeneity, the overall picture is quite grim in FL.

In Section 2, we draw various interesting parallels between our formulation and the standard federated optimization setup. These parallels are particularly insightful since they provide intuition regarding why both local computations and heterogeneity can be helpful (if exploited appropriately) for the problem we study, thereby further motivating our work.

## 2 Problem Formulation

Our model is comprised of a set of arms  $\mathcal{A}$ , a set of clients  $\mathcal{C}$ , and a central server. The set  $\mathcal{A}$  is partitioned into  $N$  disjoint arm-sets, one corresponding to each client. In particular, client  $i \in \mathcal{C}$  can only sample arms from the  $i$ -th arm set, denoted by  $\mathcal{A}_i = \{a_1^{(i)}, a_2^{(i)}, \dots, a_{|\mathcal{A}_i|}^{(i)}\}$ . Throughout, we will make the mild assumption that each arm-set contains at least two arms, i.e.,  $|\mathcal{A}_i| \geq 2, \forall i \in [N]$ . We associate a notion of time with our model: in each unit of time, a client can only sample one arm. Each time client  $i$  samples an arm  $\ell \in \mathcal{A}_i$ , it observes a random reward  $X_\ell$  drawn independently of the past (actions and observations of all clients) from a distribution  $D_\ell$  (unknown to client  $i$ ) associated with arm  $\ell$ . We will use  $\hat{r}_{i,\ell}(t)$  to denote the average of the first  $t$  independent samples  $X_{i,\ell}(s), s \in [t]$  of the reward of an arm  $\ell \in \mathcal{A}_i$  as observed by client  $i$ , i.e.,  $\hat{r}_{i,\ell}(t) = (\sum_{s=1}^t X_{i,\ell}(s))/t$ .

As is quite standard [13–19], we assume that all rewards are in  $[0, 1]$ . For each arm  $\ell \in \mathcal{A}$ , let  $r_\ell$  denote the expected reward of arm  $\ell$ , i.e.,  $r_\ell = \mathbb{E}[X_\ell]$ . We will use  $\Delta_{\ell j} \triangleq r_\ell - r_j$  to represent the difference between the mean rewards of any two arms  $\ell, j \in \mathcal{A}$ . An arm with the highest expected reward in  $\mathcal{A}_i$  (resp., in  $\mathcal{A}$ ) will be called a *locally best arm* in arm-set  $i$  (resp., a *globally best arm* in  $\mathcal{A}$ ), and be denoted by  $a_i^*$  (resp., by  $a^*$ ). For simplicity of exposition, let the arms in each arm-set  $\mathcal{A}_i$  be ordered such that  $r_{a_1^{(i)}} > r_{a_2^{(i)}} > \dots > r_{a_{|\mathcal{A}_i|}^{(i)}}$ ; thus,  $a_i^* = a_1^{(i)}$ . Without loss of generality, we assume that arm 1 is the *unique* globally best arm, and that it belongs to arm-set  $\mathcal{A}_1$ , i.e.,  $a^* = a_1^{(1)} = 1$ . We can now concretely state the problem of interest.

**Problem 1. (*Federated Best-Arm Identification*)** *Given any  $\delta \in (0, 1)$ , design an algorithm that with probability at least  $1 - \delta$  terminates in finite-time, and outputs the globally best-arm  $a^*$ , where*

$$a^* = \operatorname{argmax}_{i \in \{1, \dots, N\}, \ell \in \{1, \dots, |\mathcal{A}_i|\}} \left( r_{a_\ell^{(i)}} = \mathbb{E}[X_{a_\ell^{(i)}}] \right). \quad (1)$$

In words, given a fixed confidence level  $\delta$  as input, our goal is to design an algorithm that enables the clients and the server to identify the globally best arm with probability at least  $1 - \delta$ . Following [14], we will call such an algorithm  $(0, \delta)$ -PAC. Since each client  $i \in \mathcal{C}$  is only *partially informative* (i.e., can only sample arms from its local arm-set  $\mathcal{A}_i$ ), it will not be able to identify  $a^*$  on its own, in general. Thus, any algorithm that solves Problem 1 will necessarily require communication between the clients and the server. Given that the communication-bottleneck is a major concern in FL, our focus will be to design *communication-efficient best-arm identification algorithms*. Later, in Section 5, we will consider a *robust* version of Problem 1 where the goal will be to identify the globally best arm despite the presence of adversarial clients that can act *arbitrarily*. We now draw some interesting parallels between the above formulation and the familiar federated optimization setup that is used to model and study supervised learning problems.

**Comparison with the standard FL setting:** The canonical FL setting involves solving the following optimization problem via periodic communication between the clients and the server [1–4]:

$$\min_{x \in \mathbb{R}^d} f(x), \quad \text{where } f(x) = \frac{1}{N} \sum_{i=1}^N f_i(x), \quad \text{and } f_i(x) = \mathbb{E}_{z_i \sim P_i} [L_i(x; z_i)]. \quad (2)$$

Here,  $f_i : \mathbb{R}^d \rightarrow \mathbb{R}$  is the local loss function of client  $i$ , and  $f(x)$  is the global objective function. Let  $x_i^* = \operatorname{argmin}_{x \in \mathbb{R}^d} f_i(x)$ , and  $x^* = \operatorname{argmin}_{x \in \mathbb{R}^d} f(x)$ . Under *statistical heterogeneity*,  $P_i \neq P_j$  in general for distinct clients  $i$  and  $j$ , and hence,  $f_i(x)$  and  $f_j(x)$  may have different minima. A typical FL algorithm that solves for  $x^*$  operates in rounds. Between two successive communication rounds, each client performs local training on its own data using stochastic gradient descent (SGD) or a

---

**Algorithm 1** Local Successive Elimination at Client  $i$ 

---

1: **Parameters:**  $\{\alpha_i(t), \beta_i(t)\}$ , where

$$\alpha_i(t) = \sqrt{\log(4|\mathcal{C}||\mathcal{A}_i|t^2/\delta)}/t; \beta_i(t) = c\alpha_i(t)$$

2: **Set:**  $t = 1, \mathcal{S}_i(t) = \mathcal{A}_i$

3: Sample each arm  $\ell \in \mathcal{S}_i(t)$  once and update empirical mean  $\hat{r}_{i,\ell}(t)$  of arm  $\ell$

4: Let  $\hat{r}_{i,\max}(t) = \max_{\ell \in \mathcal{S}_i(t)} \hat{r}_{i,\ell}(t)$

5: Eliminate sub-optimal arms:  $\mathcal{S}_i(t+1) = \{\ell \in \mathcal{S}_i(t) : \hat{r}_{i,\max}(t) - \hat{r}_{i,\ell}(t) < \beta_i(t)\}$

6: If  $|\mathcal{S}_i(t+1)| > 1$ , then set  $t = t+1$  and Go To Line 3. Else, **Output**  $a_i = S_i(t+1)$

---

variant thereof. As a consequence of such local training, popular FL algorithms such as **FedAvg** [2] and **FedProx** [77] exhibit a “client-drift phenomenon”: the iterates of each client  $i$  drift-off towards the minimizer  $x_i^*$  of its own local loss function  $f_i(x)$  [6–11]. Under statistical heterogeneity, there is no apparent relationship between  $x_i^*$  and  $x^*$ , and these points may potentially be far-off from each other. Thus, client-drift can cause **FedAvg** and **FedProx** to exhibit poor convergence.<sup>2</sup> More recent algorithms such as **SCAFFOLD** [11] and **FedLin** [80] try to mitigate the client-drift phenomenon by employing techniques such as variance-reduction and gradient-tracking. However, even these more refined algorithms are unable to demonstrate any quantifiable benefits of performing multiple local steps under arbitrary statistical heterogeneity.<sup>3</sup>

Coming back to the MAB problem,  $a_i^*$  can be thought of as the direct analogue of  $x_i^*$  in the standard FL setting. We make two key observations.

- **Observation 1:** Unlike the standard FL setting, there is an immediate relationship between the global optimum, namely the globally best arm  $a^*$ , and the locally best arms  $\{a_i^*\}_{i \in \mathcal{C}}$ : *the globally best arm is the arm with the highest expected reward among the locally best arms.*
- **Observation 2:** Suppose distributions  $D_{\ell_1}$  and  $D_{\ell_2}$  corresponding to arms  $\ell_1 \in \mathcal{A}_i$  and  $\ell_2 \in \mathcal{A}_j$  are “very different” (in an appropriate sense). Intuition dictates that this should make the task of identifying the arm with the higher mean easier, i.e., *statistical heterogeneity among distributions of arms corresponding to different clients should facilitate best-arm identification.*

The first observation, although simple and rather obvious, has the following important implication. Suppose each client  $i$  does pure exploration locally to identify the best arm  $a_i^*$  in its arm-set  $\mathcal{A}_i$ . Clearly, such local computations contribute towards the overall task of identifying the globally best arm (as they help eliminate sub-optimal arms), regardless of the level of heterogeneity in the distributions of arms across arm-sets. *Thus, we immediately note a major distinction between the standard federated supervised learning setting and the federated best-arm identification problem: for the former, local steps can potentially hurt under statistical heterogeneity; for the latter, we can exploit problem structure to design local computations that always contribute towards the global task.* In fact, aligning with the second observation, we will show later in Section 4 that statistical heterogeneity of arm distributions across clients can assist in minimizing communication. Building on the insights from this section, we now proceed to develop a federated best-arm identification algorithm that solves Problem 1.

---

<sup>2</sup>By poor convergence, we mean that these algorithms either converge slowly to the true global minimum, or converge to an incorrect point (potentially fast). See [8], [9], and [80] for a detailed discussion on this topic.

<sup>3</sup>Indeed, in [80], the authors show that there exist simple instances involving two clients with quadratic loss functions, for which, performing multiple local steps yields no improvement in the convergence rate even when gradient-tracking is employed.

### 3 Federated Successive Elimination

In this section, we develop an algorithm for solving Problem 1. Since minimizing communication is one of our primary concerns, let us start by asking what a client can achieve *on its own* without interacting with the server. To this end, consider the following example to help build intuition.

**Example:** Suppose there are just 2 clients with associated arm-sets  $\mathcal{A}_1 = \{1, 2\}$  and  $\mathcal{A}_2 = \{3, 4\}$ . The mean rewards are  $r_1 = 0.9, r_2 = 0.3, r_3 = 0.8,$  and  $r_4 = 0.7$ . From classical MAB theory, we know that to identify the best arm among two arms with gap in mean rewards  $\Delta$ , each of the arms needs to be sampled  $O(1/\Delta^2)$  times. Thus, after sampling each of the arms in  $\mathcal{A}_1$  for  $O(1/\Delta_{12}^2)$  number of times (resp., arms in  $\mathcal{A}_2$  for  $O(1/\Delta_{34}^2)$  times), client 1 (resp., client 2) will be able to identify arm 1 (resp., arm 3) as the locally best-arm in  $\mathcal{A}_1$  (resp.,  $\mathcal{A}_2$ ) with high probability. However, at this stage, arm 1 will likely have been sampled much fewer times than arm 3 (since  $1/\Delta_{12}^2 \approx 3,$  and  $1/\Delta_{34}^2 = 100$ ), and hence, client 1’s estimate of arm 1’s true mean will be much coarser than client 2’s estimate of arm 3’s true mean. The main message here is that *simply identifying locally best arms from each arm-set and then naively comparing between them may result in rejection of the globally best-arm*. Thus, we need a more informed strategy that accounts for imbalances in sampling across arm-sets. With this in mind, we develop Fed-SEL.

---

#### Algorithm 2 Federated Successive Elimination (Fed-SEL)

---

##### Phase I: Identifying locally best arms

- 1: Each client  $i \in \mathcal{C}$  runs Algo. 1 on  $\mathcal{A}_i$ . If Algo. 1 terminates at client  $i$ , then client  $i$  reports  $\{a_i, \hat{r}_{i,a_i}(\bar{t}_i), \bar{t}_i\}$  to the server
- 

##### Phase II: Identifying the globally best arm

- 2: **for**  $k \in \{0, 1, \dots\}$  **do**
  - Server initializes  $\bar{\Psi}(0) = \mathcal{C}$  and does:
- 3: For each active client  $i \in \bar{\Psi}(k)$ , set  $\tilde{r}_i(k) = \hat{r}_{i,a_i}(\bar{t}_i)$  if  $k = 0$ ; else if  $k \geq 1$ , decode  $\tilde{r}_i(k) = \text{Dec}_{i,k}(\sigma_i(k))$ , and compute

$$\tilde{r}_i^{(U)}(k) = \tilde{r}_i(k) + 2\alpha_i(\bar{t}_i + kH); \quad \tilde{r}_i^{(L)}(k) = \tilde{r}_i(k) - 2\alpha_i(\bar{t}_i + kH)$$

- 4: Compute  $\tilde{r}_{\max}^{(L)}(k) = \max_{i \in \bar{\Psi}(k)} \tilde{r}_i^{(L)}(k)$
  - 5: Update active client set:  $\bar{\Psi}(k+1) = \{i \in \bar{\Psi}(k) : \tilde{r}_{\max}^{(L)}(k) < \tilde{r}_i^{(U)}(k)\}$
  - 6: **If**  $|\bar{\Psi}(k+1)| > 1$ , broadcast the threshold  $\tilde{r}_{\max}^{(L)}(k)$ . **Else**, output  $\bar{a} = a_{\bar{\Psi}(k+1)}$  and terminate
    - Each active client  $i \in \bar{\Psi}(k)$  does:
  - 7: **If**  $\tilde{r}_{\max}^{(L)}(k)$  is received from server *and*  $\tilde{r}_{\max}^{(L)}(k) < \tilde{r}_i^{(U)}(k)$ , then sample arm  $a_i$   $H$  times and transmit  $\sigma_i(k+1) = \text{Enc}_{i,k+1}(\hat{r}_{i,a_i}(\bar{t}_i + (k+1)H))$  to server. **Else**, deactivate
  - 8: **end for**
- 

**Description of Fed-SEL:** Our proposed algorithm Fed-SEL, outlined in Algo. 2, and depicted in Fig. 1, involves two distinct phases. In Phase I, each client  $i \in \mathcal{C}$  runs a local successive elimination sub-routine (Algo. 1) in parallel on its arm-set  $\mathcal{A}_i$  to identify the locally best arm  $a_i^*$  in  $\mathcal{A}_i$ . Note that this requires no communication. Phase II then involves periodic encoded communication between the clients and the server to identify the best arm among the locally best arms. We assume that each client  $i$  knows the parameters  $|\mathcal{C}|, |\mathcal{A}_i|$ , and  $\delta$ , and that the server knows  $|\mathcal{C}|, \{|\mathcal{A}_i|\}_{i \in \mathcal{C}}$ , and  $\delta$ . Additionally, we assume that the communication period  $H$  is known to both the clients and the server. We now describe each of the main components of Fed-SEL in detail.

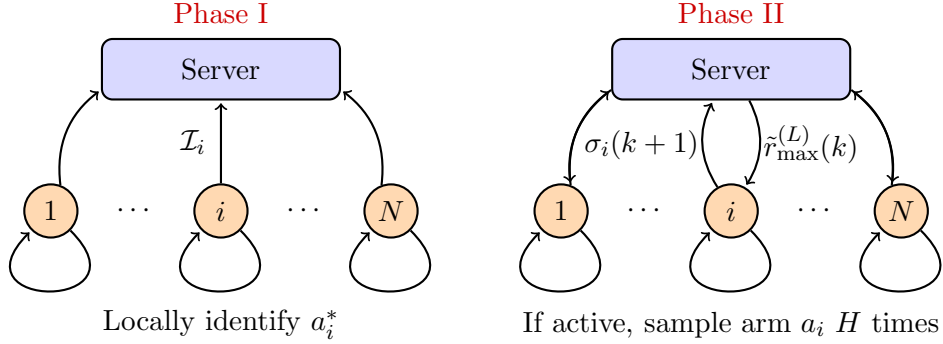


Figure 1: Illustration of Fed-SEL. **(Left)** In Phase I, each client  $i$  performs local computations (represented by self-loops) to identify the best arm  $a_i^*$  in  $\mathcal{A}_i$ . It then transmits  $\mathcal{I}_i = \{a_i, \hat{r}_{i,a_i}(\bar{t}_i), \bar{t}_i\}$  to the server. **(Right)** In round  $k$  of Phase II, the server broadcasts a threshold  $\tilde{r}_{\max}^{(L)}(k)$  that is used by each client  $i$  to determine if it is active. If active, it performs  $H$  local sampling steps, and transmits an encoded version  $\sigma_i(k+1)$  of the estimate  $\hat{r}_{i,a_i}(\bar{t}_i + (k+1)H)$  of arm  $a_i$ 's mean.

**Phase I:** Algo. 1 is based on the successive elimination technique [14], and proceeds in epochs. In each epoch  $t$ , client  $i$  maintains a set of active arms  $\mathcal{S}_i(t) \subseteq \mathcal{A}_i$ . Each active arm  $\ell \in \mathcal{S}_i(t)$  is sampled once, and its empirical mean  $\hat{r}_{i,\ell}(t)$  is then updated. For the next epoch, an arm is retained in the active set if and only if its empirical mean is not much lower than the empirical mean of the arm with the highest empirical mean in epoch  $t$ ; see line 5 of Algo. 1. This process continues till only one arm is left, at which point Algo. 1 terminates at client  $i$  and outputs the last arm  $a_i$ . Let  $\bar{t}_i$  represent the final epoch count when Algo. 1 terminates at client  $i$ . It is easy to then see that arm  $a_i$ , the output of Algo. 1 (if it terminates) at client  $i$ , is sampled  $\bar{t}_i$  times at the end of Phase I. The quantities  $\{\alpha_i(t), \beta_i(t)\}$  associated with arm-set  $\mathcal{A}_i$  are given by line 1 of Algo. 1. Here,  $c$  is a flexible parameter we introduce to control the accuracy of client  $i$ 's estimate  $\hat{r}_{i,a_i}(\bar{t}_i)$ .

**Phase II:** The goal of Phase II is to identify the best arm from the set  $\{a_i\}_{i \in \mathcal{C}}$ . Comparing arms across clients entails communication via the server. Since such communication is costly, we consider *periodic exchanges* between the clients and the server as in standard FL algorithms. To further reduce communication, such exchanges are *encoded* using a novel *adaptive encoding* strategy that we will explain shortly. Specifically, Phase II operates in rounds, where in each round  $k \in \{0, 1, \dots\}$ , the server maintains a set of *active clients*  $\bar{\Psi}(k)$ : a client  $i$  is deemed active if and only if arm  $a_i$  is still under contention for being the globally best arm. To decide upon the latter, the server aggregates the encoded information received from clients to compute the threshold  $\tilde{r}_{\max}^{(L)}(k)$  (lines 3 and 4 of Algo. 2). Computing this threshold requires some care to account for (i) imbalances in sampling across arm-sets, and (ii) errors introduced due to encoding (quantization). Having computed  $\tilde{r}_{\max}^{(L)}(k)$ , the server executes line 5 of Algo. 2 to eliminate sub-optimal arms from the set  $\{a_i\}_{i \in \bar{\Psi}(k)}$ . It then updates the active client set and broadcasts  $\tilde{r}_{\max}^{(L)}(k)$  if  $|\bar{\Psi}(k+1)| > 1$ .

Only the arms that remain in  $\{a_i\}_{i \in \bar{\Psi}(k+1)}$  require further sampling by the corresponding clients. Each client  $i \in \bar{\Psi}(k)$  uses the threshold  $\tilde{r}_{\max}^{(L)}(k)$  to determine whether it should sample  $a_i$  any further (i.e., it checks whether to remain active or not). If the conditions in line 7 of Algo. 2 pass, then client  $i$  samples arm  $a_i$   $H$  times, and transmits an encoded version  $\sigma_i(k+1)$  of  $\hat{r}_{i,a_i}(\bar{t}_i + (k+1)H)$  to the server. Here,  $H$  is the communication period; if  $H = 1$ , then the active clients interact with the server at each time-step in Phase II. Since we aim to reduce the number of communication rounds,



we instead let each client perform multiple local sampling steps before communicating with the server, i.e.,  $H > 1$  for our setting.<sup>4</sup> Phase II lasts till the active client set reduces to a singleton, at which point the locally best arm corresponding to the last active client is deemed to be the globally best arm by the server. We now describe our encoding-decoding strategy.

**Adaptive Encoding Strategy:** At the end of round  $k - 1$ , each active client  $i \in \bar{\Psi}(k)$  encodes  $\hat{r}_{i,a_i}(\bar{t}_i + kH)$  using a quantizer with range  $[0, 1]$  and bit-precision level  $B_i(k)$ , where

$$B_i(k) = \left\lceil \log_2 \left( \frac{1}{\alpha_i(\bar{t}_i + kH)} \right) \right\rceil. \quad (3)$$

Specifically, the interval  $[0, 1]$  is quantized into  $2^{B_i(k)}$  bins, and each bin is assigned a unique binary symbol. Since all rewards belong to  $[0, 1]$ ,  $\hat{r}_{i,a_i}(\bar{t}_i + kH)$  falls in one of the bins in  $[0, 1]$ . Let  $\sigma_i(k)$  be the binary representation of the index of this bin. Compactly, we use  $\text{Enc}_{i,k}(\hat{r}_{i,a_i}(\bar{t}_i + kH)) = \sigma_i(k)$  to represent the above encoding operation at client  $i$ . The key idea we employ in our encoding technique described above is that of *adaptive quantization*: the bit precision level  $B_i(k)$  is *successively refined* in each round based on the confidence interval  $\alpha_i(\bar{t}_i + kH)$ .

For correct decoding, we assume that the server is aware of the encoding strategies at the clients, and the parameters  $|\mathcal{C}|$ ,  $\{|\mathcal{A}_i|\}_{i \in \mathcal{C}}$ ,  $H$ , and  $\delta$ . Based on this information, and the fact that at the end of Phase I, each client  $i$  transmits  $\bar{t}_i$  to the server, note that the server can *exactly* compute  $\alpha_i(\bar{t}_i + kH)$ , and hence  $B_i(k)$  for each  $i \in \mathcal{C}$ . Let  $\tilde{r}_i(k)$  be the center of the bin associated with  $\sigma_i(k)$ . Then, we compactly represent the decoding operation at the server (corresponding to client  $i$ ) as  $\text{Dec}_{i,k}(\sigma_i(k)) = \tilde{r}_i(k)$ . Using  $\tilde{r}_i(k)$  as a proxy for  $\hat{r}_{i,a_i}(\bar{t}_i + kH)$ , the server computes upper and lower estimates, namely  $\tilde{r}_i^{(U)}(k)$  and  $\tilde{r}_i^{(L)}(k)$ , of the mean of arm  $a_i$  in line 3 of Algo. 2.

This completes the description of **Fed-SEL**. A few remarks are now in order.

**Remark 1.** *Fed-SEL is communication-efficient by design: Phase I requires incurs only one round of communication (at the end of the phase) while Phase II involves periodic communication. Moreover, in each round  $k \in \{1, 2, \dots\}$  of Phase II, every client transmits encoded information about the empirical mean of only one arm from its arm-set. We note here that the encoding strategy needs to be designed carefully so as to prevent the globally best-arm from being eliminated due to quantization errors; our proposed approach takes this into consideration.*

**Remark 2.** *In our current approach, the clients transmit information without encoding only once, at the end of Phase I (i.e., when  $k = 0$ ). This is done to inform the server of the sample counts  $\{\bar{t}_i\}_{i \in \mathcal{C}}$  needed for correct decoding during Phase II.<sup>5</sup> Below, we outline a simple way to bypass the need for transmitting the parameters  $\{\bar{t}_i\}_{i \in \mathcal{C}}$ .*

*At the end of Phase I, suppose the first time the server receives information from client  $i$  is at time-step  $t'_i$ .<sup>6</sup> Since client  $i$  can sample only one arm at each time-step, the maximum number of epochs at client  $i$  during Phase I is  $t'_i/|\mathcal{A}_i| \triangleq \tilde{t}_i$ . Thus,  $\tilde{t}_i$  provides a lower bound on  $\bar{t}_i$  - the number of times arm  $a_i$  is sampled during Phase I. Using  $\tilde{t}_i$  now in place of  $\bar{t}_i$ , one can employ more conservative confidence intervals  $\alpha_i(\tilde{t}_i + kH)$  both in line 3 of **Fed-SEL**, and also for encoding. While this strategy will also lead to identification of the best arm, it will in general require more rounds to terminate relative to **Fed-SEL**. As a final comment, using  $O(\log_2 |\mathcal{A}_i|)$  bits, one can encode the arm label  $a_i$  as well. Based on the above discussion, it is easy to see that with minor modifications to our current algorithm, one can effectively encode all transmissions from the clients to the server.*

<sup>4</sup>The local sampling steps in line 7 of **Fed-SEL** can be viewed as the analog of performing multiple local gradient-descent steps in standard federated optimization algorithms.

<sup>5</sup>Recall that in the encoding operation at the end of round  $k - 1$ , the bit-precision  $B_i(k)$  is a function of  $\alpha_i(\bar{t}_i + kH)$ . Other than  $\bar{t}_i$ , all other terms featuring in  $\alpha_i(\bar{t}_i + kH)$  are known a priori to the server.

<sup>6</sup>Here, we implicitly assume that there are no transmission delays, and that the server is equipped with the means to keep track of time.

## 4 Performance Guarantees for Fed-SEL

### 4.1 A Notion of Heterogeneity

In Section 2, we provided an intuitive argument that statistical heterogeneity of arm distributions across different arm-sets should facilitate the task of identifying the globally best arm. To make this argument precise, however, we need to fill in two important missing pieces: (i) defining an appropriate notion of heterogeneity for our problem, and (ii) formally establishing that our proposed algorithm Fed-SEL benefits from the notion of heterogeneity so defined. In this section, we address both these objectives. We start by introducing the concept of *arm-heterogeneity indices* that will play a key role in our subsequent analysis of Fed-SEL.

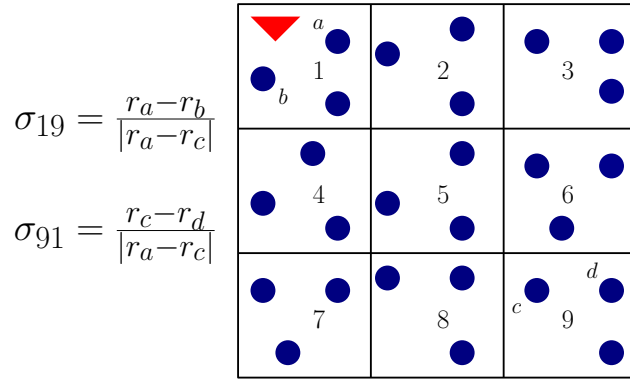


Figure 2: Illustration of Definition 1 via a target detection example. An environment is partitioned into 9 sub-regions (arm-sets) with 3 sensors (arms) in each sub-region. The goal is to detect a target based on the sensor readings. For this purpose, an agent (client) is assigned to each sub-region. The target is depicted by the red triangle and the sensors are depicted by the blue circles. Arms  $a$  and  $b$  (resp., arms  $c$  and  $d$ ) are the two top arms (i.e., closest to the target) in sub-region 1 (resp., in sub-region 9). The arm-heterogeneity indices  $\{\sigma_{19}, \sigma_{91}\}$  for  $\mathcal{A}_1$  and  $\mathcal{A}_9$  are as shown in the figure.

**Definition 1. (Arm-Heterogeneity Index)** For any two arm-sets  $\mathcal{A}_i$  and  $\mathcal{A}_j$  such that  $r_{a_1^{(i)}} \neq r_{a_1^{(j)}}$ , the arm-heterogeneity index  $\sigma_{ij}$  is defined as follows:

$$\sigma_{ij} = \frac{r_{a_1^{(i)}} - r_{a_2^{(i)}}}{|r_{a_1^{(i)}} - r_{a_1^{(j)}}|}. \quad (4)$$

With Definition 1, we aim to capture the intuitive idea that *distributions of arms within an arm-set are more similar than those across arm-sets*. The arm-heterogeneity indices we have introduced quantify the extent to which the above statement is true. In words,  $\sigma_{ij}$  measures the closeness between the two best arms within  $\mathcal{A}_i$  relative to the closeness between the best arms in each of the sets  $\mathcal{A}_i$  and  $\mathcal{A}_j$ ; here, “closeness” between two arms is quantified by the difference in their means.<sup>7</sup> Together, the index pair  $\{\sigma_{ij}, \sigma_{ji}\}$  captures the level of dissimilarity or heterogeneity between arm-sets  $\mathcal{A}_i$  and  $\mathcal{A}_j$ ; smaller values of these indices reflect more heterogeneity. To see why the latter is true, notice that both  $\sigma_{ij}$  and  $\sigma_{ji}$  scale inversely with the gap in mean rewards of the two best arms in  $\mathcal{A}_i$  and  $\mathcal{A}_j$ . Thus, if these arms have “well-separated” mean rewards - an intuitive measure of heterogeneity - then this would lead to small values of  $\sigma_{ij}$  and  $\sigma_{ji}$ .

<sup>7</sup>Recall that we use  $a_1^{(i)}$  to denote the top arm in  $\mathcal{A}_i$ , i.e., the arm with the highest mean reward in  $\mathcal{A}_i$ .

**How does heterogeneity arise in practice?** Our definition of heterogeneity is motivated by scenarios where the partitioning of arms into disjoint arm-sets is not arbitrary, but rather adheres to some structure. Consider for instance the target detection application we discussed in the Introduction; see Fig. 2 for an illustration. For this setting, sensors closer to the target have higher readings than those farther away. As a result, one should expect the readings of sensor  $a$  to be more similar to those of sensor  $b$  (in the same sub-region) as compared to those of sensor  $c$  in sub-region 9. This *spatial heterogeneity* effect is precisely captured by low values of the arm-heterogeneity index  $\sigma_{19}$ . One can similarly argue that spatial heterogeneity leads to low values of  $\sigma_{91}$ .

## 4.2 Main Result and Discussion

Our first main result, Theorem 1, shows that Fed-SEL is  $(0, \delta)$ -PAC, and provides a detailed characterization of its sample- and communication-complexity. To state the result cleanly, we will use  $i_1$  as a shorthand for  $a_1^{(i)}$ , and employ the following notations:

$$\bar{c}_i = \sqrt{\frac{4|\mathcal{C}||\mathcal{A}_i|}{\delta}}, \quad \Delta_i^* = r_{a_1^{(i)}} - r_{a_2^{(i)}}, \quad \text{and} \quad \Delta^* = r_1 - \max_{i \in \mathcal{C} \setminus \{1\}} r_{i_1}.$$

Thus,  $\Delta^*$  is the difference in the mean rewards of the two best arms in  $\mathcal{A}$ . We remind the reader here that the best arm  $a^* = 1$  belongs to the arm-set  $\mathcal{A}_1$  of client 1.

**Theorem 1.** *Suppose Fed-SEL is run with  $\beta_i(t) = 8\alpha_i(t), \forall i \in \mathcal{C}$ . Then with probability at least  $1 - \delta$ , the following statements hold.*

- **(Consistency)** Fed-SEL terminates in finite-time and outputs the globally best arm, i.e., we have  $\bar{a} = a^*$ .
- **(Communication-Complexity)** The number of rounds  $R_i$  a client  $i \in \mathcal{C} \setminus \{1\}$  remains active during Phase II is bounded above as follows:  $R_i \leq \max\{R_{i1}, R_{1i}\}$ , where

$$R_{i1} \leq \frac{1}{H} \left( \frac{128}{(\Delta_i^*)^2} (\sigma_{i1}^2 - 1) \log \frac{128\bar{c}_i}{\Delta_{1i1}^2} + \frac{256}{(\Delta_i^*)^2} \log \sigma_{i1} + O \left( \frac{1}{\Delta_{1i1}^2} \log \log \frac{\bar{c}_i}{\Delta_{1i1}^2} \right) \right) + 1, \text{ if } \sigma_{i1} > 1, \quad (5)$$

and  $R_{i1} = 1$  if  $\sigma_{i1} \leq 1$ . Similarly,

$$R_{1i} \leq \frac{1}{H} \left( \frac{128}{(\Delta_1^*)^2} (\sigma_{1i}^2 - 1) \log \frac{128\bar{c}_1}{\Delta_{1i1}^2} + \frac{256}{(\Delta_1^*)^2} \log \sigma_{1i} + O \left( \frac{1}{\Delta_{1i1}^2} \log \log \frac{\bar{c}_1}{\Delta_{1i1}^2} \right) \right) + 1, \text{ if } \sigma_{1i} > 1, \quad (6)$$

and  $R_{1i} = 1$  if  $\sigma_{1i} \leq 1$ . The number of communication rounds  $R$  during Phase II is then given by  $R = \max_{i \in \mathcal{C} \setminus \{1\}} R_i$ ; client 1 always remains active till Fed-SEL terminates, i.e.,  $R_1 = R$ . Moreover, each client  $i \in \mathcal{C}$  transmits at most

$$O \left( \log_2 \left( \frac{1}{\Delta^*} \right) \right)$$

bits of information in each round  $k \in \{1, \dots, R_{i1} - 1\}$ .

- **(Sample-Complexity)** The total number of arm-pulls made by client  $i$  is bounded above by

$$\underbrace{O \left( \sum_{\ell \in \mathcal{A}_i \setminus \{i_1\}} \frac{\log \left( \frac{|\mathcal{C}||\mathcal{A}_i|}{\delta \Delta_{i_1 \ell}^2} \right)}{\Delta_{i_1 \ell}^2} \right)}_{\text{Phase-I arm-pulls}} + \underbrace{HR_i}_{\text{Phase-II arm-pulls}}. \quad (7)$$

Since each client can only sample arms from its own arm-set, it is apparent that at least one round of communication is necessary for solving Problem 1, in general. The next result - an immediate corollary of Theorem 1 - identifies heterogeneous regimes where one round of communication is also sufficient.<sup>8</sup>

**Corollary 1.** *Suppose Fed-SEL is run with  $\beta_i(t) = 8\alpha_i(t), \forall i \in \mathcal{C}$ . Additionally, suppose*

$$\max_{i \in \mathcal{C} \setminus \{1\}} \{\sigma_{1i}, \sigma_{i1}\} \leq 1.$$

*Then, with probability at least  $1 - \delta$ , Fed-SEL terminates after just one round of communication between the clients and the server, and outputs the globally best arm, i.e.,  $\bar{a} = a^*$ .*

The proof of Theorem 1 is deferred to Appendix A. In what follows, we discuss several important implications of Theorem 1.

- **(Benefit of Heterogeneity)** From the first two dominant terms in the expressions for  $R_{i1}$  and  $R_{1i}$  in equations (5) and (6), respectively, observe that lower values of the arm-heterogeneity indices translate to fewer communication rounds. Moreover, fewer rounds imply fewer arm-pulls (and hence, lower sample-complexity) during Phase II, as is evident from Eq. (7). Since low values of arm-heterogeneity indices correspond to high heterogeneity, we conclude that *heterogeneity helps reduce both the sample- and communication-complexity of Fed-SEL.*

- **(Benefit of Local Steps)** Even if the level of heterogeneity is low in a given instance, one can reduce the number of communication rounds of Fed-SEL by increasing the communication period  $H$ ; see equations (5) and (6). Based on these equations and (7), we note that  $H$  shows up as an additive term in the upper-bound on the number of arm-pulls made during Phase II. The key implication of this fact is that *one can ramp up  $H$  to significantly reduce communication and get away with a modest price in terms of additional sample-complexity.* The main underlying reason for this stems from the observation that with Fed-SEL, more local steps always help by providing better arm-mean estimates.

- **(Benefit of Collaboration)** We have already discussed how heterogeneity assists in reducing the sample-complexity of Phase II. To focus on the number of arm-pulls made by a client  $i$  during Phase I, let us now turn our attention to the first term in Eq. (7). The summation in this term is over the local arm-set  $\mathcal{A}_i$  of client  $i$ , as opposed to the global arm-set  $\mathcal{A}$ . Since  $|\mathcal{A}_i|$  can be much smaller than  $|\mathcal{A}|$  in practice, the benefit of collaboration is immediate: each client needs to explore only a small subset of the global arm-set, i.e., *the task of exploration is distributed among clients.*

- **(Number of Bits Exchanged)** Since  $\Delta^*$  is the gap in mean rewards of the two best arms (i.e., the two arms that are hardest to distinguish) in the global arm set  $\mathcal{A}$ , the quantity  $\log_2(1/\Delta^*)$  essentially captures the complexity of a given instance of the MAB problem. Intuitively, one should expect that the amount of information that needs to be exchanged to solve a given instance of Problem 1 should scale with the complexity of that instance, i.e., with  $\log_2(1/\Delta^*)$ . From Theorem 1, we note that our adaptive encoding strategy formalizes this intuition to a certain extent.

**Remark 3.** *From the above discussion, note that although our algorithm itself is oblivious to both the level of heterogeneity and the complexity of a given instance, the sample- and communication-complexity bounds we obtain naturally adapt to these defining features of the instance, i.e., our guarantees are instance-dependent.*

---

<sup>8</sup>When  $\sigma_{i1} = 1$ , we would ideally like the upper-bound on  $R_{i1}$  to evaluate to 1. Setting  $\sigma_{i1} = 1$  in (5) causes the first two dominant terms in the upper-bound for  $R_{i1}$  to vanish, as desired. The lower order term that remains is an artifact of our analysis.

We now briefly comment on the main steps in the proof of Theorem 1.

**Proof Sketch for Theorem 1:** We first define a “good event”  $\mathcal{E}$  with measure at least  $1 - \delta$ , on which, the empirical means of all arms are within their respective confidence intervals at all times. On this event  $\mathcal{E}$ , we show using standard arguments that at the end of Phase I, each client  $i$  successfully identifies the locally best arm  $a_i^*$  in  $\mathcal{A}_i$ , i.e.,  $a_i = a_i^* = a_1^{(i)}$ . At this stage, to exploit heterogeneity, we require the locally best arm from each arm-set to be “well-explored”. Accordingly, we establish that  $|\hat{r}_{i,a_i}(\bar{t}_i) - r_{a_1^{(i)}}| \leq c/(c-2)^2 \Delta_i^*$ ,  $\forall i \in \mathcal{C}$ , where  $c$  is the flexible design parameter in line 1 of Algo. 1. The next key piece of our analysis pertains to relating the amount of exploration that has already been done during Phase I, to that which remains to be done during Phase II for eliminating all arms in the set  $\{a_1^{(i)}\}_{i \in \mathcal{C} \setminus \{1\}}$ ; this is precisely where the arm-heterogeneity indices show up in the analysis. To complete the above step, we derive new upper and lower bounds on the Lambert  $W$  function that may be of independent interest; see Lemma 5 in Appendix A.1.

Before moving on to the next section, we discuss how certain aspects of our algorithm and analysis can be potentially improved.

**Potential Refinements:** As is most often the case, there is considerable room for improvement. First, the dependence on the number of arms within the logarithmic terms of our sample-complexity bounds can be improved using a variant of the Median Elimination Algorithm [13]. Second, note from (7) that for each client  $i$ , its sample-complexity during Phase I scales inversely with the local arm-gaps  $\Delta_{i_1 \ell}$ , where  $\ell \in \mathcal{A}_i \setminus \{i_1\}$ , and  $i_1 = a_1^{(i)}$  is the locally best-arm in  $\mathcal{A}_i$ . Ideally, however, we would like to achieve sample-complexity bounds that scale inversely with  $\Delta_{1\ell}$ , where  $1 \in \mathcal{A}_1$  is the globally best arm. For all clients other than client 1 (i.e., clients who cannot sample arm 1 directly), this would necessarily require communication via the server. Thus, while we save on communication during Phase I, the price paid for doing so is a larger sample-complexity.

One may thus consider the alternative of merging Phases I and II. Our reasons, however, for persisting with separate phases are as follows. First, in our current approach, a client starts communicating when the best arm in its arm-set has already been well-explored. This, in turn, considerably reduces the workload for Phase II, and helps reduce communication significantly (see Corollary 1 for instance). The second important reason is somewhat more subtle: as we explain in the next section, having two separate phases aids the task of tolerating adversarial clients.

## 5 Robust Federated Best-Arm Identification

In this section, we will focus on the following question: *Can we still hope to identify the best-arm when certain clients act adversarially?* To formally answer this question, we will consider a Byzantine adversary model [81, 82] where adversarial clients have complete knowledge of the model, can collude among themselves, and act arbitrarily. Let us start by building some intuition. Suppose the client associated with the arm-set containing the globally best arm  $a^*$  is adversarial. Since this is the only client that can sample  $a^*$ , there is not much hope of identifying the best arm in this case. This simple observation reveals the need for *information-redundancy: we need multiple clients to sample each arm-set to account for adversarial behaviour*. Accordingly, with each arm-set  $\mathcal{A}_j$ , we now associate a group of clients  $\mathcal{C}_j$ ; thus,  $\mathcal{C} = \cup_{j \in [N]} \mathcal{C}_j$ .<sup>9</sup> We will denote the set of honest clients by  $\mathcal{H}$ , the set of adversarial clients by  $\mathcal{J}$ , and allow at most  $f$  clients to be adversarial in each client-group, i.e.,  $|\mathcal{C}_j \cap \mathcal{J}| \leq f, \forall j \in [N]$ . We will also assume that the server knows  $f$ .

Simply restricting the number of adversaries in each client-group is not enough to identify the best arm; there are additional challenges that need to be dealt with. To see this, consider a scenario

---

<sup>9</sup>From now on, we will use the index  $j$  to refer to arm-sets and client-groups, and the index  $i$  to refer to clients.

where an arm  $\ell \in \mathcal{A}_j$  has not been adequately explored by an honest client  $i \in \mathcal{C}_j$ . Due to the stochastic nature of our model, client  $i$ , although honest, may have a poor estimate of arm  $\ell$ 's true mean reward at this stage. How can we distinguish this scenario from one where an adversarial client in  $\mathcal{C}_j$  deliberately transmits a completely incorrect estimate of arm  $\ell$ ? In this context, one natural idea is to enforce each honest client to start transmitting information only when it has sufficiently explored every arm in its arm-set. This further justifies the rationale behind having two separate phases in Fed-SEL. In what follows, we describe a robust variant of Fed-SEL.

---

**Algorithm 3** Robust Federated Successive Elimination (Robust Fed-SEL)

---

**Phase I:** *Identifying locally best arms*

- 1: For each  $j \in [N]$ , every client  $i \in \mathcal{C}_j \cap \mathcal{H}$  runs Algo. 1 on  $\mathcal{A}_j$  with  $\beta_j(t) = 8\alpha_j(t)$ . If Algo. 1 terminates at client  $i$ , then client  $i$  reports  $\{a_i, \hat{r}_{i,a_i}(\bar{t}_i), \bar{t}_i\}$  to the server
- 

**Phase II:** *Identifying the globally best arm*

- 2: For each  $j \in [N]$ , server identifies a representative arm for  $\mathcal{A}_j$  as  $\bar{a}_j = \text{Maj\_Vot}(\{a_i\}_{i \in \mathcal{C}'_j})$ , where  $\mathcal{C}'_j \subseteq \mathcal{C}_j$  are the first  $(2f + 1)$  clients from  $\mathcal{C}_j$  to report to the server at the end of Phase I
- 3: Let  $\bar{\mathcal{C}}_j = \{i \in \mathcal{C}'_j : a_i = \bar{a}_j\}$  be those clients in  $\mathcal{C}'_j$  that report  $\bar{a}_j$ , and define  $f_j = f - |\mathcal{C}'_j| + |\bar{\mathcal{C}}_j|$
- 4: **for**  $k \in \{0, 1, \dots\}$  **do**
  - Server initializes  $\bar{\Psi}(0) = [N]$  and does:
    - 5: For each active client-group  $j \in \bar{\Psi}(k)$ , decode  $\tilde{r}_i(k)$  and compute the pair  $\{\tilde{r}_i^{(U)}(k), \tilde{r}_i^{(L)}(k)\}$  exactly as in Fed-SEL for every client  $i \in \bar{\mathcal{C}}_j$
    - 6: For each  $j \in \bar{\Psi}(k)$ , compute *robust* upper and lower estimates of the mean reward of  $\bar{a}_j$ :

$$\bar{r}_j^{(U)}(k) = \text{Trim}(\{\tilde{r}_i^{(U)}(k)\}_{i \in \bar{\mathcal{C}}_j}, f_j) \text{ and } \bar{r}_j^{(L)}(k) = \text{Trim}(\{\tilde{r}_i^{(L)}(k)\}_{i \in \bar{\mathcal{C}}_j}, f_j)$$

- 7: Compute  $\bar{r}_{\max}^{(L)}(k) = \max_{j \in \bar{\Psi}(k)} \bar{r}_j^{(L)}(k)$
  - 8: Update active client groups:  $\bar{\Psi}(k+1) = \{j \in \bar{\Psi}(k) : \bar{r}_{\max}^{(L)}(k) < \bar{r}_j^{(U)}(k)\}$
  - 9: **If**  $|\bar{\Psi}(k+1)| > 1$ , broadcast a binary  $N \times 1$  vector  $\mathbf{d}_k$  that encodes active client-groups. **Else**, output  $\bar{a} = \bar{a}_{\bar{\Psi}(k+1)}$  and terminate
    - For each  $j \in [N]$ , each client  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$  does:
      - 10: **If**  $\mathbf{d}_k$  is received from server and  $\mathbf{d}_k(j) = 1$ , then sample arm  $a_i$   $H$  times, encode  $\hat{r}_{i,a_i}(\bar{t}_i + (k+1)H)$  as in Fed-SEL, and transmit the encoding  $\sigma_i(k+1)$  to the server. **Else**, deactivate
      - 11: **end for**
- 

**Description of Robust Fed-SEL:** To identify  $a^*$  despite adversaries, we propose Robust Fed-SEL, and outline its main steps in Algorithm 3. Like the Fed-SEL algorithm, Robust Fed-SEL also involves two separate phases. During Phase I, in each client group  $\mathcal{C}_j, j \in [N]$ , every honest client  $i \in \mathcal{C}_j \cap \mathcal{H}$  independently runs Algo. 1 on  $\mathcal{A}_j$ , and then transmits  $\{a_i, \hat{r}_{i,a_i}(\bar{t}_i), \bar{t}_i\}$  to the server exactly as in Fed-SEL. The key operations performed by the server as follows.

**1) Majority Voting to Identify Representative Arms:** At the onset of Phase II, the server waits till it has heard from at least  $(2f + 1)$  clients of each client group before it starts acting. Next, for each arm-set  $\mathcal{A}_j$ , the server aims to compute a *representative locally best arm*  $\bar{a}_j$ . Note that adversarial clients in  $\mathcal{C}_j$  may collude and decide to report an arm to the server that has true mean much lower than that of the locally best arm  $a_1^{(j)}$  in  $\mathcal{A}_j$ . Nonetheless, we should expect every honest client  $i \in \mathcal{C}_j$  to output  $a_i = a_1^{(j)}$  with high probability. Thus, if honest clients are in majority in  $\mathcal{C}_j$ , a simple majority voting operation (denoted by Maj\_Vot) at the server should

suffice to reveal the best arm in  $\mathcal{A}_j$  with high probability; this is precisely the rationale behind line 2. During the remainder of Phase II, for each client group  $j \in [N]$ , the server only uses the information of the subset  $\bar{\mathcal{C}}_j \subset \mathcal{C}_j$  of clients who reported the representative arm  $\bar{a}_j$ .

**2) Trimming to Compute Robust Arm-Statistics:** Phase II involves periodic communication and evolves in rounds. In each round  $k$ , the server maintains a set  $\bar{\Psi}(k)$  of active client-groups: group  $j$  is deemed active if  $\bar{a}_j$  is still under contention for being the globally best arm. To decide whether group  $j$  should remain active, the server performs certain trimming operations that we next describe. Given a non-negative integer  $e$ , and a set  $\{m_i\}_{i \in [K]}$  of  $K$  scalars where  $K \geq 2e + 1$ , let  $\text{Trim}(\{m_i\}_{i \in [K]}, e)$  denote the operation of removing the highest  $e$  and lowest  $e$  numbers from  $\{m_i\}_{i \in [K]}$ , and then returning any one of the remaining numbers. Based on this  $\text{Trim}$  operator, the server computes *robust upper and lower estimates, namely  $\bar{r}_j^{(U)}(k)$  and  $\bar{r}_j^{(L)}(k)$ , of the mean of the representative arm  $\bar{a}_j, \forall j \in \bar{\Psi}(k)$* ; see line 6.<sup>10</sup> Using these robust estimates, the server subsequently updates the set of active client-groups by executing lines 7-8.

If group  $j$  remains active, then the server needs more refined arm estimates of arm  $\bar{a}_j$  from  $\bar{\mathcal{C}}_j$ . Accordingly, the server sends out a binary  $N \times 1$  vector  $\mathbf{d}_k$  that encodes the active client-groups:  $\mathbf{d}_k(j) = 1$  if and only if  $j \in \bar{\Psi}(k + 1)$ . Upon receiving  $\mathbf{d}_k$ , each honest client  $i$  in  $\bar{\mathcal{C}}_j$  checks if  $\mathbf{d}_k(j) = 1$ . If so, it samples arm  $a_i = \bar{a}_j$   $H$  times, encodes  $\hat{r}_{i, a_i}(\bar{l}_i + (k + 1)H)$  as in **Fed-SEL**, and transmits the encoding  $\sigma_i(k + 1)$  to the server. If  $\mathbf{d}_k$  is not received from the server, or if  $\mathbf{d}_k(j) = 0$ , client  $i$  does nothing further. This completes the description of **Robust Fed-SEL**.

**Comments on Algorithm 3:** Before providing formal guarantees for **Robust Fed-SEL**, a few comments are in order. First, note that unlike **Fed-SEL** where the server transmitted a threshold to enable the clients to figure out whether they are active or not, the server instead transmits a  $N \times 1$  vector  $\mathbf{d}_k$  in **Robust Fed-SEL**. The reason for this is as follows. Even if  $\bar{r}_{\max}^{(L)}(k) \geq \bar{r}_j^{(U)}(k)$  for some client-group  $j$ , it may very well be that  $\bar{r}_{\max}^{(L)}(k) < \bar{r}_i^{(U)}(k)$  for some  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$ . Thus, even if the server is able to eliminate client group  $j$  on its end, simply broadcasting the threshold  $\bar{r}_{\max}^{(L)}(k)$  may not be enough for every client in  $\bar{\mathcal{C}}_j \cap \mathcal{H}$  to realize that it no longer needs to sample arm  $\bar{a}_j$ . Note that for line 10 of **Robust Fed-SEL** to make sense, we assume that every client is aware of the client-group to which it belongs.

During Phase II, for each group  $j$ , since the server only uses the information of clients in  $\bar{\mathcal{C}}_j$ , we have not explicitly specified actions for honest clients in  $\mathcal{C}_j \setminus \bar{\mathcal{C}}_j$  to avoid cluttering the exposition. A potential specification is as follows. Suppose the first time an honest client  $i \in \mathcal{C}_j$  hears from the server, it is still in the process of completing Phase I. Since in this case client  $i$  cannot belong to  $\bar{\mathcal{C}}_j$ , it deactivates and does nothing further. In all other cases, client  $i \in \mathcal{C}_j$  acts exactly as specified in line 10 of **Robust Fed-SEL**.

The main result of this section is as follows.

**Theorem 2.** *Suppose  $|\mathcal{C}_j \cap \mathcal{J}| \leq f$  and  $|\mathcal{C}_j| \geq 3f + 1, \forall j \in [N]$ . Then, **Robust Fed-SEL** is  $(0, \delta)$ -PAC.*

**Discussion:** Under information-redundancy, i.e., under the condition that “enough” clients sample every arm in each arm-set ( $|\mathcal{C}_j| \geq 3f + 1, \forall j \in [N]$ ), Theorem 2 shows that **Robust Fed-SEL** is an adversarially-robust  $(0, \delta)$ -PAC algorithm. Despite the flurry of research on tolerating adversaries in distributed optimization/supervised learning [26–35], there is no analogous result (that we know of) in the context of MAB problems. Theorem 2 fills this gap. The sample- and communication-complexity bounds for **Robust Fed-SEL** are of the same order as that of **Fed-SEL**;

---

<sup>10</sup>The trimming parameter  $f_j$  is set based on the observation that clients in  $\mathcal{C}'_j$  who report an arm other than  $\bar{a}_j$  are adversarial with high probability.

hence, we do not explicitly state them here. For a discussion on this topic, see Appendix B where we provide the proof of Theorem 2.

**Remark 4.** *The type of redundancy assumption we make in Theorem 2 is extremely standard in the literature on tolerating worst-case adversarial models. Similar assumptions are made in [26–35] where typically less than half of the agents (clients) are allowed to be adversarial. The reason why we allow at most 1/3 of the clients in each client-group to be adversarial (as opposed to at most 1/2) is because some adversaries may choose not to transmit anything at all. In this case, the server may never acquire enough information from  $\mathcal{C}_j$  to compute  $\bar{a}_j$ . Now consider a slightly weaker adversary model where every client transmits when it is supposed to, but the information being sent can be arbitrarily corrupted for clients under attack. For this setting, our results go through identically under the weaker assumption that  $|\mathcal{C}_j \cap \mathcal{J}| \leq f$  and  $|\mathcal{C}_j| \geq 2f + 1, \forall j \in [N]$ .*

We now investigate how the ideas developed herein can be extended to the peer-to-peer setting.

## 6 Robust Peer-to-Peer Best-Arm Identification

---

**Algorithm 4** Robust Networked Successive Elimination protocol at each Client  $i \in \mathcal{C}_j \cap \mathcal{H}$

---

**Step 1:** *Updating  $\mathbf{I}_i(j)$  and  $\mathbf{R}_i(j)$*

- 1: Run Algo. 1 on  $\mathcal{A}_j$  with  $\beta_j(t) = 6\alpha_j(t)$ . If Algo. 1 terminates, set  $\mathbf{I}_i(j) = a_i$ ,  $\mathbf{R}_i(j) = \hat{r}_{i,a_i}(\bar{t}_i)$ , and transmit  $\mathbf{I}_i(j), \mathbf{R}_i(j)$  to each  $k \in \mathcal{N}_i^+$

**Step 2:** *Updating  $\mathbf{I}_i(\ell)$  and  $\mathbf{R}_i(\ell), \forall \ell \in [N] \setminus \{j\}$*

- 2: Let  $\mathcal{N}'_{i,\ell} \subseteq \mathcal{N}_i$  be the first  $(2f + 1)$  in-neighbors of  $i$  that report about arm-set  $\ell$
- 3: Set  $\mathbf{I}_i(\ell) = \text{Maj\_Vot} \left( \{\mathbf{I}_k(\ell)\}_{k \in \mathcal{N}'_{i,\ell}} \right)$ . Let  $\bar{\mathcal{N}}_{i,\ell} = \{k \in \mathcal{N}'_{i,\ell} : \mathbf{I}_k(\ell) = \mathbf{I}_i(\ell)\}$
- 4: Set  $\mathbf{R}_i(\ell) = \text{Trim} \left( \{\mathbf{R}_k(\ell)\}_{k \in \bar{\mathcal{N}}_{i,\ell}}, f_{i,\ell} \right)$ , where  $f_{i,\ell} = f - |\mathcal{N}'_{i,\ell}| + |\bar{\mathcal{N}}_{i,\ell}|$ . Transmit  $\mathbf{I}_i(\ell), \mathbf{R}_i(\ell)$  to each  $k \in \mathcal{N}_i^+$

**Step 3:** *Identifying globally best arm*

- 5: Let  $p_i^* = \text{argmax}_{\ell \in [N]} \mathbf{R}_i[\ell]$ . **Output  $\mathbf{I}_i(p_i^*)$ .**
- 

In this section, we consider the challenging setting of best-arm identification over a fully-distributed network in the presence of adversarial clients. Formally, the network is described by a static directed graph  $\mathcal{G} = (\mathcal{V}, \mathcal{D})$ , where the vertex set  $\mathcal{V}$  is composed of the set of clients, i.e.,  $\mathcal{V} = \mathcal{C} = \cup_{j \in [N]} \mathcal{C}_j$ , and  $\mathcal{D}$  represents the set of edges between them. An edge  $(u, v) \in \mathcal{D}$  indicates that client  $u$  can directly transmit information to client  $v$ . The set of all in-neighbors (resp., out-neighbors) of client  $u$  is defined as  $\mathcal{N}_u = \{v \in \mathcal{V} : (v, u) \in \mathcal{D}\}$  (resp.,  $\mathcal{N}_u^+ = \{v \in \mathcal{V} : (u, v) \in \mathcal{D}\}$ ).

Our **goal** is to design a distributed algorithm that ensures that each honest client  $i \in \mathcal{H}$  can eventually identify the globally best arm  $a^*$ , despite the actions of the adversarial clients  $\mathcal{J}$ . Clearly, we need some upper-bound on the number of adversaries for the problem to make sense. To this end, we will consider an  $f$ -local Byzantine adversary model [83–86] where there are at most  $f$  adversarial clients in the in-neighborhood of each honest client, i.e.,  $|\mathcal{N}_i \cap \mathcal{J}| \leq f, \forall i \in \mathcal{H}$ . Recall that clients in  $\mathcal{C}_j$  can only sample arms from  $\mathcal{A}_j$ . Thus, to acquire information about some other arm-set  $\mathcal{A}_\ell$  (say), we need information to diffuse from the clients in  $\mathcal{C}_\ell$  to clients in  $\mathcal{C}_j$ . If there is a small cut (bottleneck) in the graph separating  $\mathcal{C}_\ell$  from  $\mathcal{C}_j$ , then such a cut can be infested by adversaries, thereby preventing the flow of information necessary for identifying  $a^*$ . From this intuitive reasoning, note that we need sufficient disjoint paths linking different client groups to



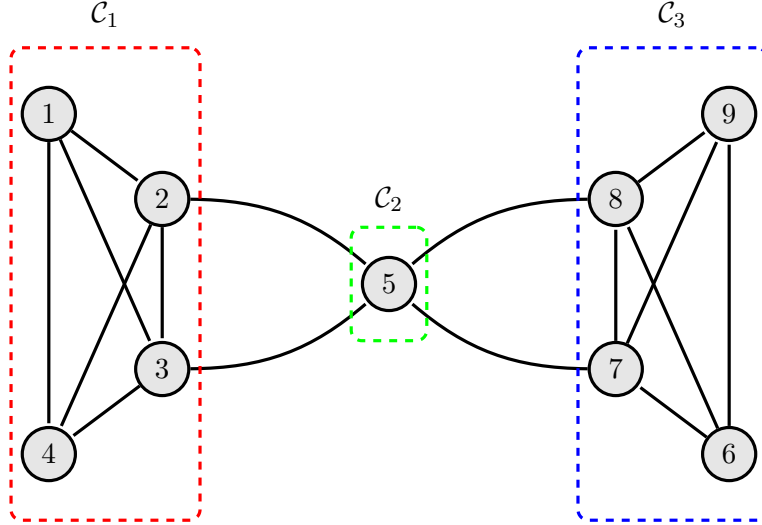


Figure 3: The figure shows a network composed of three client groups. Clients in group  $j$  can sample arms in arm-set  $\mathcal{A}_j$ , where  $j \in \{1, 2, 3\}$ . Given that client 5 is the only source of information for arms in  $\mathcal{A}_2$ , if it acts adversarially, then all other clients have no way of learning about the arms in  $\mathcal{A}_2$ . This highlights the need for information-redundancy. Moreover, since clients in  $\mathcal{C}_1$  and  $\mathcal{C}_3$  can interact only via client 5, if client 5 is adversarial, then it can create a bottleneck between  $\mathcal{C}_1$  and  $\mathcal{C}_3$ , making it impossible for clients in  $\mathcal{C}_1$  (resp.,  $\mathcal{C}_3$ ) to learn about arms in arm-set  $\mathcal{A}_3$  (resp.,  $\mathcal{A}_1$ ). This highlights the need for redundant paths in the network. The *strong-robustness* property in Definition 2 captures both the requirements described above.

have any hope of tackling adversaries. The key graph-theoretic property that we will exploit in this context is a notion of *strong-robustness* [86] adapted to our setting.

**Definition 2. (Strong-robustness w.r.t.  $\mathcal{C}_j$ )** Given any positive integer  $r$ , the graph  $\mathcal{G}$  is said to be strongly  $r$ -robust w.r.t. the client group  $\mathcal{C}_j$  if every non-empty subset  $\mathcal{B} \subseteq \mathcal{V} \setminus \{\mathcal{C}_j\}$  contains at least one client  $i$  with  $r$  in-neighbors outside  $\mathcal{B}$ , i.e., for every non-empty set  $\mathcal{B} \subseteq \mathcal{V} \setminus \{\mathcal{C}_j\}$ ,  $\exists i \in \mathcal{B}$  such that  $|\mathcal{N}_i \setminus \mathcal{B}| \geq r$ .

We note that if  $\mathcal{G}$  is strongly  $(3f+1)$ -robust w.r.t.  $\mathcal{C}_j$ , then  $|\mathcal{C}_j| \geq 3f+1$  (follows by considering  $\mathcal{B} = \mathcal{V} \setminus \{\mathcal{C}_j\}$ ). Thus, the above definition implicitly captures information-redundancy. At the same time, it ensures that there is adequate redundancy in the network-structure to reliably disseminate information about arm-set  $\mathcal{A}_j$  from  $\mathcal{C}_j$  to the rest of the network. We now develop an algorithm, namely Algorithm 4, that leverages the strong-robustness property in Definition 2.

**Description of Algorithm 4:** Each client  $i$  maintains two  $N$ -dimensional vectors  $\mathbf{I}_i$  and  $\mathbf{R}_i$ . The  $m$ -th entry of  $\mathbf{I}_i$ , namely  $\mathbf{I}_i(m)$ , stores client  $i$ 's estimate of the locally best arm in arm-set  $\mathcal{A}_m$ , and  $\mathbf{R}_i(m)$  stores an estimate of the corresponding arm's mean. Consider a client  $i \in \mathcal{C}_j \cap \mathcal{H}$ . Since client  $i$  can sample all arms in  $\mathcal{A}_j$ , it updates  $\mathbf{I}_i(j)$  and  $\mathbf{R}_i(j)$  on its own by running Algo. 1 on  $\mathcal{A}_j$ . To learn about the best arm in all arm-sets other than  $\mathcal{A}_j$ , client  $i$  needs to interact with its neighbors, some of whom might be adversarial. For each  $\ell \in [N] \setminus \{j\}$ , the strategy employed by client  $i$  to update  $\mathbf{I}_i(\ell)$  and  $\mathbf{R}_i(\ell)$  is very similar to that of the server in Robust Fed-SEL. In short,  $i$  collects information about  $\mathcal{A}_\ell$  from its neighbors, performs majority voting to decide  $\mathbf{I}_i(\ell)$  (line 3), and does a trimming operation to compute  $\mathbf{R}_i(\ell)$  (line 4). For this operation, client  $i$  only uses

the estimates of those neighbors that report the majority-voted arm  $\mathbf{I}_i(\ell)$ . Once all components of  $\mathbf{I}_i$  and  $\mathbf{R}_i$  have been populated, client  $i$  executes line 5 to identify the globally best arm.

We now present the main result of this section.

**Theorem 3.** *Suppose the following arm-heterogeneity condition holds:*

$$\max_{j \in [N] \setminus \{1\}} \{\sigma_{1j}, \sigma_{j1}\} \leq 1. \quad (8)$$

Moreover, suppose the adversary model is  $f$ -local, and  $\mathcal{G}$  is strongly  $(3f + 1)$ -robust w.r.t.  $\mathcal{C}_j, \forall j \in [N]$ . Then, with probability at least  $1 - \delta$ , Algorithm 4 terminates in finite-time, and outputs  $\mathbf{I}_i(p_i^*) = a^*, \forall i \in \mathcal{H}$ .

**Discussion:** The above result is the only one we know of that provides any guarantees against adversarial attacks for MAB problems over general networks. While the strong-robustness property in Definition 2 has been exploited earlier in the context of resilient distributed state estimation [86] and hypothesis-testing/inference [87], Theorem 3 reveals that the same graph-theoretic property ends up playing a crucial role for the best-arm identification problem we consider here.<sup>11</sup>

To isolate the subtleties and challenges associated with tolerating adversaries over a general network, we worked under an arm-heterogeneity assumption. This assumption effectively bypasses the need for Phase II, i.e., the sample-complexity of a client is given by the number of arm-pulls it makes during Phase I. We believe that the ideas presented in this paper can be extended in an appropriate way to study the networked setting *without* making the arm-heterogeneity assumption.

## 7 Simulations

In this section, we verify some of our theoretical findings based on a simple simulation example. Motivated by distributed target tracking applications in cooperative learning [20–23], we consider a target detection problem involving 9 static sensors distributed over an environment. The environment is divided into 3 sub-regions with 3 sensors and a mobile robot in each sub-region. Each mobile robot can access the sensor readings of the sensors in its sub-region, and transmit information to a central controller. The goal is to detect a static target located in one of the sub-regions. We model this setting by mapping sensors to arms, sub-regions to arm-sets, and robots to clients. Each sensor displays binary readings: sensors that are located closer to the target have a higher probability of displaying 1. Thus, finding the “best arm” in this setting corresponds to localizing the target, and a robot visiting a sensor corresponds to a client sampling an arm. Ideally, to minimize battery life of the robots, we would like to locate the target with a small number of visits (sample-complexity), and only a few information exchanges with the controller (communication-complexity).

We assume the target is located in the first sub-region that corresponds to arm-set  $\mathcal{A}_1$ . Arm rewards (sensor readings) are drawn from Bernoulli distributions with means  $\{0.9, 0.9 - 0.05\sigma, 0.1\}$ ,  $\{0.85, 0.8, 0.3\}$ , and  $\{0.7, 0.6, 0.5\}$  associated with the arm-sets (sub-regions)  $\mathcal{A}_1, \mathcal{A}_2$ , and  $\mathcal{A}_3$ , respectively. Based on the definition of arm-heterogeneity indices in Definition 1, the parameter  $\sigma \in [1, 15]$  captures the level of heterogeneity in the example we have described above. In particular, it is easy to verify that for each  $i \in [3]$ ,

$$r_{a_1}^{(i)} - r_{a_2}^{(i)} \leq \sigma |r_{a_1}^{(i)} - r_{a_1}^{(j)}|, \forall j \in [3] \setminus \{i\}. \quad (9)$$

---

<sup>11</sup>By making a connection to bootstrap percolation theory, the authors in [86] show that the strong-robustness property in Definition 2 can be checked in polynomial time.

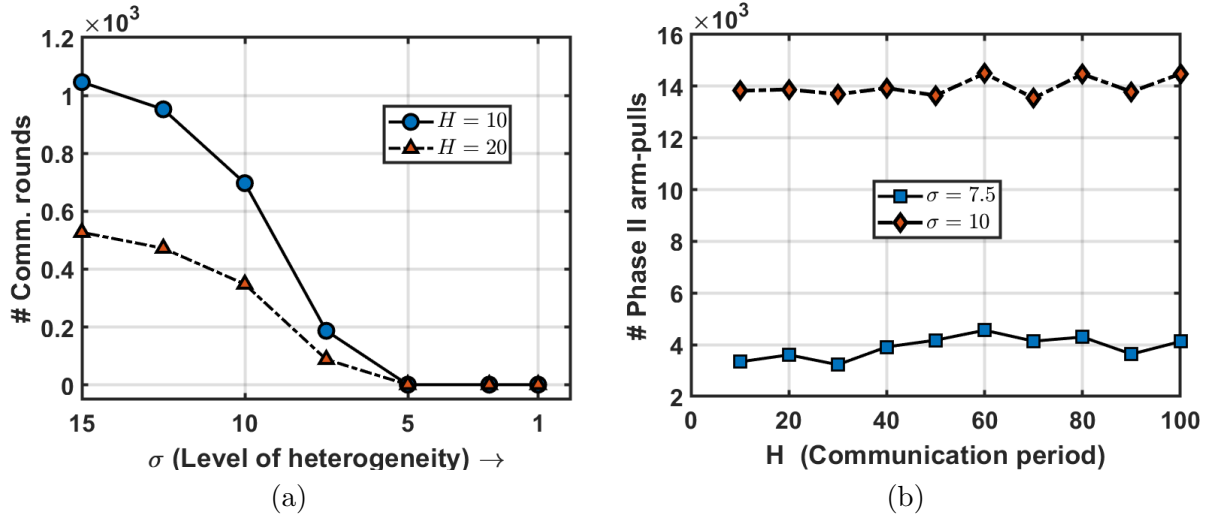


Figure 4: Illustration of the performance of Fed-SEL. (a) Plot of the number of communication rounds versus the level of heterogeneity  $\sigma$ . (b) Plot of the number of Phase II arm-pulls versus the communication period  $H$ .

For the example we have considered, arm-heterogeneity has a very natural spatial interpretation: *sensors located closer to the target output higher readings than those farther away*. Note that arm 1 in  $\mathcal{A}_1$  is the best arm, and hence closest to the target. With confidence parameter  $\delta = 0.1$ , we implement a simplified version of Fed-SEL (Algo. 2) without any adaptive quantization. We vary the level of heterogeneity by tuning  $\sigma$ , and for each value of  $\sigma$ , perform 50 trials.

**Main Observations:** Below, we discuss our experimental findings from Figure 4.

- From Fig. 4(a), we note a distinct trend: the number of communication rounds decrease with an increase in the level of arm-heterogeneity  $\sigma$ , exactly as suggested by Theorem 1. Moreover, when  $\sigma = 1$ , Fed-SEL terminates in 1 round, complying with Corollary 1. We conclude that Fed-SEL *does indeed benefit from the notion of heterogeneity that we have introduced*.
- From Fig. 4(a), we note that doubling the communication period  $H$  roughly halves the number of communication rounds, as expected. Importantly, increasing the communication period  $H$  does not cause the number of arm-pulls in Phase II to increase much; see Fig. 4(b). These observations once again align with our theoretical findings in Theorem 1: *since local computations always help for Fed-SEL, one can significantly reduce communication by increasing  $H$ , and at the same time, not pay much of an additional price in terms of sample-complexity*.

## 8 Conclusion and Future Work

We introduced and studied the federated best-arm identification problem where clients with disjoint arm-sets collaborate via a server to identify the globally best arm with the highest mean pay-off. We defined a novel notion of arm-heterogeneity for this problem, and proved that when such heterogeneity is exploited in a principled way, one can significantly reduce the number of communication rounds required to identify the best arm (with high probability). Our finding is unique in that it reveals that unlike federated supervised learning, heterogeneity can provably help for best-arm identification problems in stochastic multi-armed bandits. Finally, a key advantage of our algorithmic approach is that it is robust to worst-case adversarial actions.

There are several interesting avenues of future research, as we discuss below.

- First, we aim to derive lower-bounds on the sample-complexity for the problem we have introduced. This will shed light on the tightness of our results in Theorem 1, and inform the need for potentially better algorithms. It would also be interesting to explore alternate formulations where the communication budget is fixed a priori, and the goal is to maximize the probability of identifying the best arm subject to such a budget.
- Perhaps the most interesting direction is to see if the ideas developed in this paper have immediate extensions to general Markov decision processes (MDP’s) in the context of reinforcement learning. Specifically, we wish to investigate the trade-offs between communication-complexity and performance measures such as PAC bounds or regret for distributed sequential decision-making problems.
- Another important direction is to explore related settings where statistical heterogeneity can provably help. Our conjecture is that for problems where the globally optimal parameter belongs to a *finite set*, one can establish results similar to those in this paper by using *elimination-based* techniques. In this context, one ripe candidate problem is that of hypothesis-testing/statistical inference where the true hypothesis belongs to a finite set of possible hypotheses. Recent results have shown that for peer-to-peer versions of this problem [88, 89], one can design principled approaches to significantly reduce communication.
- Some recent works have shown that by either considering alternate problem formulations (such as personalized variants of FL [90–93]), or by drawing on ideas from representation learning [94], one can train meaningful statistical models in FL, despite statistical heterogeneity. It may be worthwhile to investigate whether such ideas apply to the setting we consider. Finally, while we studied robustness to adversarial actions, it would also be interesting to explore robustness to other challenges, such as distribution shifts [95], or stragglers [80, 96, 97].

## References

- [1] Jakub Konečný, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.
- [2] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agueria y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, pages 1273–1282. PMLR, 2017.
- [3] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konečný, Stefano Mazzocchi, H Brendan McMahan, et al. Towards federated learning at scale: System design. *arXiv preprint arXiv:1902.01046*, 2019.
- [4] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Keith Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977*, 2019.
- [5] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3):50–60, 2020.

- [6] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang. On the convergence of fedavg on non-iid data. *arXiv preprint arXiv:1907.02189*, 2019.
- [7] Grigory Malinovskiy, Dmitry Kovalev, Elnur Gasanov, Laurent Condat, and Peter Richtarik. From local sgd to local fixed-point methods for federated learning. In *International Conference on Machine Learning*, pages 6692–6701. PMLR, 2020.
- [8] Zachary Charles and Jakub Konečný. On the outsized importance of learning rates in local update methods. *arXiv preprint arXiv:2007.00878*, 2020.
- [9] Zachary Charles and Jakub Konečný. Convergence and accuracy trade-offs in federated learning and meta-learning. In *International Conference on Artificial Intelligence and Statistics*, pages 2575–2583. PMLR, 2021.
- [10] Reese Pathak and Martin J Wainwright. Fedsplit: An algorithmic framework for fast federated optimization. *arXiv preprint arXiv:2005.05238*, 2020.
- [11] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*, pages 5132–5143. PMLR, 2020.
- [12] Don Kurian Dennis, Tian Li, and Virginia Smith. Heterogeneity for the win: One-shot federated clustering. *arXiv preprint arXiv:2103.00697*, 2021.
- [13] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Pac bounds for multi-armed bandit and markov decision processes. In *International Conference on Computational Learning Theory*, pages 255–270. Springer, 2002.
- [14] Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(6), 2006.
- [15] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil’ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439. PMLR, 2014.
- [16] Jean-Yves Audibert, Sébastien Bubeck, and Rémi Munos. Best arm identification in multi-armed bandits. In *COLT*, pages 41–53, 2010.
- [17] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, 2011.
- [18] Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *NIPS-Twenty-Sixth Annual Conference on Neural Information Processing Systems*, 2012.
- [19] Kevin Jamieson and Robert Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2014.
- [20] Volkan Cevher, Marco F Duarte, and Richard G Baraniuk. Distributed target localization via spatial sparsity. In *2008 16th European Signal Processing Conference*, pages 1–5. IEEE, 2008.

- [21] Antonio Franchi, Paolo Stegagno, Maurizio Di Rocco, and Giuseppe Oriolo. Distributed target localization and encirclement with a multi-robot system. *IFAC Proceedings Volumes*, 43(16):151–156, 2010.
- [22] Vladimir Savic, Henk Wymeersch, and Santiago Zazo. Belief consensus algorithms for fast distributed target tracking in wireless sensor networks. *Signal Processing*, 95:149–160, 2014.
- [23] Philip Dames, Pratap Tokekar, and Vijay Kumar. Detecting, localizing, and tracking an unknown number of moving targets using a team of mobile robots. *The International Journal of Robotics Research*, 36(13-14):1540–1553, 2017.
- [24] Shahin Shahrampour, Alexander Rakhlin, and Ali Jadbabaie. Distributed detection: Finite-time analysis and impact of network topology. *IEEE Transactions on Automatic Control*, 61(11):3256–3268, 2015.
- [25] Robert M Corless, Gaston H Gonnet, David EG Hare, David J Jeffrey, and Donald E Knuth. On the lambertw function. *Advances in Computational mathematics*, 5(1):329–359, 1996.
- [26] Yudong Chen, Lili Su, and Jiaming Xu. Distributed statistical machine learning in adversarial settings: Byzantine gradient descent. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(2):1–25, 2017.
- [27] Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui, and Julien Stainer. Machine learning with adversaries: Byzantine tolerant gradient descent. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 118–128, 2017.
- [28] Dong Yin, Yudong Chen, Ramchandran Kannan, and Peter Bartlett. Byzantine-robust distributed learning: Towards optimal statistical rates. In *International Conference on Machine Learning*, pages 5650–5659. PMLR, 2018.
- [29] Lingjiao Chen, Zachary Charles, Dimitris Papailiopoulos, et al. Draco: Robust distributed training via redundant gradients. *arXiv preprint arXiv:1803.09877*, 2018.
- [30] Dan Alistarh, Zeyuan Allen-Zhu, and Jerry Li. Byzantine stochastic gradient descent. *arXiv preprint arXiv:1803.08917*, 2018.
- [31] Cong Xie, Oluwasanmi Koyejo, and Indranil Gupta. Generalized byzantine-tolerant sgd. *arXiv preprint arXiv:1802.10116*, 2018.
- [32] Liping Li, Wei Xu, Tianyi Chen, Georgios B Giannakis, and Qing Ling. Rsa: Byzantine-robust stochastic aggregation methods for distributed learning from heterogeneous datasets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 1544–1551, 2019.
- [33] Avishek Ghosh, Justin Hong, Dong Yin, and Kannan Ramchandran. Robust federated learning in a heterogeneous environment. *arXiv preprint arXiv:1906.06629*, 2019.
- [34] Avishek Ghosh, Raj Kumar Maity, Swanand Kadhe, Arya Mazumdar, and Kannan Ramchandran. Communication efficient and byzantine tolerant distributed learning. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2545–2550. IEEE, 2020.
- [35] Avishek Ghosh, Raj Kumar Maity, and Arya Mazumdar. Distributed newton can communicate less and resist byzantine workers. *arXiv preprint arXiv:2006.08737*, 2020.

- [36] Eshcar Hillel, Zohar Karnin, Tomer Koren, Ronny Lempel, and Oren Somekh. Distributed exploration in multi-armed bandits. *arXiv preprint arXiv:1311.0800*, 2013.
- [37] Chao Tao, Qin Zhang, and Yuan Zhou. Collaborative learning with limited interaction: Tight bounds for distributed exploration in multi-armed bandits. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 126–146. IEEE, 2019.
- [38] Keqin Liu and Qing Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE Transactions on Signal Processing*, 58(11):5667–5681, 2010.
- [39] Dileep Kalathil, Naumaan Nayyar, and Rahul Jain. Decentralized learning for multiplayer multiarmed bandits. *IEEE Transactions on Information Theory*, 60(4):2331–2345, 2014.
- [40] Soumya Kar, H Vincent Poor, and Shuguang Cui. Bandit problems in networks: Asymptotically efficient distributed allocation rules. In *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pages 1771–1778. IEEE, 2011.
- [41] Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. Distributed cooperative decision-making in multiarmed bandits: Frequentist and bayesian algorithms. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 167–172. IEEE, 2016.
- [42] Peter Landgren, Vaibhav Srivastava, and Naomi Ehrich Leonard. Distributed cooperative decision making in multi-agent multi-armed bandits. *Automatica*, 125:109445, 2021.
- [43] Shahin Shahrampour, Alexander Rakhlin, and Ali Jadbabaie. Multi-armed bandits in multi-agent networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2786–2790. IEEE, 2017.
- [44] Swapna Buccapatnam, Jian Tan, and Li Zhang. Information sharing in distributed stochastic bandits. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2605–2613. IEEE, 2015.
- [45] Ravi Kumar Kolla, Krishna Jagannathan, and Aditya Gopalan. Collaborative learning of stochastic bandits over a social network. *IEEE/ACM Transactions on Networking*, 26(4):1782–1795, 2018.
- [46] Yuanhao Wang, Jiachen Hu, Xiaoyu Chen, and Liwei Wang. Distributed bandit learning: Near-optimal regret with efficient communication. *arXiv preprint arXiv:1904.06309*, 2019.
- [47] Abishek Sankararaman, Ayalvadi Ganesh, and Sanjay Shakkottai. Social learning in multi agent multi armed bandits. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 3(3):1–35, 2019.
- [48] David Martínez-Rubio, Varun Kanade, and Patrick Rebeschini. Decentralized cooperative stochastic bandits. *arXiv preprint arXiv:1810.04468*, 2018.
- [49] Abhimanyu Dubey et al. Kernel methods for cooperative multi-agent contextual bandits. In *International Conference on Machine Learning*, pages 2740–2750. PMLR, 2020.
- [50] Abhimanyu Dubey and Alex Pentland. Differentially-private federated linear bandits. *arXiv preprint arXiv:2010.11425*, 2020.
- [51] Anusha Lalitha and Andrea Goldsmith. Bayesian algorithms for decentralized stochastic bandits. *arXiv preprint arXiv:2010.10569*, 2020.

- [52] Ronshee Chawla, Abishek Sankararaman, Ayalvadi Ganesh, and Sanjay Shakkottai. The gossiping insert-eliminate algorithm for multi-agent bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 3471–3481. PMLR, 2020.
- [53] Ronshee Chawla, Abishek Sankararaman, and Sanjay Shakkottai. Multi-agent low-dimensional linear bandits. *arXiv preprint arXiv:2007.01442*, 2020.
- [54] Avishek Ghosh, Abishek Sankararaman, and Kannan Ramchandran. Collaborative learning and personalization in multi-agent stochastic linear bandits. *arXiv preprint arXiv:2106.08902*, 2021.
- [55] Mridul Agarwal, Vaneet Aggarwal, and Kamyar Azizzadenesheli. Multi-agent multi-armed bandits with limited communication. *arXiv preprint arXiv:2102.08462*, 2021.
- [56] Zhaowei Zhu, Jingxuan Zhu, Ji Liu, and Yang Liu. Federated bandit: A gossiping approach. In *Abstract Proceedings of the 2021 ACM SIGMETRICS/International Conference on Measurement and Modeling of Computer Systems*, pages 3–4, 2021.
- [57] Chengshuai Shi, Cong Shen, and Jing Yang. Federated multi-armed bandits with personalization. In *International Conference on Artificial Intelligence and Statistics*, pages 2917–2925. PMLR, 2021.
- [58] Kwang-Sung Jun, Lihong Li, Yuzhe Ma, and Xiaojin Zhu. Adversarial attacks on stochastic bandits. *arXiv preprint arXiv:1810.12188*, 2018.
- [59] Fang Liu and Ness Shroff. Data poisoning attacks on stochastic bandits. In *International Conference on Machine Learning*, pages 4042–4050. PMLR, 2019.
- [60] Thodoris Lykouris, Vahab Mirrokni, and Renato Paes Leme. Stochastic bandits robust to adversarial corruptions. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 114–122, 2018.
- [61] Anupam Gupta, Tomer Koren, and Kunal Talwar. Better algorithms for stochastic bandits with adversarial corruptions. In *Conference on Learning Theory*, pages 1562–1578. PMLR, 2019.
- [62] Ilija Bogunovic, Andreas Krause, and Jonathan Scarlett. Corruption-tolerant gaussian process bandit optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 1071–1081. PMLR, 2020.
- [63] Evrard Garcelon, Baptiste Roziere, Laurent Meunier, Jean Tarbouriech, Olivier Teytaud, Alessandro Lazaric, and Matteo Pirotta. Adversarial attacks on linear contextual bandits. *arXiv preprint arXiv:2002.03839*, 2020.
- [64] Ilija Bogunovic, Arpan Losalka, Andreas Krause, and Jonathan Scarlett. Stochastic linear bandits robust to adversarial attacks. In *International Conference on Artificial Intelligence and Statistics*, pages 991–999. PMLR, 2021.
- [65] Daniel Vial, Sanjay Shakkottai, and R Srikant. Robust multi-agent multi-armed bandits. In *Proceedings of the Twenty-second International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pages 161–170, 2021.



- [66] Sebastian U Stich. Local sgd converges fast and communicates little. *arXiv preprint arXiv:1805.09767*, 2018.
- [67] Jianyu Wang and Gauri Joshi. Cooperative sgd: A unified framework for the design and analysis of communication-efficient sgd algorithms. *arXiv preprint arXiv:1808.07576*, 2018.
- [68] Artin Spiridonoff, Alex Olshevsky, and Ioannis Ch Paschalidis. Local sgd with a communication overhead depending only on the number of workers. *arXiv preprint arXiv:2006.02582*, 2020.
- [69] Amirhossein Reisizadeh, Aryan Mokhtari, Hamed Hassani, Ali Jadbabaie, and Ramtin Pedarsani. Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization. In *International Conference on Artificial Intelligence and Statistics*, pages 2021–2031. PMLR, 2020.
- [70] Farzin Haddadpour, Mohammad Mahdi Kamani, Mehrdad Mahdavi, and Viveck Cadambe. Local sgd with periodic averaging: Tighter analysis and adaptive synchronization. In *Advances in Neural Information Processing Systems*, pages 11082–11094, 2019.
- [71] Farzin Haddadpour, Mohammad Mahdi Kamani, Aryan Mokhtari, and Mehrdad Mahdavi. Federated learning with compression: Unified analysis and sharp guarantees. In *International Conference on Artificial Intelligence and Statistics*, pages 2350–2358. PMLR, 2021.
- [72] Blake Woodworth, Kumar Kshitij Patel, Sebastian U Stich, Zhen Dai, Brian Bullins, H Brendan McMahan, Ohad Shamir, and Nathan Srebro. Is local sgd better than minibatch sgd? *arXiv preprint arXiv:2002.07839*, 2020.
- [73] Ahmed Khaled, Konstantin Mishchenko, and Peter Richtárik. First analysis of local gd on heterogeneous data. *arXiv preprint arXiv:1909.04715*, 2019.
- [74] Ahmed Khaled, Konstantin Mishchenko, and Peter Richtárik. Tighter theory for local sgd on identical and heterogeneous data. In *International Conference on Artificial Intelligence and Statistics*, pages 4519–4529. PMLR, 2020.
- [75] Farzin Haddadpour and Mehrdad Mahdavi. On the convergence of local descent methods in federated learning. *arXiv preprint arXiv:1910.14425*, 2019.
- [76] Anastasia Koloskova, Nicolas Loizou, Sadra Boreiri, Martin Jaggi, and Sebastian U Stich. A unified theory of decentralized sgd with changing topology and local updates. *arXiv preprint arXiv:2003.10422*, 2020.
- [77] Anit Kumar Sahu, Tian Li, Maziar Sanjabi, Manzil Zaheer, Ameet Talwalkar, and Virginia Smith. On the convergence of federated optimization in heterogeneous networks. *arXiv preprint arXiv:1812.06127*, 3, 2018.
- [78] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *arXiv preprint arXiv:2007.07481*, 2020.
- [79] Eduard Gorbunov, Filip Hanzely, and Peter Richtárik. Local sgd: Unified theory and new efficient methods. In *International Conference on Artificial Intelligence and Statistics*, pages 3556–3564. PMLR, 2021.

- [80] Aritra Mitra, Rayana Jaafar, George J Pappas, and Hamed Hassani. Achieving linear convergence in federated learning under objective and systems heterogeneity. *arXiv preprint arXiv:2102.07053*, 2021.
- [81] Marshall Pease, Robert Shostak, and Leslie Lamport. Reaching agreement in the presence of faults. *Journal of the ACM (JACM)*, 27(2):228–234, 1980.
- [82] Danny Dolev, Nancy A Lynch, Shlomit S Pinter, Eugene W Stark, and William E Weihl. Reaching approximate agreement in the presence of faults. *Journal of the ACM (JACM)*, 33(3):499–516, 1986.
- [83] C.Y. Koo. Broadcast in radio networks tolerating Byzantine adversarial behavior. In *Proc. of the ACM Symp. on Principles of Distributed Comp.*, pages 275–282. ACM, 2004.
- [84] A. Pelc and D. Peleg. Broadcasting with locally bounded Byzantine faults. *Info. Proc. Letters*, 93(3):109–115, 2005.
- [85] H. J. LeBlanc, H. Zhang, X. Koutsoukos, and S. Sundaram. Resilient asymptotic consensus in robust networks. *IEEE Journal on Selected Areas in Comm.*, 31(4):766–781, 2013.
- [86] Aritra Mitra and Shreyas Sundaram. Byzantine-resilient distributed observers for LTI systems. *Automatica*, 108:108487, 2019.
- [87] Aritra Mitra, John A Richards, and Shreyas Sundaram. A new approach to distributed hypothesis testing and non-Bayesian learning: Improved learning rate and Byzantine-resilience. *IEEE Transactions on Automatic Control*, 2020.
- [88] Aritra Mitra, John A Richards, and Shreyas Sundaram. A communication-efficient algorithm for exponentially fast non-Bayesian learning in networks. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 8347–8352. IEEE, 2019.
- [89] Aritra Mitra, John A Richards, Saurabh Bagchi, and Shreyas Sundaram. Distributed inference with sparse and quantized communication. *IEEE Transactions on Signal Processing*, 2021.
- [90] Yihan Jiang, Jakub Konečný, Keith Rush, and Sreeram Kannan. Improving federated learning personalization via model agnostic meta learning. *arXiv preprint arXiv:1909.12488*, 2019.
- [91] Sen Lin, Guang Yang, and Junshan Zhang. A collaborative learning framework via federated meta-learning. In *2020 IEEE 40th International Conference on Distributed Computing Systems (ICDCS)*, pages 289–299. IEEE, 2020.
- [92] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*, 2020.
- [93] Filip Hanzely, Slavomír Hanzely, Samuel Horváth, and Peter Richtárik. Lower bounds and optimal algorithms for personalized federated learning. *arXiv preprint arXiv:2010.02372*, 2020.
- [94] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. *arXiv preprint arXiv:2102.07078*, 2021.
- [95] Amirhossein Reisizadeh, Farzan Farnia, Ramtin Pedarsani, and Ali Jadbabaie. Robust federated learning: The case of affine distribution shifts. *arXiv preprint arXiv:2006.08907*, 2020.

- [96] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in Neural Information Processing Systems*, 33, 2020.
- [97] Amirhossein Reisizadeh, Isidoros Tziotis, Hamed Hassani, Aryan Mokhtari, and Ramtin Pedarsani. Straggler-resilient federated learning: Leveraging the interplay between statistical accuracy and system heterogeneity. *arXiv preprint arXiv:2012.14453*, 2020.
- [98] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994.
- [99] Ioannis Chatzigeorgiou. Bounds on the lambert function and their application to the outage analysis of user cooperation. *IEEE Communications Letters*, 17(8):1505–1508, 2013.

## A Proof of Theorem 1

The goal of this section is to prove Theorem 1. Throughout the proof, to simplify notation, we will use  $i_1$  as a shorthand to represent the best arm  $a_1^{(i)}$  in arm-set  $\mathcal{A}_i$ . Also, for any arm  $\ell \in \mathcal{A}_i$ , we will use  $T_\ell$  to denote the number of pulls made to arm  $\ell$  by client  $i$  during the course of operation of Fed-SEL. Before diving into the formal argument, we first outline the main steps of the proof.

- **Step 1:** The first step of our analysis is to define a “good event”  $\mathcal{E}$ , and show that it occurs with probability at least  $1 - \delta$ , where  $\delta \in (0, 1)$  is the given confidence parameter. On this good event, the empirical estimates of all arms are within appropriate confidence intervals (as given by the  $\alpha$ 's) of the true means. See Lemma 1.
- **Step 2:** On  $\mathcal{E}$ , we show that the local subroutine (Algo. 1) at each client  $i \in \mathcal{C}$  terminates in finite-time, and outputs the locally best-arm  $a_1^{(i)}$  in  $\mathcal{A}_i$ , i.e.,  $a_i = a_1^{(i)}, \forall i \in \mathcal{C}$ . This step follows from standard arguments [13, 14]; we provide a proof only for completeness. See Lemma 2.
- **Step 3:** The next key step of our analysis is to bound the error of each client  $i$ 's estimate  $\hat{r}_{i,a_i}(\bar{t}_i)$  of the true mean  $r_{i_1}$ , at the end of Phase I, as a function of the local arm-gap  $\Delta_i^* = r_{i_1} - r_{i_2}$ . This analysis, carried out in Lemma 3, reveals that the locally best arm from each arm-set is “well-explored” at the end of Phase I.
- **Step 4:** The final piece of the analysis is to relate the amount of exploration that has been done during Phase I, to that which remains to be done during Phase II, for eliminating each client in the set  $\mathcal{C} \setminus \{1\}$ . Specifically, for each client  $i \in \mathcal{C} \setminus \{1\}$ , we seek a lower bound on the number of times arm  $i_1$  has been sampled during Phase I, and an upper bound on the number of times it still needs to be sampled during Phase II, before being deemed sub-optimal. See Lemma 4. To complete this step, we derive new bounds on the Lambert  $W$  function in Lemma 5 of Appendix A.1. These bounds may be of independent interest.

We now proceed to establish each of the above steps, starting with Step 1.

**Lemma 1.** *Consider the event  $\mathcal{E}$  defined as follows:*

$$\mathcal{E} \triangleq \bigcap_{i \in \mathcal{C}} \bigcap_{\ell \in \mathcal{A}_i} \bigcap_{t \in [T_\ell]} \{|\hat{r}_{i,\ell}(t) - r_\ell| \leq \alpha_i(t)\}. \quad (10)$$

*Then, it holds that  $\mathbb{P}(\mathcal{E}) \geq 1 - \delta$ .*

*Proof.* We can assume that the sequence of rewards for each arm is drawn before the algorithm is initiated. That way, for any  $t \in \{1, 2, \dots, \infty\}$ , the empirical estimate of the mean of an arm  $\ell$  after  $t$  pulls is well defined, even if it has not actually been pulled  $t$  times by a client. Applying the

union bound gives us

$$\begin{aligned}
\mathbb{P}(\mathcal{E}^c) &\leq \sum_{i \in \mathcal{C}} \sum_{\ell \in \mathcal{A}_i} \sum_{t=1}^{\infty} \mathbb{P}(\{|\hat{r}_{i,\ell}(t) - r_\ell| > \alpha_i(t)\}) \\
&\leq \sum_{i \in \mathcal{C}} \sum_{\ell \in \mathcal{A}_i} \sum_{t=1}^{\infty} 2 \exp(-2t\alpha_i^2(t)) \\
&\leq \sum_{i \in \mathcal{C}} \sum_{\ell \in \mathcal{A}_i} \sum_{t=1}^{\infty} \frac{\delta}{2|\mathcal{C}||\mathcal{A}_i|t^2} \\
&\leq \sum_{i \in \mathcal{C}} \sum_{\ell \in \mathcal{A}_i} \frac{\delta}{2|\mathcal{C}||\mathcal{A}_i|} \left(1 + \int_1^{\infty} \frac{1}{z^2} dz\right) \\
&= \sum_{i \in \mathcal{C}} \sum_{\ell \in \mathcal{A}_i} \frac{\delta}{|\mathcal{C}||\mathcal{A}_i|} = \delta.
\end{aligned} \tag{11}$$

The second step above follows from an application of Hoeffding's inequality [98].  $\square$

For Step 2, we have the following lemma.

**Lemma 2.** *Suppose Fed-SEL is run with  $\beta_i(t) = c\alpha_i(t)$ , where  $c \geq 2$ . Then, on the event  $\mathcal{E}$ , for each client  $i \in \mathcal{C}$ , Algo. 1 terminates in finite-time and outputs  $a_i = a_1^{(i)}$ .*

*Proof.* We prove the result for the case when  $c = 2$ ; the same reasoning applies when  $c > 2$ . Fix a client  $i \in \mathcal{C}$ . We first argue that on  $\mathcal{E}$ , the locally best arm  $a_1^{(i)}$  in  $\mathcal{A}_i$  never gets eliminated when Algo. 1 is run. To see this, observe that on  $\mathcal{E}$ , the following holds in each epoch  $t$ :

$$\hat{r}_{i,i_1}(t) \geq r_{i_1} - \alpha_i(t).$$

At the same time, for any other arm  $\ell \in \mathcal{S}_i(t) \setminus \{i_1\}$ , the following also holds on  $\mathcal{E}$ :

$$\hat{r}_{i,\ell}(t) \leq r_\ell + \alpha_i(t) < r_{i_1} + \alpha_i(t).$$

Thus, on  $\mathcal{E}$ , we have that

$$\hat{r}_{i,\ell}(t) - \hat{r}_{i,i_1}(t) < 2\alpha_i(t) = \beta_i(t).$$

From line 5 of Algo. 1, it is then clear that  $a_1^{(i)}$  does not get eliminated on  $\mathcal{E}$ . We now argue that every arm  $\ell \in \mathcal{A}_i \setminus \{i_1\}$  eventually gets eliminated. In particular, we claim that

$$\alpha_i(t) \leq \frac{\Delta_{i_1\ell}}{4}$$

is a sufficient condition to eliminate arm  $\ell$ . To see this, suppose the above condition holds. We then have

$$\begin{aligned}
r_{i_1} - \alpha_i(t) &\geq r_\ell + \alpha_i(t) + 2\alpha_i(t) \\
&\stackrel{\mathcal{E}}{\implies} \hat{r}_{i,i_1}(t) \geq \hat{r}_{i,\ell}(t) + 2\alpha_i(t) = \hat{r}_{i,\ell}(t) + \beta_i(t) \\
&\implies \hat{r}_{i,\max}(t) \geq \hat{r}_{i,\ell}(t) + \beta_i(t).
\end{aligned} \tag{12}$$

The claim then follows from line 5 of Algo. 1. This completes the proof.  $\square$

The next lemma provides a bound on the confidence intervals at the end of Phase I.

**Lemma 3.** *Suppose Fed-SEL is run with  $\beta_i(t) = c\alpha_i(t)$ , where  $c > 2$ . Then, on the event  $\mathcal{E}$ , the following holds for every client  $i \in \mathcal{C}$ :*

$$\alpha_i(\bar{t}_i) < \frac{c}{(c-2)^2} \Delta_i^*, \quad (13)$$

where  $\Delta_i^* = r_{i_1} - r_{i_2}$ .

*Proof.* Fix a client  $i \in \mathcal{C}$ . From Lemma 2, we know that every arm  $\ell \in \mathcal{A}_i \setminus \{i_1\}$  eventually gets eliminated when client  $i$  runs Algo. 1 on  $\mathcal{A}_i$ . We split our subsequent analysis into two cases depending on *when* the second best arm  $i_2$  in  $\mathcal{A}_i$  gets eliminated during Phase I.

**Case 1:** Suppose arm  $i_2$  is eliminated in the last epoch  $\bar{t}_i$ . On event  $\mathcal{E}$ , the arm with the largest empirical mean in  $\mathcal{S}_i(\bar{t}_i)$  must be  $i_1$  (from Lemma 2). Thus, we must have

$$\begin{aligned} \hat{r}_{i,i_1}(\bar{t}_i) &\geq \hat{r}_{i,i_2}(\bar{t}_i) + c\alpha_i(\bar{t}_i) \\ \xrightarrow{\mathcal{E}} r_{i_1} + \alpha_i(\bar{t}_i) &\geq \hat{r}_{i,i_2}(\bar{t}_i) + c\alpha_i(\bar{t}_i) \\ \xrightarrow{\mathcal{E}} r_{i_1} + \alpha_i(\bar{t}_i) &\geq r_{i_2} - \alpha_i(\bar{t}_i) + c\alpha_i(\bar{t}_i) \\ \implies \alpha_i(\bar{t}_i) &\leq \frac{\Delta_i^*}{(c-2)}. \end{aligned} \quad (14)$$

**Case 2:** Suppose arm  $i_2$  is eliminated in some epoch  $t < \bar{t}_i$ . For this to happen on  $\mathcal{E}$ , it must be that  $i_1 = \operatorname{argmax}_{\ell \in \mathcal{S}_i(t)} \hat{r}_{i,\ell}(t)$ . This is easy to verify. Moreover, note that there must exist some arm  $\ell \in \mathcal{A}_i \setminus \{i_1 \cup i_2\}$  that does not get eliminated in epoch  $t$ , but gets eliminated in epoch  $\bar{t}_i$ . Thus, the following inequalities must hold simultaneously:

$$\hat{r}_{i,i_1}(t) - \hat{r}_{i,i_2}(t) \geq c\alpha_i(t); \quad \hat{r}_{i,i_1}(t) - \hat{r}_{i,\ell}(t) < c\alpha_i(t), \quad (15)$$

implying  $\hat{r}_{i,\ell}(t) > \hat{r}_{i,i_2}(t)$ . Since we are on  $\mathcal{E}$ , this further implies that

$$r_\ell + \alpha_i(t) > r_{i_2} - \alpha_i(t) \implies 2\alpha_i(t) > \Delta_{i_2\ell}. \quad (16)$$

At the same time, since arm  $i_2$  gets eliminated in epoch  $t$ , we must also have

$$\alpha_i(t) \leq \frac{\Delta_i^*}{(c-2)}, \quad (17)$$

following the analysis of Case 1. Combining Eq. (16) with the above bound yields

$$\begin{aligned} \Delta_{i_2\ell} &< 2 \frac{\Delta_i^*}{(c-2)} \\ \implies r_{i_2} - r_{i_1} + r_{i_1} - r_\ell &< 2 \frac{\Delta_i^*}{(c-2)} \\ \implies \Delta_{i_1\ell} &< \frac{c}{(c-2)} \Delta_i^*. \end{aligned} \quad (18)$$

Finally, since arm  $\ell$  gets eliminated in the last epoch  $\bar{t}_i$ , an argument identical to the Case 1 analysis reveals that

$$\alpha_i(\bar{t}_i) \leq \frac{\Delta_{i_1\ell}}{(c-2)}. \quad (19)$$

Combining the above inequality with the one in Eq. (18), we obtain

$$\alpha_i(\bar{t}_i) < \frac{c}{(c-2)^2} \Delta_i^*. \quad (20)$$

This completes the analysis for Case 2. The claim of the lemma then follows immediately from equations (14) and (20).  $\square$

Before proceeding to our next result, we need to first take a brief detour and familiarize ourselves with the Lambert  $W$  function that will show up in our subsequent analysis. Accordingly, we start by noting that the Lambert  $W$  function is defined to be the multi-valued inverse of the function  $f(w) = we^w$ , where  $w \in \mathbb{C}$  is any complex number [25].<sup>12</sup> Specifically, the Lambert  $W$  function, denoted by  $W(x)$ , has the following defining property:

$$W(x)e^{W(x)} = x, \text{ where } x \in \mathbb{C}. \quad (21)$$

When  $x$  is real and belongs to  $[-1/e, 0)$ ,  $W(x)$  can take on two possible values. The branch satisfying  $W(x) \geq -1$  is called the *principal branch* and is denoted by  $W_0(x)$ . The other branch with values satisfying  $W(x) \leq -1$  is denoted by  $W_{-1}(x)$ . The two branches meet at the point  $x = -1/e$ , where they both take on the value of  $-1$ . For a depiction of these branches, see Fig. 5.

We now state and prove the final ingredient required for establishing Theorem 1.

**Lemma 4.** *For  $t \in [1, \infty)$ , define  $\alpha(t) \triangleq \sqrt{2 \log(\bar{c}t)}/t$ , where  $\bar{c} > e$ . Given  $\Delta \in (0, 1)$  and  $\sigma > 1$  such that  $\sigma\Delta \leq 1$ , suppose  $\tilde{t}$  and  $t'$  satisfy:*

$$\alpha(\tilde{t}) = \frac{\sigma\Delta}{8}, \text{ and } \alpha(t') = \frac{\Delta}{8}.$$

Then, we have

$$t' - \tilde{t} \leq \left( \frac{128}{\Delta^2} \log \frac{128\bar{c}}{\Delta^2} - \frac{128}{\sigma^2\Delta^2} \log \frac{128\bar{c}}{\sigma^2\Delta^2} \right) + \left( \frac{512}{\Delta^2} \log \log \frac{128\bar{c}}{\Delta^2} - \frac{32}{\sigma^2\Delta^2} \log \log \frac{128\bar{c}}{\sigma^2\Delta^2} \right). \quad (22)$$

*Proof.* Our goal is to leverage the bounds on the  $W_{-1}(x)$  branch that we derived in Lemma 5. To do so, we need to first transform the equations concerning  $\alpha(\tilde{t})$  and  $\alpha(t')$  into an appropriate form. Accordingly, simple manipulations yield:

$$\left( -\frac{\sigma^2\Delta^2\tilde{t}}{128} \right) \exp \left( -\frac{\sigma^2\Delta^2\tilde{t}}{128} \right) = -\frac{\sigma^2\Delta^2}{128\bar{c}}; \quad \left( -\frac{\Delta^2 t'}{128} \right) \exp \left( -\frac{\Delta^2 t'}{128} \right) = -\frac{\Delta^2}{128\bar{c}}. \quad (23)$$

Based on the conditions of the lemma, it is easy to see that both  $-\frac{\sigma^2\Delta^2}{128\bar{c}}$  and  $-\frac{\Delta^2}{128\bar{c}}$  belong to the interval  $[-1/e^2, 0)$ . Using the defining property of the Lambert  $W$  function in Eq. (21), we claim:

$$\tilde{t} = -\frac{128}{\sigma^2\Delta^2} W_{-1} \left( -\frac{\sigma^2\Delta^2}{128\bar{c}} \right); \quad t' = -\frac{128}{\Delta^2} W_{-1} \left( -\frac{\Delta^2}{128\bar{c}} \right). \quad (24)$$

We need to now argue why the above solutions for  $\tilde{t}$  and  $t'$  involve the  $W_{-1}(x)$  branch, and not the principal branch  $W_0(x)$ . To see why, note that  $W_0(x) \geq ex, \forall x \in [-1/e^2, 0)$ . This in turn follows by observing that:

$$W_0(x) \geq -1 \implies e \geq e^{-W_0(x)} \implies ex \leq xe^{-W_0(x)} = W_0(x),$$

where we used  $x < 0$  and Eq. (21). Now suppose the solution for  $\tilde{t}$  in Eq. (24) involves  $W_0(x)$  instead of  $W_{-1}(x)$ . Using  $W_0(x) \geq ex$  then leads to the following:

$$\tilde{t} = -\frac{128}{\sigma^2\Delta^2} W_0 \left( -\frac{\sigma^2\Delta^2}{128\bar{c}} \right) \leq \frac{e}{\bar{c}} < 1,$$

since  $\bar{c} > e$ . This leads to a contradiction as  $\alpha(t)$  is only defined for  $t \in [1, \infty)$ . A similar argument can be made for the solution involving  $t'$ . We have thus justified (24).

<sup>12</sup>Here, and in what follows, we use  $e$  to denote the exponential function.

We can now directly appeal to the bounds in Lemma 5 to obtain

$$W_{-1} \left( -\frac{\sigma^2 \Delta^2}{128\bar{c}} \right) \leq \log \left( \frac{\sigma^2 \Delta^2}{128\bar{c}} \right) - \frac{1}{4} \log \log \frac{128\bar{c}}{\sigma^2 \Delta^2}; \quad W_{-1} \left( -\frac{\Delta^2}{128\bar{c}} \right) \geq \log \left( \frac{\Delta^2}{128\bar{c}} \right) - 4 \log \log \frac{128\bar{c}}{\Delta^2}. \quad (25)$$

Using the above bounds in conjunction with the expressions for  $\tilde{t}$  and  $t'$  in Eq. (24) immediately leads to the claim of the lemma.  $\square$

We are now in position to complete the proof of Theorem 1.

*Proof. (Theorem 1)* We first argue that Fed-SEL is consistent, i.e., it outputs the globally best arm. To this end, we claim that on the event  $\mathcal{E}$  in Eq. (10), client 1 always remains active during Phase II (recall that  $a^* = 1 \in \mathcal{A}_1$ ). To see why this is true, fix any client  $i \in \mathcal{C}$ , and note that based on our encoding-decoding strategy,

$$\begin{aligned} |\tilde{r}_i(k) - \hat{r}_{i,a_i}(\bar{t}_i + kH)| &\leq \frac{1}{2} \frac{1}{2^{B_i(k)}} \\ &\leq \frac{\alpha_i(\bar{t}_i + kH)}{2}, \forall k \geq 1, \end{aligned} \quad (26)$$

where we used the expression for  $B_i(k)$  in Eq. (3). The above bound holds trivially when  $k = 0$ , since  $\tilde{r}_i(0) = \hat{r}_{i,a_i}(\bar{t}_i)$ . Next, observe that at any round  $k$ ,

$$\begin{aligned} \tilde{r}_i^{(L)}(k) &= \tilde{r}_i(k) - 2\alpha_i(\bar{t}_i + kH) \\ &\stackrel{(26)}{\leq} \hat{r}_{i,a_i}(\bar{t}_i + kH) + \frac{\alpha_i(\bar{t}_i + kH)}{2} - 2\alpha_i(\bar{t}_i + kH) \\ &\stackrel{(a)}{=} \hat{r}_{i,i_1}(\bar{t}_i + kH) - \frac{3}{2}\alpha_i(\bar{t}_i + kH) \\ &\stackrel{\mathcal{E}}{\leq} r_{i_1} + \alpha_i(\bar{t}_i + kH) - \frac{3}{2}\alpha_i(\bar{t}_i + kH) \\ &< r_{i_1}. \end{aligned} \quad (27)$$

Here, (a) follows from Lemma 2 where we showed that on  $\mathcal{E}$ ,  $a_i = a_1^{(i)} = i_1, \forall i \in \mathcal{C}$ . Using the same reasoning as above, we can show that

$$\tilde{r}_i^{(U)}(k) > r_{i_1}, \forall i \in \mathcal{C}. \quad (28)$$

From (27) and (28), we conclude that on  $\mathcal{E}$ ,

$$\tilde{r}_1^{(U)}(k) > r_1 \text{ and } \tilde{r}_i^{(L)}(k) < r_{i_1}, \forall i \in \mathcal{C} \setminus \{1\}.$$

Based on lines 4 and 5 of Fed-SEL, it is easy to then see that client 1 always remains in the active-client set  $\bar{\Psi}(k)$  on the event  $\mathcal{E}$ .

We now show that every client  $i \in \mathcal{C} \setminus \{1\}$  eventually gets removed from the active-client set. In particular, for a client  $i \in \mathcal{C} \setminus \{1\}$  to get removed, we claim that the following is a set of sufficient conditions:

$$\begin{aligned} \alpha_i(\bar{t}_i + kH) &\leq \frac{\Delta_{1i_1}}{8}, \text{ and} \\ \alpha_1(\bar{t}_1 + kH) &\leq \frac{\Delta_{1i_1}}{8}. \end{aligned} \quad (29)$$



Suppose both the above conditions hold. We then have

$$\begin{aligned}
\tilde{r}_i^{(U)}(k) &= \tilde{r}_i(k) + 2\alpha_i(\bar{t}_i + kH) \\
&\stackrel{(26)}{\leq} \hat{r}_{i,a_i}(\bar{t}_i + kH) + \frac{\alpha_i(\bar{t}_i + kH)}{2} + 2\alpha_i(\bar{t}_i + kH) \\
&\stackrel{\varepsilon}{=} \hat{r}_{i,i_1}(\bar{t}_i + kH) + \frac{5}{2}\alpha_i(\bar{t}_i + kH) \\
&\stackrel{\varepsilon}{\leq} r_{i_1} + \alpha_i(\bar{t}_i + kH) + \frac{5}{2}\alpha_i(\bar{t}_i + kH) \\
&\stackrel{(29)}{\leq} r_{i_1} + \frac{7}{16}\Delta_{1i_1}.
\end{aligned} \tag{30}$$

Similarly, we have

$$\tilde{r}_1^{(L)}(k) \geq r_1 - \frac{7}{16}\Delta_{1i_1}. \tag{31}$$

From (30) and (31), we note that  $\tilde{r}_1^{(L)}(k) > \tilde{r}_i^{(U)}(k)$ , which is sufficient for client  $i$  to get removed from the active-client set (based on line 5 of **Fed-SEL**). Thus, we have argued that on  $\mathcal{E}$ , **Fed-SEL** will eventually terminate with  $\bar{\Psi}(k+1) = \{1\}$ , for some round  $k$ . The claim that **Fed-SEL** is  $(0, \delta)$ -PAC then follows immediately by noting that on  $\mathcal{E}$ ,  $a_1 = a^* = 1$ .

We now derive bounds on the communication-complexity of **Fed-SEL**. For any  $i \in \mathcal{C} \setminus \{1\}$ , let  $R_{i1}$  (resp.,  $R_{1i}$ ) be the first round such that  $\alpha_i(\bar{t}_i + kH) \leq \Delta_{1i_1}/8$  (resp.,  $\alpha_1(\bar{t}_1 + kH) \leq \Delta_{1i_1}/8$ ). Based on the fact that  $\alpha_i(t)$  is monotonically decreasing for all  $i \in \mathcal{C}$  (this is easy to verify), and the above analysis, client  $i$  remains active for at most  $R_i = \max\{R_{i1}, R_{1i}\}$  rounds. To bound  $R_{i1}$ , we start by noting that with  $c = 8$  in Lemma 3,

$$\alpha_i(\bar{t}_i) < \frac{2}{9}\Delta_i^* < \frac{\Delta_i^*}{8} = \frac{\sigma_{i1}\Delta_{1i_1}}{8},$$

where for the last step we used the definition of the arm-heterogeneity index  $\sigma_{i1}$  in Eq. (4). Clearly, if  $\sigma_{i1} \leq 1$ , then  $R_{i1} = 1$ . Now suppose  $\sigma_{i1} > 1$ . In this case, we have

$$R_{i1} \leq \left\lceil \frac{1}{H} \left( \left\lceil \alpha_i^{-1} \left( \frac{\Delta_{1i_1}}{8} \right) \right\rceil - \left\lfloor \alpha_i^{-1} \left( \frac{\sigma_{i1}\Delta_{1i_1}}{8} \right) \right\rfloor \right) \right\rceil. \tag{32}$$

We now apply Lemma 4 with

$$\alpha(t) = \alpha_i(t), \Delta = \Delta_{1i_1}, \sigma = \sigma_{i1}, \text{ and } \bar{c} = \bar{c}_i = \sqrt{\frac{4|\mathcal{C}||\mathcal{A}_i|}{\delta}}.$$

Since  $|\mathcal{A}_i| \geq 2$ , we have  $\bar{c}_i > e$ . Moreover,  $\sigma_{i1}\Delta_{1i_1} = \Delta_i^* \leq 1$ . Thus, Lemma 4 is indeed applicable. From (22) and (32), we then have

$$R_{i1} \leq \frac{1}{H} \left( \underbrace{\left( \frac{128}{\Delta_{1i_1}^2} \log \frac{128\bar{c}_i}{\Delta_{1i_1}^2} - \frac{128}{\sigma_{i1}^2 \Delta_{1i_1}^2} \log \frac{128\bar{c}_i}{\sigma_{i1}^2 \Delta_{1i_1}^2} \right)}_{T_1} + T_2 \right) + 1,$$

where

$$T_2 = \frac{512}{\Delta_{1i_1}^2} \log \log \frac{128\bar{c}_i}{\Delta_{1i_1}^2} - \frac{32}{\sigma_{i1}^2 \Delta_{1i_1}^2} \log \log \frac{128\bar{c}_i}{\sigma_{i1}^2 \Delta_{1i_1}^2} + 2 = O \left( \frac{1}{\Delta_{1i_1}^2} \log \log \frac{\bar{c}_i}{\Delta_{1i_1}^2} \right).$$

Using  $\sigma_{i1}\Delta_{1i_1} = \Delta_i^*$  to simplify the expression for  $T_1$ , we obtain

$$T_1 = \frac{128}{(\Delta_i^*)^2} (\sigma_{i1}^2 - 1) \log \frac{128\bar{c}_i}{\Delta_{1i_1}^2} + \frac{256}{(\Delta_i^*)^2} \log \sigma_{i1}.$$

Combining all the above pieces together yields the upper bound on  $R_{i1}$  in Eq. (5). Following exactly the same reasoning as above, one can derive the upper bound on  $R_{1i}$  in Eq. (6).

The total number of communication rounds  $R$  of **Fed-SEL** equals the number of rounds it takes to eliminate every client in the set  $\mathcal{C} \setminus \{1\}$ . Thus,  $R = \max_{i \in \mathcal{C} \setminus \{1\}} R_i$ . Regarding the claim about the number of bits exchanged, note from the definition of  $R_{i1}$  that for every round  $k \leq R_{i1}$ ,

$$\alpha_i(\bar{t}_i + kH) \geq \frac{\Delta_{1i_1}}{8} \geq \frac{\Delta_i^*}{8},$$

where we used the definition of  $\Delta_i^*$ . From the expression for  $B_i(k)$  in Eq. (3), we then have

$$B_i(k) \leq \log_2 \left( \frac{1}{\alpha_i(\bar{t}_i + kH)} \right) + 1 \leq \log_2 \left( \frac{8}{\Delta_i^*} \right) + 1 = O \left( \log_2 \left( \frac{1}{\Delta_i^*} \right) \right).$$

This establishes the communication-complexity of **Fed-SEL**.

Finally, we derive the sample-complexity bounds for **Fed-SEL** in Eq. (7). For any client  $i$ , the total number of arm-pulls it makes is the sum of the number of pulls during Phases I and II. During Phase II, client  $i$  only samples arm  $a_i = i_1$  from its arm-set on the event  $\mathcal{E}$ . Moreover, it keeps sampling  $i_1$  for the entire duration it remains in the active-client set. Since  $i_1$  is sampled  $H$  times in each round, and client  $i$  remains active for at most  $R_i$  rounds, the number of arm-pulls made by client  $i$  during Phase II is  $HR_i$ . Since client 1 remains active till the termination of **Fed-SEL**, we have  $R_1 = R$ .

For Phase I, recall from Lemma 2 that an arm  $\ell \in \mathcal{A}_i \setminus \{i_1\}$  gets removed from the set  $\mathcal{S}_i(t)$  when

$$\alpha_i(t) \leq \frac{\Delta_{i_1\ell}}{4},$$

which holds for

$$t = O \left( \frac{\log \left( \frac{|\mathcal{C}||\mathcal{A}_i|}{\delta\Delta_{i_1\ell}} \right)}{\Delta_{i_1\ell}^2} \right).$$

The above discussion immediately leads to the sample-complexity bound in Eq. (7). This completes the proof of Theorem 1.  $\square$

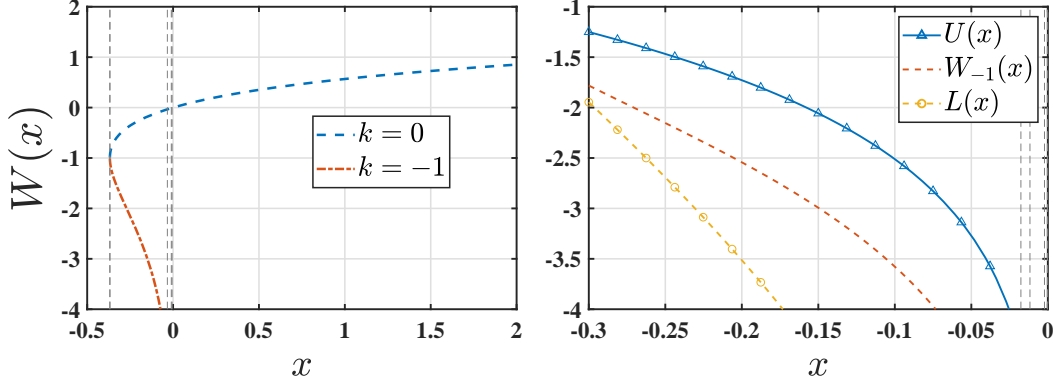


Figure 5: **(Left)** Illustration of the two branches of the Lambert  $W$  function. **(Right)** Illustration of the bounds in Lemma 5. In this figure,  $U(x)$  and  $L(x)$  are the upper and lower bounds, respectively, on  $W_{-1}(x)$  that we derive in Lemma 5; see Eq. (33).

### A.1 Bounds on the Lambert $W$ function

In this section, we will derive new bounds on the Lambert  $W$  function. To derive our bounds, we will rely on the following two facts: (i)  $W_{-1}(x) < -1, \forall x \in [-1/e^2, 0)$ , and (ii)  $-4 < W_{-1}(-1/e^2) < -3$ . The first fact follows readily by noting that  $W_{-1}(-1/e) = -1$ , and that  $W_{-1}(x)$  is a strictly decreasing function on  $(-1/e, 0)$ ; see [25]. For the second fact, one can readily use any standard mathematical programming software (such as MATLAB) to verify that  $W_{-1}(-1/e^2) \approx -3.1462$ . Alternatively, one can use the bounds derived in [99] to arrive at the second fact. We have the following result. For a pictorial depiction of our bounds, see Fig. 5.

**Lemma 5.** For  $x \in [-1/e^2, 0)$ , we have

$$\log(-x) - 4 \log(-\log(-x)) \leq W_{-1}(x) \leq \log(-x) - \frac{1}{4} \log(-\log(-x)). \quad (33)$$

*Proof.* We start by defining a function  $f_\gamma(x)$  parameterized by  $\gamma$ :

$$f_\gamma(x) = \log(-x) - \gamma \log(-\log(-x)) - W_{-1}(x). \quad (34)$$

Using  $f'_\gamma(x)$  to denote the derivative of  $f_\gamma(x)$  w.r.t.  $x$ , we obtain

$$\begin{aligned} f'_\gamma(x) &= \frac{1}{x} \left[ 1 - \frac{\gamma}{\log(-x)} - \frac{W_{-1}(x)}{1 + W_{-1}(x)} \right] \\ &= \frac{1}{x} \left[ \frac{\log(-x) - \gamma(1 + W_{-1}(x))}{\log(-x)(1 + W_{-1}(x))} \right]. \end{aligned} \quad (35)$$

Here, we used the fact that for  $x \in [-1/e^2, 0)$ , the derivative of  $W_{-1}(x)$  is given by [25]

$$\frac{W_{-1}(x)}{x(1 + W_{-1}(x))}.$$

Since  $W_{-1}(x) < -1$  for  $x \in [-1/e^2, 0)$ , the denominator of the expression for  $f'_\gamma(x)$  in Eq. (35) is negative. To study the behavior of the numerator, we consider another  $\gamma$ -parameterized function, namely  $g_\gamma(x)$ , as follows:

$$g_\gamma(x) = \log(-x) - \gamma(1 + W_{-1}(x)). \quad (36)$$

Using  $g'_\gamma(x)$  to denote the derivative of  $g_\gamma(x)$  w.r.t.  $x$ , we obtain

$$g'_\gamma(x) = \frac{1 + (1 - \gamma)W_{-1}(x)}{x(1 + W_{-1}(x))}. \quad (37)$$

**Establishing the lower bound:** For proving the lower bound on  $W_{-1}(x)$  in (33), we set  $\gamma = 4$  in Eq. (37) and note that both the numerator and the denominator of the resulting expression are positive; this once again follows from the fact that  $W_{-1}(x) < -1$  for  $x \in [-1/e^2, 0)$ . We conclude:  $g'_4(x) > 0, \forall x \in [-1/e^2, 0)$ . Moreover, setting  $\gamma = 4$  in (36), we obtain

$$\begin{aligned} g_4(-1/e^2) &= \log(1/e^2) - 4(1 + W_{-1}(-1/e^2)) \\ &= -6 - 4W_{-1}(-1/e^2) \\ &> -6 + 12 > 0, \end{aligned} \quad (38)$$

where we used  $W_{-1}(-1/e^2) < -3$ . We have thus shown that  $g_4(x) > 0, \forall x \in [-1/e^2, 0)$ . Since the numerator of  $f'_4(x)$  in Eq. (35) is precisely  $g_4(x)$ , we conclude that  $f'_4(x) < 0, \forall x \in [-1/e^2, 0)$ . At the same time, plugging in  $\gamma = 4$  in (34) yields:

$$\begin{aligned} f_4(-1/e^2) &= \log(1/e^2) - 4\log(-\log(1/e^2)) - W_{-1}(-1/e^2) \\ &= -2 - 4\log(2) - W_{-1}(-1/e^2) \\ &< 2 - 4\log(2) < 0, \end{aligned} \quad (39)$$

where we used  $W_{-1}(-1/e^2) > -4$ . Combining all the above pieces together, we conclude that  $f_4(x) < 0, \forall x \in [-1/e^2, 0)$ . This immediately leads to the desired lower bound.

**Establishing the upper bound:** The proof of the upper bound proceeds in a very similar way. We start by noting that

$$g'_{1/4}(x) = \frac{1 + 3/4W_{-1}(x)}{x(1 + W_{-1}(x))}. \quad (40)$$

For  $x \in [-1/e^2, 0)$ , while the denominator of the fraction on the R.H.S. of the above expression is positive, the numerator is negative. The latter claim follows from the following facts: (i)  $W_{-1}(-1/e^2) < -3$ , and (ii)  $W_{-1}(x)$  is monotonically decreasing. Thus,  $g'_{1/4}(x) < 0, \forall x \in [-1/e^2, 0)$ . Moreover,

$$\begin{aligned} g_{1/4}(-1/e^2) &= \log(1/e^2) - \frac{1}{4}(1 + W_{-1}(-1/e^2)) \\ &= -\frac{9}{4} - \frac{1}{4}W_{-1}(-1/e^2) \\ &< 0, \end{aligned} \quad (41)$$

where we used  $W_{-1}(-1/e^2) > -4$ . We have thus shown that  $g_{1/4}(x) < 0, \forall x \in [-1/e^2, 0)$ . Since the numerator of  $f'_{1/4}(x)$  is  $g_{1/4}(x)$ , it follows that  $f'_{1/4}(x) > 0, \forall x \in [-1/e^2, 0)$ . Finally, observe

$$\begin{aligned} f_{1/4}(-1/e^2) &= \log(1/e^2) - \frac{1}{4}\log(-\log(1/e^2)) - W_{-1}(-1/e^2) \\ &= -2 - \frac{1}{4}\log(2) - W_{-1}(-1/e^2) \\ &> 1 - \frac{1}{4}\log(2) > 0, \end{aligned} \quad (42)$$

where we used  $W_{-1}(-1/e^2) < -3$ . We thus conclude that  $f_{1/4}(x) > 0, \forall x \in [-1/e^2, 0)$ . The upper bound on  $W_{-1}(x)$  in Eq. (33) follows immediately. This concludes the proof.  $\square$

## B Proof of Theorem 2

In this section, we analyze **Robust Fed-SEL**, namely Algorithm 2. Recall that  $\mathcal{H}$  and  $\mathcal{J}$  represent the sets of honest and adversarial clients, respectively. To proceed, let  $T_{i,\ell}$  denote the total number of times an arm  $\ell \in \mathcal{A}_j$  is pulled by an honest client  $i \in \mathcal{C}_j \cap \mathcal{H}$  during the execution of **Robust Fed-SEL**. We then have the following analog of Lemma 1.

**Lemma 6.** *Consider the event  $\mathcal{E}_{\mathcal{H}}$  defined as follows:*

$$\mathcal{E}_{\mathcal{H}} \triangleq \bigcap_{j \in [N]} \bigcap_{i \in \mathcal{C}_j \cap \mathcal{H}} \bigcap_{\ell \in \mathcal{A}_j} \bigcap_{t \in [T_{i,\ell}]} \{|\hat{r}_{i,\ell}(t) - r_{\ell}| \leq \alpha_j(t)\}. \quad (43)$$

Then, it holds that  $\mathbb{P}(\mathcal{E}_{\mathcal{H}}) \geq 1 - \delta$ .

*Proof.* Following the same reasoning as in Lemma 1,

$$\begin{aligned} \mathbb{P}(\mathcal{E}_{\mathcal{H}}^c) &\leq \sum_{j \in [N]} \sum_{i \in \mathcal{C}_j \cap \mathcal{H}} \sum_{\ell \in \mathcal{A}_j} \sum_{t=1}^{\infty} \mathbb{P}(\{|\hat{r}_{i,\ell}(t) - r_{\ell}| > \alpha_j(t)\}) \\ &\leq \sum_{j \in [N]} \sum_{i \in \mathcal{C}_j \cap \mathcal{H}} \sum_{\ell \in \mathcal{A}_j} \sum_{t=1}^{\infty} 2 \exp(-2t\alpha_j^2(t)) \\ &\leq \sum_{j \in [N]} \sum_{i \in \mathcal{C}_j \cap \mathcal{H}} \sum_{\ell \in \mathcal{A}_j} \sum_{t=1}^{\infty} \frac{\delta}{2|\mathcal{C}| |\mathcal{A}_j| t^2} \\ &\leq \sum_{j \in [N]} \sum_{i \in \mathcal{C}_j \cap \mathcal{H}} \sum_{\ell \in \mathcal{A}_j} \frac{\delta}{2|\mathcal{C}| |\mathcal{A}_j|} \left(1 + \int_1^{\infty} \frac{1}{z^2} dz\right) \\ &= \sum_{j \in [N]} \sum_{i \in \mathcal{C}_j \cap \mathcal{H}} \sum_{\ell \in \mathcal{A}_j} \frac{\delta}{|\mathcal{C}| |\mathcal{A}_j|} \\ &= \frac{|\mathcal{C} \cap \mathcal{H}|}{|\mathcal{C}|} \delta \leq \delta, \end{aligned} \quad (44)$$

which leads to the desired result.  $\square$

The next lemma provides the main intermediate result we need to prove Theorem 2.

**Lemma 7.** *Suppose **Robust Fed-SEL** is run by every honest client. Moreover, suppose  $|\mathcal{C}_j \cap \mathcal{J}| \leq f$  and  $|\mathcal{C}_j| \geq 3f + 1, \forall j \in [N]$ . On event  $\mathcal{E}_{\mathcal{H}}$ , the following then hold:*

(i)  $\bar{a}_j = j_1, \forall j \in [N]$ .

(ii) In very round  $k$ ,

$$\bar{r}_j^{(U)}(k) \in \text{Conv}\left(\{\tilde{r}_i^{(U)}(k)\}_{i \in \bar{\mathcal{C}}_j \cap \mathcal{H}}\right); \quad \bar{r}_j^{(L)}(k) \in \text{Conv}\left(\{\tilde{r}_i^{(L)}(k)\}_{i \in \bar{\mathcal{C}}_j \cap \mathcal{H}}\right), \forall j \in [N], \quad (45)$$

where we use  $\text{Conv}(\Omega)$  to represent the convex hull of a set of points  $\Omega$ .

*Proof.* Fix a client group  $j \in [N]$ . Based on Lemma 2, note that when each client  $i \in \mathcal{C}_j \cap \mathcal{H}$  runs Algo. 1 on  $\mathcal{A}_j$ , the local subroutine will eventually terminate on event  $\mathcal{E}_{\mathcal{H}}$ , and client  $i$  will output  $a_i = j_1$ , i.e., each honest client  $i$  in  $\mathcal{C}_j$  will correctly identify the locally best arm  $j_1$  in  $\mathcal{A}_j$  at the end of Phase I. Since  $|\mathcal{C}_j \cap \mathcal{J}| \leq f$  and  $|\mathcal{C}_j| \geq 3f + 1$ , at least  $2f + 1$  clients from  $\mathcal{C}_j$  are guaranteed to

report to the server at the end of Phase I. Among them, since at most  $f$  can be adversarial, observe that the majority voting operation in line 2 of **Robust Fed-SEL** yields  $\bar{a}_j = j_1$ . This establishes the first claim of the lemma.

For the second claim, let us start by noting that on  $\mathcal{E}_{\mathcal{H}}$ , the number of adversaries in the set  $\bar{\mathcal{C}}_j$  is at most  $f_j = f - |\mathcal{C}'_j| + |\bar{\mathcal{C}}_j|$ . This follows from the fact that on  $\mathcal{E}_{\mathcal{H}}$ , any client  $i \in \mathcal{C}'_j$  who reports an arm other than  $j_1$  must be adversarial, i.e., on  $\mathcal{E}_{\mathcal{H}}$ ,  $\mathcal{C}'_j \setminus \bar{\mathcal{C}}_j \subseteq \mathcal{J}$ . Moreover, observe that

$$\begin{aligned}
2f_j + 1 &= 2(f - |\mathcal{C}'_j| + |\bar{\mathcal{C}}_j|) + 1 \\
&= 2(f - (2f + 1) + |\bar{\mathcal{C}}_j|) + 1 \\
&= (|\bar{\mathcal{C}}_j| - (2f + 1)) + |\bar{\mathcal{C}}_j| \\
&\leq (|\mathcal{C}'_j| - (2f + 1)) + |\bar{\mathcal{C}}_j| \\
&= |\bar{\mathcal{C}}_j|.
\end{aligned} \tag{46}$$

With these basic observations in place, we are now ready to analyze the **Trim** operation outlined in line 6 of **Robust Fed-SEL**. To establish the claim that

$$\bar{r}_j^{(U)}(k) \in \text{Conv} \left( \{\tilde{r}_i^{(U)}(k)\}_{i \in \bar{\mathcal{C}}_j \cap \mathcal{H}} \right),$$

suppose  $\text{Trim} \left( \{\tilde{r}_i^{(U)}(k)\}_{i \in \bar{\mathcal{C}}_j}, f_j \right) = \tilde{r}_w^{(U)}(k)$ , where  $w \in \bar{\mathcal{C}}_j$ . If  $w \in \mathcal{H}$ , then the claim holds trivially. Now suppose  $w$  is adversarial, i.e.,  $w \in \mathcal{J}$ . Since  $|\bar{\mathcal{C}}_j \cap \mathcal{J}| \leq f_j$ , and  $|\bar{\mathcal{C}}_j| \geq 2f_j + 1$ , there must exist honest clients  $u, v \in \bar{\mathcal{C}}_j \cap \mathcal{H}$ , such that  $\tilde{r}_v^{(U)}(k) \leq \tilde{r}_w^{(U)}(k) \leq \tilde{r}_u^{(U)}(k)$ . This establishes the claim about  $\bar{r}_j^{(U)}(k)$  in Eq. (45); the claim regarding  $\bar{r}_j^{(L)}(k)$  follows the exact same argument.  $\square$

The above lemma tells us that with high probability, (i) the server will correctly identify the locally best arm from each arm-set, and (ii) the thresholds used for eliminating client-groups are “not too corrupted” by the adversaries. We now show how these two properties feature in the proof of Theorem 2.

*Proof.* (**Theorem 2**) Throughout the proof, we will condition on the event  $\mathcal{E}_{\mathcal{H}}$  defined in Eq. (43). Based on Lemma 6, recall that  $\mathcal{E}_{\mathcal{H}}$  occurs with probability at least  $1 - \delta$ . As in the proof of Theorem 1, we start by arguing that client group 1 always remains active (recall that  $a^* = 1 \in \mathcal{A}_1$ ). To see why this is true, fix any  $j \in [N]$ , and observe that for each  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$  and every round  $k$ ,

$$\begin{aligned}
\tilde{r}_i^{(L)}(k) &= \tilde{r}_i(k) - 2\alpha_j(\bar{t}_i + kH) \\
&\stackrel{(a)}{\leq} \hat{r}_{i, \bar{a}_j}(\bar{t}_i + kH) + \frac{\alpha_j(\bar{t}_i + kH)}{2} - 2\alpha_j(\bar{t}_i + kH) \\
&\stackrel{(b)}{=} \hat{r}_{i, j_1}(\bar{t}_i + kH) - \frac{3}{2}\alpha_j(\bar{t}_i + kH) \\
&\stackrel{\mathcal{E}_{\mathcal{H}}}{\leq} r_{j_1} + \alpha_j(\bar{t}_i + kH) - \frac{3}{2}\alpha_j(\bar{t}_i + kH) \\
&< r_{j_1}.
\end{aligned} \tag{47}$$

In the above steps, (a) follows from (i) the fact that by definition of the set  $\bar{\mathcal{C}}_j$ , every client  $i$  in  $\bar{\mathcal{C}}_j$  reports  $a_i = \bar{a}_j$ , and (ii) by using the bound on the quantization error in Eq. (26). Based on Lemma 7, the majority voted arm  $\bar{a}_j$  is  $j_1$  on event  $\mathcal{E}_{\mathcal{H}}$ ; this leads to (b). Using the same reasoning, we can show that for each  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$  and every round  $k$ ,

$$\tilde{r}_i^{(U)}(k) > r_{j_1} \tag{48}$$

on event  $\mathcal{E}_{\mathcal{H}}$ . Appealing to (45) from Lemma 7, and using (47) and (48), we conclude that on  $\mathcal{E}_{\mathcal{H}}$ ,

$$\bar{r}_1^{(U)}(k) > r_1; \quad \bar{r}_j^{(L)}(k) < r_{j_1}, \forall j \in [N] \setminus \{1\}.$$

From lines 7 and 8 of **Robust Fed-SEL**, we immediately note that client-group 1 never gets eliminated. To show that **Robust Fed-SEL** is consistent, it remains to argue that on  $\mathcal{E}_{\mathcal{H}}$ , every client-group  $j \in [N] \setminus \{1\}$  eventually gets eliminated. We claim that for group  $j \in [N] \setminus \{1\}$  to get eliminated, the following is a set of sufficient conditions:

$$\begin{aligned} \alpha_j(\bar{t}_i + kH) &\leq \frac{\Delta_{1j_1}}{8}, \forall i \in \bar{\mathcal{C}}_j \cap \mathcal{H}, \text{ and} \\ \alpha_1(\bar{t}_i + kH) &\leq \frac{\Delta_{1j_1}}{8}, \forall i \in \bar{\mathcal{C}}_1 \cap \mathcal{H}. \end{aligned} \quad (49)$$

Suppose the above conditions hold. Then, based on the same reasoning we used to arrive at Eq. (30), we can show that for each  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$ ,

$$\bar{r}_i^{(U)}(k) \leq r_{j_1} + \frac{7}{16}\Delta_{1j_1}, \quad (50)$$

and for each  $i \in \bar{\mathcal{C}}_1 \cap \mathcal{H}$ ,

$$\bar{r}_i^{(L)}(k) \geq r_1 - \frac{7}{16}\Delta_{1j_1}. \quad (51)$$

Appealing once again to the second claim of Lemma 7, and using (50) and (51), we conclude that if the conditions in (49) hold, then

$$\bar{r}_1^{(L)}(k) \geq r_1 - \frac{7}{16}\Delta_{1j_1} \text{ and } \bar{r}_j^{(U)}(k) \leq r_{j_1} + \frac{7}{16}\Delta_{1j_1}, \quad (52)$$

implying  $\bar{r}_1^{(L)}(k) > \bar{r}_j^{(U)}(k)$ . From line 8 of **Robust Fed-SEL**, this is sufficient for client-group  $j$  to get eliminated. Thus, **Robust Fed-SEL** will eventually terminate on  $\mathcal{E}_{\mathcal{H}}$  with  $\bar{\Psi}(k+1) = \{1\}$ , for some round  $k$ . Consistency then follows from Lemma 7 by noting that on  $\mathcal{E}_{\mathcal{H}}$ ,  $\bar{a}_1 = a^* = 1$ .

The sample- and communication-complexity bounds of **Robust Fed-SEL** are of the same order as that of **Fed-SEL**, and can be derived in exactly the same way as in Theorem 1. We now briefly explain how this can be done. Suppose we want to compute an upper bound on the number of rounds client group  $j \in [N] \setminus \{1\}$  remains active. Start by noting that at the end of Phase I, the following bound applies *uniformly* to every  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$ :

$$\alpha_j(\bar{t}_i) < \frac{2}{9}\Delta_j^* < \frac{\sigma_{j_1}\Delta_{1j_1}}{8}.$$

This follows from Lemma 3. For each  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$ , let  $R_{i1}^{(j)}$  be the first round such that

$$\alpha_j(\bar{t}_i + kH) \leq \frac{\Delta_{1j_1}}{8}.$$

We can obtain an upper bound on  $R_{i1}^{(j)}$  for each  $i \in \bar{\mathcal{C}}_j \cap \mathcal{H}$  exactly as in Theorem 1, and show that

$$\max_{i \in \bar{\mathcal{C}}_j \cap \mathcal{H}} R_{i1}^{(j)} \leq \frac{1}{H} \left( \frac{128}{(\Delta_j^*)^2} (\sigma_{j_1}^2 - 1) \log \frac{128\bar{c}_j}{\Delta_{1j_1}^2} + \frac{256}{(\Delta_j^*)^2} \log \sigma_{j_1} + O \left( \frac{1}{\Delta_{1j_1}^2} \log \log \frac{\bar{c}_j}{\Delta_{1j_1}^2} \right) \right) + 1, \text{ if } \sigma_{j_1} > 1,$$

and  $\max_{i \in \bar{\mathcal{C}}_j \cap \mathcal{H}} R_{i1}^{(j)} = 1$ , if  $\sigma_{j_1} \leq 1$ . With  $R_{j1} \triangleq \max_{i \in \bar{\mathcal{C}}_j \cap \mathcal{H}} R_{i1}^{(j)}$ , observe that the expression for  $R_{j1}$  exactly resembles that for  $R_{i1}$  in Eq. (5) of Theorem 1.<sup>13</sup> Proceeding as we did in Theorem 1, we can now derive an upper bound on the number of rounds group  $j$  remains active. The rest of the proof mirrors that of Theorem 1, and hence we omit details.  $\square$

<sup>13</sup>Recall that in the context of Theorem 1, each client group contained only one agent, and hence the index  $i$  was used for both client groups and agents.

## C Proof of Theorem 3

In this section, we will analyze the peer-to-peer setting described in Section 6. As before, we need a “good event” on which we will base our reasoning. For the setting we consider here, the event  $\mathcal{E}_{\mathcal{H}}$  described in Eq. (43) will serve such a purpose. Recall that  $\mathbb{P}(\mathcal{E}_{\mathcal{H}}) \geq 1 - \delta$ .

Consider any honest client  $i \in \mathcal{C}_j$ . The argument we present next applies identically to honest clients in other client groups as well. Since  $i \in \mathcal{C}_j$ , running Algo. 1 on  $\mathcal{A}_j$  with  $\beta_j(t) = 6\alpha_j(t)$  yields

$$a_i = j_1; |\hat{r}_{i,a_i}(\bar{t}_i) - r_{j_1}| = |\hat{r}_{i,j_1}(\bar{t}_i) - r_{j_1}| \leq \alpha_j(\bar{t}_i) < \frac{c}{(c-2)^2} \Delta_j^* = \frac{3}{8} \Delta_j^* \quad (53)$$

on event  $\mathcal{E}_{\mathcal{H}}$ . The above claims follow directly from Lemma’s 2 and 3. Thus, based on Line 1 of Algo. 4, we have  $\mathbf{I}_i(j) = j_1$ , and  $\mathbf{R}_i(j) = \hat{r}_{i,j_1}(\bar{t}_i)$ , where  $\hat{r}_{i,j_1}(\bar{t}_i)$  satisfies the bound in Eq. (53).

Now for client  $i$  to identify the globally best arm, it must eventually receive accurate information about every other arm-set  $\ell \in [N] \setminus \{j\}$  from its neighbors, despite the presence of adversaries in the network. The next lemma shows how the strong-robustness property in Definition 2 contributes to this cause.

**Lemma 8.** *Suppose Algo. 4 is run by every  $i \in \mathcal{H}$ . Moreover, suppose the adversary model is  $f$ -local, and  $\mathcal{G}$  is strongly-robust w.r.t.  $\mathcal{C}_j, \forall j \in [N]$ . Consider any  $j \in [N]$ . Then, on the event  $\mathcal{E}_{\mathcal{H}}$ , every client  $i \in \mathcal{V} \setminus \mathcal{C}_j$  receives information about arm-set  $j$  from at least  $2f + 1$  neighbors in  $\mathcal{N}_i$ .*

*Proof.* We prove this lemma by contradiction. Fix any  $j \in [N]$ . Suppose that on the event  $\mathcal{E}_{\mathcal{H}}$ , there exists a non-empty set  $\mathcal{B} \subseteq \mathcal{V} \setminus \mathcal{C}_j$  of clients who never receive information about arm-set  $j$  from  $2f + 1$  neighbors. Based on the strong-robustness property in Def. 2, there exists a client  $i \in \mathcal{B}$  with at least  $3f + 1$  neighbors in  $\mathcal{V} \setminus \mathcal{B}$ . Consider the set  $\mathcal{Q}_{i,j} = \mathcal{N}_i \cap \{\mathcal{V} \setminus \mathcal{B}\}$ . Based on the strong-robustness property and the  $f$ -local assumption,  $|\mathcal{Q}_{i,j} \cap \mathcal{H}| \geq 2f + 1$ . Now a client  $k$  in  $\mathcal{Q}_{i,j} \cap \mathcal{H}$  can either belong to  $\mathcal{C}_j$  or not. If  $k \in \mathcal{Q}_{i,j} \cap \mathcal{H}$  belongs to  $\mathcal{C}_j$ , then on the event  $\mathcal{E}_{\mathcal{H}}$ , Algo. 1 will terminate at client  $k$ . Thus, client  $k$  will eventually transmit  $\{\mathbf{I}_k(j), \mathbf{R}_k(j)\}$  to every out-neighbor in  $\mathcal{N}_k^+$ , including client  $i$ . Even if  $k \in \mathcal{Q}_{i,j} \cap \mathcal{H}$  does not belong to  $\mathcal{C}_j$ , since  $k \notin \mathcal{B}$ , client  $k$  must have received information about arm-set  $j$  from  $2f + 1$  in-neighbors. Thus, based on lines 2-4 of Algo. 4, it must transmit  $\{\mathbf{I}_k(j), \mathbf{R}_k(j)\}$  to client  $i$  at some point of time. We conclude that on the event  $\mathcal{E}_{\mathcal{H}}$ , client  $i$  will eventually receive information about arm-set  $j$  from every neighbor in  $\mathcal{Q}_{i,j} \cap \mathcal{H}$ . This leads to the desired contradiction and concludes the proof.  $\square$

Our next goal is to prove that on  $\mathcal{E}_{\mathcal{H}}$ , for each  $j \in [N]$ , every honest client  $i \in \mathcal{V} \setminus \mathcal{C}_j$  is able to correctly identify the best arm  $j_1$  in  $\mathcal{A}_j$ , and obtain a reasonable estimate of its mean  $r_{j_1}$ . We establish these properties in the next lemma.

**Lemma 9.** *Suppose Algo. 4 is run by every  $i \in \mathcal{H}$ . Moreover, suppose the adversary model is  $f$ -local, and  $\mathcal{G}$  is strongly-robust w.r.t.  $\mathcal{C}_j, \forall j \in [N]$ . On event  $\mathcal{E}_{\mathcal{H}}$ , the following then hold  $\forall j \in [N]$ :*

$$(i) \mathbf{I}_i(j) = j_1, \forall i \in \mathcal{H}.$$

$$(ii) |\mathbf{R}_i(j) - r_{j_1}| \leq (3/8)\Delta_j^*, \forall i \in \mathcal{H}.$$

*Proof.* Fix any  $j \in [N]$ . If  $i \in \mathcal{C}_j \cap \mathcal{H}$ , then each of the above claims hold based on the arguments we presented at the beginning of this section; see Eq. (53). Thus, we only need to prove these claims for clients in  $\{\mathcal{V} \setminus \mathcal{C}_j\} \cap \mathcal{H}$ .

To do so, let us first associate a notion of “activation” with Algo. 4. For each client  $i \in \mathcal{C}_j \cap \mathcal{H}$ , we say that it gets activated when the local subroutine (Algo. 1) terminates at client  $i$ . For each  $i \in \{\mathcal{V} \setminus \mathcal{C}_j\} \cap \mathcal{H}$ , we say it gets activated upon receiving information about arm-set  $j$  from  $2f + 1$



neighbors in  $\mathcal{N}_i$ . From Lemma's 2 and 8, observe that all honest clients eventually get activated on the event  $\mathcal{E}_{\mathcal{H}}$ . Accordingly, on the event  $\mathcal{E}_{\mathcal{H}}$ , let  $\tau_1 < \tau_2 < \dots < \tau_K$  denote the activation times for clients in  $\{\mathcal{V} \setminus \mathcal{C}_j\} \cap \mathcal{H}$ .<sup>14</sup> Moreover, let  $\mathcal{H}_{\tau_s}^{(j)}$  represent those clients in  $\{\mathcal{V} \setminus \mathcal{C}_j\} \cap \mathcal{H}$  that get activated at time-step  $\tau_s$ ,  $s \in [K]$ .

We prove the result by inducting on the activation times. For the base case, consider any client  $i$  in the set  $\mathcal{H}_{\tau_1}^{(j)}$ . Recall that  $\mathcal{N}'_{i,j} \subseteq \mathcal{N}_i$  are the first  $2f + 1$  neighbors who report about arm-set  $j$  to client  $i$ . From the rules of Algo. 4, notice that an honest client  $k$  transmits  $\{\mathbf{I}_k(j), \mathbf{R}_k(j)\}$  only when it gets activated. Thus, it must be that  $\mathcal{N}'_{i,j} \cap \mathcal{H} \subseteq \mathcal{C}_j \cap \mathcal{H}$ . Moreover, since the adversary model is  $f$ -local,  $|\mathcal{N}'_{i,j} \cap \mathcal{H}| \geq f + 1$ . We conclude that at least  $f + 1$  honest clients in  $\mathcal{C}_j$  report arm  $j_1$  to client  $i$ . Based on the majority voting operation in line 3 of Algo. 4, it follows that  $\mathbf{I}_i(j) = j_1$ .

Observe also that on the event  $\mathcal{E}_{\mathcal{H}}$ , any client  $k \in \mathcal{N}'_{i,j}$  that reports an arm other than arm  $j_1$  to client  $i$  must be adversarial. Following the same reasoning as in Lemma 7, we can then establish that (i)  $|\bar{\mathcal{N}}_{i,j}| \geq 2f_{i,j} + 1$ , where  $f_{i,j} = f - |\mathcal{N}'_{i,j}| + |\bar{\mathcal{N}}_{i,j}|$ , and  $\bar{\mathcal{N}}_{i,j}$  represents those clients in  $\mathcal{N}'_{i,j}$  that report arm  $j_1$ ; and (ii)  $|\bar{\mathcal{N}}_{i,j} \cap \mathcal{H}| \leq f_{i,j}$ . Using the same arguments used to arrive at Eq. (45), we can then show that

$$\mathbf{R}_i(j) \in \text{Conv} \left( \{\mathbf{R}_k(j)\}_{k \in \bar{\mathcal{N}}_{i,j} \cap \mathcal{H}} \right) \implies \mathbf{R}_i(j) \in \text{Conv} \left( \{\mathbf{R}_k(j)\}_{k \in \mathcal{C}_j \cap \mathcal{H}} \right), \quad (54)$$

since  $\bar{\mathcal{N}}_{i,j} \cap \mathcal{H} \subseteq \mathcal{C}_j \cap \mathcal{H}$ . For client  $i$ , claim (ii) of the lemma then follows immediately from the fact that on event  $\mathcal{E}_{\mathcal{H}}$ ,  $|\mathbf{R}_k(j) - r_{j_1}| \leq (3/8)\Delta_j^*$ ,  $\forall k \in \mathcal{C}_j \cap \mathcal{H}$ . This completes the base case.

For the induction hypothesis, suppose for all  $i \in \bigcup_{s \in [m]} \mathcal{H}_{\tau_s}^{(j)}$ ,  $m \in [K - 1]$ , it holds that  $\mathbf{I}_i(j) = j_1$ , and

$$\mathbf{R}_i(j) \in \text{Conv} \left( \{\mathbf{R}_k(j)\}_{k \in \mathcal{C}_j \cap \mathcal{H}} \right). \quad (55)$$

Now consider a client  $i \in \mathcal{H}_{\tau_{m+1}}^{(j)}$ . Based on the rules of Algo. 4, and the definition of activation times, it must be that

$$\mathcal{N}'_{i,j} \cap \mathcal{H} \subseteq \{\mathcal{C}_j \cap \mathcal{H}\} \cup \left\{ \bigcup_{s \in [m]} \mathcal{H}_{\tau_s}^{(j)} \right\}.$$

Using the above fact,  $|\mathcal{N}'_{i,j} \cap \mathcal{H}| \geq f + 1$ , the induction hypothesis, and the same reasoning as the base case, it is easy to see that  $\mathbf{I}_i(j) = j_1$ , and that  $\mathbf{R}_i(j)$  satisfies the inclusion in Eq. (55). This completes the induction hypothesis and the proof.  $\square$

We are now ready to complete the proof of Theorem 3.

*Proof. (Theorem 3)* To complete the proof of Theorem 3, it remains to argue that for each honest client  $i$ ,  $\mathbf{R}_i(1) > \mathbf{R}_i(j)$ ,  $\forall j \in [N] \setminus \{1\}$ . By way of contradiction, suppose there exists some client  $i \in \mathcal{H}$  and some  $j \in [N] \setminus \{1\}$ , such that  $\mathbf{R}_i(j) \geq \mathbf{R}_i(1)$ . We then have

$$\begin{aligned} \mathbf{R}_i(j) &\geq \mathbf{R}_i(1) \\ &\stackrel{(a)}{\implies} r_{j_1} + \frac{3}{8}\Delta_j^* \geq r_1 - \frac{3}{8}\Delta_1^* \\ &\implies \Delta_{1j_1} \leq \frac{3}{8}(\Delta_1^* + \Delta_j^*) \\ &\stackrel{(b)}{\implies} \Delta_{1j_1} \leq \frac{3}{8}(\sigma_{1j} + \sigma_{j1}) \Delta_{1j_1} \\ &\stackrel{(c)}{\implies} \Delta_{1j_1} \leq \frac{3}{4}\Delta_{1j_1}, \end{aligned} \quad (56)$$

<sup>14</sup>Note that multiple clients may get activated at the same time-step.

which leads to the desired contradiction. Here, (a) follows from the second claim of Lemma 9, (b) follows from the definition of arm-heterogeneity indices in Eq. (4), and (c) follows from the assumption on arm-heterogeneity in Eq. (8). This completes the proof.  $\square$