# Differentially Private Federated Combinatorial Bandits with Constraints

Sambhav Solanki (✉), Samhita Kanaparthy, Sankarshan Damle, and Sujit Gujar

Machine Learning Lab, International Institute of Information Technology (IIIT), Hyderabad
{sambhav.solanki, s.v.samhita, sankarshan.damle}@research.iiit.ac.in, sujit.gujar@iiit.ac.in

**Abstract.** There is a rapid increase in the cooperative learning paradigm in online learning settings, i.e., *federated learning* (FL). Unlike most FL settings, there are many situations where the agents are competitive. Each agent would like to learn from others, but the part of the information it shares for others to learn from could be sensitive; thus, it desires its *privacy*. This work investigates a group of agents working concurrently to solve similar combinatorial bandit problems while maintaining quality constraints. Can these agents collectively learn while keeping their sensitive information confidential by employing differential privacy? We observe that communicating can reduce the *regret*. However, differential privacy techniques for protecting sensitive information makes the data noisy and may deteriorate than help to improve regret. Hence, we note that it is essential to decide *when to communicate* and *what shared data to learn* to strike a functional balance between regret and privacy. For such a federated combinatorial MAB setting, we propose a Privacy-preserving Federated Combinatorial Bandit algorithm, `P-FCB`. We illustrate the efficacy of `P-FCB` through simulations. We further show that our algorithm provides an improvement in terms of regret while upholding quality threshold and meaningful privacy guarantees.

**Keywords:** Combinatorial Multi-armed Bandits · Differential Privacy · Federated Learning.

## 1 Introduction

A large portion of the manufacturing industry follows the Original Equipment Manufacturer (OEM) model. In this model, companies (or aggregators) that design the product usually procure components required from an available set of OEMs. Foundries like TSMC, UMC, and GlobalFoundries handle the production of components used in a wide range of smart electronic offerings [1]. We also observe a similar trend in the automotive industry [2].

However, aggregators are required to maintain minimum *quality* assurance for their products while maximizing their revenue. Hence, they must judicially procure the components with desirable quality and cost from the OEMs. For this, aggregators should learn the quality of components provided by an OEM. OEM businesses often have numerous agents engaged in procuring the same or similar components. In such a setting, one can employ *online learning* where multiple aggregators, referred henceforth as *agents*, cooperate to learn the qualities [8, 24]. Further, decentralized (or federated) learning is gaining traction for large-scale applications [20, 33].

In general, an agent needs to procure and utilize the components from different OEMs (referred to as *producers*) to learn their quality. This learning is similar to the exploration and exploitation problem, popularly known as *Multi-armed Bandit* (MAB) [13, 15]. It needs sequential interactions between sets of producers and the learning agent. Further, we associate qualities, costs, and capacities with the producers for each agent. We model this as a combinatorial multi-armed bandit (CMAB) [5] problem with assured qualities [15]. Our model allows the agents to maximize their revenues by communicating their history of procurements to have better estimations of the qualities. Since the agents can benefit from sharing their past quality realizations, we consider them engaged in a *federated* learning process. Federated MAB often improves performance in terms of *regret* incurred per agent [16, 25][1].

Such a federated exploration/exploitation paradigm is not just limited to selecting OEMs. It is useful in many other domains such as stocking warehouse/distribution centres, flow optimization, and product recommendations on e-commerce websites [21, 27]. However, agents are competitive; thus, engaging in federated learning is not straightforward. Agents may not be willing to share their private experiences since that could negatively benefit them. For example, sharing the exact procurement quantities of components specific to certain products can reveal the market/sales projections. Thus, we desire (or many times even it is necessary) to maintain privacy when engaged in federated learning. This paper aims to design a privacy-preserving algorithm for federated CMAB with quality assurances.

**Our Approach and Contributions.** Privacy concerns for sensitive information pose a significant barrier to adopting federated learning. To preserve the privacy of such information, we employ the strong notion of *differential privacy* (DP) [9]. Note that naive approaches (e.g., Laplace or Gaussian Noise Mechanisms [10]) to achieve DP for CMAB may come at a high privacy cost or outright perform worse than non-federated solutions. Consequently, the primary challenge is carefully designing methods to achieve DP that provide meaningful privacy guarantees while performing significantly better than its non-federated counterpart.

To this end, we introduce `P-FCB`, a Privacy-preserving Federated Combinatorial Bandit algorithm. `P-FCB` comprises a novel communication algorithm among agents, while each agent is learning the qualities of the producers to cooperate in the learning process. Crucially in `P-FCB`, the agent only communicates within a specific time frame – since it is not beneficial to communicate in (i) earlier

[1] Regret is the deviation of utility gained while engaging in learning from the utility gained if the mean qualities were known.

rounds (estimates have high error probability) or (ii) later rounds (value added by communicating is minimal). While communicating in each round reduces per agent regret, it results in a high privacy loss. `P-FCB` strikes an effective balance between learning and privacy loss by limiting the number of rounds in which agents communicate. Moreover, to ensure the privacy of the shared information, the agents add calibrated noise to sanitize the information a priori. `P-FCB` also uses error bounds generated for UCB exploration [3] to determine if shared information is worth learning. We show that `P-FCB` allows the agents to minimize their regrets while ensuring strong privacy guarantees through extensive simulations.

In recent times, research has focused on the intersection of MAB and DP [19, 32]. Unlike `P-FCB`, these works have limitations to single-arm selections. To the best of our knowledge, this paper is the first to simultaneously study federated CMAB with assured quality and privacy constraints. In addition, as opposed to other DP and MAB approaches [8, 12], we consider the sensitivity of attributes specific to a producer-agent set rather than the sensitivity of general observations. In summary, our contributions in this work are as follows:

1. We provide a theoretical analysis of improvement in terms of regret in a non-private homogeneous federated CMAB setting (Theorem 1, Section 4).
2. We show that employing privacy techniques naively is not helpful and has information leak concerns (Claim 1, Section 5.2).
3. We introduce `P-FCB` to employ privacy techniques practically (Algorithm 1). `P-FCB` includes selecting the information that needs to be perturbed and defining communication rounds to provide strong privacy guarantees. The communicated information is learned selectively by using error bounds around current estimates. Selective communication helps minimize regret.
4. `P-FCB`'s improvement in per agent regret even in a private setting compared to individual learning is empirically validated through extensive simulations (Section 6).

## 2 Related Work

*Multi-armed bandits* (MAB) and their variants are a well studied class of problems [3, 6, 15, 17, 22, 23] that tackle the exploration vs. exploitation trade-off in online learning settings. While the classical MAB problem [3, 28] assumes single arm pull with stochastic reward generation, our work deals with combinatorial bandits (CMAB) [5, 11, 26, 31], whereby the learning agent pulls a subset of arms. We remark that our single-agent (non-federated) MAB formulation is closely related to the MAB setting considered in [7], but the authors there do not consider federated learning.

*Federated MAB.* Many existing studies address the MAB problem in a federated setting but restrict themselves to single-arm pulls. The authors in [24, 25] consider a federated extension of the stochastic single player MAB problem, while Huang et al. [14] considers the linear contextual bandit in a federated setting. Kim et al. [16] specifically considers the federated CMAB setting. However, none of these works address privacy.

*Privacy-preserving MAB.* The authors in [19, 32] consider a differentially private MAB setting for a single learning agent, while the works in [4, 18] consider differentially private federated MAB setting. However, these works focus only on the classical MAB setting, emphasising the communication bottlenecks. There also exists works that deal with private and federated setting for the contextual bandit problem [8, 12]. However, they do not consider pulling subsets of arms. Further, Hannun et al. [12] consider privacy over the context, while Dubey and Pentland [8] consider privacy over context and rewards. Contrarily, this paper considers privacy over the procurement strategy used.

To the best of our knowledge, we are the first to propose a solution for combinatorial bandits (CMAB) in a federated setting with the associated privacy concerns.

## 3 Preliminaries

In this section, we formally describe the combinatorial multi-armed bandit setting and its federated extension. We also define differential privacy in our context.

### 3.1 Federated Combinatorial Multi Armed Bandits

We consider a combinatorial MAB (CMAB) setting where there are $[m]$ producers and $[n]$ agents. Each producer $i \in [m]$ has a cost $k_{ij}$ and capacity $c_{ij}$ for every agent $j \in [n]$. At any round $t \in \{1, 2, \ldots, T\}$, agents procure some quantity of goods from a subset of producers under given constraint(s). We denote the procurement of an agent $j$ by $\mathbf{s}_j = (l_{1j}, l_{2j}, \ldots, l_{mj})$ where $l_{ij} \in [0, k_{ij}]$ is the quantity procured from producer $i$.

*Qualities.* Each agent observes a quality realisation for each unit it procured from producers. Since the quality of a single unit of good may not be easily identifiable, we characterize it as a Bernoulli random variable. The expected realisation of a unit procured from a producer $i$ is referred to as its quality, $q_i$. In other words, $q_i$ denotes the probability with which a procured unit of good from producer $i$ will have a quality realisation of one. While the producer's cost and capacity vary across agents, the quality values are indifferent based on agents.

*Regret.* We use $r_{ij}$ to denote expected utility gain for the agent $j$ by procuring a single unit from producer $i$, where $r_{ij} = \rho q_i - c_{ij}$ (where $\rho > 0$, is a proportionality constant). Further, the expected revenue for a procurement vector $\mathbf{s}_j$, is given by $r_{\mathbf{s}_j} = \sum_{i \in [m]} l_{ij} r_{ij}$.

The goal for the agent is to maximise its revenue, under given constraints. We consider a constraint of maintaining a minimum expected quality threshold $\alpha$ (quality constraint), for our setting. To measure the performance of an a given algorithm $A$, we use the notion of regret which signifies the deviation of the algorithm from the procurement set chosen by an Oracle when mean qualities are known. For any round $t \in \{1, 2, \ldots, T\}$, we use the following to denote the regret for agent $j$ given an algorithm $A$,

$$\mathcal{R}^t_{Aj} = \begin{cases} r_{\mathbf{s}^*_{\mathbf{j}}} - r_{\mathbf{s}^t_{Aj}}, & \textit{if } s^t_{Aj} \textit{ satisfies the quality constraint} \\ L, & \textit{otherwise} \end{cases}$$

where $\mathbf{s}^*_{\mathbf{j}}$ denotes the procurement set chosen by an Oracle, with the mean qualities known. $\mathbf{s}^t_A$ is the set chosen by the algorithm $A$ in round $t$. $L = \max_{r_{\mathbf{s}}}(r_{\mathbf{s}^*_{\mathbf{j}}} - r_{\mathbf{s}})$ is a constant that represents the maximum regret one can acquire. The overall regret for algorithm $A$ is given by $\mathcal{R}_A = \sum_{j \in [n]} \sum_{t \in [T]} \mathcal{R}^t_{Aj}$.

*Federated Regret Ratio (FRR).* We introduce FRR to help quantify the reduction in regret brought on by engaging in federated learning. FRR is the ratio of the regret incurred by an agent via a federated learning algorithm $A$ over agent's learning individually via a non-federated algorithm $NF$, i.e., $FRR = \frac{\mathcal{R}_A}{\mathcal{R}_{NF}}$. We believe, $FRR$ is a comprehensive indicator of the utility gained by engaging in federated learning, compared to direct regret, since it presents a normalised value and performance comparison over different data sets/algorithms is possible.

Observe that, $FRR \approx 1$ indicates that there is not much change in terms of regret by engaging in federated learning. If $FRR > 1$, it is detrimental to engage in federated learning, whereas if $FRR < 1$, it indicates a reduction in regret. When $FRR \approx 0$, there is almost complete reduction of regret in federated learning.

In our setting, we consider that agents communicate with each other to improve their regret. But in general, agents often engage in a competitive setting, and revealing true procurement values can negatively impact them. For instance, knowing that a company has been procuring less than their history can reveal their strategic plans, devalue their market capital, hinder negotiations etc. We give a formalisation of the notion of privacy used in our setting in the next subsection.

### 3.2 Differential Privacy (DP)

As opposed to typical federated models, we assume that the agents in our setting may be competing. Thus, agents will prefer the preservation of their sensitive information. Specifically, consider the history of procurement quantities $\mathbf{H}_{ij} = (l^t_{ij})_{t \in [T]}$ for any producer $i \in [m]$ is private to agent $j$. To preserve the privacy of $\mathbf{H}_{ij}$ while having meaningful utilitarian gains, we use the concept of Differential Privacy (DP). We tweak the standard DP definition in [9, 10] for our setting. For this, let $\mathbf{S}_j = (\mathbf{s}^t_j)_{t \in [T]}$ be complete history of procurement vectors for agent $j$.

**Definition 1 (Differential Privacy).** *In a federated setting with $n \geq 2$ agents, a combinatorial MAB algorithm $A = (A_j)^n_{j=1}$ is said to be $(\epsilon, \delta, n)-$differentially private if for any $u, v \in [n], s.t., u \neq v$, any $t_o$, any set of adjacent histories $\mathbf{H}_{iu} = (l^t_{iu})_{t \in [T]}, \mathbf{H}^{'}_{iu} = (l^t_{iu})_{t \in [T] \setminus \{t_o\}} \cup \bar{l}^{t_o}_{iu}$ for producer $i$ and any complete*

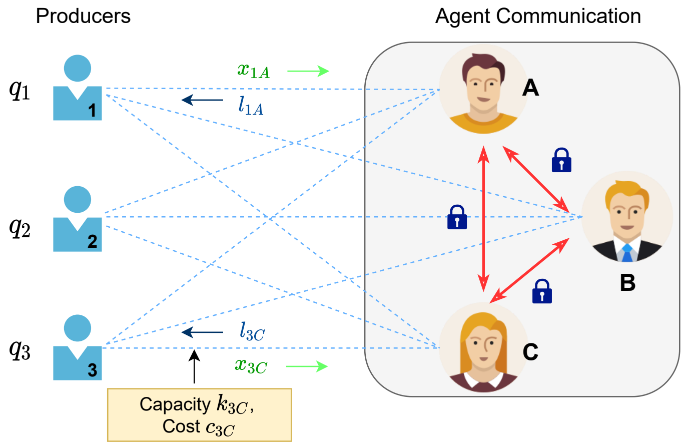


Fig. 1: Overview of the communication model for P-FCB: Agents interact with producers as part of the exploration and exploitation process. Agents also communicate among themselves to learn the qualities of producers. However, they share noisy data to maintain the privacy of their sensitive information.

*history of procurement vector* $\mathbf{S}_v$,

$$\Pr(A_v(\mathbf{H}_{iu}) \in \mathbf{S}_v) \leq e^{\epsilon} \Pr(A_v(\mathbf{H}_{iu}^{'}) \in \mathbf{S}_v) + \delta$$

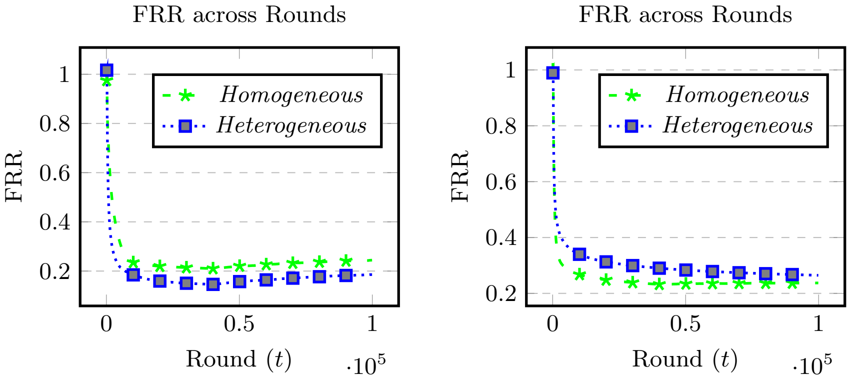
Our concept of DP in a federated CMAB formalizes the idea that the selection of procurement vectors by an agent is insusceptible to any single element $l_{ij}^t$ from another agent's procurement history. Note that the agents are not insusceptible to their own histories here.

Typically, the "$\epsilon$" parameter is referred to as the *privacy budget*. The *privacy loss* variable $\mathcal{L}$ is often useful for the analysis of DP. More formally, given a randomised mechanism $\mathcal{M}(\cdot)$ and for any output $o$, the privacy loss variable is defined as,

$$\mathcal{L}^{o}_{\mathcal{M}(\mathbf{H})||\mathcal{M}(\mathbf{H}')} = \ln\left(\frac{\Pr[\mathcal{M}(\mathbf{H}) = o]}{\Pr[\mathcal{M}(\mathbf{H}') = o]}\right). \tag{1}$$

*Gaussian Noise Mechanism* [10]. To ensure DP, often standard techniques of adding noise to values to be communicated are used. The Gaussian Noise mechanism is a popular mechanism for the same. Formally, a randomised mechanism $\mathcal{M}(x)$ satisfies $(\epsilon, \delta)$-DP if the agent communicates $\mathcal{M}(x) \triangleq x + \mathcal{N}\left(0, \frac{2\Delta(x)^2 \ln(1.25/\delta)}{\epsilon^2}\right)$. Here, $x$ is the private value to be communicated with *sensitivity* $\Delta(x)$, and $\mathcal{N}(0, \sigma^2)$ the Gaussian distribution with mean zero and variance $\sigma^2$.

In summary, Figure 1 provides an overview of the model considered. Recall that we aim to design a differentially private algorithm for federated CMAB with assured qualities. Before this, we first highlight the improvement in regret



(a) $c_{ij}, q_i \sim U[0,1]$

(b) $c_{ij}, q_i \sim \mathcal{N}(\alpha, 0.1)$

Fig. 2: Comparing $FRR$ values for Homogeneous and Heterogeneous Federated CMAB ($n = 10$, $m = 30$)

using the federated learning paradigm. Next, we discuss our private algorithm, `P-FCB`, in Section 5.

## 4 Non-private Federated Combinatorial Multi-armed Bandits

We now demonstrate the advantage of federated learning in CMAB by highlighting the reduction in regret incurred compared to agents learning individually. We first categorize Federated CMAB into the following two settings: (i) *homogeneous*: where the capacities and costs for producers are the same across agents, and (ii) *heterogeneous*: where the producer's capacity and cost varies across agents.

**Homogeneous Setting.** The core idea for single-agent learning in CMAB involves using standard $UCB$ exploration [3]. We consider an Oracle that uses the $UCB$ estimates to return an optimal selection subset. In this paper, we propose that to accelerate the learning process and for getting *tighter* error bound for quality estimations, the agents communicate their observations with each other in every round. In a homogeneous setting, this allows all agents to train a shared model locally without a central planner since the Oracle algorithm is considered deterministic. We present the formal algorithm in the extended version [29]. It's important to note that in such a setting, each agent has the same procurement history and the same expected regret.

Further, the quality constraint guarantees for the federated case follow trivially from the single agent case ([7, Theorem 2]). Additionally, in Theorem 1, we prove that the upper bound for regret incurred by each agent is $\mathcal{O}(\frac{\ln(nT)}{n})$; a significant improvement over $\mathcal{O}(\ln T)$ regret the agent will incur when playing individually. The formal proof is provided in the extended version [29].

**Theorem 1.** *For Federated CMAB in a homogeneous setting with $n$ agents, if the qualities of producers satisfy $\gamma$-seperatedness, then the individual regret incurred by each of the agents is bounded by $\mathcal{O}(\frac{\ln(nT)}{n})$.*

**Heterogeneous Setting.** In real-world, the agents may not always have the same capacities. For such a heterogeneous setting, the regret analysis is analytically challenging. For instance, we can no longer directly use Hoeffding's inequality, needed for proving Theorem 1, since the procurement histories will differ across agents. Still, the intuition for regret reduction from cooperative learning carries over.

Even in a heterogeneous setting, communicating the observations allows the agent to converge their quality estimations to the mean faster and provide tighter error bounds. Even with shared quality estimates, Oracle may return different procurement vectors for different agents based on different capacities. Thus, a weighted update in estimation is essential, and the procurement vector would also need to be communicated.

We empirically demonstrate that using federated learning in heterogeneous setting shows similar $FRR$ (ratio of regret incurred in federated setting compared to non federated setting) trend compared to homogeneous setting, over 100000 rounds for two scenarios: (i) Costs and qualities are sampled from uniform distributions, i.e. $c_{ij} \sim U[0,1]$, $q_i \sim U[0,1]$, (ii) Costs and qualities are sampled from normal distributions around the quality threshold, i.e., $c_{ij} \sim \mathcal{N}(\alpha, 0.1)$, $q_i \sim \mathcal{N}(\alpha, 0.1)$.

Fig. 2 depicts the results. From Fig. 2 we observe that the trend for both homogeneous and heterogeneous settings are quite similar. This shows that, similar to the homogeneous setting, employing federated learning reduces regret even in the heterogeneous setting.

## 5 P-FCB: Privacy-preserving Federated Combinatorial Bandit

From Section 3.2, recall that we identify the procurement history of an agent-producer pair as the agent's sensitive information. We believe that the notion of DP w.r.t. the agent-producer procurement history is reasonable. A differentially private solution ensures that the probability with which other agents can distinguish between an agent's adjacent procurement histories is upper bounded by the privacy budget $\epsilon$.

Section Outline: In this section, we first argue that naive approaches for DP are not suitable due to their lack of meaningful privacy guarantees. Second, we show that all attributes dependent on the sensitive attribute must be sanitised before sharing to preserve privacy. Third, we define a privacy budget algorithm scheme. Fourth, we formally introduce P-FCB including a selective learning procedure. Last, we provide the $(\epsilon, \delta)$-DP guarantees for P-FCB.

### 5.1 Privacy budget and Regret Trade-off

Additive noise mechanism (e.g., Gaussian Noise mechanism [10]) is a popular technique for ensuring $(\epsilon, \delta)$-DP. To protect the privacy of an agent's procurement history within the DP framework, we can build a naive algorithm for heterogeneous federated CMAB setting by adding noise to the elements of the procurement vectors being communicated in each round.

However, such a naive approach does not suitably satisfy our privacy needs. Using the Basic Composition theorem [10], which adds the $\epsilon$s and $\delta$s across queries, it is intuitive to see that communicating in every round results in a high overall $\epsilon$ value which may not render much privacy protection in practice [30]. Consider the agents interacting with the producers for $10^6$ rounds. Let $\epsilon = 10^{-2}$ for each round they communicate the perturbed values. Using Basic Composition, we can see that the overall privacy budget will be bounded by $\epsilon = 10^4$, which is practically not acceptable. The privacy loss in terms of overall $\epsilon$ grows at worst linearly with the number of rounds.

It is also infeasible to solve this problem merely by adding more noise (reducing $\epsilon$ per round) since if the communicated values are too noisy, they can negatively affect the estimates. This will result in the overall regret increasing to a degree that it may be better to not cooperatively learn. To overcome this challenge, we propose to decrease the number of rounds in which agents communicate information.

Secondly, if the sample size for the local estimates is too small, noise addition can negatively effect the regret incurred. On the other hand, if the sample size of local estimate is too large, the local estimate will have tight error bounds and deviating from the local estimate too much may result in the same.

**When to Learn.** Based on the above observations, we propose the following techniques to strike an effective trade-off between the privacy budget and regret.

1. To limit the growth of $\epsilon$ over rounds, we propose that communication happens only when the current round number is equal to a certain threshold (denoted by $\tau$) which doubles in each communication round. Thus, there are only $\log(T)$ communications rounds, where density of communication rounds decrease over rounds.
2. We propose to communicate only for a specific interval of rounds, i.e., for each round $t \in [\underline{t}, \bar{t}]$. *No* communication occurs outside these rounds. This ensures that agent communication only happens in rounds when it is useful and not detrimental.

### 5.2 Additional Information Leak with Actual Quality Estimates and Noisy Weights

It is also important to carefully evaluate the way data is communicated every round since it may lead to privacy leaks. For example, consider that all agents communicate their local estimates of the producer qualities and perturbation of the total number of units procured from each producer to arrive at the estimation. We now formally analyse the additional information leak in this case.

```
Procedure 1 CheckandUpdate($W, \tilde{w}, Y, \tilde{y}, \omega_1, \omega_2, n, t$)
$\hat{q} \longleftarrow \frac{Y}{W}$
if $\frac{\tilde{y}}{\tilde{w}} \in \left[\hat{q} - \omega_1\sqrt{\frac{3ln(nt)}{2W}}, \hat{q} + \omega_1\sqrt{\frac{3ln(nt)}{2W}}\right]$ then
    $W \longleftarrow W + \omega_2\tilde{w}$
    $Y \longleftarrow Y + \omega_2\tilde{y}$
end if
return $W, Y$
```

W.l.o.g. our analysis is for any arbitrarily picked producer $i \in [m]$ and agent $j \in [n]$. As such, we omit the subscripts “$i$" for producer and “$j$" for the agent. We first set up the required notations as follows.

Notations: Consider $\hat{q}^t, W^t$ as *true* values for the empirical estimate of quality and total quantity procured till the round $t$ (not including $t$). Next, let $\tilde{W}^t$ denote *noisy* value of $W^t$ (with the noise added using any additive noise mechanism for DP [10]). We have $w^t$ as the quantity procured in round $t$. Last, let $\hat{q}^{obsv_t}$ denote the quality estimate based on just round $t$. Through these notations, we can compute $\hat{q}^{t+1}$ for the successive round $t+1$ as follows: $\hat{q}^{t+1} = \frac{W^t \times \hat{q}^t + w^t \times \hat{q}^{obsv_t}}{W^t + w^t}$.

**Claim 1** *Given $\hat{q}^t, W^t, \tilde{W}^t, w^t$ and $\hat{q}^{obsv_t}$, the privacy loss variable $\mathcal{L}$ is not defined if $\hat{q}^t$ is also not perturbed.*

We present the formal proof in the extended version [29]. With Claim 1, we show that $\epsilon$ may not be bounded even after sanitising the sensitive data due to its dependence on other non-private communicated data. This is due to the fact that the local mean estimates are a function of the procurement vectors and the observation vectors. Thus, it becomes insufficient to just perturb the quality estimates.

We propose that whenever communication happens, only procurement and observation values based on rounds since last communication are shared. Additionally, to communicate weighted quality estimates, we use the Gaussian Noise mechanism to add noise to *both* the procurement values and realisation values. The sensitivity ($\Delta$) for noise sampling is equal to the capacity of the producer-agent pair.

### 5.3 Privacy Budget Allocation

Since the estimates are more sensitive to noise addition when the sample size is smaller, we propose using monotonically decreasing privacy budget for noise generation. Formally, let total privacy budget be denoted by $\epsilon$ with $(\epsilon^1, \epsilon^2, \ldots)$ corresponding to privacy budgets for communication rounds $(1, 2, \ldots)$. Then, we have $\epsilon^1 > \epsilon^2 > \ldots$. Specifically, we denote $\epsilon^z$ as the privacy budget in the $z^{th}$ communication round, where $\epsilon^z \longleftarrow \frac{\epsilon}{2 \times \log(T)} + \frac{\epsilon}{2^{z+1}}$.

**Algorithm 1** P-FCB

```
Inputs : Total rounds $T$, Quality threshold $\alpha$, $\epsilon$, $\delta$, Cost set $\{\mathbf{c}_j\} = \{(c_{i,j})_{i\in[m]}\}$, Capacity set $\{\mathbf{k}_j\} = \{(k_{i,j})_{i\in[m]}\}$, Start round $\underline{t}$, Stop round $\bar{t}$
/* Initialisation Step */
$t \longleftarrow 0$, $\tau \longleftarrow 1$
$[\forall i \in [m], \forall j \in [n]]$ Initialise total and uncommunicated procurement $(W_{i,j}, w_{i,j})$ and realisations $(Y_{i,j}, y_{i,j})$
while $t \le \frac{3ln(yT)}{2n\zeta^2}$ (Pure Explore Phase) do
    for all the agents $j \in [n]$ do
        Pick procurement vector $\mathbf{s}_j^t = (1)^m$ and observe quality realisations $\mathbf{X}^t_{\mathbf{s}_j^t,j}$.
        $[\forall i \in [m]]$ Update $W_{i,j}^{t+1}, w_{i,j}^{t+1}, Y_{i,j}^{t+1}, y_{i,j}^{t+1}$ using Eq. 2
        if $t \in [\underline{t}, \bar{t}]$ and $t \ge \tau$ then                ▷ Communication round
            $[\forall i \in [m]]$ Calculate $\tilde{w}_{i,j}, \tilde{y_{i,j}}$ according to Eq. 3,4
            for each agent $z \in [n]/j$ do
                Send $\{\tilde{w}_{i,j}, \tilde{y}_{i,j}\}$ to agent $z$
                $[\forall i \in [m]]$ $W_{i,z}^{t+1}, Y_{i,z}^{t+1} \longleftarrow$ CheckandUpdate($W_{i,z}^{t+1}, \tilde{w}_{i,j}, Y_{i,z}^{t+1}, \tilde{y}_{i,j}, .$)
            end for
            $[\forall i \in [m]]$ $w_{i,j}^{t+1} \longleftarrow 0$, $y_{i,j}^{t+1} \longleftarrow 0$
            $\tau \longleftarrow 2 \times \tau$
        end if
        Update quality estimate
        $t \longleftarrow t + 1$
    end for
end while
while $t \le T$, $\forall j \in [n]$ (Explore-Exploit Phase) do
    $[\forall i \in [m]]$ Calculate the upper confidence bound of quality estimate, $(\hat{q}_{i,j}^t)^+$
    Pick procurement vector using $\mathbf{s}_j^t = \mathbf{Oracle}((\hat{q}_{i,j}^t)^+, \mathbf{c}_j, \mathbf{k}_j, .)$ and observe its realisations $\mathbf{X}^t_{\mathbf{s}_j^t,j}$.
    $[\forall i \in [m]]$ Update $W_{i,j}^{t+1}, w_{i,j}^{t+1}, Y_{i,j}^{t+1}, y_{i,j}^{t+1}$ using Eq. 2
    if $t \in [\underline{t}, \bar{t}]$ and $t \ge \tau$ then                ▷ Communication round
        $[\forall i \in [m]]$ Calculate $\tilde{w}_{i,j}, \tilde{y_{i,j}}$ according to Eq. 3,4
        for each agent $z \in [n]/j$ do
            Send $\{\tilde{w}_{i,j}, \tilde{y}_{i,j}\}$ to agent $z$
            $[\forall i \in [m]]$ $W_{i,z}^{t+1}, Y_{i,z}^{t+1} \longleftarrow$ CheckandUpdate($W_{i,z}^{t+1}, \tilde{w}_{i,j}, Y_{i,z}^{t+1}, \tilde{y}_{i,j}, .$)
        end for
        $[\forall i \in [m]]$ $w_{i,j}^{t+1} \longleftarrow 0$, $y_{i,j}^{t+1} \longleftarrow 0$
        $\tau \longleftarrow 2 \times \tau$
    end if
    Update quality estimate
    $t \longleftarrow t + 1$
end while
```

### 5.4 P-FCB: Algorithm

Based on the feedback from the analysis made in previous subsections, we now present a private federated CMAB algorithm for the heterogeneous setting, namely P-FCB. Algorithm 1 formally presents P-FCB. Details follow.

**Algorithm 1 Outline.** The rounds are split into two phases. During the initial pure exploration phase (Lines 6-22), the agents explore all the producers by procuring evenly from all of them. The length of the pure exploration phase is carried over from the non-private algorithm. In this second phase (Lines 23-38), explore-exploit, the agents calculate the $UCB$ for their quality estimates. Then the Oracle is used to provide a procurement vector based on the cost, capacity, $UCB$ values as well as the quality constraint ($\alpha$). Additionally, the agents communicate their estimates as outlined in Sections 5.1 and 5.2. The agents update their quality estimates at the end of each round using procurement and observation values (both local and communicated), Lines 19 and 36.

$$\begin{gathered} w_{i,j}^{t+1} \longleftarrow w_{i,j}^{t} + l_{i,j}^{t} \ ; \ W_{i,j}^{t+1} \longleftarrow W_{i,j}^{t} + l_{i,j}^{t} \\ y_{i,j}^{t+1} \longleftarrow y_{i,j}^{t} + x_{i,j}^{t} \ ; \ Y_{i,j}^{t+1} \longleftarrow Y_{i,j}^{t} + x_{i,j}^{t} \\ q_{i,j}^{t+1} \longleftarrow \frac{Y_{i,j}^{t+1}}{W_{i,j}^{t+1}} \end{gathered} \tag{2}$$

**Noise Addition.** From Section 5.2, we perturb both uncommunicated procurement and realization values for each agent-producer pair using the Gaussian Noise mechanism. Formally, let $w_{i,j}^t, y_{i,j}^t$ be the uncommunicated procurement and realization values. Then $\tilde{w}_{i,j}, \tilde{y}_{i,j}$ are communicated, which are calculated using the following privatizer,

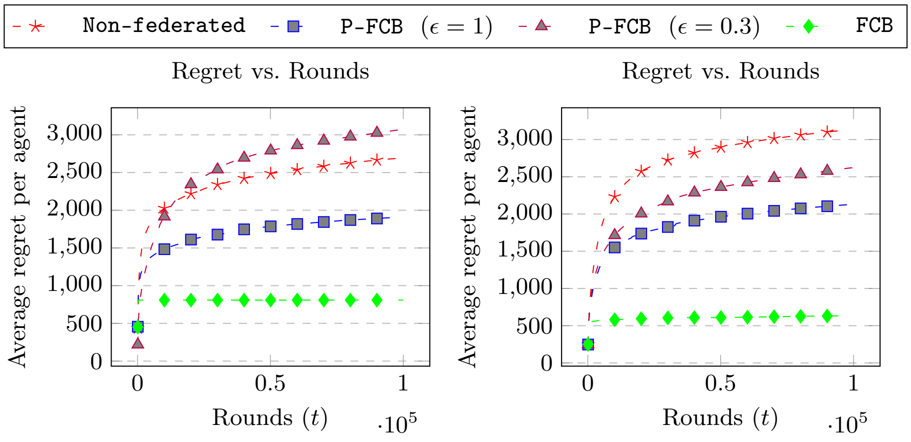
$$\tilde{w}_{i,j} = w_{i,j}^t + \mathcal{N}(0, \frac{2k_{i,j}^2 \log(1.25/\delta)}{(\epsilon^z)^2}) \tag{3}$$

$$\tilde{y}_{i,j} = y_{i,j}^t + \mathcal{N}(0, \frac{2k_{i,j}^2 \log(1.25/\delta)}{(\epsilon^z)^2}) \tag{4}$$

where $\epsilon^z$ is the privacy budget corresponding to the $z^{th}$ communication round.

**What to Learn.** To minimise the regret incurred, we propose that the agents selectively choose what communications to learn from. Weighted confidence bounds around local estimates are used to determine if a communication round should be learned from. Let $\xi_{i,j}^t = \sqrt{\frac{3ln(t)}{2\sum_{z\in\{1,2,\ldots,t\}} l_{i,j}^z}}$ denote the confidence interval agent $j$ has w.r.t. local quality estimate of producer $i$. Then, the agents only selects to learn from a communication if $\hat{q}_{i,j}^t - \omega_1 \xi_{i,j}^t < q_{(communicated)i,j} < \hat{q}_{i,j}^t + \omega_1 \xi_{i,j}^t$ where $\omega_1$ is a weight factor and $q_{(communicated)i,j} = \frac{\tilde{y}_{i,j}}{\tilde{w}_{i,j}}$.

The local observations are weighed more compared to communicated observations for calculating overall estimates. Specifically, $\omega_2 \in [0,1]$ is taken as the weighing factor for communicated observations.



(a) $c_{ij}, q_i \sim U[0, 1]$

(b) $c_{ij}, q_i \sim \mathcal{N}(\alpha = 0.4, 0.2)$

Fig. 3: `EXP1`: Regret Comparison across rounds ($n = 10$, $m = 30$)

### 5.5 `P-FCB`: $(\epsilon, \delta)$-DP Guarantees

In each round, we perturb the values being communicated by adding Gaussian noises satisfying $(\epsilon', \delta')$-DP to them. It is a standard practice for providing DP guarantees for group sum queries. Let $\mathcal{M}$ be a randomised mechanism which outputs the sum of values for a database input $d$ using Gaussian noise addition. Since Oracle is deterministic, each communication round can be considered a post-processing of $\mathcal{M}$ whereby subset of procurement history is the the database input. Thus making individual communication rounds satisfy $(\epsilon', \delta')$-DP.

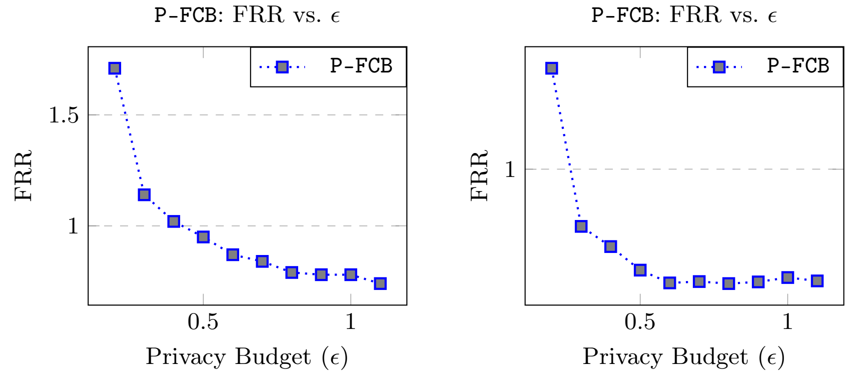
The distinct subset of procurement histories used in each communication round can be considered as independent DP mechanisms. Using the Basic Composition theorem, we can compute the overall $(\epsilon, \delta)$-DP guarantee. In `P-FCB`, we use a target privacy budget, $\epsilon$, to determine the noise parameter $\sigma$ in each round based on Basic composition. Thus, this can be leveraged as a tuning parameter for privacy/regret optimisation.

## 6 Experimental Results

In this section, we compare `P-FCB` with non-federated and non-private approaches for the combinatorial bandit (CMAB) setting with constraints. We first explain the experimental setup, then note our observations and analyze the results obtained.

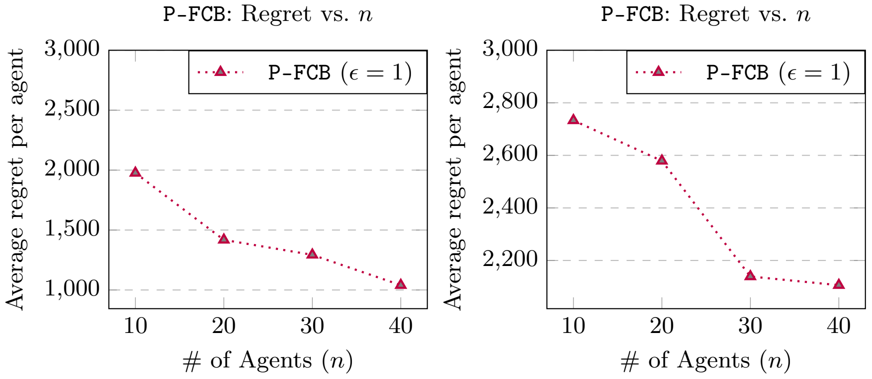
### 6.1 Setup

For our setting, we generate costs and qualities for the producers from: (a) uniform distributions, i.e., $q_i, c_{ij} \sim U[0, 1]$ (b) normal distributions, i.e., $q_i, c_{ij} \sim$



(a) $c_{ij}, q_i \sim U[0,1]$ (b) $c_{ij}, q_i \sim \mathcal{N}(\alpha = 0.4, 0.2)$

Fig. 4: EXP2: FRR for P-FCB while varying privacy budget $\epsilon$ (with $n = 10$, $m = 30$, $t = 100000$)



(a) $c_{ij}, q_i \sim U[0,1]$ (b) $c_{ij}, q_i \sim \mathcal{N}(\alpha = 0.4, 0.2)$

Fig. 5: EXP3: Average regret per agent with P-FCB by varying the number of learners $n$ (with $\epsilon = 1$, $t = 100000$)

$\mathcal{N}(\alpha, 0)$. For both cases, the capacities are sampled from a uniform distribution, $k_{ij} \sim U[1, 50]$. We use the following tuning parameters in our experiments: $\alpha = 0.4$, $\delta = 0.01$ (i.e., $\delta < 1/n$), $\underline{t} = 200$, $\overline{t} = 40000$, $\omega_1 = 0.1$, $\omega_2 = 10$. For our Oracle, we deploy the *Greedy SSA* algorithm presented in Deva et al. [7]. Further, to compare P-FCB's performance, we construct the following two *non-private* baselines:

1. `Non-Federated`. We use the single agent algorithm for subset selection under constraints proposed in Deva et al. [7]. It follows *UCB* exploration similar to P-FCB but omits any communication done with other agents.

2. FCB. This is the non-private variant of P-FCB. That is, instead of communicating $\tilde{w}_{ij}$ and $\tilde{y}_{ij}$, the true values $w_{ij}^t$ and $y_{ij}^t$ are communicated.

We perform the following experiments to measure P-FCB's performance:

- EXP1: For fixed $n = 10$, $m = 30$, we observe the regret growth over rounds ($t$) and compare it to non-federated and non-private federated settings.
- EXP2: For fixed $n = 10$, $m = 30$, we observe $FRR$ (ratio of regret incurred in federated setting compared to non federated setting) at $t = 100000$ while varying $\epsilon$ to see the regret variance w.r.t. privacy budget.
- EXP3: For fixed $\epsilon = 1$, $m = 30$, we observe average regret at $t = 100000$ for varying $n$ to study the effect of number of communicating agents.

For EXP1 and EXP2, we generate 5 instances by sampling costs and quality from both Uniform and Normal distributions. Each instance is simulated 20 times and we report the corresponding average values across all instances. Likewise for EXP3, instances with same producer quality values are considered with costs and capacities defined for different numbers of learners. For each instance, we average across 20 simulations.

### 6.2 Results

- EXP1. P-FCB shows significant improvement in terms of regret (Fig. 3) at the cost of relatively low privacy budget. Compared to FCB, P-FCB ($\epsilon = 1$) and Non-federated incurs 136%,233% more regret respectively for uniform sampling and 235%, 394% more regret respectively for normal sampling. This validates efficacy of P-FCB.
- EXP2. We study the performance of the algorithm with respect to privacy budget (Fig. 4). We observe that according to our expectations, the regret decreases as privacy budget is increased. This decrease in regret is sub-linear in terms of increasing $\epsilon$ values. This is because as privacy budget increases, the amount of noise in communicated data decreases.
- EXP3. We see (Fig. 5) an approximately linear decrease in per agent regret as the number of learning agents increases. This reinforces the notion of reduction of regret, suggested in Section 4, by engaging in federated learning is valid in a heterogeneous private setting.

Discussion: Our experiments demonstrate that P-FCB, through selective learning in a federated setting, is able to achieve a fair regret and privacy trade-off. P-FCB achieves reduction in regret (compared to non-federated setting) for low privacy budgets.

With regards to hyperparamters, note that lower $\omega_2$ suggests tighter bounds while selecting what to learn, implying a higher confidence in usefulness of the communicated data. Thus, larger values for $\omega_1$ can be used if $\omega_2$ is decreased. In general, our results indicate that it is optimal to maintain the value $\omega_1 \cdot \omega_2$ used in our experiments. Also, the communication start time, should be such that the sampled noise is at-least a magnitude smaller than the accumulated uncommunicated data (e.g., $\underline{t} \approx 200$). This is done to ensure that the noisy data is not detrimental to the learning process.

The DP-ML literature suggests a privacy budget $\epsilon < 1$ [30]. From Fig. 4, we note that P-FCB performs well within this privacy budget. While our results achieve a fair regret and privacy trade-off, in future, one can further fine tune these hyperparameters through additional experimentation and/or theoretical analysis.

## 7 Conclusion and Future Work

This paper focuses on learning agents which interact with the same set of producers (“arms") and engage in federated learning while maintaining privacy regarding their procurement strategies. We first looked at a non-private setting where different producers' costs and capacities were the same across all agents and provided theoretical guarantees over optimisation due to federated learning. We then show that extending this to a heterogeneous private setting is non-trivial, and there could be potential information leaks. We propose P-FCB which uses *UCB* based exploration while communicating estimates perturbed using Gaussian method to ensure differential privacy. We defined a communication protocol and a selection learning process using error bounds. This provided a meaningful balance between regret and privacy budget. We empirically showed notable improvement in regret compared to individual learning, even for considerably small privacy budgets.

Looking at problems where agents do not share exact sets of producers but rather have overlapping subsets of available producers would be an interesting direction to explore. It is also possible to extend our work by providing theoretical upper bounds for regret in a differentially private setting. In general, we believe that the idea of when to learn and when not to learn from others in federated settings should lead to many interesting works.

## References

1. Foundry model, \url{https://en.wikipedia.org/w/index.php?title=Foundry_model&oldid=1080269386}
2. Original equipment manufacturer, \url{https://en.wikipedia.org/w/index.php?title=Original_equipment_manufacturer&oldid=1080228401}
3. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. Machine Learning **47**, 235–256 (2004)
4. Chen, S., Tao, Y., Yu, D., Li, F., Gong, B., Cheng, X.: Privacy-preserving collaborative learning for multiarmed bandits in iot. IEEE Internet of Things Journal **8**(5), 3276–3286 (2021)
5. Chen, W., Wang, Y., Yuan, Y.: Combinatorial multi-armed bandit: General framework and applications. In: ICML. PMLR (17–19 Jun 2013)
6. Chiusano, F., Trovò, F., Carrera, G.D., Boracchi, Restelli, M.: Exploiting history data for nonstationary multi-armed bandit. In: ECML/PKDD (2021)

7. Deva, A., Abhishek, K., Gujar, S.: A multi-arm bandit approach to subset selection under constraints. p. 1492–1494. AAMAS '21, AAMAS (2021)
8. Dubey, A., Pentland, A.: Differentially-private federated linear bandits. Advances in Neural Information Processing Systems **33**, 6003–6014 (2020)
9. Dwork, C.: Differential privacy. In: Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II. p. 1–12. ICALP'06 (2006)
10. Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. **9**(3–4), 211–407 (aug 2014)
11. Gai, Y., Krishnamachari, B., Jain, R.: Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. In: DySPAN 2010 (2010)
12. Hannun, A.Y., Knott, B., Sengupta, S., van der Maaten, L.: Privacy-preserving contextual bandits. CoRR **abs/1910.05299** (2019), `http://arxiv.org/abs/1910.05299`
13. Ho, C.J., Jabbari, S., Vaughan, J.W.: Adaptive task assignment for crowdsourced classification. In: ICML. pp. 534–542 (2013)
14. Huang, R., Wu, W., Yang, J., Shen, C.: Federated linear contextual bandits. In: Advances in Neural Information Processing Systems. vol. 34. Curran Associates, Inc. (2021)
15. Jain, S., Gujar, S., Bhat, S., Zoeter, O., Narahari, Y.: A quality assuring, cost optimal multi-armed bandit mechanism for expertsourcing. Artificial Intelligence **254**, 44–63 (01 2018)
16. Kim, T., Bae, S., Lee, J., Yun, S.: Accurate and fast federated learning via combinatorial multi-armed bandits. CoRR (2020), `https://arxiv.org/abs/2012.03270`
17. Li, L., Chu, W., Langford, J., Schapire, R.E.: A contextual-bandit approach to personalized news article recommendation. In: international conference on World wide web (2010)
18. Li, T., Song, L.: Privacy-preserving communication-efficient federated multi-armed bandits. IEEE Journal on Selected Areas in Communications **40**(3), 773–787 (2022)
19. Malekzadeh, M., Athanasakis, D., Haddadi, H., Livshits, B.: Privacy-preserving bandits. In: Proceedings of Machine Learning and Systems. vol. 2, pp. 350–362 (2020)
20. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. PMLR (2017)
21. Mehta, D., Yamparala, D.: Policy gradient reinforcement learning for solving supply-chain management problems. In: Proceedings of the 6th IBM Collaborative Academia Research Exchange Conference (I-CARE) on I-CARE 2014. p. 1–4 (2014)
22. Roy, K., Zhang, Q., Gaur, M., Sheth, A.: Knowledge infused policy gradients with upper confidence bound for relational bandits. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 35–50. Springer (2021)
23. Saber, H., Saci, L., Maillard, O.A., Durand, A.: Routine Bandits: Minimizing Regret on Recurring Problems. In: ECML-PKDD 2021. Bilbao, Spain (Sep 2021)
24. Shi, C., Shen, C.: Federated multi-armed bandits. Proceedings of the AAAI Conference on Artificial Intelligence **35**(11), 9603–9611 (May 2021)
25. Shi, C., Shen, C., Yang, J.: Federated multi-armed bandits with personalization. In: Proceedings of The 24th International Conference on Artificial Intelligence and Statistics. pp. 2917–2925 (2021)

26. Shweta, J., Sujit, G.: A multiarmed bandit based incentive mechanism for a subset selection of customers for demand response in smart grids. Proceedings of the AAAI Conference on Artificial Intelligence **34**(02), 2046–2053 (Apr 2020)
27. Silva, N., Werneck, H., Silva, T., Pereira, A.C., Rocha, L.: Multi-armed bandits in recommendation systems: A survey of the state-of-the-art and future directions. Expert Systems with Applications **197**, 116669 (2022)
28. Slivkins, A.: Introduction to multi-armed bandits. CoRR **abs/1904.07272** (2019), `http://arxiv.org/abs/1904.07272`
29. Solanki, S., Kanaparthy, S., Damle, S., Gujar, S.: Differentially private federated combinatorial bandits with constraints (2022), `https://arxiv.org/abs/2206.13192`
30. Triastcyn, A., Faltings, B.: Federated learning with bayesian differential privacy. In: 2019 IEEE International Conference on Big Data (Big Data). pp. 2587–2596. IEEE (2019)
31. Wang, S., Chen, W.: Thompson sampling for combinatorial semi-bandits. In: Proceedings of the 35th International Conference on Machine Learning. pp. 5114–5122 (2018)
32. Zhao, H., Xiao, M., Wu, J., Xu, Y., Huang, H., Zhang, S.: Differentially private unknown worker recruitment for mobile crowdsensing using multi-armed bandits. IEEE Transactions on Mobile Computing (2021)
33. Zheng, Z., Zhou, Y., Sun, Y., Wang, Z., Liu, B., Li, K.: Applications of federated learning in smart cities: recent advances, taxonomy, and open challenges. Connection Science (2021)