Online Allocation with Differential Privacy

Jianyi Yang*
University of Houston
jyang66@uh.edu

Xingyu Zhou
Wayne State University
xingyu.zhou@wayne.edu

Mostafa Mushsharat
University of Houston
mmushsha@cougarnet.uh.edu

Shaolei Ren
University of California, Riverside
shaolei@ucr.edu

Abstract

We study private online allocation problems, where allocation decisions must satisfy differential privacy (DP) to protect sensitive user information while optimizing performance under limited resources. We first formalize differential privacy notions suited to different privacy requirements in online allocation and establish the fundamental performance limits of any algorithm under these definitions. Based on these insights, we propose PPOA, a privacy-preserving meta algorithm for online allocation, and establish sufficient conditions under which PPOA preserves Joint DP (JDP) or Local DP (LDP) while achieving asymptotically near-optimal performance. Building on these conditions, PPOA can be instantiated into a variety of concrete algorithms. In particular, we present the specific designs, which include PPOA-DMD that preserves JDP and LDP and PPOA-FTRL that preserves JDP. We analyze their performance under both adversarial and stochastic settings and characterize the fundamental trade-offs between DP and allocation performance. Finally, we demonstrate the superior performance of the proposed algorithms via numerical experiments on AI model routing in battery-powered edge systems.

1 Introduction

Online allocation, with applications such as resource-constrained computing [31, 46], online advertising [41, 9], and inventory management [22, 10], is a fundamental problem where an action $x_t \in \mathcal{R}^d$ is decided for the user at round $t \in [T]$ with a reward function $f_t : \mathcal{X}_t \to \mathcal{R}_+$, and a resource consumption function $b_t : \mathcal{X}_t \to \mathcal{R}_+^m$ where $\mathcal{R}_+$ represents the positive real domain and m is the number of resources. The objective is to maximize the cumulative reward while ensuring the element-wise constraints given the resource budget $B \in \mathcal{R}_+^m$ over T rounds, formulated below.

$$\mathrm{OPT}(\mathbf{y}) \coloneqq \max_{x_{1:T}} \sum_{t=1}^{T} f_t(x_t) \ \text{ s.t., } \ \sum_{t=1}^{T} b_t(x_t) \le B \ \text{ and } \ x_t \in \mathcal{X}_t, \ \ \forall t \in [T] \tag{1}$$

The presence of budget constraints inherently couples the decisions across different requests, so the decision for a request contains the information of other requests. This coupling induces a resource-constraint-based side channel through which an attacker can infer other users' contextual information by observing the actions. The privacy risks are further illustrated by application examples in Section 2.

Differentially private resource allocation has been studied both for problems with specific privacy requirements [15, 24] and for more general settings aimed at achieving algorithmic performance

*Correspondence to <jyang66@uh.edu>

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

guarantees [14, 28]. Among the latter, [14] propose an *offline* joint differential private algorithm for *convex* objectives that guarantees expected performance over the algorithm's randomness, while [28] develop an online joint differential private algorithm under a *random* arrival model. However, neither work addresses the adversarial online setting with worst-case competitive performance guarantees that hold uniformly over all context sequences, with failure probabilities vanishing as the time horizon grows. This paper aims to answer the question:

What are the fundamental limits of online allocation under different DP requirements, and how can we achieve meaningful performance guarantees in the adversarial setting under these DP constraints?

Answering this question requires solving the following challenges. **(1) Understanding the performance limits of privacy requirements.** Online allocation admits multiple privacy notions under different settings, making it essential to characterize the fundamental performance limits imposed by these definitions. **(2) Private coordination under global constraints.** Online allocation requires inter-user coordination through shared information, such as remaining budgets or dual variables, to enforce global resource constraints and optimize overall performance. However, such coordination can leak sensitive user information, and the way it is privatized critically affects performance. Currently, there is a lack of a unified meta-framework that privatizes coordination variables while provably maintaining feasibility and achieving strong allocation performance. **(3) Adversarial performance guarantee under DP.** Maintaining competitive performance under privacy constraints is particularly challenging for adversarial context sequences, as the randomness introduced by DP mechanisms must be controlled to obtain guarantees with vanishing failure probabilities. Directly applying private online learning methods, such as noisy Dual Mirror Descent [14], within such a framework fails to provide adversarial guarantees, as unbounded privatization noise can accumulate over time and lead to large performance gaps. We make the following primary contributions.

- We introduce the standard DP, Joint DP, and Local DP notions into online allocation problems and characterize the fundamental tension between DP and allocation performance by establishing best-achievable performance. Our analysis reveals that no algorithm can achieve near-optimal performance guarantees under (ϵ, δ)-DP (Theorem 1), and we characterize the DP-performance trade-offs by establishing the best-achievable competitive ratio under JDP and LDP (Theorem 2).

- We propose a meta framework for online allocation with JDP or LDP (PPOA; Algorithm 1) and establish sufficient conditions under which PPOA preserves (ϵ, δ)-JDP or (ϵ, δ)-LDP (Theorem 3), satisfies the resource constraints with any prescribed high probability, and achieves near-optimal performance simultaneously in both adversarial and i.i.d. stochastic settings (Theorem 4). PPOA can be instantiated by incorporating different DP designs that satisfy the conditions.

- We design two instantiations of PPOA, including PPOA-DMD (Algorithm 2) which preserves both (ϵ, δ)-JDP and (ϵ, δ)-LDP (Theorem 5) and the PPOA-FTRL (Algorithm 3) specifically for (ϵ, δ)-JDP preservation (Theorem 7). They are proved to achieve asymptotic optimal performance with different DP-performance trade-offs (Theorem 8,Theorem 6). They are empirically evaluated by sufficient ablations and baseline comparisons on the battery-powered edge computing application.

2 Application Examples

We summarize two applications here while leaving their details and an additional example about *budget-aware tool calls in agent service* in Section I.

Task Offloading for Battery-Powered Edge Computing In battery-powered edge computing systems, IoT devices handle tasks from different users under limited energy budgets and decide in real time to offload computation to cloud or other edge servers to balance latency, performance, and energy consumption [31, 39, 59]. However, the shared energy constraint introduces a side channel: the decision for one request can implicitly reveal information about prior requests. This creates privacy risks: by observing the offloading decisions for its own tasks, a malicious user can infer the information of other users. Thus, privacy-preserving offloading algorithms are needed to protect user information while maintaining high performance.

Energy Management in Multi-Tenant Data Centers A multi-tenant data center is a decentralized system where the tenants are co-located in one data center and supported by shared resources but decide their own computing configurations locally [29, 18, 16]. During peak hours, the multi-tenant data center must reply on a shared energy storage system for computation, which determines a shared

energy budget for all the tenants. During these hours, each tenant determines its configurations based on its workload characteristics while adhering to a shared global energy constraint. Enforcing this constraint requires servers to exchange coordination information, which can create side channels that leak sensitive information about other tenants' workloads. To mitigate such risks, tenants must privatize their shared coordination information, motivating the local DP problem in resource allocation.

3 Related Work

Online Allocation Online allocation has been widely studied. Some works consider the stochastic settings where the inputs are drawn from a distribution [19, 35, 4, 8]. Also, online algorithms have been proposed to guarantee the worst-case performance given adversarial inputs [7, 53, 52, 36]. Recently, algorithms are designed to achieve the best performances in both adversarial and stochastic settings [7, 12]. The privacy risks of online allocation are often overlooked by these methods while we design algorithms to achieve near-optimal performance even under DP guarantees.

Differential Private Online Algorithms Differential private algorithms have been studied for many online problems such as online learning [2, 32, 58], online bandit learning [54, 5, 11], and reinforcement learning [56, 23, 17, 6, 63, 50]. Private online algorithms are also studied under different DP definitions. Some works consider a relaxed JDP definition which provides strong enough privacy guarantee for their settings [51, 28, 6]. Moreover, LDP mechanisms [62, 49, 33, 23, 64] are proposed to privatize information shared by each local user. The study of DP for online allocation remains nascent and presents great algorithmic challenges for performance guarantee with budget constraint satisfaction.

Private Resource Allocation. Private resource allocation problems have been studied in many settings. [15] reveal the privacy risk in an offline resource-constrained networked system and provides several privatizing mechanisms. [14] develop a noisy dual mirror descent algorithm for JDP preserving offline resource allocation with convex objectives, which achieves a performance bound of $\widetilde{\mathcal{O}}(\frac{1}{\epsilon})$ with bounded constraint violations. [28] provide a jointly differentially private multiplicative weight algorithm for a packing problem with linear objective, achieving a bound of $\widetilde{\mathcal{O}}(\frac{1}{\epsilon})$. Although an online implementation is provided by [28], it requires the assumption of random arrival. While [24] studies online allocation problem, they focus on protecting against two specific inference attacks in an auction setting, and its mechanism only privatizes a threshold parameter arising from its particular algorithm design, which does not protect the underlying user context information.

4 Performance Metrics, Privacy Definitions and Performance Limits

This section introduces the privacy definitions and performance metrics for online allocation. We first introduce some notations and assumptions for the online allocation problem in (1). We denote the context for user at round t as $y_t = \{\mathcal{X}_t, f_t, b_t\}$ and $\boldsymbol{y} = [y_1, y_2 \cdots, y_T]$ which satisfy Assumption 1.

Assumption 1. We assume that the reward function and resource consumption at each round are upper bounded, i.e. $\forall t \in [T], 0 \leq f_t(x_t) \leq \bar{f}$, and $\forall t \in [T], \forall j \in [m]$, $0 \leq [b_t(x)]_j \leq \bar{b}_j$ and $\sup \|b_t(x)\|_2 \leq \bar{b}$. Also for each $j \in [m]$ the consumption upper bound $\bar{b}_j$ is no smaller than the time-averaged budget $\rho_j = B_j/T$. The action set $\mathcal{X}_t$ always includes 0 to represent a null action, with $b_t(0) = 0$ and $f_t(0) = 0$ accordingly. We *do not* assume other function assumptions like continuity, convexity or smoothness.

4.1 Performance Metrics

To guarantee DP, we need to introduce random noise in algorithm design, which makes it difficult to strictly guarantee constraint feasibility given any finite horizon. We therefore require that, the algorithms can be configured to satisfy the budget constraints with any prescribed failure probability as is defined below.

Definition 1 (Constraint feasibility w.h.p.)**.** An online allocation algorithm $\mathcal{A}$ satisfies the constraint feasibility with high probability (w.h.p.) if for any prescribed failure probability $\zeta \in (0, 1)$ and any

fixed context sequence $\boldsymbol{y} \in \mathcal{Y}^T$, with probability at least $1-\zeta$ over the randomness of algorithm $\mathcal{A}$, the actions $\{x_1, \cdots, x_T\}$ of algorithm $\mathcal{A}$ satisfy the constraint $\sum_{t=1}^{T} b_t(x_t) \leq B$.

The performance metrics for online allocation algorithms under both stochastic and adversarial settings are defined for algorithms that satisfy the constraint feasibility w.h.p. (Definition 1). In the stochastic setting, the user contexts $\boldsymbol{y}$ are drawn from a distribution $\mathcal{P}_T$. Under this setting, the objective is to minimize the expected regret expressed as $\mathbb{E}\left[\mathrm{OPT}(\boldsymbol{y}) - F_T^{\mathcal{A}}(\boldsymbol{y})\right]$ where $F_T^{\mathcal{A}}(\boldsymbol{y}) = \sum_{t=1}^{T} f_t(x_t)$ is the total reward given an algorithm $\mathcal{A}$. We mainly express the theoretical results for the stationary stochastic setting where the context of each user y_t is drawn independently from the same distribution $\mathcal{P}$, but the results can be generalized to non-stationary setting without much effort.

More importantly, we focus on **worst-case performance in the adversarial setting**, seeking guarantees that hold uniformly over all context sequences, with failure probabilities over the algorithm's randomness vanishing as the time horizon grows. This formulation accommodates arbitrary context sequences while accounting for the stochastic behavior introduced by privacy mechanisms. Such performance guarantees are particularly relevant to safety-critical applications [43, 27, 48]. In this setting, worst-case performance is typically evaluated using the competitive ratio [41, 21, 53, 52], defined as the minimum ratio between the algorithm reward and the optimal offline reward over context sequences and the algorithm randomness, i.e., $\inf \frac{F_T^{\mathcal{A}}(\boldsymbol{y})}{\mathrm{OPT}(\boldsymbol{y})}$. However, guaranteeing the competitive ratio becomes extremely challenging due to the randomness introduced by DP. Thus, we relax the competitive ratio to *asymptotic competitive ratio in probability* defined below to illustrate meaningful guarantee.

Definition 2 (Asymptotic competitive ratio in probability)**.** A random algorithm $\mathcal{A}$ has an asymptotic competitive ratio of at least $1/\alpha$ in probability, for a fixed $\alpha \geq 1$, if, for every $\kappa > 0$,

$$\lim_{T\to\infty} \sup_{\boldsymbol{y}\in\mathcal{Y}^T} \Pr_{\mathcal{A}}\left[\frac{\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y})}{T} > \kappa\right] = 0,$$

where the probability is over the algorithm's internal randomness.

The definition of asymptotic competitive ratio in probability adapts the notion of convergence in probability [45] to competitive performance. An algorithm with a positive asymptotic competitive ratio in probability achieves the corresponding competitive guarantee up to a sublinear additive error uniformly over all input sequences, with a failure probability that vanishes as the time horizon grows.

4.2 Privacy Definitions and Performance Limits

The actions of online allocation (1), if observed by malicious attackers, may disclose sensitive contextual information about rewards and resource consumption, thereby introducing potential privacy risks. The privacy requirements are formally specified by the following privacy notions.

Differential Privacy The standard (ϵ, δ)-DP defined below is commonly adopted.

Definition 3 ($(\epsilon,\delta)-$DP)**.** For any $\epsilon \geq 0$, $\delta \in (0,1)$, an algorithm $\mathcal{A} : \mathcal{Y}^T \to \mathcal{X}^T$, where $\mathcal{Y}^T = \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T$ and $\mathcal{X}^T = \mathcal{X}_1 \times \cdots \times \mathcal{X}_T$ are compositions of T-round inputs and outputs, is (ϵ, δ)-differential private if for any pair of neighboring sequence $\boldsymbol{y}, \boldsymbol{y}' \in \mathcal{Y}^T$ that differ in the input of one user t, and any subset of $O \subset \mathcal{X}^T$, $\Pr(\mathcal{A}(\boldsymbol{y}) \in O) \leq e^{\epsilon} \cdot \Pr(\mathcal{A}(\boldsymbol{y}') \in O) + \delta$.

DP with a smaller ϵ requires the outputs of two neighboring sequences are less distinguishable, thereby providing a stronger privacy guarantee. In addition, allowing $\delta \in (0,1)$ relaxes this requirement by permitting the privacy guarantee to fail with probability at most δ. However, (ϵ, δ)-DP significantly restricts the variation of actions in response to user contexts, making it impossible to achieve near-optimal reward performance. This is substantiated by the following Theorem.

Theorem 1. *For any $\epsilon > 0, 0 < \delta < 1$, **no** online allocation algorithm that preserves (ϵ, δ)-DP can achieve a sublinear regret. In addition, the asymptotic competitive ratio in probability (Definition 2) of an (ϵ, δ)-DP algorithm is at most δ.*

The proof of Theorem 1 are deferred to Appendix A. Intuitively, the standard (ϵ, δ)-DP restricts the online allocation algorithm to adapt its actions in response to context variations, which incur an unavoidable regret at each round. In the scenarios like the constructed hard instances of the proof,

the identity of the rewarding action encodes the user's private context. Consequently, DP limits how reliably the algorithm can select that rewarding action, leading to linear expected regret for fixed privacy parameters. Moreover, the asymptotic competitive ratio in probability is at most δ, which is typically chosen to be close to zero to guarantee effective privacy in real applications [58, 30, 11]. If it is required to preserve pure DP with $\delta = 0$, no online allocation algorithm can achieve positive competitive ratio even asymptotically.

Joint Differential Privacy In many practical scenarios, the standard (ϵ, δ)-DP is unnecessarily stringent because the action x_t assigned to user t is typically not directly observable by other users. For example, in battery-powered edge computing systems, a centralized server determines the resource allocation for each request without revealing these decisions to others. Each user can observe only the configuration of its own request and has no access to the allocations of other users. It is also reasonable to assume that x_t is observable only to the allocator and the user at round t in many other examples (Section 2). Under this setting, it is not necessary to privatize the action output x_t itself.

Nevertheless, privacy leakage can still arise through resource-based side channels that implicitly couple the actions of different users via shared resource constraints. Since users operate under a shared resource budget, the allocation decision for one user may reveal information of previous users through the remaining budget status or coordination variables. An adversary can send queries to the allocator acting as a normal user, thereby inferring sensitive information about others from its own observed allocations. To mitigate this risk, we adopt (ϵ, δ)-joint differential privacy (JDP, Definition 4), which protects against privacy leakage induced by resource-constraint-based side channels. Compared to standard (ϵ, δ)-DP, (ϵ, δ)-JDP provides a meaningful relaxation by leveraging the fact that the action of a benign user t is not observable by other users or attackers. This relaxed notion enables stronger reward guarantees while still preserving privacy against information leakage through shared resource constraints.

Definition 4 ($(\epsilon, \delta)-$Joint DP (JDP))**.** For any $\epsilon \geq 0$, $\delta \in (0, 1)$, an algorithm $\mathcal{A} : \mathcal{Y}^T \to \mathcal{X}^T$ is (ϵ, δ)-jointly differential private if for any two context sequences $\boldsymbol{y}, \boldsymbol{y}' \in \mathcal{Y}^T$ that differ in the input of a user t, and any subset $O \subset \mathcal{X}_{-t}$, $\Pr(\mathcal{A}(\boldsymbol{y})_{-t} \in O) \leq e^\epsilon \cdot \Pr(\mathcal{A}(\boldsymbol{y}')_{-t} \in O) + \delta$, where $\mathcal{X}_{-t} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \mathcal{X}_{t-1} \times \mathcal{X}_{t+1} \times \cdots \mathcal{X}_T$ is the output space excluding the output space of user t and $\mathcal{A}(\boldsymbol{y})_{-t} = [\mathcal{A}(\boldsymbol{y})_1, \mathcal{A}(\boldsymbol{y})_2, \cdots \mathcal{A}(\boldsymbol{y})_{t-1}, \mathcal{A}(\boldsymbol{y})_{t+1}, \cdots, \mathcal{A}(\boldsymbol{y})_T]$ is the algorithm output space excluding that of user t.

An upper bound of the competitive ratio for online allocation with (ϵ, δ)-JDP is provided in the next Theorem. It demonstrates that the adversarial performance limits of online allocation degrades as the (ϵ, δ)-JDP requirement becomes more stringent (More discussions in Appendix A.3).

Theorem 2. *For any $\epsilon > 0, 0 < \delta < 1$, no $(\epsilon, \delta)-$JDP preserving online allocation algorithm $\mathcal{A}$ can achieve a competitive ratio $(\inf_{\boldsymbol{y} \in \mathcal{Y}^T} \frac{F_T^{\mathcal{A}}(\boldsymbol{y})}{\mathrm{OPT}(\boldsymbol{y})})$ higher than $1 - \frac{1-\delta}{e^\epsilon + 1}$.*

Local Differential Privacy Both standard (ϵ, δ)-DP and (ϵ, δ)-JDP are designed for centralized settings, where a trusted allocator determines the resource allocation. However, in many scenarios (Section 2), allocation decisions are made locally at the user side. In this setting, each user does not share its raw allocation action with others, and thus there is no need to privatize the action itself. Nevertheless, users must share coordination signals (e.g., remaining budget states) to enforce globally shared resource constraints, which may still reveal sensitive information.

To protect privacy in this decentralized setting, we introduce local differential privacy (LDP) [23, 62, 33, 64] to regulate the sharing of coordination information. Without directly privatizing the action outputs, (ϵ, δ)-LDP provides a more relaxed notion than standard (ϵ, δ)-DP, allowing greater flexibility in achieving reward guarantees. Moreover, (ϵ, δ)-LDP implies (ϵ, δ)-JDP if the local mechanism $\mathcal{M}$ is equivalently operated by a centralized allocator, as formalized in Lemma 13.

Definition 5 ($(\epsilon, \delta)-$Local DP (LDP))**.** For any $\epsilon > 0$, $\delta \in (0, 1)$, a local privacy preserving mechanism $\mathcal{M} : \mathcal{Y} \to \mathcal{Z}$, where $\mathcal{Y}$ is the local context space and $\mathcal{Z}$ is the space of shared coordination information to other users, is (ϵ, δ)-locally differential private if for any adjacent contexts $y, y' \in \mathcal{Y}$, and any subset of $O \subset \mathcal{Z}$, $\Pr(\mathcal{M}(y) \in O) \leq e^\epsilon \cdot \Pr(\mathcal{M}(y') \in O) + \delta$.

Since (ϵ, δ)-LDP implies (ϵ, δ)-JDP when a locally privacy-preserving algorithm is equivalently implemented by a centralized allocator, as formalized in Lemma 13, online allocation under (ϵ, δ)-LDP does not admit a higher competitive ratio upper bound than (ϵ, δ)-JDP.

5 Meta Algorithm

5.1 Meta Algorithm Design

To preserve $(\epsilon, \delta)-$JDP or $(\epsilon, \delta)-$LDP, the meta algorithm in Algorithm 1 performs an action selection (primal update) for user t based on a clean context y_t, a privatized dual variable μ_t, and a privatized remaining budget $\tilde{B}_t$. The action is pre-selected by balancing the immediate reward optimization with the budget saving for potentially higher future rewards with the privatized μ_t as the balancing weight, i.e. $\hat{x}_t = \arg\max_{x \in \mathcal{X}_t} \left\{ f_t(x) - \mu_t^\top b_t(x) \right\}$. Next in Line 4 of Algorithm 1, the allocator selects the action x_t by projecting the pre-selected action $\hat{x}_t$ to ensure the budget constraint satisfaction. This projection must guarantee the budget consumption $b_t(x_t)$ at each round t does not exceed the lower bound of the true remaining budget $B_t = B_1 - \sum_{\tau=1}^{t-1} b_\tau(x_\tau)$, which is estimated by subtracting the gap bound $B_{\text{gap},t}$ between the privatized and clean budgets from the privatized remaining budget $\tilde{B}_t$.

Algorithm 1 PPOA: Meta Algorithm for Privacy Preserving Online Allocation

Require: Initialize virtual budget $\tilde{B}_1 = B_1 = \rho T$, dual variable μ_1, and step size $\eta > 0$.

1: **for** $t = 1$ to T **do**
2: *Primal update (Line 3-4)*
3: Pre-select action: $\hat{x}_t = \arg\max_{x \in \mathcal{X}_t} \left\{ f_t(x) - \mu_t^\top b_t(x) \right\}$
4: If $b_t(\hat{x}_t) \leq \tilde{B}_t - B_{\text{gap,t}} \mathbf{1}_m$, $x_t = \hat{x}_t$; Else, $x_t = 0$
Private budget & dual update (Line 5-6)
5: Differential private remaining budget update: $\tilde{B}_{t+1} = \mathcal{M}_B(\{b_\tau(x_\tau)\}_{\tau \in [t]}, \epsilon, \delta)$ (Joint DP) or $\tilde{B}_{t+1} = \mathcal{M}_B(\tilde{B}_t, b_t(x_t), \epsilon, \delta)$ (Local DP)
6: Differential private dual update: $\mu_{t+1} = \mathcal{M}_D(\{b_\tau(x_\tau)\}_{\tau \in [t]}, \epsilon, \delta)$ (Joint DP) or $\mu_{t+1} = \mathcal{M}_D(\mu_t, b_t(x_t), \epsilon, \delta)$ (Local DP)
7: **end for**

To preserve $(\epsilon, \delta)-$JDP or $(\epsilon, \delta)-$LDP, Algorithm 1 explicitly performs differential private mechanisms for the remaining budget and the Lagrangian dual as shown in Line 5-6 because the two variables are dependent on the context information of previous users. If we only require to preserve JDP, the mechanisms have access to all previous clean budget consumptions. However, to preserve LDP, the remaining budget and the Lagrangian dual are updated locally and can only access the privatized variables from previous rounds. Next, we present the conditions to guarantee the privacy notions, budget constraints, and allocation performance by PPOA.

5.2 Conditions for Privacy Preservation

In online allocation, each action x_t is a function of the clean context y_t and the intermediate variables $\tilde{B}_t$ and μ_t, so Billboard Lemma 12 directly implies the following theorem about the joint or local DP guarantees of Algorithm 1 as indicated by the theorem below (proof in Appendix C.1).

Theorem 3. *If the compositions of the remaining budget $(\tilde{B}_1, \cdots, \tilde{B}_T)$ and the dual variables $(\mu_1, \cdots, \mu_T)$ satisfy $(\epsilon, \delta)-$DP, Algorithm 1 satisfies $(\epsilon, \delta)-$JDP if it is implemented as a central privacy preserving algorithm, and $(\epsilon, \delta)-$LDP if it is implemented as a local privacy preserving algorithm.*

Theorem 3 shows that as long as the mechanisms $\mathcal{M}_B$ and $\mathcal{M}_D$ in Algorithm 1 are $(\epsilon, \delta)-$differential private, then Algorithm 1 which maps user contexts to actions is $(\epsilon, \delta)-$ JDP or LDP.

5.3 Conditions for Performance Guarantee

Under the $(\epsilon, \delta)-$ JDP or LDP by Theorem 3, we derive the conditions to satisfy the budget constraints and guarantee asymptotic performance in both adversarial and i.i.d. stochastic settings.

Theorem 4. *The following statements hold for Algorithm 1 with prescribed parameter $\zeta \in (0, 1)$.*
*(1) **Constraint feasibility.** Algorithm 1 satisfies the constraint feasibility with high probability at least $1 - \zeta$ as Definition 1 if, for every fixed sequence $\boldsymbol{y} \in \mathcal{Y}^T$, it holds with probability at least $1 - \zeta$ over the randomness of $\mathcal{A}$ that $\forall t \in [T], \|\tilde{B}_t - B_t\|_\infty \leq B_{\text{gap},t}$.*
*(2) **Performance guarantee.** Under the conditions in (1), Algorithm 1 guarantees an asymptotic competitive ratio of at least $1/\alpha$, $\alpha \geq 1$ in probability as Definition 2 and a sublinear expected regret if $B_{\text{gap},t} \leq \Gamma(T, \zeta)$ and for every fixed sequence $\boldsymbol{y} \in \mathcal{Y}^T$, with probability at least $1 - \zeta$ over the randomness of $\mathcal{A}$, $\forall \tau \in [T], \forall \mu \in \mathcal{D}, \sum_{t=1}^{\tau} \big(w_t(\mu_t) - w_t(\mu) \big) \leq$*

$\Gamma(T,\zeta)$, *where* $\Gamma(T,\zeta)$ *is deterministic and the algorithm uses* $\zeta = \zeta_T \in (0,1)$ *satisfying* $\lim_{T\to\infty}\zeta_T = 0, T\zeta_T = O(1), \lim_{T\to\infty}\frac{\Gamma(T,\zeta_T)}{T} = 0$, *and the dual region* $\mathcal{D} \subset \mathcal{R}_+^m$ *contains* $\{\mathbf{0}, \frac{\bar{f}}{\alpha\rho_1}\mathbf{e}_1, \cdots, \frac{\bar{f}}{\alpha\rho_m}\mathbf{e}_m, \frac{\bar{f}}{\rho_1}\mathbf{e}_1, \cdots, \frac{\bar{f}}{\rho_m}\mathbf{e}_m\}$ *where* $\mathbf{e}_j \in \mathcal{R}^m$ *has 1 at* $j-$*th entry and 0 at the others.*

The proof is in Appendix C.2. The asymptotic competitive ratio $\frac{1}{\alpha}$ in probability, achieved by Algorithm 1 is optimal for online allocation with adversarial inputs, even in the absence of JDP or LDP requirements [9]. Therefore, Theorem 4 shows that, when its two conditions are satisfied, PPOA asymptotically approaches the optimal performance in both adversarial and i.i.d. stochastic settings. The first condition requires that the reserved budget cap $B_{\text{gap},T}$ in Line 4 of Algorithm 1 upper bounds the discrepancy between the privatized remaining budget and its clean counterpart, and increases in a sublinear way with T. This condition guarantees at each round the selected action satisfies the constraints with arbitrarily high probability. The second condition requires that the dual update in Line 6 of Algorithm 1 ensures convergence of the dual objective relative to any fixed dual variable in the set $\mathcal{D}$. The resulting convergence rate directly determines the primal regret performance.

In practice, context distributions may vary over time, and contexts may exhibit temporal dependence. The asymptotic competitive ratio in probability of PPOA established in Theorem 4 continues to hold in this setting, since its failure-probability bound is uniform over all context sequences. However, sublinear expected regret is not guaranteed without additional assumptions, as discussed in Section C.2.4.

By the conditions in Theorem 3 and Theorem 4, the private online allocation can be converted to a problem of designing differential private online learning algorithms with bounded convergence error. In this way, PPOA provides a unifying interface for a wide range of existing and future privacy-preserving online learning designs to meet different privacy and performance requirements.

6 Instantiated Algorithms and Performance Analysis

In this section, we instantiate PPOA (Algorithm 1) to obtain PPOA-DMD, which preserves both JDP and LDP, and PPOA-FTRL, which preserves only JDP, and provide performance analysis to characterize the privacy-performance trade-offs.

6.1 PPOA-DMD: Algorithm for JDP and LDP

PPOA-DMD is presented in Algorithm 2, which can be implemented in either centralized or decentralized allocation systems. In a centralized system, each arriving user uploads its raw context to the centralized allocator, and the allocator selects an action for the user by Line 2 in Algorithm 2. Next, to protect the user context information, the centralized allocator updates the remaining budget and the dual variable in a differential private way by Line 3-5 in Algorithm 2 and uses them to select action for the next user.

Algorithm 2 PPOA-DMD

Require: Initialize virtual budget $\tilde{B}_1 = B_1 = \rho T$, dual variable μ_1, noise variance σ^2, step size $\eta > 0$.

1: **for** $t = 1$ to T **do**
2: Select an action by Line 2-4 in Algorithm 1.
3: Privatize the budget consumption: $\tilde{b}_t(x_t) = b_t(x_t) + n_t$, $n_t \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.
4: Update remaining budget $\tilde{B}_{t+1} = \tilde{B}_t - \tilde{b}_t(x_t)$.
5: Update the dual variable by private DMD in (2).
6: **end for**

In a decentralized system, each user at round t makes the allocation decision locally by Line 2 in Algorithm 2 without uploading or sharing its raw context and selected action. Then, the user only disseminates the privatized budget consumption $\tilde{b}_t(x_t)$ in Line 3 of Algorithm 2 to other users. The user at the next round $t+1$ updates the remaining budget and dual variable by Line 3-5 in Algorithm 2 based on the privatized consumption $\tilde{b}_t(x_t)$ and use them to select an action locally.

Algorithm 2 needs a dual learning procedure to update the dual variable in Line 5. Directly applying Dual Mirror Descent (DMD) in [44] updates the dual variable as $\mu_{t+1} = \text{argmin}_{\mu\geq 0}(\rho - \tilde{b}_t(x_t))^\top\mu + \frac{1}{\eta}V_h(\mu,\mu_t)$ where $V_h(\mu,\mu_t) = h(\mu) - h(\mu_t) - \nabla h(\mu_t)^\top(\mu-\mu_t)$ denotes the Bregman divergence associated with a 1-strongly convex function $h(\mu)$. However, noise in $\tilde{b}_t(x_t)$ accumulates in the

DMD process, which induces a linear term in the convergence bound of the dual objective, violating Condition (2) in Theorem 4. To this end, PPOA-DMD employs a *projected* DMD update, i.e.

$$\mu_{t+1} = \arg \min_{\mu \geq 0,\ \|\mu\|_2 \leq \mu_{\max}} (\rho - \tilde{b}_t(x_t))^\top \mu + \frac{1}{\eta} V_h(\mu, \mu_t), \tag{2}$$

which constrains the dual variable norm by a finite hyperparameter $\mu_{\max} > 0$. This projection stabilizes the dual learning process by avoiding unreasonably large dual variable due to unbounded noise injection, which is essential for preserving the worst-case performance.

Analysis for PPOA-DMD Theorem 5 (proof in Appendix D.1) specifies the way to set the noise standard deviation σ to preserve $(\epsilon, \delta)-$JDP and $(\epsilon, \delta)-$LDP.

Theorem 5. *Choosing* $\sigma = \frac{\sqrt{2}\bar{b}(\sqrt{\ln(1/\delta)}+\sqrt{\ln(1/\delta)+\epsilon})}{\epsilon}$*, PPOA-DMD (Algorithm 2) preserves* $(\epsilon, \delta)-$*LDP in decentralized settings and* $(\epsilon, \delta)-$*JDP in centralized settings.*

Next, we prove that the privacy mechanism in PPOA-DMD satisfies the conditions in Theorem 4 (proof in Appendix D.2), thus asymptotically approaching the optimal solution for both i.i.d. stochastic and adversarial settings.

Theorem 6. *With* σ *set in Theorem 5, the budget gap as* $B_{\text{gap},t} = \sigma\sqrt{2t \ln \frac{2mT}{\zeta}}$*, the step size as* $\eta = \mathcal{O}(\frac{\epsilon}{\sqrt{T \ln(1/\delta)}})$*, dual norm bound as* $\mu_{\max} \geq \max_j \frac{\bar{f}}{\rho_j}$*, and the failure probability as* $\zeta = \zeta_T = \frac{1}{T^\vartheta + 1}$ *given a constant* $\vartheta \geq 1$*, PPOA-DMD satisfies the conditions in Theorem 4 with* $\Gamma(T, \zeta_T) = \widetilde{\mathcal{O}}\left(\frac{\mu_{\max}\sqrt{T \log(1/\delta)}}{\epsilon}\right)$ *where* $\widetilde{\mathcal{O}}$ *absorbs the terms scaling with* $\log(T)$*, so it achieves an asymptotic competitive ratio in probability (Definition 2) of at least* $1/\alpha$ *with* $\alpha = \sup_{j \in [m]} \frac{\bar{b}_j}{\rho_j}$ *and a sublinear expected regret under i.i.d. arrivals in the order of* $\widetilde{\mathcal{O}}\left(\frac{\mu_{\max}\sqrt{T \log(1/\delta)}}{\epsilon}\right)$.

Theorem 6 shows that PPOA-DMD with (ϵ, δ)-JDP/LDP has regret scaling with $\sqrt{\log(1/\delta)}/\epsilon$, illustrating a privacy-performance trade-off: higher DP levels (smaller ϵ or δ) lead to larger regrets. Furthermore, selecting a smaller $\mu_{\max}$ in projected DMD (2) leads to smaller regret bounds since a smaller $\mu_{\max}$ reduces the impact of noise more effectively. However, $\mu_{\max}$ must not be overly small to ensure PPOA-DMD can search a proper dual variable. Theorem 4 shows that $\mu_{\max}$ should be at least $\max_j \frac{\bar{f}}{\rho_j}$. This is because with less budget (smaller ρ_j) or larger reward bound $\bar{f}$, a proper dual variable must be larger to encourage resource saving or match the reward scaling.

6.2 PPOA-FTRL: Enhanced Algorithm for JDP

In this section, we show that an enhanced instantiation of PPOA can achieve a better privacy–performance trade-off when only JDP is required in a centralized setting. Intuitively, the advantage of the JDP-preserving centralized setting stems from the presence of a trusted allocator with access to users' raw contexts. By PPOA, the decision at each round does not depend on individual historical data, but rather on aggregated information from past users, such as the remaining budget and dual variables for resource coordination. Unlike the LDP-preserving decentralized setting, where information aggregation is based on privatized local information, the centralized allocator can leverage raw data to construct more accurate aggregations. As a result, to achieve the same level of differential privacy, centralized mechanisms (e.g., [32]) require significantly less noise than LDP-based aggregation. Therefore, when only JDP is required, we aim to exploit centralized aggregation in PPOA to enable substantially improved allocation performance.

Motivated by this goal, we develop PPOA-FTRL (Algorithm 3 in Appendix B). The mirror descent update in PPOA-DMD depends on the dual update of the last round, so it cannot utilize the centralized information aggregation. Comparably, PPOA-FTRL relies on Follow the Regularized Leader (FTRL) [44] to update the dual variable directly based on the aggregated budget consumption. It privatizes cumulative budget consumption $\sum_{\tau=1}^{t} b_\tau(x_\tau)$ as s_t by the tree aggregation strategy [32] which injects much less noise than directly perturbing the cumulative consumption to preserve the same level of (ϵ, δ)-JDP. Based on the privatized cumulative consumption s_t, we can replace Line 5-6

in Algorithm 1 by the remaining budget update as $\tilde{B}_{t+1} = B_1 - s_t$ and the private dual update as $\mu_{t+1} = \arg\min_{\mu \geq 0}(\rho t - s_t)^\top \mu + \frac{1}{\eta} h(\mu)$, where $h(\mu)$ is a 1-strongly convex regularization function. Further details about PPOA-FTRL are provided in Section B.

Analysis for PPOA-FTRL The $(\epsilon, \delta)-$JDP preservation with the selection of noise standard decision is justified below (proof in Appendix E.1).

Theorem 7. *PPOA-FTRL (Algorithm 3) preserves $(\epsilon, \delta)-$JDP if the standard deviation of the Gaussian noise added to each node of the tree is set as $\sigma = \frac{\sqrt{2\bar{b}}\sqrt{\lceil \log_2(T) \rceil}(\sqrt{\ln(1/\delta)} + \sqrt{\ln(1/\delta) + \epsilon})}{\epsilon}$.*

Next, based on the conditions in Theorem 4 and the convergence of FTRL with the tree aggregation mechanism, we can get the performance guarantee in the next theorem (proof in Appendix E.2).

Theorem 8. *With σ set in Theorem 7, the budget gap as $B_{\text{gap},t} = \sigma\sqrt{2\lceil \log_2(t) \rceil \ln \frac{2mT}{\zeta}}$, the step size as $\eta = \mathcal{O}(\sqrt{\frac{\epsilon}{T\sqrt{\log_2(T)\ln(1/\delta)}}})$, and the failure probability as $\zeta = \zeta_T = \frac{1}{T^\vartheta + 1}$ given a constant $\vartheta \geq 1$, PPOA-FTRL satisfies the conditions in Theorem 4 with $\Gamma(T, \zeta_T) = \widetilde{\mathcal{O}}\left(\frac{\sqrt{T\sqrt{\log(1/\delta)}}}{\sqrt{\epsilon}}\right)$ where $\widetilde{\mathcal{O}}$ absorbs the terms scaling with $\log(T)$, so it achieves an asymptotic competitive ratio in probability (Definition 2) of at least $1/\alpha$ with $\alpha = \sup_{j \in [m]} \frac{\bar{b}_j}{\rho_j}$ and a sublinear expected regret under i.i.d. arrivals in the order of $\widetilde{\mathcal{O}}\left(\frac{\sqrt{T\sqrt{\log(1/\delta)}}}{\sqrt{\epsilon}}\right)$.*

We can find that the regret bounds of PPOA-FTRL scale with $\widetilde{\mathcal{O}}(\sqrt{\sqrt{\log(1/\delta)/\epsilon}})$ which is the squared root of the privacy-related scaling (Theorem 6) achieved by PPOA-DMD. This indicates that within the regime of stringent privacy requirements (e.g. $\epsilon < 1$ and δ is close to zero), PPOA-FTRL provides significantly better allocation performance than PPOA-DMD. This benefit stems from the FTRL-based aggregation, which introduces much less Gaussian noise to the aggregated information.

7 Numerical Results

We consider a request routing problem for a battery-powered edge AI system, which enables time-sensitive edge AI services [61, 37, 39, 55]. Beyond this, the numerical evaluation on the application of online biding in advertising [34, 3] is provided in Appendix I. The edge AI service can deploy multiple models on an edge data center with finite energy budget in its battery. As user requests arrive sequentially, the system must decide in real time which model should be selected to serve the request. The experiments are based on the RouterBench dataset [26] and setup details are given in the Appendix G.2.

Baselines. We compare PPOA-DMD and PPOA-FTRL against the following baselines. **Non-private baselines**: • Greedy: greedy selects the constraint-feasible action that directly maximizes the reward $f_t(x)$. • DMD: the dual mirror descent algorithm [7] with $h(\mu)$ being the squared norm function. • FTRL: update the dual variable with follow the regularized leader [44] with $h(\mu)$ being the squared norm function. **Private baselines**: • DPA: a centralized online allocation algorithm preserving standard $(\epsilon, \delta)-$DP in Definition 3 by additionally applying Gaussian mechanism to reward and consumption functions of PPOA-FTRL. • NDMD: Noisy DMD [14] adapted to satisfy the constraint feasibility in the same way as PPOA without applying the projection in 2. • NFTRL: Directly applying noisy FTRL without applying tree aggregation (TA) [13] for consumption aggregation. • Pri-DOWM: Algorithm 2 in [28] adapted to satisfy the constraint feasibility in the same way as PPOA.

Baseline Comparison We evaluate both the expected and adversarial performance of the algorithms in Table 1. All methods exhibit zero budget violation on the dataset, ensuring a fair comparison. As expected, DMD and FTRL achieve the best performance without privacy constraints. More importantly, when requiring both JDP and LDP, PPOA-DMD consistently outperforms the direct method NDMD, demonstrating the effectiveness of the dual projection in (2) for mitigating the impact

Algorithms	Non-private algorithms			Private algorithms					
	Greedy	DMD	FTRL	DPA	NDMD	NFTRL	Pri-DOWM	**PPOA-DMD**	**PPOA-FTRL**
AVG	0.710	**0.919**	0.914	0.591±0.0001	0.778±0.001	0.726±0.001	0.775± 0.0006	0.801±0.001	**0.879**±0.0002
CR	0.632	**0.876**	0.867	0.514±0.003	0.635±0.015	0.597±0.019	0.652±0.022	0.665±0.023	**0.834**±0.005

Table 1: The expected (AVG) and adversarial (CR) performance of algorithms. Among the private algorithms, DPA admits $(\epsilon, \delta)-$DP, NDMD and PPOA-DMD admit $(\epsilon, \delta)-$JDP or $(\epsilon, \delta)-$LDP, NFTRL and PPOA-FTRL admit $(\epsilon, \delta)-$JDP, all with $\epsilon = 10$ and $\delta = 0.01$.

of noise in the dual learning process. When only JDP is required, PPOA-FTRL achieves significantly better performance than NFTRL, NDMD, Pri-DOWM, and PPOA-DMD, as its tree-aggregation mechanism more effectively leverages raw historical contexts. Finally, DPA enforces the strongest DP notion, so despite operating in a centralized way, it exhibits the worst performance among the private baselines, which empirically supports the impossibility result in Theorem 1.

Privacy-Performance Trade-off Figure 1 reports the competitive ratio as a function of the privacy level ϵ, illustrating the privacy–performance trade-off. As expected, looser privacy requirements (larger ϵ) lead to improved reward performance across all methods. More importantly, among algorithms that satisfy both JDP and LDP, PPOA-DMD achieves a consistently better trade-off curve than NDMD, owing to the dual projection in (2), which stabilizes the dual updates under noise injection. Furthermore, when only JDP is required, PPOA-FTRL attains a significantly better trade-off than the other algorithms, highlighting the effectiveness of the tree-aggregation strategy for online allocation.

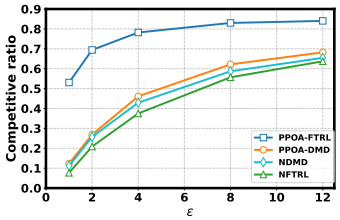


Figure 1: Competitive ratio with different ϵ under $\delta = 0.01$.

Results of ablation studies on privacy levels, budget reservation setups in PPOA, dual norm cap in PPOA-DMD, and budget B are available in Appendix G.3.

8 Conclusion

In this paper, we study the problem of online allocation with DP. We first formally analyze the performance limits for different DP definitions. We then propose a meta-algorithm PPOA and establish sufficient conditions under which PPOA preserves JDP and LDP, satisfies the budget constraint with any prescribed high probability, and achieves asymptotically optimal performance. We instantiate PPOA to JDP/LDP-preserving PPOA-DMD and an enhanced JDP-preserving PPOA-FTRL with concrete performance analysis. Finally, we validate our theoretical findings through numerical studies.
Limitations and future work. While the work establishes the first performance bounds for online allocation under DP definitions, they are not proved to be the best bounds in terms of the scaling w.r.t. privacy levels. There is still room to improve the bounds especially in an almost-sure asymptotic or a non-asymptotic way. Also, it is interesting to explore meaningful performance guarantee under the most strict DP notion.
Broader impacts. The results in this work provide a algorithmic foundation to improve the privacy protection in wide online resource allocation systems.

Acknowledgement

We thank the NeurIPS reviewers for their insightful comments during the rebuttal process, which have helped us improve our work. Jianyi Yang was supported in part by the University of Houston Energy Transition Institute Seed Grant Program. Shaolei Ren were supported in part by the NSF under grants NSF-2449990 and NSF-2326598. Xingyu Zhou is supproted in part by the NSF under grants NSF CAREER-2441519 and NSF-2312835.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[2] Naman Agarwal and Karan Singh. The price of differential privacy for online learning. In *International Conference on Machine Learning*, pages 32–40. PMLR, 2017.

[3] Gagan Aggarwal, Ashwinkumar Badanidiyuru, Santiago R Balseiro, Kshipra Bhawalkar, Yuan Deng, Zhe Feng, Gagan Goel, Christopher Liaw, Haihao Lu, Mohammad Mahdian, et al. Auto-bidding and auctions in online advertising: A survey. *ACM SIGecom Exchanges*, 22(1):159–183, 2024.

[4] Lin An, Andrew A Li, Benjamin Moseley, and Gabriel Visotsky. Best of many in both worlds: Online resource allocation with predictions under unknown arrival model. *arXiv preprint arXiv:2402.13530*, 2024.

[5] Achraf Azize and Debabrota Basu. When privacy meets partial information: A refined analysis of differentially private bandits. *Advances in Neural Information Processing Systems*, 35:32199–32210, 2022.

[6] Shaojie Bai, Lanting Zeng, Chengcheng Zhao, Xiaoming Duan, Mohammad Sadegh Talebi, Peng Cheng, and Jiming Chen. Private online learning in adversarial mdps: Full-information and bandit. In *Seventeenth European Workshop on Reinforcement Learning*.

[7] Santiago Balseiro, Haihao Lu, and Vahab Mirrokni. Dual mirror descent for online allocation problems. In *International Conference on Machine Learning*, pages 613–628. PMLR, 2020.

[8] Santiago Balseiro, Haihao Lu, and Vahab Mirrokni. Regularized online allocation problems: Fairness and beyond. In *International Conference on Machine Learning*, pages 630–639. PMLR, 2021.

[9] Santiago R Balseiro and Yonatan Gur. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

[10] Siddhartha Banerjee, Chamsi Hssaine, and Sean R Sinclair. Online fair allocation of perishable resources. *ACM SIGMETRICS Performance Evaluation Review*, 51(1):55–56, 2023.

[11] Debabrota Basu, Christos Dimitrakakis, and Aristide Tossou. Differential privacy for multi-armed bandits: What is it and what is its cost? *arXiv preprint arXiv:1905.12298*, 2019.

[12] Matteo Castiglioni, Andrea Celli, and Christian Kroer. Online learning with knapsacks: the best of both worlds. In *International Conference on Machine Learning*, pages 2767–2783. PMLR, 2022.

[13] T-H Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24, 2011.

[14] Du Chen and Geoffrey A Chua. Noisy dual mirror descent: A near optimal algorithm for jointly-dp convex resource allocation. *Advances in Neural Information Processing Systems*, 37:91723–91755, 2024.

[15] Joann Qiongna Chen, Tianhao Wang, Zhikun Zhang, Yang Zhang, Somesh Jha, and Zhou Li. Differentially private resource allocation. In *Proceedings of the 39th Annual Computer Security Applications Conference*, pages 772–786, 2023.

[16] Niangjun Chen, Xiaoqi Ren, Shaolei Ren, and Adam Wierman. Greening multi-tenant data center demand response. *ACM SIGMETRICS Performance Evaluation Review*, 43(2):36–38, 2015.

[17] Sayak Ray Chowdhury and Xingyu Zhou. Differentially private regret minimization in episodic markov decision processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6375–6383, 2022.

[18] Xiangming Dai, Jason Min Wang, and Brahim Bensaou. Energy-efficient virtual machines scheduling in multi-tenant data centers. *IEEE Transactions on Cloud Computing*, 4(2):210–221, 2015.

[19] Nikhil R Devanur, Kamal Jain, Balasubramanian Sivan, and Christopher A Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. *Journal of the ACM (JACM)*, 66(1):1–41, 2019.

[20] Apple Differential Privacy Team. Learning with privacy at scale, 2017.

[21] Matthew Fahrbach, Zhiyi Huang, Runzhou Tao, and Morteza Zadimoghaddam. Edge-weighted online bipartite matching. *Journal of the ACM*, 69(6):1–35, 2022.

[22] Brandon Fain, Kamesh Munagala, and Nisarg Shah. Fair allocation of indivisible public goods. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 575–592, 2018.

[23] Evrard Garcelon, Vianney Perchet, Ciara Pike-Burke, and Matteo Pirotta. Local differential privacy for regret minimization in reinforcement learning. *Advances in Neural Information Processing Systems*, 34:10561–10573, 2021.

[24] Jianxiong Guo and Weili Wu. Differential privacy-based online allocations towards integrating blockchain and edge computing. *arXiv preprint arXiv:2101.02834*, 2021.

[25] Justin Hsu, Zhiyi Huang, Aaron Roth, Tim Roughgarden, and Zhiwei Steven Wu. Private matchings and allocations. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 21–30, 2014.

[26] Qitian Jason Hu, Jacob Bieker, Xiuyu Li, Nan Jiang, Benjamin Keigwin, Gaurav Ranganath, Kurt Keutzer, and Shriyash Kaustubh Upadhyay. Routerbench: A benchmark for multi-llm routing system. *arXiv preprint arXiv:2403.12031*, 2024.

[27] Wenzhen Huang, Tong Li, Yuting Cao, Zhe Lyu, Yanping Liang, Li Yu, Depeng Jin, Junge Zhang, and Yong Li. Safe-nora: Safe reinforcement learning-based mobile network resource allocation for diverse user demands. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 885–894, 2023.

[28] Zhiyi Huang and Xue Zhu. Near optimal jointly private packing algorithms via dual multiplicative weight update. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 343–357. SIAM, 2018.

[29] Mohammad A Islam, Xiaoqi Ren, Shaolei Ren, Adam Wierman, and Xiaorui Wang. A market approach for handling power emergencies in multi-tenant data center. In *2016 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pages 432–443. IEEE, 2016.

[30] Prateek Jain, Pravesh Kothari, and Abhradeep Thakurta. Differentially private online learning. In *Conference on Learning Theory*, pages 24–1. JMLR Workshop and Conference Proceedings, 2012.

[31] Hongbo Jiang, Xingxia Dai, Zhu Xiao, and Arun Iyengar. Joint task offloading and resource allocation for energy-constrained mobile edge computing. *IEEE Transactions on Mobile Computing*, 22(7):4000–4015, 2022.

[32] Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu. Practical and private (deep) learning without sampling or shuffling. In *International Conference on Machine Learning*, pages 5213–5225. PMLR, 2021.

[33] Chonghua Liao, Jiafan He, and Quanquan Gu. Locally differentially private reinforcement learning for linear mixture markov decision processes. In *Asian Conference on Machine Learning*, pages 627–642. PMLR, 2023.

[34] Hairen Liao, Lingxiao Peng, Zhenchuan Liu, and Xuehua Shen. ipinyou global rtb bidding algorithm competition dataset. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pages 1–6, 2014.

[35] Luofeng Liao, Yuan Gao, and Christian Kroer. Nonstationary dual averaging and online fair allocation. *Advances in Neural Information Processing Systems*, 35:37159–37172, 2022.

[36] Qiulin Lin, Yanfang Mo, Junyan Su, and Minghua Chen. Competitive online optimization with multiple inventories: A divide-and-conquer approach. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(2):1–28, 2022.

[37] Zehong Lin, Suzhi Bi, and Ying-Jun Angela Zhang. Optimizing ai service placement and resource allocation in mobile edge intelligence systems. *IEEE Transactions on Wireless Communications*, 20(11):7257–7271, 2021.

[38] Tengxiao Liu, Zifeng Wang, Jin Miao, I Hsu, Jun Yan, Jiefeng Chen, Rujun Han, Fangyuan Xu, Yanfei Chen, Ke Jiang, et al. Budget-aware tool-use enables effective agent scaling. *arXiv preprint arXiv:2511.17006*, 2025.

[39] Quyuan Luo, Shihong Hu, Changle Li, Guanghui Li, and Weisong Shi. Resource scheduling in edge computing: A survey. *IEEE Communications Surveys and Tutorials*, 23(4):2131–2165, 2021.

[40] H Brendan McMahan. A survey of algorithms and analysis for adaptive online learning. *Journal of Machine Learning Research*, 18(90):1–50, 2017.

[41] Aranyak Mehta et al. Online matching and ad allocation. *Foundations and Trends® in Theoretical Computer Science*, 8(4):265–368, 2013.

[42] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.

[43] Md Noor-A-Rahim, Zilong Liu, Haeyoung Lee, GG Md Nawaz Ali, Dirk Pesch, and Pei Xiao. A survey on resource allocation in vehicular networks. *IEEE transactions on intelligent transportation systems*, 23(2):701–721, 2020.

[44] Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

[45] Hossein Pishro-Nik. *Introduction to probability, statistics, and random processes*. Kappa Research, LLC, Blue Bell, PA, USA, 2014.

[46] Daler Rakhmatov and Sarma Vrudhula. Energy management for battery-powered embedded systems. *ACM Transactions on Embedded Computing Systems (TECS)*, 2(3):277–324, 2003.

[47] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, 2019.

[48] Hong Ren, Cunhua Pan, Yansha Deng, Maged Elkashlan, and Arumugam Nallanathan. Resource allocation for secure urllc in mission-critical iot scenarios. *IEEE Transactions on Communications*, 68(9):5793–5807, 2020.

[49] Wenbo Ren, Xingyu Zhou, Jia Liu, and Ness B Shroff. Multi-armed bandits with local differential privacy. *arXiv preprint arXiv:2007.03121*, 2020.

[50] Alexandre Rio, Merwan Barlier, Igor Colin, and Albert Thomas. Differentially private deep model-based reinforcement learning. *arXiv preprint arXiv:2402.05525*, 2024.

[51] Roshan Shariff and Or Sheffet. Differentially private contextual linear bandits. *Advances in Neural Information Processing Systems*, 31, 2018.

[52] Bo Sun, Lin Yang, Mohammad Hajiesmaili, Adam Wierman, John CS Lui, Don Towsley, and Danny HK Tsang. The online knapsack problem with departures. In *Abstract Proceedings of the 2023 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, pages 59–60, 2023.

[53] Bo Sun, Ali Zeynali, Tongxin Li, Mohammad Hajiesmaili, Adam Wierman, and Danny HK Tsang. Competitive algorithms for the online multiple knapsack problem with application to electric vehicle charging. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 4(3):1–32, 2020.

[54] Aristide Tossou and Christos Dimitrakakis. Algorithms for differentially private multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[55] Shreshth Tuli, Fatemeh Mirhakimi, Samodha Pallewatta, Giuliano Zawad, Bahman Javadi, Feng Yan, Rajkumar Buyya, and Nicholas R. Jennings. Ai augmented edge and fog computing: Trends and challenges. *Journal of Network and Computer Applications*, 216:103648, 2023.

[56] Giuseppe Vietri, Borja Balle, Akshay Krishnamurthy, and Steven Wu. Private reinforcement learning with pac and regret guarantees. In *International Conference on Machine Learning*, pages 9754–9764. PMLR, 2020.

[57] Grant Wilkins, Srinivasan Keshav, and Richard Mortier. Offline energy-optimal llm serving: Workload-based energy models for llm inference on heterogeneous systems. *ACM SIGENERGY Energy Informatics Review*, 4(5):113–119, 2024.

[58] Zheng Xu, Yanxiang Zhang, Galen Andrew, Christopher A Choquette-Choo, Peter Kairouz, H Brendan McMahan, Jesse Rosenstock, and Yuanbo Zhang. Federated learning of gboard language models with differential privacy. *arXiv preprint arXiv:2305.18465*, 2023.

[59] Lichao Yang, Heli Zhang, Ming Li, Jun Guo, and Hong Ji. Mobile edge computing empowered energy efficient task offloading in 5g. *IEEE Transactions on Vehicular Technology*, 67(7):6398–6409, 2018.

[60] Xiang Yue, Huseyin Inan, Xuechen Li, Girish Kumar, Julia McAnallen, Hoda Shajari, Huan Sun, David Levitan, and Robert Sim. Synthetic text generation with differential privacy: A simple and practical recipe. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1321–1342, 2023.

[61] Yanwei Yue, Guibin Zhang, Boyang Liu, Guancheng Wan, Kun Wang, Dawei Cheng, and Yiyan Qi. Masrouter: Learning to route llms for multi-agent systems. *arXiv preprint arXiv:2502.11133*, 2025.

[62] Kai Zheng, Tianle Cai, Weiran Huang, Zhenguo Li, and Liwei Wang. Locally differentially private (contextual) bandits learning. *Advances in Neural Information Processing Systems*, 33:12300–12310, 2020.

[63] Xingyu Zhou. Differentially private reinforcement learning with linear function approximation. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(1):1–27, 2022.

[64] Xingyu Zhou and Komo Wei ZHANG. Locally private and robust multi-armed bandits. *Advances in Neural Information Processing Systems*, 37:7362–7399, 2024.

A Proofs and Discussions on Performance Limits of DP Definitions

A.1 Proof of Theorem 1

Lower bound of expected regret We prove the theorem by establishing an example below. Select a constant $K \geq 2$. At every round t, the action set is $\mathcal{X} = \{0, 1, \ldots, K\}$. Let S_t be a random variable uniformly distributed over $\{1, \ldots, K\}$ and $S_1, \cdots, S_T$ are independent and identically distributed. The reward and cost functions are defined as

$$f_t(x) = \mathbf{1}\{x = S_t\}, \qquad b_t(x) = b \cdot \mathbf{1}\{x = S_t\}, \tag{3}$$

where $b \geq \rho$ is a constant. The total budget is $B = \rho \cdot T$ with $\rho \in (0, 1)$. In the constructed example, the null action $0 \in \mathcal{X}$ has zero reward and cost, i.e. $f_t(0) = 0, b_t(0) = 0$. The rewards and costs are non-negative and upper bounded by one. Thus, the constructed example satisfies the considered problem setting in Section 4.

Let $n_T = \left\lfloor \frac{\rho T}{b} \right\rfloor$. A feasible policy chooses the rewarding action $x_t = S_t$ for the first n_T rounds and the null action $x_t = 0$ thereafter, obtaining total reward n_T. Since every unit of reward costs exactly b, any solution to the problem earns at most n_T units of reward. Consequently, for any context sequence $\boldsymbol{y}$ in the construction, the optimal reward is

$$OPT(\boldsymbol{y}) = n_T = \left\lfloor \frac{\rho T}{b} \right\rfloor. \tag{4}$$

Next, we bound the expected reward given an algorithm $\mathcal{A}$ with DP guarantee. Denote the probability that the algorithm $\mathcal{A}$ selects the $i-$th action ($i \in \{0, 1, \ldots, K\}$) given $S_t = j$ ($j \in \{1, \ldots, K\}$) at round t as

$$p_{t,i|j} = \Pr(x_t = i \mid S_t = j) = \mathbb{E}_{\Pr(S_{-t})}\left[\Pr(x_t = i \mid S_t = j, S_{-t})\right],$$

where these probabilities average over the algorithm's internal randomness and the contexts at all other rounds $S_{-t} = \{S_1, \cdots, S_{t-1}, S_{t+1}, S_T\}$. Due to the independence of the contexts across time, $\forall j \in \{1, \ldots, K\}, \Pr(S_{-t} \mid S_t = j) = \Pr(S_{-t})$. Given two sequences that differ only in user t's context $S_t = i$ and $S_t = j$, applying DP (Definition 3) to the observation of x_t and averaging over the context sequence S_{-t}, it holds that

$$p_{t,i|i} \leq e^{\epsilon} p_{t,i|j} + \delta, \qquad \forall i \neq j. \tag{5}$$

Summing (5) over all pairs (i, j) with $i, j \in \{1, \ldots, K\}$ and $i \neq j$, we have

$$\begin{aligned}
(K-1)\sum_{i=1}^{K} p_{t,i|i} &\leq e^{\epsilon} \sum_{j=1}^{K} \sum_{i=1, i\neq j}^{K} p_{t,i|j} + K(K-1)\delta \\
&\leq e^{\epsilon} \sum_{j=1}^{K} (1 - p_{t,j|j}) + K(K-1)\delta \\
&= e^{\epsilon} \left(K - \sum_{i=1}^{K} p_{t,i|i} \right) + K(K-1)\delta,
\end{aligned} \tag{6}$$

where the second inequality holds because $\sum_{i=1, i\neq j}^{K} p_{t,i|j} = 1 - p_{t,j|j} - p_{t,0|j} \leq 1 - p_{t,j|j}$. Since S_t is uniformly drawn from $\{1, \ldots, K\}$, following (6), we have

$$\mathbb{E}_{\mathcal{A},\boldsymbol{y}}[f_t(x_t)] = \frac{1}{K} \sum_{i=1}^{K} p_{t,i|i} \leq \frac{e^{\epsilon} + (K-1)\delta}{e^{\epsilon} + K - 1} \coloneqq q_K. \tag{7}$$

Thus, the total expected reward is bounded as

$$\mathbb{E}_{\mathcal{A},\boldsymbol{y}}[F_T^{\mathcal{A}}(\boldsymbol{y})] \leq q_K T. \tag{8}$$

Thus, the expected regret is lower bounded as

$$\mathbb{E}_{\mathcal{A},\boldsymbol{y}}\left[OPT(\boldsymbol{y}) - F_T^{\mathcal{A}}(\boldsymbol{y})\right] \geq \left(\frac{\rho}{b} - q_K\right) T - 1$$

Since $q_K = \frac{e^{\epsilon} + (K-1)\delta}{e^{\epsilon} + K - 1} = \delta + \frac{e^{\epsilon}(1-\delta)}{e^{\epsilon} + K - 1}$ and $K \geq 2$, it holds that $\delta \leq q_K < 1$, and $q_K \to \delta$ when the action set size approaches infinite ($K \to \infty$). Thus, there exist instances with $\frac{\rho}{b} > q_K$ such that the expected regret increases at least linearly, i.e. $\mathbb{E}_{\mathcal{A},\boldsymbol{y}}\left[OPT(\boldsymbol{y}) - F_T^{\mathcal{A}}(\boldsymbol{y})\right] \geq \kappa T$ given a fixed $\kappa > 0$. Thus, no algorithm can simultaneously guarantee DP and sublinear expected regret.

Upper bound of asymptomatic competitive ratio Next, we prove the upper bound of asymptomatic competitive ratio in probability by contradiction using the hard example in (3). If the there exists an algorithm with DP guarantee and a asymptomatic competitive ratio in probability $\frac{1}{\alpha} > \delta$, then by Definition 2, it holds for every $\kappa > 0$ that,

$$\lim_{T\to\infty} \sup_{\boldsymbol{y}\in\mathcal{Y}^T} \Pr_{\mathcal{A}} \left[\frac{\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y})}{T} > \kappa \right] = 0. \tag{9}$$

Since $\lim_{K\to\infty} q_K \to \delta < \frac{1}{\alpha}$ by the definition of q_K in (7), we can find instances with finite K such that $\alpha q_K < 1$. Choose the rewarding cost in the constructed instance as $b = \min\left\{1, \frac{2\rho}{1+\alpha q_K}\right\}$. This construction ensures that $\rho < b \leq 1$, $\frac{\rho}{b} \geq \frac{1+\alpha q_K}{2}$. Denote $\kappa' = \frac{1-\alpha q_K}{2} > 0$. Following (4) and (8), it holds for any $\alpha > 0$ that

$$\begin{aligned} \mathbb{E}_{\mathcal{A},\boldsymbol{y}}\left[OPT(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y})\right] &\geq \left\lfloor \frac{\rho T}{b} \right\rfloor - \alpha q_K T \\ &\geq \left(\frac{\rho}{b} - \alpha q_K\right) T - 1 \\ &\geq \frac{1-\alpha q_K}{2} T - 1 = \kappa' T - 1. \end{aligned} \tag{10}$$

Since rewards are nonnegative, we have $\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y}) \leq \mathrm{OPT}(\boldsymbol{y}) \leq T$ for every sequence $\boldsymbol{y}$. Therefore, for every $\kappa > 0$, it holds that

$$\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y}) \leq \kappa T + T \cdot \mathbf{1}\left\{\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y}) > \kappa T\right\}.$$

Taking expectations over the algorithm $\mathcal{A}$ and the sequence $\boldsymbol{y}$, we have

$$\kappa' T - 1 \leq \mathbb{E}_{\mathcal{A},\boldsymbol{y}}\left[OPT(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y})\right] \leq \kappa T + T \Pr_{\mathcal{A},\boldsymbol{y}}\left[\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y}) > \kappa T\right].$$

Choosing the tolerance in (9) as $\kappa = \kappa'/2$, we have

$$\begin{aligned} \lim_{T\to\infty} \sup_{\boldsymbol{y}\in\mathcal{Y}^T} \Pr_{\mathcal{A}} \left[\frac{\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y})}{T} > \frac{\kappa'}{2} \right] &\geq \lim_{T\to\infty} \Pr_{\mathcal{A},\boldsymbol{y}} \left[\frac{\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y})}{T} > \frac{\kappa'}{2} \right] \\ &= \lim_{T\to\infty} \Pr_{\mathcal{A},\boldsymbol{y}}\left[\mathrm{OPT}(\boldsymbol{y}) - \alpha F_T^{\mathcal{A}}(\boldsymbol{y}) > \kappa T\right] \geq \lim_{T\to\infty} (\frac{\kappa'}{2} - \frac{1}{T}) = \frac{\kappa'}{2} > 0, \end{aligned} \tag{11}$$

where the first inequality holds since the supremum over inputs is at least the average over their distribution.

Therefore, (11) contradicts with the definition of asymptomatic competitive ratio in probability of at least $1/\alpha$ in Definition 2, and the asymptomatic competitive ratio in probability of any algorithm is at most δ.

A.2 Proof of Theorem 2

Proof. We create a scenario to establish the upper bound of the competitive ratio as below. The considered scenario only has two rounds, and the action set for each round is $\mathcal{X} = \{0, 1\}$. We assume the budget consumption function is $b_t(x) = x$ and the total budget is $B = 1$. The scenario has two possible reward sequences which differ on the reward of the first round. The first sequence $\mathcal{S}_1$ is $f_1(x) = (1+k)\Delta\mathbb{I}(x)$, $f_2(x) = \Delta\mathbb{I}(x)$ and occurs with a probability of ρ. The second sequence $\mathcal{S}_2$ is $f_1 = 0$, $f_2(x) = \Delta\mathbb{I}(x)$ and occurs with a probability of $1 - \rho$.

Due to the budget constraint in the constructed scenario, the feasible action sequences includes $\boldsymbol{x}^{(1)} = \{x_1 = 1, x_2 = 0\}$, $\boldsymbol{x}^{(2)} = \{x_1 = 0, x_2 = 0\}$ and $\boldsymbol{x}^{(3)} = \{x_1 = 0, x_2 = 1\}$. We can calculate the competitive ratios of all three feasible action sequences. Let the probability of $\mathcal{S}_1$ be $\Pr(\mathcal{S}_1) = p$ and the probability of $\mathcal{S}_2$ be $\Pr(\mathcal{S}_2) = 1 - p$, the expected competitive ratio can be

expressed as

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{y}\sim \mathrm{Pr}_{1:T}}\left[\frac{F_T^{\mathcal{A}}(\boldsymbol{y})}{\mathrm{OPT}(\boldsymbol{y})}\right] &= p\left(\Pr(\boldsymbol{x}^{(1)} \mid \mathcal{S}_1) + \frac{1}{1+k}\Pr(\boldsymbol{x}^{(3)} \mid \mathcal{S}_1)\right) + (1-p)\Pr(\boldsymbol{x}^{(3)} \mid \mathcal{S}_2) \\
&= p\Pr(\boldsymbol{x}^{(1)} \mid \mathcal{S}_1) + \frac{p}{1+k}\Big(1 - \Pr(x_2 = 0 \mid \mathcal{S}_1)\Big) + (1-p)\Big(1 - \Pr(x_2 = 0 \mid \mathcal{S}_2)\Big) \\
&\le 1 - \frac{kp}{1+k} + \frac{kp}{1+k}\Pr(x_2 = 0 \mid \mathcal{S}_1) - (1-p)\Pr(x_2 = 0 \mid \mathcal{S}_2) \\
&\le 1 - \frac{kp}{1+k} + \frac{kp}{1+k}\Pr(x_2 = 0 \mid \mathcal{S}_1) - (1-p)\left(e^{-\epsilon}\Pr(x_2 = 0 \mid \mathcal{S}_1) - e^{-\epsilon}\delta\right) \\
&\le 1 - \frac{kp}{1+k} + (1-p)e^{-\epsilon}\delta,
\end{aligned} \tag{12}$$

where the first inequality holds because $\Pr(\boldsymbol{x}^{(1)} \mid \mathcal{S}_1) \le \Pr(x_2 = 0 \mid \mathcal{S}_1)$, the second inequality holds by the definition of $(\epsilon, \delta)-$JDP, and last inequality holds by choosing $p \le \frac{e^{-\epsilon}}{k/(1+k)+e^{-\epsilon}}$ and the fact $\Pr(x_2 = 0 \mid \mathcal{S}_1) \in [0, 1]$.

Choosing $p = \frac{e^{-\epsilon}}{k/(1+k)+e^{-\epsilon}}$ which minimizes the competitive ratio, the upper bound of the expected competitive ratio is $1 - \frac{1-\delta}{e^{\epsilon}+1+\frac{1}{k}}$, which approaches $1 - \frac{1-\delta}{e^{\epsilon}+1}$ when $k \to \infty$. Thus, $1 - \frac{1-\delta}{e^{\epsilon}+1}$ is the upper bound of competitive ratio in the worst case. □

A.3 Discussions on the Performance Limits of JDP or LDP

Theorem 2 provides a competitive ratio upper bound for any online allocation algorithm that satisfies $(\epsilon, \delta)-$JDP. The same upper bound also applies to any online allocation algorithm that satisfies $(\epsilon, \delta)-$LDP. The reason is that (ϵ, δ)-LDP implies (ϵ, δ)-JDP when a locally privacy-preserving algorithm is equivalently implemented by a centralized allocator, as formalized in Lemma 13. Consequently, online allocation under (ϵ, δ)-LDP cannot admits a better competitive ratio upper bound than (ϵ, δ)-JDP.

In fact, (ϵ, δ)-JDP (Definition 4) and (ϵ, δ)-LDP (Definition 5) provide privacy guarantees for central and local privacy-preserving models, respectively, while being more relaxed than standard (ϵ, δ)-DP. JDP and LDP allow online algorithms greater flexibility in selecting actions in response to users' contexts. Thus, while standard (ϵ, δ)-DP cannot guarantee any competitive ratio larger than δ which is close to 0, it is possible to guarantee meaningful performance bounds for online allocation with JDP or LDP. Interestingly, even as ϵ and δ approach zero, the competitive ratio upper bound remains $\frac{1}{2}$, which matches the competitive ratio upper bound of a greedy algorithm that selects action at each round by maximizing the reward [41].

Nevertheless, a fundamental tension between privacy preservation and allocation performance persists due to resource-constraint-based side channels: higher privacy levels can lead to lower competitive ratio upper bound (Theorem 2). This fundamental tension is illustrated by the JDP-preserving online allocation example in the proof of Theorem 2 in Section A.2. The example is understood as below. The optimal strategy under the first situation $\mathcal{S}_1$ is to select $x_1 = 1$. In contrast, under the second situation $\mathcal{S}_2$, the optimal strategy is to select $x_1 = 0$ and reserve the budget to select $x_2 = 1$. JDP requires protecting the reward information of f_1 even when the second action x_2 is observable. However, if an attacker observes $x_2 = 1$, they can infer that $x_1 = 0$ due to the budget constraint, which implies a high probability that the underlying situation is $\mathcal{S}_2$ (i.e., $f_1(x) = 0$). This privacy leakage arises from a resource-constraint-based side channel. To mitigate such leakage, a JDP-preserving algorithm is forced to adopt suboptimal strategies, thereby sacrificing allocation performance. This example establishes a competitive ratio upper bound for (ϵ, δ)-JDP preserving online allocation in Theorem 2.

B More Details about PPOA-FTRL

PPOA-FTRL is provided in Algorithm 3. PPOA-FTRL makes primal decisions in the same manner as Lines 2–4 of Algorithm 1. It then performs privacy-preserving updates of the remaining budget

and the dual variables. In the non-private setting, a centralized online learning method based on *Follow the Regularized Leader* (FTRL) [44] can be applied to update the dual variable by minimizing the regularized cumulative dual objective: $\mu_{t+1} = \arg\min_{\mu>0} \left(\sum_{\tau=1}^{t} w_\tau(\mu)\right) + \frac{1}{\eta}h(\mu)$, where $w_\tau(\mu) = (\rho - b_\tau(x_\tau))^\top \mu$ denotes the dual objective at round τ, and $h(\mu)$ is a 1-strongly convex regularization function. We find that both the FTRL dual update and the remaining budget update $B_{t+1} = B_1 - \sum_{\tau=1}^{t} b_\tau(x_\tau)$ depend on previous users' information through the cumulative budget consumption $\sum_{\tau=1}^{t} b_\tau(x_\tau)$. Thus, we privatize this consumption sum as s_t in Line 5 of Algorithm 1 using the tree-aggregation mechanism [32]. Substituting s_t for the clean cumulative consumption yields the privatized dual update

$$\mu_{t+1} = \arg\min_{\mu\geq 0}(\rho t - s_t)^\top \mu + \frac{1}{\eta}h(\mu), \tag{13}$$

and the privatized remaining budget is $\tilde{B}_{t+1} = B_1 - s_t$.

Algorithm 3 PPOA-FTRL

Require: Initialize virtual budget $\tilde{B}_1 = B_1 = \rho T$, dual variable μ_1, noise variance σ^2, step size $\eta > 0$.

1: Initialize a binary tree $\mathcal{T} \leftarrow \mathsf{InitializeTree}(T, \sigma^2)$.
2: **for** $t = 1$ to T **do**
3: Select an action by Line 2-4 in Algorithm 1.
4: Add new budget consumption to the tree: $\mathcal{T} \leftarrow \mathsf{AddToTree}(\mathcal{T}, t, b_t(x_t))$.
5: Get privatized sum of the consumption via Tree Aggregation: $s_t = \mathsf{GetSum}(\mathcal{T}, t)$.
6: Update remaining budget $\tilde{B}_{t+1} = B_1 - s_t$.
7: Update the dual variable by private FTRL in Eqn. (13).
8: **end for**

The tree-aggregation mechanism efficiently leverages clean historical information to compute a privatized cumulative budget consumption while avoiding redundant noise injection. In Line 1 of Algorithm 3, the function $\mathsf{InitializeTree}(T, \sigma^2)$ constructs a binary tree $\mathcal{T}$ with $2^{\lceil \log T \rceil}$ leaf nodes. Each node in the tree is independently initialized by sampling from a Gaussian distribution $\mathcal{N}(0, \sigma^2 \boldsymbol{I})$, which ensures that the value stored at every node of $\mathcal{T}$ is differentially private. After selecting the action x_t at round t, the function $\mathsf{AddToTree}(\mathcal{T}, t, b_t(x_t))$ adds the resource consumption $b_t(x_t)$ to all nodes along the path from the t-th leaf node to the root of $\mathcal{T}$. The privatized cumulative budget consumption up to round t is then obtained via the function $\mathsf{GetSum}(\mathcal{T}, t)$, which aggregates the minimum number of tree nodes required to represent the prefix sum. This design ensures that each update contributes to at most $\mathcal{O}(\lceil \log_2 t \rceil)$ noisy nodes.

The details of Tree-based Aggregation are provided below. The key functions of the tree-based aggregation scheme [32] are detailed below.

1. InitializeTree(T, σ^2): Initialize a binary tree $\mathcal{T}$ with $2^{\lceil \log T \rceil}$ leaf nodes. Each node in the tree is sampled i.i.d. from $\mathcal{N}(0, \sigma^2 \cdot \boldsymbol{I}_{m\times m})$.

2. AddToTree($\mathcal{T}, t, b_t$): Add the resource consumption $b_t(x_t)$ to the values of all the nodes on the path starting from the $t-$th leaf node to the root of $\mathcal{T}$.

3. GetSum($\mathcal{T}, t$): Let $\{u_1, u_2, \cdots, u_h\}$ be the nodes in the path from the root to the t-th leaf node. Return the private sum by the following procedure: (a) Initialize $s = 0$ and convert t into binary representation $[r_1, \cdots, r_h]$ with r_h being the smallest bit. (b) For each $j \in [h]$, if $r_j = 1$, add the value of the left sibling of node u_j to s. If u_j is the left sibling, add itself to s.

C Proofs for Theorems in Section 5

C.1 Proof of Theorem 3

Proof. We can denote the primal update in Line 2-4 of Algorithm 1 as a function $x_t = \pi_t(y_t, \mu_t, \tilde{B}_t)$. By the condition in Theorem 3, the mechanism to privatize the dual variables and remaining budgets $\{(\mu_t, \tilde{B}_t), t \in [T]\}$ are $(\epsilon, \delta)-$DP with respect to all the inputs $\{y_t, t \in [T]\}$. Thus, the conditions of Billboard Lemma 12 are satisfied. Thus, the composition of $\{x_1, \cdots, x_T\}$ preserves $(\epsilon, \delta)-$JDP.

Next, we prove that PPOA satisfies $(\epsilon, \delta)-$LDP. According to the definition of LDP (Definition 5), $(\epsilon, \delta)-$LDP holds if a local mechanism $\mathcal{M}(y)$, which maps local context to the information communicated with other users, satisfies the $(\epsilon, \delta)-$DP. In Algorithm 1, the local mechanisms are $\mathcal{M}_B$

and $\mathcal{M}_D$ which output the privatized remaining budget $\tilde{B}_{t+1}$ and the dual variable μ_{t+1}, respectively. Thus, if the composition of the remaining budgets $(\tilde{B}(1), \cdots, \tilde{B}(T))$ and dual variables $(\mu_1, \cdots, \mu_T)$ satisfies $(\epsilon, \delta)-$DP, Algorithm 1 satisfies $(\epsilon, \delta)-$LDP.

□

C.2 Proof of Theorem 4

C.2.1 Proof of constraint feasibility

By Condition (1) in Theorem 4, we have $B_t \geq \tilde{B}_t - B_{\text{gap},t}$. Thus, at each round t, by ensuring $b_t(x_t) \leq \tilde{B}_t - B_{\text{gap},t}$ with probability at least $1-\zeta$, Algorithm 1 always guarantees the constraint $\sum_{t=1}^{T} b_t(x_t) \leq B_1$ with probability at least $1-\zeta$ for any horizon T.

C.2.2 Proof for Adversarial Setting

Proof. Let τ_A be the first time slot when there exists a $j \in [m]$ such that $\bar{b}_j \geq \tilde{B}_{\tau_A+1,j} - B_{\text{gap},\tau_A+1}$, which implies that τ_A, typically termed as "stopping time", is a random variable depending on the randomness of the algorithm.

If $t \leq \tau_A$, since $x_t = \arg\max_{x\in\mathcal{X}}\{f_t(x) - \mu_t^\top b_t(x)\}$, we have for the optimal action x_t^*, $f_t(x_t^*) - \mu_t^\top b_t(x_t^*) \leq f_t(x_t) - \mu_t^\top b_t(x_t)$. Also, we have $0 \leq f_t(x_t) - \mu_t^\top b_t(x_t)$ because $f_t(0) = 0$ and $b_t(0) = 0$. Combining them, we also have the following inequality held.

$$\begin{aligned}
\alpha f_t(x_t) &= f_t(x_t) + (\alpha-1) f_t(x_t) \\
&\geq f_t(x_t) - \mu_t^\top b_t(x_t) + \mu_t^\top b_t(x_t) + (\alpha-1) f_t(x_t) \\
&\geq f_t(x_t^*) - \mu_t^\top b_t(x_t^*) + \mu_t^\top b_t(x_t) + (\alpha-1)\mu_t^\top b_t(x_t) \\
&= f_t(x_t^*) - \alpha\mu_t^\top(\rho - b_t(x_t)) - \mu_t^\top b_t(x_t^*) + \alpha\mu_t^\top\rho \\
&\geq f_t(x_t^*) - \alpha w_t(\mu_t),
\end{aligned} \tag{14}$$

where the last inequality holds because $\alpha = \sup_{j\in[m]} \frac{\bar{b}_j}{\rho_j}$ and so $\alpha\mu_t^\top\rho \geq \mu_t^\top\bar{\boldsymbol{b}} \geq \mu_t^\top b_t(x_t^*)$ where $\bar{\boldsymbol{b}} = [\bar{b}_1, \cdots, \bar{b}_m]$.

Given a stopping time τ_A, with probability at least $1-\zeta$, $\zeta \in (0,1)$, we can write the reward difference as

$$\begin{aligned}
&\sum_{t=1}^{T} f_t(x_t^*) - \alpha\sum_{t=1}^{T} f_t(x_t) \\
\leq &\sum_{t=1}^{T} f_t(x_t^*) - \alpha\sum_{t=1}^{\tau_A} f_t(x_t) \\
\leq &\sum_{t=1}^{T} f_t(x_t^*) - \sum_{t=1}^{\tau_A}(f_t(x_t^*) - \alpha w_t(\mu_t)) \\
\leq &(T-\tau_A)\bar{f} + \alpha\sum_{t=1}^{\tau_A} w_t(\mu) + \alpha\Gamma(T,\zeta)
\end{aligned} \tag{15}$$

where the second inequality holds by (14), the third inequality holds with probability at least $1-\zeta$ over the randomness of the algorithm for any $\mu \in \mathcal{D}$ by Condition (2) in Theorem 4.

If $\tau_A = T$, set $\mu = 0$. Then, we have with probability at least $1-\zeta$,

$$\sum_{t=1}^{T} f_t(x_t^*) - \alpha\sum_{t=1}^{T} f_t(x_t) \leq \alpha\Gamma(T,\zeta) \tag{16}$$

Otherwise, there exists a resource j such that $\bar{b}_j \geq \tilde{B}_{\tau_A+1,j} - B_{\text{gap},\tau_A+1}$. By the condition that $\forall t \in [T], \|\tilde{B}_t - B_t\| \leq B_{\text{gap},t} \leq \Gamma(T,\zeta)$, it holds that

$$\sum_{t=1}^{\tau_A} b_{t,j}(x_t) = B_j - B_{\tau_A+1,j} \geq B_j - \tilde{B}_{\tau_A+1,j} - B_{\text{gap},\tau_A+1} \geq T\rho_j - \bar{b}_j - 2\Gamma(T,\zeta) \tag{17}$$

Next we choose $\mu = \mu_j \mathbf{e}_j$ and we can bound the complementary slackness with probability at least $1-\zeta$ as

$$\begin{aligned}
&\alpha \sum_{t=1}^{\tau_A} w_t(\mu) = \alpha \sum_{t=1}^{\tau_A} \mu^\top (\rho - b_t(x_t)) \\
&= \alpha \sum_{t=1}^{\tau_A} \mu_j (\rho_j - b_{t,j}(x_t)) \\
&= \alpha\tau_A \mu_j \rho_j - \alpha\mu_j \sum_{t=1}^{\tau_A} b_{t,j}(x_t) \\
&\leq -\alpha(T-\tau_A)\mu_j\rho_j + \alpha\mu_j\bar{b}_j + 2\alpha\mu_j\Gamma(T,\zeta),
\end{aligned} \tag{18}$$

where the inequality holds by (17).

By choosing $\mu_j = \frac{\bar{f}}{\alpha\rho_j}$, we have

$$\alpha \sum_{t=1}^{\tau_A} w_t(\mu) \leq -(T-\tau_A)\bar{f} + \frac{\bar{b}_j}{\rho_j}\bar{f} + \frac{2\bar{f}}{\rho_j}\Gamma(T,\zeta) \tag{19}$$

Substituting the bound of $\alpha\sum_{t=1}^{\tau_A} w_t(\mu)$ back to (15), with probability of at least $1-\zeta, \zeta \in (0,1)$ over the randomness of the algorithm, it holds for any stopping time τ_A (either $\tau_A < T$ or $\tau_A = T$) that

$$\sum_{t=1}^{T} f_t(x_t^*) - \alpha \sum_{t=1}^{T} f_t(x_t) \leq \alpha\bar{f} + (\frac{2\bar{f}}{\hat{\rho}} + \alpha)\Gamma(T,\zeta), \tag{20}$$

where $\hat{\rho} = \min_{j\in[m]} \rho_j$.

Since the algorithm uses $\zeta = \zeta_T \in (0,1)$ satisfying $\lim_{T\to\infty}\zeta_T = 0, \lim_{T\to\infty}\frac{\Gamma(T,\zeta_T)}{T} = 0$, as $T \to \infty$, for any fixed sequence $\boldsymbol{y} \in \mathcal{Y}^T$, we have with probability of at least $(1-\zeta_T) \to 1$ over the randomness of the algorithm that

$$\frac{1}{T}(\sum_{t=1}^{T} f_t(x_t^*) - \alpha \sum_{t=1}^{T} f_t(x_t)) \leq \frac{\alpha\bar{f}}{T} + \frac{(\frac{2\bar{f}}{\hat{\rho}} + \alpha)\Gamma(T,\zeta_T)}{T} \to 0, \tag{21}$$

This establishes an asymptotic competitive ratio of at least $1/\alpha$ in probability, as defined in Definition 2. □

C.2.3 Proof for i.i.d. Stochastic Setting

Proof. All expectations below are over the i.i.d. contexts and the algorithm's internal randomness, unless stated otherwise. Define the expected dual objective as

$$\bar{D}(\mu \mid \mathcal{P}) = \frac{1}{T}\mathbb{E}[D(\mu \mid \boldsymbol{y})] = \mathbb{E}[f^*(\mu)] + \mu^\top\rho, \qquad f_t^*(\mu) = \max_{x\in\mathcal{X}_t}\{f_t(x) - \mu^\top b_t(x)\}.$$

Let $\xi_t = \{(y_1, v_1), \ldots, (y_t, v_t)\}$, where v_t collects the internal randomness used by the algorithm at round t, and let $\varsigma(\xi_t)$ denote the corresponding history. Since μ_t is $\varsigma(\xi_{t-1})$-measurable and y_t is independent of this history, it holds that $\mathbb{E}[f_t^*(\mu_t) \mid \varsigma(\xi_{t-1})] = \bar{D}(\mu_t \mid \mathcal{P}) - \mu_t^\top\rho$. The null action, nonnegative costs, and nonnegative dual variables ensure that $0 \leq f_t^*(\mu_t) \leq \bar{f}$.

Define the martingale

$$Z_t = \sum_{s=1}^{t} \left(\mathbb{E}[f_s^*(\mu_s) \mid \varsigma(\xi_{s-1})] - f_s^*(\mu_s)\right), Z_0 = 0.$$

Its increments are bounded in absolute value by $\bar{f}$. Use the stopping time $\tau_A \in \{0, \ldots, T\}$ from the adversarial proof: it is the first $t < T$ for which some resource j satisfies $\tilde{B}_{t+1,j} - B_{\mathrm{gap},t+1} \le \bar{b}_j$, and is T if no such time exists. Thus, $x_t = \hat{x}_t$ for every $t \le \tau_A$. Since $\tau_A \le T$, the optional stopping theorem gives $\mathbb{E}[Z_{\tau_A}] = 0, |Z_{\tau_A}| \le T\bar{f}$. For $t \le \tau_A$, the primal update satisfies

$$\begin{aligned} f_t(x_t) &= f_t^*(\mu_t) + \mu_t^\top b_t(x_t) \\ &= \mathbb{E}[f_t^*(\mu_t) \mid \varsigma(\xi_{t-1})] + \mu_t^\top \rho + \mu_t^\top (b_t(x_t) - \rho) + f_t^*(\mu_t) - \mathbb{E}[f_t^*(\mu_t) \mid \varsigma(\xi_{t-1})] \\ &= \bar{D}(\mu_t \mid \mathcal{P}) - w_t(\mu_t) - (Z_t - Z_{t-1}). \end{aligned} \tag{22}$$

By weak duality, for every fixed $\mu \in \mathbb{R}_+^m$, we have

$$\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] \le T\bar{D}(\mu \mid \mathcal{P}).$$

Because this inequality holds for every μ, the deterministic quantity $\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})]/T$ also lower bounds $\bar{D}(\mu_t \mid \mathcal{P})$ for every realized μ_t. Using nonnegative rewards after τ_A and $\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] \le T\bar{f}$, we obtain, for each realization,

$$\begin{aligned} &\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] - F_T^{\mathcal{A}}(\boldsymbol{y}) \\ \le &\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] - \sum_{t=1}^{\tau_A} \bar{D}(\mu_t \mid \mathcal{P}) + \sum_{t=1}^{\tau_A} w_t(\mu_t) + Z_{\tau_A} \\ \le &\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] - \tau_A \bar{D}(\bar{\mu}_A \mid \mathcal{P}) + \sum_{t=1}^{\tau_A} w_t(\mu_t) + Z_{\tau_A} \\ \le &\left(1 - \frac{\tau_A}{T}\right) \mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] + \sum_{t=1}^{\tau_A} w_t(\mu_t) + Z_{\tau_A} \\ \le &(T - \tau_A)\bar{f} + \sum_{t=1}^{\tau_A} w_t(\mu_t) + Z_{\tau_A}, \end{aligned} \tag{23}$$

where $\bar{\mu}_A = \frac{1}{\tau_A} \sum_{t=1}^{\tau_A} \mu_t$, the second inequality holds because the dual objective $\bar{D}(\bar{\mu}_A \mid \mathcal{P})$ is convex, the third inequality holds by weak duality, and the last inequality holds .

Let $\mathcal{E}$ be the event on which Condition (1) and Condition (2) of Theorem 4 both hold. The union bound gives $\Pr(\mathcal{E}^c) \le 2\zeta$: the bound holds for every fixed context sequence and therefore also after averaging over the i.i.d. contexts. On $\mathcal{E}$, Condition (2) and (23) imply that for every $\mu \in \mathcal{D}$,

$$\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] - F_T^{\mathcal{A}}(\boldsymbol{y}) \le (T - \tau_A)\bar{f} + \sum_{t=1}^{\tau_A} w_t(\mu) + \Gamma(T, \zeta) + Z_{\tau_A}. \tag{24}$$

If $\tau_A = T$, choose $\mu = \mathbf{0}$. Otherwise, there exists a resource j that satisfies $\bar{b}_j \ge \tilde{B}_{\tau_A+1,j} - B_{\mathrm{gap},\tau_A+1}$. Choose such a resource j and set $\mu = (\bar{f}/\rho_j)\mathbf{e}_j \in \mathcal{D}$. On $\mathcal{E}$, we have by Condition (1) and (2) that $B_{\tau_A+1,j} \le \tilde{B}_{\tau_A+1,j} + B_{\mathrm{gap},\tau_A+1} \le \bar{b}_j + 2\Gamma(T, \zeta)$, and hence it holds that

$$\sum_{t=1}^{\tau_A} b_{t,j}(x_t) = T\rho_j - B_{\tau_A+1,j} \ge T\rho_j - \bar{b}_j - 2\Gamma(T, \zeta).$$

It follows that

$$\begin{aligned} \sum_{t=1}^{\tau_A} w_t(\mu) &= \sum_{t=1}^{\tau_A} \mu^\top (\rho - b_t(x_t)) \\ &= \tau_A \mu_j \rho_j - \mu_j \sum_{t=1}^{\tau_A} b_{t,j}(x_t) \\ &\le -(T - \tau_A)\bar{f} + \frac{\bar{b}_j}{\rho_j}\bar{f} + \frac{2\bar{f}}{\rho_j}\Gamma(T, \zeta), \end{aligned} \tag{25}$$

where the inequality holds by the bound of $\sum_{t=1}^{\tau_A} b_{t,j}(x_t)$ and the choice of μ.

Combining both cases of τ_A in (24), we have on $\mathcal{E}$,

$$\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] - F_T^{\mathcal{A}}(\boldsymbol{y}) \le \alpha \bar{f} + \left(\frac{2\bar{f}}{\hat{\rho}} + 1\right) \Gamma(T, \zeta) + Z_{\tau_A}, \tag{26}$$

where $\alpha = \max_{j\in[m]} \bar{b}_j/\rho_j$ and $\hat{\rho} = \min_{j\in[m]} \rho_j > 0$.

On $\mathcal{E}^c$, the difference between expected reward and the algorithm reward is unconditionally upper bounded due to the nonnegative rewards: $\mathbb{E}[\mathrm{OPT}(\boldsymbol{y})] - F_T^{\mathcal{A}}(\boldsymbol{y}) \le T\bar{f}$. Taking expectations over the event $\mathcal{E}$ and $\mathcal{E}^c$, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{OPT}(\boldsymbol{y}) - F_T^{\mathcal{A}}(\boldsymbol{y})\right] \\
\le& \alpha\bar{f} + \left(\frac{2\bar{f}}{\hat{\rho}} + 1\right) \Gamma(T, \zeta) + \mathbb{E}[Z_{\tau_A}\mathbf{1}_{\mathcal{E}}] + T\bar{f}\,\Pr(\mathcal{E}^c) \\
=& \alpha\bar{f} + \left(\frac{2\bar{f}}{\hat{\rho}} + 1\right) \Gamma(T, \zeta) - \mathbb{E}[Z_{\tau_A}\mathbf{1}_{\mathcal{E}^c}] + T\bar{f}\,\Pr(\mathcal{E}^c) \\
\le& \alpha\bar{f} + \left(\frac{2\bar{f}}{\hat{\rho}} + 1\right) \Gamma(T, \zeta) + 2T\bar{f}\zeta,
\end{aligned} \tag{27}$$

where the first equality holds because $\mathbb{E}[Z_{\tau_A}] = \mathbb{E}[Z_{\tau_A}\mathbf{1}_{\mathcal{E}}] + \mathbb{E}[Z_{\tau_A}\mathbf{1}_{\mathcal{E}^c}] = 0$, and the last inequality holds by $|Z_{\tau_A}| \le T\bar{f}$.

Finally, configure the algorithm with $\zeta = \zeta_T$ as in Condition (2). Since $\zeta_T \to 0$, $T\zeta_T = O(1)$ and $\Gamma(T, \zeta_T)/T \to 0$, it follows by (27) that

$$\mathbb{E}\left[\mathrm{OPT}(\boldsymbol{y}) - F_T^{\mathcal{A}}(\boldsymbol{y})\right] \le \alpha\bar{f} + \left(\frac{2\bar{f}}{\hat{\rho}} + 1\right) \Gamma(T, \zeta) + 2T\bar{f}\zeta_T = o(T),$$

thus establishing the sublinear expected regret. □

C.2.4 Discussion on the Extension to the Non-stationary Setting

A more general stochastic input model is the non-stationary setting, in which the context distribution varies over time, i.e. the input distributions in $\mathcal{P}_{1:T} = \{\mathcal{P}_1, \ldots, \mathcal{P}_T\}$ are not necessarily identical. One way to quantify the degree of non-stationarity is through the mean total variance deviation

$$\mathrm{TV}(\mathcal{P}_{1:T}) = \sum_{t=1}^{T} \mathbb{D}_{\mathrm{TV}}\left(\mathcal{P}_t,\, \bar{\mathcal{P}}\right), \qquad \bar{\mathcal{P}} := \frac{1}{T}\sum_{t=1}^{T} \mathcal{P}_t, \tag{28}$$

where $\mathbb{D}_{\mathrm{TV}}$ denotes the total variation distance between two distributions. A larger value of $\mathrm{TV}(\mathcal{P}_{1:T})$ indicates a more rapidly changing context distribution, making it harder for an online algorithm to track the distribution shift.

Without additional assumptions, [7] show that online allocation algorithms can incur linear expected regret in the non-stationary setting even in the absence of privacy constraints. Therefore, under the considered DP notions, one may not expect sublinear regret either.

Nevertheless, the analysis for the i.i.d. stochastic setting provides a useful starting point for extending performance bounds to non-stationary stochastic settings by adapting arguments similar to those in Theorem 3 of [7]. In particular, their analysis shows that non-stationarity contributes an additional term in the expected regret that scales with $\mathrm{TV}(\mathcal{P}_{1:T})$.

D Proofs for Theorems in Section 6.1

D.1 Proof of Theorem 5

Proof. The Global Sensitivity of the consumption function is bounded as

$$\Delta_b = \sup_{\mathcal{D}_t, \mathcal{D}_t'} \|b_t(\mathcal{D}_t) - b_t'(\mathcal{D}_t')\|_2 = \sup_{b_t, b_t', x_t, x_t'} \|b_t(x_t) - b_t'(x_t')\|_2 \le 2\bar{b}, \tag{29}$$

where the inequality holds by triangle inequality. Thus, by Gaussian mechanism for RDP in Lemma 14, the consumption mechanism $\tilde{b}_t(x_t) = b_t(x_t) + n_t$ satisfies $(\alpha, \frac{2\alpha\bar{b}^2}{\sigma^2})$-RDP. By Lemma 17, the composition $(\tilde{b}_1, \cdots, \tilde{b}_{t+1})$ satisfies $(\alpha, \frac{2\alpha\bar{b}^2}{\sigma^2})$-RDP.

Since the budget dynamic and the dual mirror descent are all post processing, the composition of dual variables $(\mu_1, \cdots, \mu_T)$ and remaining budgets $(\tilde{B}_1, \cdots, \tilde{B}_T)$ both satisfy $(\alpha, \frac{2\alpha\bar{b}^2}{\sigma^2})$-RDP. By Lemma 15, the composition of variables $((\mu_1, \tilde{B}_1), \cdots, (\mu_T, \tilde{B}_T))$ satisfies $(\frac{2\alpha\bar{b}^2}{\sigma^2} + \frac{\ln(1/\delta)}{\alpha-1}, \delta)$-DP with $\delta \in (0,1)$ and $\alpha > 1$. Optimally choosing $\alpha = 1 + \sqrt{\frac{\ln(1/\delta)\sigma^2}{2\bar{b}^2}}$, the composition satisfies $(\frac{2\bar{b}^2}{\sigma^2} + 2\sqrt{\frac{2\bar{b}^2}{\sigma^2}\ln(1/\delta)}, \delta)$-DP. Choosing $\sigma = \frac{\sqrt{2}\bar{b}(\sqrt{\ln(1/\delta)} + \sqrt{\ln(1/\delta)+\epsilon})}{\epsilon}$, we can guarantee $(\epsilon, \delta)-$DP, which proves the condition in Theorem 3. □

D.2 Proof of Theorem 6

We prove the two conditions in Theorem 4 are satisfied by PPOA-DMD.

D.2.1 Proof of Condition (1)

Proof. The true remaining budget at round t can be expressed as $B_t = B_1 - \sum_{s=1}^{t-1} b_s(x_s)$ and the corrupted remaining budget is expressed as $\tilde{B}_t = B_1 - \sum_{s=1}^{t-1} \tilde{b}_s(x_s) = B_1 - \sum_{s=1}^{t-1}(b_s(x_s) + n_s)$.

Let $v_t = \sum_{i=1}^{t} n_i$. Observing that $\{n_s\}_{s\in[t]} \in \mathcal{N}(0, \sigma^2)$ are independent Gaussian random variables, we have $v_t \sim \mathcal{N}(0, t\sigma^2 \boldsymbol{I})$. By Lemma 18 and union bounds, given a resource $j \in [m]$, it holds $\forall t \in [T]$ with probability at least $1-\zeta$ that,

$$|v_{t,j}| \leq \sigma\sqrt{2t\ln\frac{2T}{\zeta}}, \tag{30}$$

By union bound, it holds with probability at least $1-\zeta$ that

$$\|B_t - \tilde{B}_t\|_\infty = \|v_t\|_\infty \leq \sigma\sqrt{2t\ln\frac{2mT}{\zeta}} = B_{\text{gap},t}. \tag{31}$$

□

D.2.2 Proof of Condition (2)

To prove Condition (2), we first prove the convergence of projected mirror descent in Lemma 9 and the concentration of $\sum_{t=1}^{T} \mu_t^\top n_t$ given a norm-bounded μ_t in Lemma 10.

Lemma 9. *If μ_t is updated as $\mu_{t+1} = \arg\min_{\mu:\|\mu\|_2\leq\mu_{\max}} \tilde{g}_t^\top \mu + \frac{1}{\eta}V_h(\mu, \mu_t)$ where $\tilde{g}_t = g_t + n_t, n_t \sim \mathcal{N}(0, \sigma\boldsymbol{I})$ and $\|g_t\|_2 \leq G$ for T rounds, then for any $\mu : \|\mu\|_2 \leq \mu_{\max}$, it holds for $\tilde{w}_t(\mu_t) = \tilde{g}_t^\top \mu_t$ with probability at least $1-\zeta, \zeta \in (0,1)$ that*

$$\sum_{t=1}^{T} \tilde{w}_t(\mu_t) - \tilde{w}_t(\mu) \leq \eta\left(G^2 + m\sigma^2\right)T + \frac{1}{\eta}V_h(\mu, \mu_1) + \eta\sigma^2\sqrt{mT\ln(1/\zeta)} + \eta\sigma^2\ln(1/\zeta). \tag{32}$$

Proof. Since $\sum_{t=1}^{T}\|n_t/\sigma\|^2$ is a χ^2-distribution with mT degrees of freedom, by Laurent-Massart bound (Lemma 20), we have $\Pr(\sum_{t=1}^{T}\|n_t/\sigma\|^2 \geq mT + 2\sqrt{mTx} + 2x) \leq \exp(-x)$ and further implies with probability at least $1-\zeta$, we have $\|n_t\|^2 \leq mT\sigma^2 + 2\sigma^2\sqrt{mT\ln(1/\zeta)} + 2\sigma^2\ln(1/\zeta)$. Thus, It holds with probability at least $1-\zeta$ for the perturbed sug-gradient that

$$\begin{aligned}\sum_{t=1}^{T}\|\tilde{g}_t\|_2^2 = \sum_{t=1}^{T}\|g_t + n_t\|_2^2 &\leq 2\sum_{t=1}^{T}\left(\|g_t\|_2^2 + \|n_t\|_2^2\right) \\ &\leq 2T\left(G^2 + m\sigma^2\right) + 4\sigma^2\sqrt{mT\ln(1/\zeta)} + 2\sigma^2\ln(1/\zeta)\end{aligned} \tag{33}$$

Replace g_t in Lemma 21 with $\tilde{g}_t$, it holds with probability at least $1-\zeta$ that

$$\begin{aligned}&\sum_{t=1}^{T} \tilde{w}_t(\mu_t) - \tilde{w}_t(\mu) \\ \leq &\eta T\left(G^2 + m\sigma^2\right) + \frac{1}{\eta} V_h(\mu, \mu_1) + 2\eta\sigma^2\sqrt{mT\ln(1/\zeta)} + \eta\sigma^2\ln(1/\zeta)\end{aligned} \tag{34}$$

□

Lemma 10. *If $\forall t \in [T]$, $\|\mu_t\|_2 \leq \mu_{\max}$. It holds with probability at least $1-\zeta, \zeta \in (0,1)$ that*

$$\sum_{t=1}^{T} \mu_t^\top n_t \geq -\mu_{\max}\sigma\sqrt{2T\ln(1/\zeta)} \tag{35}$$

Proof. Define a sequence $\xi_t = \{(y_1, n_1), \cdots, (y_t, n_t)\}$ and let $\mathcal{F}_t$ be the sigma-algebra generated by ξ_t, then $\{\mathcal{F}_t\}_{t\in[T]}$ is a Filtration. Therefore, μ_t is $\mathcal{F}_{t-1}$-measurable and $n_t \sim \mathcal{N}(0, \sigma^2 I)$ are independent across t. Then we have

$$\mathbb{E}[\mu_t^\top n_t \mid \mathcal{F}_{t-1}] = \mu_t^\top \mathbb{E}[n_t \mid \mathcal{F}_{t-1}] = 0.$$

Thus, conditional on $\mathcal{F}_{t-1}$, $\mu_t^\top n_t$ is Gaussian with mean 0 and variance

$$\mathrm{Var}(\mu_t^\top n_t \mid \mathcal{F}_{t-1}) = \sigma^2\|\mu_t\|_2^2 \leq \sigma^2\mu_{\max}^2.$$

Hence, by Gaussian Moment Generation function, for every $\lambda \in \mathbb{R}$,

$$\mathbb{E}\Big[e^{\lambda\mu_t^\top n_t} \mid \mathcal{F}_{t-1}\Big] = \exp\left(\frac{\lambda^2\sigma^2\|\mu_t\|_2^2}{2}\right) \leq \exp\left(\frac{\lambda^2\sigma^2\mu_{\max}^2}{2}\right).$$

Let $W_t = \sum_{s=1}^{t} \mu_s^\top n_s$. It holds for any $t \in [T]$ that

$$\mathbb{E}\big[e^{\lambda W_t}\big] = \mathbb{E}\Big[e^{\lambda W_{t-1}}\mathbb{E}\Big[e^{\lambda\mu_t^\top n_t} \mid \mathcal{F}_{t-1}\Big]\Big] \leq \exp\left(\frac{\lambda^2\sigma^2\mu_{\max}^2}{2}\right)\mathbb{E}\big[e^{\lambda W_{t-1}}\big] \tag{36}$$

Applying (36) iteratively given $W_0 = 0$, we have

$$\mathbb{E}\big[e^{\lambda W_T}\big] \leq \exp\left(\frac{\lambda^2 T\sigma^2\mu_{\max}^2}{2}\right).$$

Therefore, for any $u > 0$ and any $\lambda > 0$, Chernoff's bound yields

$$\Pr(W_T \leq -u) \leq e^{-\lambda u}\mathbb{E}\big[e^{-\lambda W_T}\big] \leq \exp\left(-\lambda u + \frac{\lambda^2 T\sigma^2\mu_{\max}^2}{2}\right).$$

Optimizing over λ at $\lambda = \frac{u}{T\sigma^2\mu_{\max}^2}$ gives

$$\Pr(W_T \leq -u) \leq \exp\left(-\frac{u^2}{2T\sigma^2\mu_{\max}^2}\right).$$

Taking $u = \sigma\mu_{\max}\sqrt{2T\ln(1/\zeta)}$ shows that, with probability at least $1-\zeta$,

$$\sum_{t=1}^{T} \mu_t^\top n_t \geq -\sigma\mu_{\max}\sqrt{2T\ln(1/\zeta)}.$$

□

Proof of Theorem 6

Proof. We are ready to prove the convergence bound of the projected private DMD in Algorithm 2. Denote the privatized dual objective as $\tilde{w}_t(\mu) = \tilde{g}_t^\top \mu$ where $\tilde{g}_t = \rho - b_t(x_t) + n_t$ and $n_t \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Let $\mu_1, \cdots, \mu_T$ be the dual variables given by the projected private DMD. The clean dual objective can be expressed as

$$\begin{aligned}
&\sum_{t=1}^{T}[w_t(\mu_t) - w_t(\mu)] \\
=&\sum_{t=1}^{T}[\tilde{w}_t(\mu_t) - \tilde{w}_t(\mu)] - \sum_{t=1}^{T} n_t\mu_t + \sum_{t=1}^{T} n_t\mu \\
\leq&\eta\left(G^2 + m\sigma^2\right)T + \frac{1}{\eta}V_h(\mu, \mu_1) + 2\eta\sigma^2\sqrt{mT\ln(3/\zeta)} + \eta\sigma^2\ln(3/\zeta) \\
&+ \mu_{\max}\sigma\sqrt{2T\ln(3/\zeta)} + \mu\sigma\sqrt{2T\ln\frac{6mT}{\zeta}}
\end{aligned} \tag{37}$$

where the inequality holds by Lemma 9, Lemma 10, and (31) and union bounds. Choosing $\eta = \sqrt{\frac{V_h(\mu,\mu_1)}{T(G^2+m\sigma^2)}}$, it holds that

$$\begin{aligned}
&\sum_{t=1}^{T}[w_t(\mu_t) - w_t(\mu)] \\
\leq&2\sqrt{V_h(\mu, \mu_1)\left(G^2 + m\sigma^2\right)T} + \mu_{\max}\sigma\sqrt{2T\ln(3/\zeta)} \\
&+ \mu\sigma\sqrt{2T\ln\frac{6mT}{\zeta}} + \sqrt{V_h(\mu, \mu_1)}\left(2\sigma\sqrt{m\ln(3/\zeta)} + \sigma\frac{\ln(3/\zeta)}{\sqrt{T}}\right) \\
=&\mathcal{O}(\sqrt{T\ln(T/\zeta)}\sigma\mu_{\max}) = \mathcal{O}(\sqrt{T\ln(T/\zeta)}\mu_{\max}\frac{\sqrt{\log(1/\delta)}}{\epsilon}),
\end{aligned} \tag{38}$$

where the last equality holds by the choice of $\sigma = \frac{\sqrt{2\bar{b}}(\sqrt{\ln(1/\delta)}+\sqrt{\ln(1/\delta)+\epsilon})}{\epsilon}$.

We define $\Gamma(T, \zeta) = \mathcal{O}(\sqrt{T\ln(T/\zeta)}\mu_{\max}\frac{\sqrt{\log(1/\delta)}}{\epsilon})$ By (31) in the proof of condition (1), we have

$$B_{\mathrm{gap},T} = \sigma\sqrt{2T\ln\frac{2mT}{\zeta}} \leq \mathcal{O}(\sqrt{T\ln(T/\zeta)}\mu_{\max}\frac{\sqrt{\log(1/\delta)}}{\epsilon}),$$

thus satisfying the requirement of Condition (2) on $B_{\mathrm{gap},T}$.

Moreover, by choosing $\zeta = \zeta_T = \frac{1}{T^\vartheta+1}$ given a constant $\vartheta \geq 1$, we have

$$\Gamma(T, \zeta_T) = \mathcal{O}(\sqrt{T\ln(T)}\mu_{\max}\frac{\sqrt{\log(1/\delta)}}{\epsilon}) = \widetilde{\mathcal{O}}(\sqrt{T}\mu_{\max}\frac{\sqrt{\log(1/\delta)}}{\epsilon}).$$

Therefore, conditions in (2) of Theorem 4 are satisfied.

□

E Proofs for Theorems in Section 6.2

E.1 Proof of Theorem 7

Proof. By Billboard Lemma (Lemma 12), if $(s_1, \cdots, s_T)$ satisfies $(\epsilon, \delta)-$ differential privacy, then the composition $(\mu_1, \cdots, \mu_T)$ satisfies $(\epsilon, \delta)-$ differential privacy, and the composition $(x_1, \cdots, x_T)$ satisfies $(\epsilon, \delta)-$JDP. Thus, we only need to prove the condition that $(s_1, \cdots, s_T)$ satisfies $(\epsilon, \delta)-$ differential privacy.

The Global Sensitivity of the consumption function is bounded as

$$\Delta_b = \sup_{\mathcal{D}_t, \mathcal{D}_t'} \|b_t(\mathcal{D}_t) - b_t'(\mathcal{D}_t')\|_2 = \sup_{b_t, b_t', x_t, x_t'} \|b_t(x_t) - b_t'(x_t')\|_2 \leq 2\bar{b}, \tag{39}$$

where the inequality holds by triangle inequality. Thus, by the tree initialization procedure in Algorithm 3, each node in the tree satisfies $(\alpha, \frac{2\alpha\bar{b}^2}{\sigma^2})-$RDP by Lemma 14. Since each leaf node affects at most $\lceil \log_2(T) \rceil$ nodes which locate at the path from the leaf node to the root of the tree, by Lemma 16 (Proposition 1 in [42]), the composition of all the nodes in the tree $\mathcal{T}$ satisfies $(\alpha, \frac{2\alpha\lceil \log_2(T) \rceil \bar{b}^2}{\sigma^2})-$RDP. Since $(s_1, \cdots, s_T)$ is obtained by the post-processing on the tree nodes, it also satisfies $(\alpha, \frac{2\alpha\lceil \log_2(T) \rceil \bar{b}^2}{\sigma^2})-$RDP. This implies $(s_1, \cdots, s_T)$ satisfies $(\frac{2\alpha\lceil \log_2(T) \rceil \bar{b}^2}{\sigma^2} + \frac{\ln(1/\delta)}{\alpha-1}, \delta)$-DP with $\delta \in (0,1)$ and $\alpha > 1$ and $(x_1, \cdots, x_T)$ satisfies $(\frac{2\alpha\lceil \log_2(T) \rceil \bar{b}^2}{\sigma^2} + \frac{\ln(1/\delta)}{\alpha-1}, \delta)$-JDP by Billboard Lemma 12. Optimally choosing $\alpha = 1 + \sqrt{\frac{\ln(1/\delta)\sigma^2}{2\lceil \log_2(T) \rceil \bar{b}^2}}$, the composition $(x_1, \cdots, x_T)$ satisfies $(\frac{2\lceil \log_2(T) \rceil \bar{b}^2}{\sigma^2} + 2\sqrt{\frac{2\lceil \log_2(T) \rceil \bar{b}^2}{\sigma^2} \ln(1/\delta)}, \delta)$-JDP. Choosing $\sigma = \frac{\sqrt{2}\lceil \log_2(T) \rceil \bar{b}(\sqrt{\ln(1/\delta)} + \sqrt{\ln(1/\delta)+\epsilon})}{\epsilon}$, we can guarantee $(\epsilon, \delta)-$JDP. □

E.2 Proof of Theorem 8

Theorem 8 holds if the conditions in Theorem 4 are satisfied, so we separately prove the two conditions.

E.2.1 Proof of Condition (1)

Proof. The clean remaining budget at round t can be expressed as $B_t = B_1 - \sum_{t=1}^{T} b_t(x_t)$. Let $\mathcal{M}_t$ be the set of nodes that need to be aggregated by the $\mathrm{GetSum}(\mathcal{T}, t)$ function in the tree-based aggregation scheme. We have the number of aggregated nodes is $|\mathcal{M}_t| \leq \lceil \log_2(t) \rceil$. Thus, the privatized remaining budget is expressed as $\tilde{B}_t = B_1 - s_t = B_1 - \sum_{\tau=1}^{t-1}(b_\tau(x_\tau)) - \sum_{i \in \mathcal{M}_t} n_i = B_t - \sum_{i \in \mathcal{M}_t} n_i$ where n_i is the noise added to the tree node i.

Let $v_t = \sum_{i \in \mathcal{M}_t} n_i$.Observing that the noise added to each node in $\mathcal{M}_t$ is an independent Gaussian variable $\mathcal{N}(0, \sigma^2 \boldsymbol{I})$, we have $v_t \sim \mathcal{N}(0, \log_2(t)\sigma^2 \boldsymbol{I})$. Given any $j \in [m]$, by Lemma 18 and union bounds, given a resource $j \in [m]$, it holds $\forall t \in [T]$ with probability at least $1 - \zeta$ that,

$$|v_{t,j}| \leq \sigma \sqrt{2 \lceil \log_2(t) \rceil \ln \frac{2T}{\zeta}}, \tag{40}$$

By union bound, it holds with probability at least $1 - \zeta$ that

$$\|B_t - \tilde{B}_t\|_\infty = \|v_t\|_\infty \leq \sigma \sqrt{2 \lceil \log_2(t) \rceil \ln \frac{2mT}{\zeta}} = B_{\mathrm{gap},t}. \tag{41}$$

□

E.2.2 Proof of Condition (2)

Proof. We first prove that if μ_t is updated by perturbed FTRL as $\mu_{t+1} = \arg\min_\mu (\sum_{\tau=1}^{t} g_\tau)^\top \mu + v_t^\top \mu + \frac{1}{\eta} h(\mu)$ where $\forall t \in [T], \|g_t\| \leq G$, then for any μ, it holds for $w_t(\mu_t) = g_t^\top \mu_t$ that

$$\sum_{t=1}^{T} w_t(\mu_t) - w_t(\mu) \leq \eta(\frac{1}{2} G^2 T + G \sum_{t=1}^{T} \|v_t\|) + \frac{1}{\eta}(h(\mu) - \min_{\beta > 0} h(\beta)), \tag{42}$$

Denote the non-perturbed FTRL update as $\mu'_{t+1} = \arg\min_\mu (\sum_{\tau=1}^{t} g_\tau)^\top \mu + \frac{1}{\eta} h(\mu)$. Applying Lemma 23 with $\phi_1(\mu) = \eta(\sum_{\tau=1}^{t} g_\tau)^\top \mu + h(\mu)$ and $\phi_2(\mu) = \eta(\sum_{\tau=1}^{t} g_\tau)^\top \mu + h(\mu) + \eta v_t^\top \mu$, we have

$$\|\mu_{t+1} - \mu'_{t+1}\| \leq \eta \|v_t\|. \tag{43}$$

Thus, we have

$$\begin{aligned}\sum_{t=1}^{T} w_t(\mu_t) - w_t(\mu) = &\sum_{t=1}^{T}[g_t^\top \mu_t - g_t^\top \mu_t' + g_t^\top \mu_t' - g_t^\top \mu] \\ \leq &\eta G \sum_{t=1}^{T} \|v_t\| + \frac{\eta}{2} G^2 T + \frac{1}{\eta}(h(\beta) - \min_{\nu>0} h(\beta)),\end{aligned}$$

where the inequality holds by (43), Cauchy-Schwartz inequality and Lemma 22.

In Algorithm 3, we have $g_t = \rho - b_t(x_t)$, so $G = \max\{\rho, \bar{b} - \rho\}$. Also, we have $v_t = \sum_{i \in \mathcal{M}_t} n_i$ in Algorithm 3 where $\mathcal{M}_t$ is the set of nodes that need to be aggregated by the $\mathrm{GetSum}(\mathcal{T}, t)$ function and $|\mathcal{M}_t| \leq \lceil \log_2(t) \rceil$.

By (41), we have with probability at least $1 - \zeta$ that $\forall t \in [T]$, $\|v_t\|_2 \leq \sqrt{m}\|v_t\|_\infty \leq \sigma\sqrt{2m \lceil \log_2(T) \rceil \ln(\frac{2mT}{\zeta})}$. Thus, it holds with probability at least $1 - \zeta$ that

$$\sum_{t=1}^{T} w_t(\mu_t) - w_t(\mu) \leq \eta T(\frac{G^2}{2} + G\sigma\sqrt{2m \lceil \ln(T) \rceil \ln(\frac{2mT}{\zeta})}) + \frac{1}{\eta}(h(\mu) - \min_{\beta>0} h(\beta)). \tag{44}$$

By choosing $\eta = \sqrt{\frac{h(\mu) - \min_{\beta>0} h(\beta)}{T(\frac{G^2}{2} + G\sigma\sqrt{2m \lceil \log_2(T) \rceil \ln(\frac{2mT}{\zeta})})}}$, it holds with probability at least $1-\zeta, \zeta \in (0,1)$ that

$$\begin{aligned}&\sum_{t=1}^{T} w_t(\mu_t) - \sum_{t=1}^{T} w_t(\mu) \\ &\quad \leq \sqrt{T(\frac{G^2}{2} + G\sigma\sqrt{2m \lceil \log_2(T) \rceil \ln(\frac{2mT}{\zeta})})(h(\mu) - \min_{\beta>0} h(\beta))} = \widetilde{\mathcal{O}}(\frac{\sqrt{T\sqrt{\log(1/\delta)}}}{\sqrt{\epsilon}}),\end{aligned} \tag{45}$$

where $G = \max\{\rho, \bar{b} - \rho\}$, and $\mu \in \{0, \frac{\bar{f}}{\alpha\rho_1}\mathbf{e}_1, \cdots, \frac{\bar{f}}{\alpha\rho_m}\mathbf{e}_m\}$ where m is the number of constraints, $\mathbf{e}_j \in \{0,1\}^m$ with 1 at $j-$th entry and 0 at the other entries.

□

By (41) in the proof of Condition (1), to ensure $B_{\mathrm{gap},t} = \sigma\sqrt{2 \lceil \log_2(t) \rceil \ln \frac{2mT}{\zeta}} \leq \Gamma(T, \zeta)$ and the dual convergence bound in Condition (2), we can choose

$$\begin{aligned}\Gamma(T,\zeta) = &\sqrt{T(\frac{G^2}{2} + G\sigma\sqrt{2m \lceil \log_2(T) \rceil \ln(\frac{2mT}{\zeta})})(h(\mu) - \min_{\beta>0} h(\beta))} \\ &+ \sigma\sqrt{2 \lceil \log_2(T) \rceil \ln \frac{2mT}{\zeta}},\end{aligned} \tag{46}$$

where we set $\sigma = \frac{\sqrt{2}\bar{b}\sqrt{\lceil \log_2(T) \rceil}(\sqrt{\ln(1/\delta)} + \sqrt{\ln(1/\delta) + \epsilon})}{\epsilon}$ to preserve (ϵ, δ)-JDP.

Moreover, by choosing $\zeta = \zeta_T = \frac{1}{T^\vartheta + 1}$ given a constant $\vartheta \geq 1$, we have

$$\Gamma(T, \zeta_T) = \mathcal{O}(\sqrt{T \ln(T)} \frac{\sqrt{\log(1/\delta)}}{\epsilon}) = \widetilde{\mathcal{O}}(\frac{\sqrt{T\sqrt{\log(1/\delta)}}}{\sqrt{\epsilon}}).$$

Therefore, conditions in (2) of Theorem 4 are satisfied.

F Important Lemmas

F.1 Lemmas about privacy

Lemma 11. *Gaussian Mechanism $\mathcal{A}(\mathcal{D}) := p(\mathcal{D}) + \mathcal{N}(0, \sigma^2\mathbf{I})$, where $\sigma = \Delta\sqrt{2\log(1.25/\delta)}/\epsilon$ with $\Delta := \sup_{\mathcal{D}\sim\mathcal{D}'} \|p(\mathcal{D}) - p(\mathcal{D}')\|_2$ as the global sensitivity of $p : \mathcal{D} \to \mathcal{Z}$, is (ϵ, δ)-DP.*

Lemma 12 (Billboard Lemma [25])**.** *Suppose $\mathcal{M} : \mathcal{Y}^T \to \mathcal{Z}$ is $(\epsilon, \delta)-$DP. Consider any set of functions $\pi_t : \mathcal{Y}_t \times \mathcal{Z} \to \mathcal{X}_t, t \in [T]$ where $\mathcal{Y}_t$ is input space of user t. Then the composition $(\pi_1(\mathcal{Y}_1, \mathcal{M}(\mathcal{Y}^T)), \cdots, \pi_t(\mathcal{Y}_t, \mathcal{M}(\mathcal{Y}^T)))$ is $(\epsilon, \delta)-$ joint differential private.*

Lemma 12 shows that a JDP preserving algorithm can decide an action x_t for user t based on the true context y_t and a privatized information z from other users, using a non-privatized policy π_t. Thus, JDP-preserving algorithms offer much more flexibility for reward optimization than DP. The proposed algorithms in Section 6 preserve $(\epsilon, \delta)-$JDP.

Lemma 13 (Relationship between LDP and JDP)**.** *If each local privacy preserving mechanism $\mathcal{M}_t : \mathcal{Y}_t \to \mathcal{Z}_t$ at user t satisfies (ϵ, δ)-LDP, then an algorithm $\mathcal{A} : \mathcal{Y}^T \to \mathcal{X}^T$ which combines the same local processing and privacy preserving mechanisms $\{\mathcal{M}_t, t \in [T]\}$ at a trusted central node satisfies (ϵ, δ)-JDP.*

Lemma 13 directly holds by Billboard Lemma in Lemma 12 which establishes that a JDP preserving algorithm can select action for a user t based on the corresponding context y_t and the privatized information z from other users. This fact is useful for offloading the computation burden from local users when a trusted central node is available.

Lemma 14 ([42])**.** *Gaussian Mechanism $\mathcal{M}(\mathcal{D}) := f(\mathcal{D}) + \mathcal{N}(0, \sigma^2\mathbf{I})$ is $(\alpha, \frac{\alpha\cdot\Delta_f^2}{2\sigma^2})$-RDP where $\Delta_f := \sup_{\mathcal{D}\sim\mathcal{D}'} \|f(\mathcal{D}) - f(\mathcal{D}')\|_2$ is the global sensitivity of $f : \mathcal{D} \to \mathcal{R}$.*

Lemma 15 ([42])**.** *If $\mathcal{M}$ is $(\alpha, \epsilon)-$RDP, it is also $(\epsilon + \frac{\ln(1/\delta)}{\alpha-1}, \delta) - DP$ for any $\delta \in (0, 1)$.*

Lemma 16. *[Proposition 1 in [42]] Let $\mathcal{M}_1 : \mathcal{D} \to \mathcal{R}_1$ be $(\alpha, \epsilon_1)-$RDP and $\mathcal{M}_2 : \mathcal{D} \times \mathcal{R}_1 \to \mathcal{R}_2$ be $(\alpha, \epsilon_2) - RDP$. Let $X := \mathcal{M}_1(\mathcal{D})$ be the output of $\mathcal{M}_1$ and $Y := \mathcal{M}_2(\mathcal{D}, X)$ be the output of $\mathcal{M}_2$, then the composition (X, Y) satisfies $(\alpha, (\epsilon_1 + \epsilon_2)-$RDP.*

Lemma 17. *Let $\mathcal{M}_1 : \mathcal{D}_1 \to \mathcal{R}_1$ be $(\alpha, \epsilon_1)-$RDP and $\mathcal{M}_2 : \mathcal{D}_2 \times \mathcal{R}_1 \to \mathcal{R}_2$ be $(\alpha, \epsilon_2) - RDP$. $\mathcal{D}_1$ and $\mathcal{D}_2$ are disjoint. Let $X := \mathcal{M}_1(\mathcal{D}_1)$ be the output of $\mathcal{M}_1$ and $Y := \mathcal{M}_2(\mathcal{D}_2, X)$ be the output of $\mathcal{M}_2$, then the composition (X, Y) satisfies $(\alpha, \max\{\epsilon_1, \epsilon_2\})-$RDP.*

Proof. Let P_1 be the distribution of X, P_2 be the distribution of Y, and P is the joint distribution of (X, Y) given the dataset $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2\}$. Let $\mathcal{D}'$ be a neighboring dataset of $\mathcal{D}$, i.e. either $\mathcal{D}' = \{\mathcal{D}_1', \mathcal{D}_2\}$ or $\mathcal{D}' = \{\mathcal{D}_1, \mathcal{D}_2'\}$. Given the neighboring dataset $\mathcal{D}'$, denote P_1' be the distribution of X', P_2' be the distribution of Y', and P' is the joint distribution of (X', Y'). Then we have

$$\begin{aligned}
\exp\left[(\alpha-1)D_\alpha(P||P')\right] &= \mathbb{E}_{P'}\left[\left(\frac{P(x,y)}{P'(x,y)}\right)^\alpha\right] \\
&= \int_{\mathcal{R}_1\times\mathcal{R}_2} \left(P(x,y)\right)^\alpha \left(P'(x,y)\right)^{1-\alpha} \mathrm{d}x\mathrm{d}y \\
&= \int_{\mathcal{R}_1\times\mathcal{R}_2} \left(P_1(x)P_2(y \mid x)\right)^\alpha \left(P_1'(x)P_2'(y \mid x)\right)^{1-\alpha} \mathrm{d}x\mathrm{d}y \\
&= \int_{\mathcal{R}_1} P_1(x)^\alpha P_1'(x)^{1-\alpha} \int_{\mathcal{R}_2} \left(P_2(y \mid x)\right)^\alpha \left(P_2'(y \mid x)\right)^{1-\alpha} \mathrm{d}y\mathrm{d}x
\end{aligned} \tag{47}$$

If $\mathcal{D}' = \{\mathcal{D}_1, \mathcal{D}_2'\}$, then X and X' share the same distribution, i.e. $P_1 = P_1'$. Thus, we have

$$\begin{aligned}
\exp\left[(\alpha-1)D_\alpha(P||P')\right] &= \int_{\mathcal{R}_1} P_1(x) \int_{\mathcal{R}_2} \left(P_2(y \mid x)\right)^\alpha \left(P_2'(y \mid x)\right)^{1-\alpha} \mathrm{d}y\mathrm{d}x \\
&\le \int_{\mathcal{R}_1} P_1(x) \exp((\alpha-1)\epsilon_2)\mathrm{d}x = \exp((\alpha-1)\epsilon_2),
\end{aligned} \tag{48}$$

where the inequality holds because $\mathcal{M}_2$ is $(\alpha, \epsilon_2)-$RDP.

If $\mathcal{D}' = \{\mathcal{D}_1', \mathcal{D}_2\}$, then given the same X, the conditional distributions of Y and Y' are the same, i.e. $P_2(Y \mid X) = P_2'(Y' \mid X)$. Thus, we have

$$\begin{aligned}\exp\left[(\alpha-1)D_\alpha(P||P')\right] &= \int_{\mathcal{R}_1} P_1(x)^\alpha P_1'(x)^{1-\alpha} \int_{\mathcal{R}_2} P_2(y \mid x)\mathrm{d}y\mathrm{d}x \\ &\leq \int_{\mathcal{R}_1} P_1(x)^\alpha P_1'(x)^{1-\alpha}\mathrm{d}x = \exp((\alpha-1)\epsilon_1),\end{aligned} \tag{49}$$

where the inequality holds because $\mathcal{M}_1$ is $(\alpha, \epsilon_1)-$RDP. Therefore, we have $D_\alpha(P||P') \leq \max(\epsilon_1, \epsilon_2)$ which translates to $(\alpha, \max(\epsilon_1, \epsilon_2))-$RDP of the composition (X, Y). □

F.2 Lemmas about concentration

Lemma 18 (Sub-Gaussian Concentration)**.** *Given a sub-Gaussian variable X with $\mathbb{E}\left[e^{\lambda(X-\mathbb{E}[X])}\right] \leq \exp(\frac{\lambda^2\sigma^2}{2})$, it holds that*

$$\Pr(|X - \mathbb{E}[X]| \geq y) \leq 2\exp(\frac{-y^2}{2\sigma^2}).$$

Lemma 19 (Azuma-Hoeffding Inequality)**.** *Suppose $\{X_k : k = 0, 1, 2, 3\cdots\}$ is a martingale or super-martingale and $|X_k = X_{k-1}| \leq c_k$, then for any $y > 0$ and integer N, it holds that*

$$\Pr[|X_N - X_0| \geq y] \leq 2\exp\left(\frac{-y^2}{2\sum_{k=1}^N c_k^2}\right).$$

Lemma 20 (Laurent-Massart Bound)**.** *Let $Z_1, \cdots, Z_k$ be independent, standard Gaussian random variables, and the sum of their squares is $X = \sum_{i=1}^k Z_i^2$. It holds that*

$$\mathbb{E}[X] = k, \quad \Pr(X - k \geq 2\sqrt{kx} + 2x) \leq e^{-x}. \tag{50}$$

F.3 Lemmas on Online Learning

Lemma 21 (Convergence of DMD (Proposition 5 in [7]))**.** *If μ_t is updated as $\mu_{t+1} = \arg\min_{\mu\in\mathcal{U}} g_t^\top \mu + \frac{1}{\eta}V_h(\mu, \mu_t)$ where $\mathcal{U}$ is a convex set and $\forall t \in [T], \|g_t\|_2 \leq G$, then for any $\mu : \mathcal{U}$, it holds for $w_t(\mu_t) = g_t^\top \mu_t$ that*

$$\sum_{t=1}^T w_t(\mu_t) - w_t(\mu) \leq \frac{\eta}{2}G^2T + \frac{1}{\eta}V_h(\mu, \mu_1). \tag{51}$$

Lemma 22 (Convergence of FTRL (Extended from Corollary 7.9 in [44]))**.** *If μ_t is updated as $\mu_{t+1} = \arg\min_{\mu\in\mathcal{U}} \sum_{t=1}^T g_t^\top \mu + \frac{1}{\eta}h(\mu)$ where $\mathcal{U}$ is a convex set, $h(\mu)$ is a 1-strongly convex function w.r.t. $\|\cdot\|_2$, and $\forall t \in [T], \|g_t\|_2 \leq G$, then for any $\mu : \mathcal{U}$, it holds for $w_t(\mu_t) = g_t^\top \mu_t$ that*

$$\sum_{t=1}^T w_t(\mu_t) - w_t(\mu) \leq \frac{\eta}{2}G^2T + \frac{1}{\eta}(h(\mu) - \min_{\beta\in\mathcal{U}} h(\beta)). \tag{52}$$

Lemma 23 (Lemma 7 in [40])**.** *Let $\phi_1 : \mathcal{C} \to \mathcal{R}$ be a convex function such that $x_1 = \arg\min_{x\in\mathcal{C}} \phi_1(x)$ exists. Let ψ be a convex function such that $\phi_2(x) = \phi_1(x) + \psi(x)$ is 1-strongly convex w.r.t. norm $\|\cdot\|$. Let $x_2 = \arg\min_{x\in\mathcal{C}} \phi_2(x)$. Then, for any $g \in \partial\psi(x_1)$, we have $\|x_1 - x_2\| \leq \|g\|_*$ where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.*

G Details of Empirical Studies and Additional Numerical Results

G.1 Problem Statement

We consider a request routing problem in battery-powered edge AI systems [61, 37]. The rapid growth of AI has driven their deployment on resource-constrained edge devices to enable low-latency, real-time applications such as autonomous systems, industrial automation, and augmented reality [39, 55]. In such systems, multiple AI models with different capabilities and energy consumption

profiles are deployed on battery-powered edge servers. As user requests arrive sequentially, the system must decide in real time which model should serve each request, balancing inference quality while satisfying the energy constraint.

This setting naturally leads to an online resource allocation problem, where routing decisions must be made without knowledge of future requests while ensuring that the cumulative energy consumption satisfies a global budget constraint. The shared energy budget introduces a side channel: decisions for one request may reveal information about others through the remaining resource state, raising privacy concerns.

Resource allocation can be performed in either centralized or decentralized settings. In a *centralized* setting, edge servers are managed by a single entity, and resource allocation is performed by a trusted centralized controller with access to raw contextual information. In this case, joint differential privacy (JDP) is required to protect user information while enabling efficient coordination. In contrast, in a *decentralized* setting, edge servers may belong to different entities, and no centralized allocator exists. The servers can only share limited coordination information to satisfy global constraints, and local differential privacy (LDP) is required to protect user data. Our framework addresses both settings, enabling privacy-preserving and efficient request routing under global resource constraints.

G.2 Experiment Setups

Reward and consumption models Concretely, the objective of the problem is to maximize the overall inference quality, subject to the long term energy resource constraint. This online allocation problem has a finite horizon of T rounds. At each round $t \in \{1, \dots, T\}$, a user arrives with the prompt features y_t. The allocator must choose an action from the set $X_t = \{0, 1, \dots, M\}$ with each element representing one of M AI models or a null action 0 that represents skipping this request. Selecting a model $x_t \in \{1, \dots, M\}$ at round t yields a reward $f_t(x_t)$ which depends on the $acc(x_t, y_t) \in \{0, 1\}$ accuracy of the user prompt y_t on the selected model x_t based on an accuracy predictor. The null action gives $f_t(0) = 0$. Each action incurs a scalar resource consumption $b_t(x_t)$ The null action consumes no resource for this request, i.e., $b_t(0) = 0$. All actions are drawn from a shared energy resource with total budget $B = \rho T$, where $\rho > 0$ is the per-round budget rate. The realized actions $x_t \in X_t$ must satisfy $\sum_{t=1}^{T} b_t(x_t) \leq B$.

Dataset We use the RouterBench dataset [26], a benchmark designed for studying prompt routing across heterogeneous LLMs. Routerbench consists of 36,496 real user prompts paired with per-model evaluation outcomes.

To get the inference accuracy for incoming prompts, we use a MLP model that, given a prompt embedding and a target LLM, outputs the probability of correctness. Prompt embeddings are generated using the `all-MiniLM-L12-v2` model from the `SentenceTransformer` library [47], which provides compact semantic representations of text. The MLP predictive model has 2 hidden layers, each with width 100. The energy consumption data is generated via an empirical estimation function inspired by [57]: $b_t(j) = \alpha_{j,0}\, \tau_t^{\text{in}} + \alpha_{j,1}\, \tau_j^{\text{out}} + \alpha_{j,2}\, \tau_t^{\text{in}} \tau_j^{\text{out}}$, where τ_t^{in} is the input token length of the prompt at round t, τ_j^{out} is a model-specific output token cap, and $\alpha_{j,0}, \alpha_{j,1}, \alpha_{j,2} > 0$ are model-dependent coefficients.

The dataset contains 111,524 requests in total. We convert it into an episodic dataset which contains 440 episodes, each with 4000 rounds.

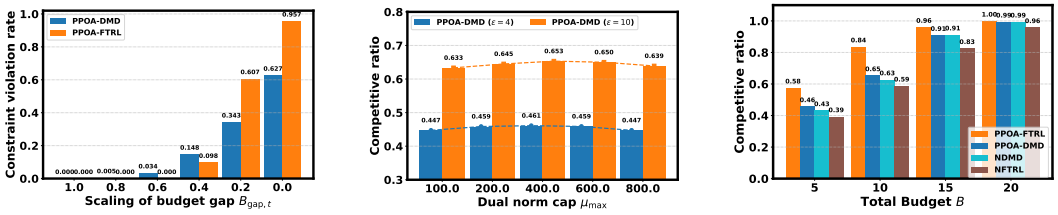
Baselines. We compare PPOA-DMD and PPOA-FTRL against the following baselines.

Non-private baselines:
• Greedy: greedy selects the constraint-feasible action that directly maximizes the reward $f_t(x)$.
• DMD: the dual mirror descent algorithm [7] with $h(\mu)$ being the squared norm function.
• FTRL: update the dual variable with follow the regularized leader [44] with $h(\mu)$ being the squared norm function.

Private baselines:
• DPA: a centralized online allocation algorithm preserving standard $(\epsilon, \delta)-$DP in Definition 3 by additionally applying Gaussian mechanism to reward and consumption functions of PPOA-FTRL.
• NDMD: Noisy DMD [14] adapted to satisfy the constraint feasibility in the same way as PPOA without applying the projection in 2.

(δ, ϵ)	CR				AVG			
	PPOA-DMD	NDMD	**PPOA-FTRL**	NFTRL	**PPOA-DMD**	NDMD	**PPOA-FTRL**	NFTRL
(0.001, 2)	*0.228*	0.206	**0.657**	0.171	*0.440*	0.377	**0.735**	0.322
(0.001, 4)	*0.419*	0.395	**0.765**	0.336	*0.624*	0.610	**0.821**	0.514
(0.001, 10)	*0.632*	0.601	**0.832**	0.567	*0.783*	0.763	**0.874**	0.705
(0.0001, 2)	*0.194*	0.174	**0.627**	0.145	*0.401*	0.347	**0.701**	0.285
(0.0001, 4)	*0.374*	0.357	**0.752**	0.307	*0.593*	0.580	**0.810**	0.480
(0.0001, 10)	*0.612*	0.581	**0.828**	0.533	*0.768*	0.748	**0.870**	0.687

Table 2: The expected (AVG) and adversarial (CR) performance of algorithms under different JDP or LDP levels. NDMD and PPOA-DMD admit both JDP and LDP, and NFTRL and PPOA-FTRL admit only JDP. Bold font indicates the best performance under JDP, and Italic font indicates the best performance under LDP.



(a) Resource constraint violation rate under different $B_{\text{gap},t}$.

(b) Competitive ratio of PPOA-DMD with $\mu_{\max}$.

(c) Competitive ratio under different budgets B.

Figure 2: Ablation studies over different budget cap scaling, dual norm cap, and total budget.

• NFTRL: Directly applying noisy FTRL without applying tree aggregation (TA) [13] for consumption aggregation.
• Pri-DOWM: Algorithm 2 in [28] adapted to satisfy the constraint feasibility in the same way as PPOA.

Performance metrics The average performance is evaluated by the average reward normalized by the average optimal reward over the sequences in the dataset, i.e. $\text{AVG} = \frac{\mathbb{E}[F_T^{\mathcal{A}}(\boldsymbol{y})]}{\mathbb{E}[OPT(\boldsymbol{y})]}$. The adversarial performance is evaluated by the empirical competitive ratio, i.e. $\text{CR} = \inf \frac{F_T^{\mathcal{A}}(\boldsymbol{y})}{OPT(\boldsymbol{y})}$.

Default experimental setting In default setting of our experiments, each round corresponds to routing a user prompt from the RouterBench dataset to one of $M = 11$ real LLMs. Each resource allocation episode has $T = 4000$ rounds. Rewards are rescaled to lie in $[0, 1]$ and the energy consumption per prompt is no greater than 0.01kWh. The battery capacity is set as 10kWh by default. The dual variable is initialized as $\mu_1 = 0$. The dual step size is fixed to $\eta = 2000$. The maximum dual norm bound in PPOA-DMD is set as $\mu_{\max} = 400$. If not stated, the privacy parameters are set as $\varepsilon = 10$, $\delta = 0.01$.

The evaluations are conducted on a labtop (Macbook Air). It takes about 5 minutes to execute PPOA for all the sequences in the dataset.

G.3 Additional Numerical Results

Ablation of Privacy Levels for Algorithms with $(\epsilon, \delta)-$JDP/LDP. In Table 2, we further evaluate the average and adversarial performance of PPOA-DMD and NDMD, which satisfy both JDP and LDP, as well as PPOA-FTRL and NFTRL, which satisfy only JDP, under different combinations of δ and ϵ. Since prior studies and practical applications [20, 1, 60, 14] commonly adopt δ that is close to zero and ϵ that lies between 1 and 10, we evaluate the algorithms on privacy levels with $\delta \in \{0.001, 0.0001\}$ and $\epsilon \in \{2, 4, 10\}$. We observe that, for all algorithms, stronger privacy requirements (smaller δ or ϵ) consistently lead to lower competitive ratios and average rewards, further illustrating the inherent trade-off between privacy and allocation performance. Moreover, across all combinations of δ and ϵ, and for both adversarial and average performance metrics, PPOA-DMD consistently outperforms NDMD under both JDP and LDP, while PPOA-FTRL outperforms NFTRL under JDP. These results provide strong empirical evidence for the effectiveness of PPOA-DMD in

(δ, ϵ)	CR				AVG			
	PPOA-DMD	NDMD	**PPOA-FTRL**	NFTRL	**PPOA-DMD**	NDMD	**PPOA-FTRL**	NFTRL
(0.001, 0.1)	0.0010	0.0010	0.0004	0.0006	0.0057	0.0057	0.0003	0.0025
(0.001, 0.2)	0.0044	0.0045	0.0032	0.0030	0.0204	0.0205	0.0007	0.0100
(0.001, 0.5)	0.087	0.083	**0.192**	0.053	*0.783*	0.763	**0.283**	0.018
(0.001, 1)	0.088	0.084	**0.470**	0.060	*0.241*	0.221	**0.567**	0.153

Table 3: The expected (AVG) and adversarial (CR) performance with strong privacy guarantee. NDMD and PPOA-DMD admit both JDP and LDP, and NFTRL and PPOA-FTRL admit only JDP. Bold font indicates the best performance under JDP, and Italic font indicates the best performance under LDP. Only scores larger than 0.1 are marked.

mitigating the impact of noise through dual projection, as well as the advantage of PPOA-FTRL in leveraging tree aggregation to achieve lower noise for the same privacy level.

Moreover, to assess the performance with very strong differential privacy, we evaluate our algorithms with privacy levels $\epsilon = \{0.1, 0.2, 0.5, 1\}$. The results are provided in Table 3. We can find that PPOA still outperforms their counterparts with the same privacy notion under small enough privacy levels ($\epsilon = 1$ and $\epsilon = 0.5$). When we continue to increase the privacy guarantee to $\epsilon = 0.2$ and $\epsilon = 0.1$, the allocation performance guarantee becomes very weak.

To examine performance under strong privacy requirements, we further evaluate the algorithms with $\delta = 0.001$ and $\epsilon \in \{0.1, 0.2, 0.5, 1\}$. As shown in Table 3, PPOA-DMD and PPOA-FTRL outperform their respective baselines under the same privacy notions in both CR and AVG at $\epsilon = 0.5$ and $\epsilon = 1$. When ϵ decreases further to 0.2 and 0.1, all methods exhibit very low allocation performance. These results illustrate the substantial utility cost of imposing stringent privacy requirements at the evaluated horizon $T = 4000$. Nevertheless, a longer horizon can lead to greater utility under strong privacy because for any fixed privacy parameters, Theorem 4 establishes that PPOA achieves vanishing expected regret under i.i.d. arrivals and an asymptotic competitive ratio of at least $1/\alpha$ in probability under adversarial arrivals.

Effects of Reserved Budget Gap The reserved budget gap $B_{\text{gap},t}$ in Line 4 of PPOA (Algorithm 1) is critical for ensuring budget feasibility. In the default setting, we set $B_{\text{gap},t}$ according to Theorem 6 for PPOA-DMD and Theorem 8 for PPOA-FTRL, so that the resource constraints are satisfied. In Figure 2a, we scale the reserved budget gap as $\vartheta \cdot B_{\text{gap},t}$ across all rounds, where $\vartheta \in \{1.0, 0.8, 0.6, 0.4, 0.2, 0.0\}$. Here, $\vartheta = 1$ corresponds to the default setting, while $\vartheta = 0$ indicates no reserved budget gap. We observe that as ϑ decreases, the resource constraint violation rate increases rapidly, demonstrating the necessity of the budget gap design in PPOA. Moreover, since PPOA-DMD injects more noise into the dual updates, it exhibits a higher violation rate than PPOA-FTRL when the reserved budget gap is sufficiently large to be effective. In contrast, when the reserved budget gap is close to zero, PPOA-FTRL shows a higher violation rate, likely because it is less conservative and tends to allocate more resources under this setting in the absence of a reserved safety margin.

Effects of $\mu_{\max}$ on PPOA-DMD $\mu_{\max}$ is a key parameter in the dual update of PPOA-DMD (Algorithm 2), used to stabilize the dual learning process under noise injection for adversarial performance guarantees. According to Theorem 6, $\mu_{\max}$ should be at least $\max_j \frac{\bar{f}}{\rho_j}$, which equals 100 in our default experimental setting. In Figure 2b, we evaluate the empirical competitive ratio of PPOA-DMD for $\mu_{\max} \in \{100, 200, 400, 600, 800\}$ under two privacy levels, $\epsilon = 4$ and $\epsilon = 10$. The results show that an overly small $\mu_{\max}$ degrades adversarial performance by overly restricting the dual domain and limiting the search for suitable dual variables. Conversely, an excessively large $\mu_{\max}$ also degrades performance, as it weakens the effect of regularizing noisy dual updates. Under this setting, $\mu_{\max} = 400$ achieves the best performance for both $\epsilon = 4$ and $\epsilon = 10$.

Effect of Total Budgets The adversarial performance of PPOA depends critically on the time-averaged budget $\rho = B/T$. We therefore evaluate the competitive ratio of PPOA and other private baselines under different choices of B with a fixed horizon $T = 4000$, as shown in Figure 2c. All algorithms achieve zero constraint violation across all budget settings on the dataset, ensuring a fair comparison. As expected, all methods achieve better performance as the total budget increases.

More importantly, under tighter budget constraints, PPOA-DMD exhibits larger improvements over NDMD. This is because in resource-constrained regimes, the impact of injected noise on performance is amplified, making the effect of dual projection more critical. A similar trend is observed when comparing PPOA-FTRL with NFTRL: under stricter budgets, the degradation caused by noise becomes more pronounced, and the superior information aggregation mechanism of PPOA-FTRL yields greater gains. Overall, these results demonstrate the robustness and superiority of the proposed algorithms across varying budget regimes.

H Empirical Studies on Online Advertising

H.1 Problem Statement

We evaluate private online allocation algorithms on a budget-constrained bidding problem over a horizon of T auctions in online advertising [3, 34]. At round t, an ad impression arrives with contextual information such as the user's device and browser, geographic region, webpage domain, and ad-slot characteristics. An advertiser must decide a bid action $x_t \in \mathcal{X}_t$ for an incoming ad opportunity at each round t. If the advertiser wins the auction for the ad opportunity, it earns the reward $f_t(x_t)$ related to the user click rate for this opportunity and pays the price $b_t(x)$ for this opportunity. The null action $x_t = 0$ skips the auction and satisfies $f_t(0) = b_t(0) = 0$. The total payment of the advertiser must not exceed a limited budget B over a horizon T.

Privacy matters because an impression's context may reveal information about the user who generated it. That context can influence later bid actions observed by other users, which creates privacy leakage. We can protect from the privacy leakage by joint differential privacy. The resulting task is to earn as much reward as possible while respecting both the budget constraint and the privacy guarantee.

H.2 Experiment Setups

Dataset We evaluate the private online allocation algorithms on the iPinYou online advertising dataset [34]. Each independent problem instance has horizon $T = 4000$, and there are 600 instance sequences in the testing set.

Reward and consumption models For each expression t, we fit a click probability model pclick_t based on the iPinYou dataset with impression features and corresponding click rate. We construct the synthetic market price sequence $\{\mathrm{price}_t\}$ as a weighted combination of the scaled slot price, slot area, and hour, plus a constant baseline, multiplied by a Gaussian noise. For each nonzero action x, the value attained from a click is set as $\mathrm{vclick}_t = 1$. The winning probability of the non-zero biding action x is modeled as $\mathrm{win}_t(x) = \mathrm{sigmoid}((x \cdot \mathrm{vclick}_t \cdot \mathrm{pclick}_t - \mathrm{price}_t)/\tau)$, where τ is half the standard deviation of the synthetic prices in the saved datasets. If the bid wins, the advertiser displays the ad and pays the modeled market price $\{\mathrm{price}_t\}$. The expected reward given an action x for impression t is $f_t(x) = \mathrm{win}_t(x) \cdot \mathrm{vclick}_t \cdot \mathrm{pclick}_t$ and expected expense is $b_t(x) = \mathrm{win}_t(x) \cdot \mathrm{pclick}_t$.

Baselines and performance metrics Baselines and performance metrics are the same as those of the main experiment in Section G.2.

Default setting In the default setting of our online advertising experiments, each round corresponds to an ad impression from the medium iPinYou dataset. The advertiser chooses among seven positive biding options and a no-bid action. We evaluate 600 episodes of $T = 4000$ rounds each. Rewards are rescaled to $[0, 1]$, and the scaled expected spend per impression is at most 0.1. The total spending budget is $B = 100$ per episode.

In PPOA algorithms, the dual variable is initialized at $\mu_1 = 10$. The dual step size is set as $\eta = 2.5$ for PPOA-DMD and $\eta = 7.5$ for PPOA-FTRL using a validation dataset. The upper bound used in the PPOA-DMD dual projection is $\mu_{\max} = 5$. Unless otherwise stated, the privacy parameters are $\varepsilon = 10$ and $\delta = 0.001$.

H.3 Numerical Results

We provide the main performance evaluation for PPOA and baselines in Table 4. Without privacy constraints, DMD and FTRL achieve the highest AVG and CR values. Among methods supporting

Algorithms	Non-private algorithms			Private algorithms					
	Greedy	DMD	FTRL	DPA	NDMD	NFTRL	Pri-DOWM	**PPOA-DMD**	**PPOA-FTRL**
AVG (10, 0.001)	0.691	0.986	0.986	0.641±0.0003	0.599±0.001	0.733±0.001	0.769±0.0006	0.832±0.001	**0.921**±0.0001
CR (10, 0.001)	0.618	0.972	0.972	0.563±0.008	0.480±0.01	0.624±0.004	0.661± 0.006	0.713±0.017	**0.882**±0.004
AVG (4, 0.001)	0.691	0.986	0.986	0.568±0.0002	0.506±0.0015	0.517±0.002	0.516±0.0018	0.637±0.002	**0.840**±0.0002
CR (4, 0.001)	0.618	0.972	0.972	0.489±0.006	0.351±0.004	0.318±0.02	0.358± 0.014	0.405±0.027	**0.782**±0.007

Table 4: The expected (AVG) and adversarial (CR) performance of algorithms in the application of online bidding. Among the private algorithms, DPA admits $(\epsilon, \delta)-$DP, NDMD and PPOA-DMD admit $(\epsilon, \delta)-$JDP or $(\epsilon, \delta)-$LDP, NFTRL and PPOA-FTRL admit $(\epsilon, \delta)-$JDP

both JDP and LDP, PPOA-DMD consistently outperforms NDMD across both privacy settings and performance metrics, supporting the benefit of dual projection in (2) for controlling the impact of privacy noise. When only JDP is required, PPOA-FTRL achieves the best performance among all private algorithms. Although all private methods experience performance degradation as ϵ decreases, PPOA-FTRL maintains its advantage in both metrics. These results are consistent with the benefit of its tree-aggregation mechanism for accurately aggregating historical resource consumption. The adversarial performance of Pri-DOWM with a multiplicative dual update drops a lot under stronger privacy preservation with less ϵ due to the lack of adversarial performance guarantee. Moreover, despite enforcing standard DP, the centralized algorithm DPA maintain a CR because it can perform centralized tree aggregation to get a good consumption estimation. The adversarial performance of Pri-DOWM, which uses a multiplicative dual update and lacks an adversarial performance guarantee, deteriorates substantially under stronger privacy requirements with smaller ϵ. Meanwhile, despite enforcing standard DP, the centralized algorithm DPA retains a CR of 0.489 at $\epsilon = 4$, outperforming several private baselines, due to the benefit of centralized tree aggregation for estimating cumulative resource consumption in DPA.

I More Details about Applications and Additional Examples

I.1 Task Offloading for Battery-Powered Edge Computing

A battery-powered edge node (mobile device, IoT gateway, or on-vehicle ECU) serves a stream of computation requests from heterogeneous users. At round t, request $y_t = \{X_t, f_t, b_t\}$ specifies an admissible offloading action set X_t, a reward f_t encoding latency / accuracy / QoS satisfaction, and an on-device energy cost b_t. Following [31, 39, 59], an action $x_t \in X_t$ jointly determines the local-versus-edge offloading ratio, the uplink transmission power, the computing-resource share, and thereby fixes both the latency profile and energy consumption $b_t(x_t)$. Subject to a finite battery B, the platform must maximize cumulative reward online, instantiating the online allocation problem in Eq. (1) of our paper. Coordination across rounds is carried by two quantities: the residual-budget state $\tilde{B}_t$ and the dual price μ_t, which together summarize the entire request history into a compact signal that governs the per-request offloading rule (Algorithm 1, Line 3).

The system admits both centralized and decentralized modes. In the *centralized* mode, a trusted edge allocator collects each arriving user's context, runs the primal–dual update on its behalf, and returns the assigned action; users observe only their own action and never see $\tilde{B}_t$ or μ_t directly. JDP preservation is required under this mode. In the *decentralized* mode, edge servers from different parties are collocated and share the same battery, and a trusted allocator does not exist: each server makes its own offloading decision locally and instead broadcasts coordination summaries (privatized budget consumption, dual updates) so that other servers can maintain consistent estimates of $\tilde{B}_t$ and μ_t and respect the shared energy budget. LDP preservation is required under this mode.

Threat Model. We consider a malicious user who participates in the system as an ordinary tenant. The attacker observes only the actions assigned to its *own* requests; it has no direct access to other users' contexts, identities, or actions. Nevertheless, because the per-request offloading rule is a function of the residual budget and the dual price, the action returned for the attacker's own request reflects the current level of resource pressure. By comparing the actions assigned to its requests against what would normally be expected for those requests, the attacker can detect whether the shared resource is being consumed faster than usual or not: if the allocation is smaller than typical—for instance, more requests are offloaded to cloud—then the residual budget is being depleted rapidly, which suggests that one or more heavy consumers are currently being served.

This signal alone does not reveal much about user information by itself. The risk arises when the attacker combines it with auxiliary information about a target user. Two representative scenarios illustrate this. First, if the attacker knows when a target user typically arrives at the system, it can submit its own requests at those moments and infer, from the allocations it receives, whether the target's jobs in that window are heavy or light, leaking task-type information about the user. Second, if the attacker knows that a target user is characterized by distinctively heavy (or distinctively light) and time-sensitive (or less time-sensitive) jobs, the attacker can probe the system continuously with its own requests and watch for sustained periods of unusually tight (or unusually loose) allocations; the timing of those periods reveals the target's arrival pattern. In both cases the attacker leverages a small amount of external knowledge to get a much more informative inference about a specific user.

The same logic applies in the decentralized deployment, with the difference that the coordination signal $\tilde{B}_t, \mu_t$ is broadcast directly and the attacker no longer needs its own action as a proxy. This is the resource-constraint-based side channel: the shared energy constraint nonlinearly couples decisions across rounds, so any quantity used to enforce that constraint becomes a transcript of the request history.

I.2 Energy Management in Multi-Tenant Data Centers.

A multi-tenant (colocation) data center co-locates servers owned and operated by independent tenants under a shared power and energy infrastructure. Crucially, the operator does not control tenant servers [29, 18, 16], so the system is *purely decentralized*: no trusted central allocator exists, and tenants must coordinate among themselves to respect a shared global energy budget. During peak hours the facility relies on a shared energy storage system whose capacity over the peak window defines this global budget B for all tenants. At round t, the workload of an arriving tenant request is captured by $y_t = \{X_t, f_t, b_t\}$, where X_t is the set of admissible task-level configurations (server count, CPU frequency, batch size, scheduling option), f_t is a performance reward (throughput, SLA satisfaction, revenue), and b_t is the induced energy draw. Tenants coordinate power reduction during peak hours through a clearing price, and our setting treats the per-tenant configuration choice $x_t \in X_t$ as the action of local resource allocation problem subject to the finite global budget B. Each tenant decides x_t locally and broadcasts coordination summaries—privatized energy consumption and dual (energy price) updates—so that the remaining tenants can maintain consistent estimates of the residual budget $\tilde{B}_t$ and the dual price μ_t. This decentralized operating mode is precisely the LDP regime of PPOA.

Threat Model. We consider a malicious tenant that participates as an ordinary occupant of the data center and runs the same protocol as everyone else but has a goal of infering other tenants' information. Because the system is decentralized, the attacker directly observes the broadcast coordination signal—the disseminated $\tilde{B}_t$ and μ_t, or equivalently the privatized per-round energy consumption used to update them. It has no access to other tenants' raw workloads, server states, or configurations, but the coordination signal itself reflects the current level of energy pressure across the facility. By comparing the trajectory of $\tilde{B}_t$ and μ_t against what would normally be expected—the budget draining faster than usual, or the dual price climbing earlier in the peak window than typical—the attacker can detect when the shared energy budget is being depleted unusually quickly, which suggests that one or more heavy-consuming tenants are currently active.

This signal alone is coarse, but it becomes informative when combined with auxiliary knowledge of a target tenant. First, if the attacker knows the time windows in which a target tenant typically runs jobs, it can monitor the coordination signal during those windows and infer whether the target's workload is heavy or light, leaking workload-type information such as training versus inference, batch versus interactive, or normal versus burst regimes. Second, if the attacker knows that the target tenant is characterized by distinctively heavy (or distinctively light) workloads, it can monitor the coordination signal continuously and watch for sustained periods of unusually fast (or unusually slow) budget depletion; the timing of those periods reveals the target's activity pattern, which can in turn be correlated with business hours, release cycles, or operational events.

I.3 Budget-Aware Tool Calls in Agent Service.

A tool-augmented agent service handles user queries by deciding, for each incoming query, which external tools to invoke (search engines, web APIs, retrieval databases) and how many tool-call rounds to spend. Recent work on budget-aware agent scaling [38] shows that effective performance under bounded resources requires explicit awareness of a long-term tool-call cost budget B, motivating an online allocation formulation. At round t, the query is captured by $y_t = \{X_t, f_t, b_t\}$, where X_t is the set of admissible tool-call strategies (tool subset, number of rounds, reasoning depth), f_t is the answer-quality reward, and b_t is the induced token cost or tool-call cost consumption. This is a *centralized* setting: a trusted agent-service operator runs the primal–dual update for each user and returns the assigned strategy, while users never see $\tilde{B}_t$ or μ_t directly. It therefore corresponds exclusively to the JDP regime of PPOA.

Threat Model. A malicious user submits queries as an ordinary user and observes only the tool-call actions assigned to its own queries—which tools were invoked, how many calls were made, the depth of the reasoning trace, etc. Because the tool-call action is a function of $\tilde{B}_t$ and μ_t, deviations from what is normally expected for those queries (such as fewer tool calls than usual or shallower retrieval) indicate that the shared cost budget is under pressure and that one or more heavy consumers are currently active. Combined with auxiliary knowledge, this becomes informative: if the attacker knows when a target user typically queries, it can probe in those windows and infer whether the target is currently issuing heavy or light queries; if the attacker knows the target is characterized by distinctively heavy (or light) queries, sustained periods of unusually constrained actions reveal the target's activity pattern. Algorithm 1 closes this side channel by preserving JDP guarantees described in our paper.

J Notation Table

Table 5: Main Notations.

Notation	Description
Online allocation model	
T, $t \in [T]$	Time horizon and round index; $[T] = \{1, \dots, T\}$.
m, $j \in [m]$	Number of resources and resource index.
$\mathcal{X}_t$, x_t	Available action set and selected action at round t.
$f_t(x)$, $b_t(x)$	Reward and resource-consumption vector for action x at round t.
$\bar{f}$, $\bar{b}_j$, $\bar{b}$	Uniform bounds satisfying $0 \le f_t(x) \le \bar{f}$, $0 \le b_{t,j}(x) \le \bar{b}_j$, and $\|b_t(x)\|_2 \le \bar{b}$.
y_t, $\boldsymbol{y}$	User context $y_t = (\mathcal{X}_t, f_t, b_t)$ and context sequence $\boldsymbol{y} = (y_1, \dots, y_T)$.
$\mathcal{Y}_t$, $\mathcal{Y}^T$	Round-t context space and sequence space $\mathcal{Y}^T = \prod_{t=1}^{T} \mathcal{Y}_t$.
B, ρ	Total resource-budget vector and per-round budget vector, with $B = T\rho$.
B_t	True remaining budget before round t: $B_t = B - \sum_{s=1}^{t-1} b_s(x_s)$.
$\hat{\rho}$	Smallest per-round resource budget, $\hat{\rho} = \min_{j\in[m]} \rho_j > 0$.
$\mathcal{A}$, $F_T^{\mathcal{A}}(\boldsymbol{y})$	Online algorithm and its total reward, $F_T^{\mathcal{A}}(\boldsymbol{y}) = \sum_{t=1}^{T} f_t(x_t)$.
$\mathrm{OPT}(\boldsymbol{y})$	Optimal reward achievable with full knowledge of $\boldsymbol{y}$.
α	$\alpha = \max_{j\in[m]} \frac{\bar{b}_j}{\rho_j}$; the corresponding competitive ratio is $1/\alpha$.
Privacy and budget estimation	
ϵ, δ	Multiplicative privacy parameter and additive slack in the (ϵ, δ) privacy inequality.
ζ, ζ_T	Failure-probability parameter for the performance bounds and its horizon-dependent choice.
σ, σ^2	Gaussian noise standard deviation and variance per coordinate.
$\tilde{b}_t(x_t)$	Privatized consumption: $\tilde{b}_t(x_t) = b_t(x_t) + n_t$, with $n_t \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_m)$.
s_t	Privatized cumulative consumption through round t, released by tree aggregation.
$\tilde{B}_t$	Privatized estimate of the remaining budget before round t.
$B_{\mathrm{gap},t}$	Reserved budget margin bounding $\|\tilde{B}_t - B_t\|_\infty$ on the budget-estimation event.
$\mathcal{M}_B$, $\mathcal{M}_D$	Privacy-preserving mechanisms for updating the remaining budget and dual variable.

Continued on the next page

Table 5 (continued)

Notation	Description
Primal and dual updates	
μ_t, μ	Nonnegative dual variable used at round t and a dual comparator.
$\mathcal{D}$, $\mu_{\max}$	Dual region and Euclidean norm bound used in the projected DMD update.
$\hat{x}_t$	Pre-selected action before the budget check.
$w_t(\mu)$	Linear dual loss, $w_t(\mu) = \mu^\top(\rho - b_t(x_t))$.
$f_t^*(\mu)$	Maximized Lagrangian reward, $f_t^*(\mu) = \max_{x \in \mathcal{X}_t}\{f_t(x) - \mu^\top b_t(x)\}$.
$D(\mu \mid \boldsymbol{y})$	Offline dual objective, $D(\mu \mid \boldsymbol{y}) = \sum_{t=1}^T f_t^*(\mu) + T\mu^\top\rho$.
η, h	Dual-update step size and strongly convex regularization function.
$V_h(\mu, \nu)$	Bregman divergence, $V_h(\mu, \nu) = h(\mu) - h(\nu) - \nabla h(\nu)^\top(\mu - \nu)$.
$\mathbf{e}_j$	j-th coordinate vector in $\mathbb{R}^m$.
Stochastic analysis	
$\mathcal{P}$, $\mathcal{P}_t$	Common context distribution in the i.i.d. setting and marginal distribution of y_t in the general non-stationary setting.
$\bar{\mathcal{P}}$	Average marginal distribution, $\bar{\mathcal{P}} = \frac{1}{T}\sum_{t=1}^T \mathcal{P}_t$.
$\bar{D}(\mu \mid \mathcal{P})$	Expected dual objective per round, $\bar{D}(\mu \mid \mathcal{P}) = \mathbb{E}_{y\sim\mathcal{P}}[f_y^*(\mu)] + \mu^\top\rho$.
v_t, ξ_t	Internal randomness used at round t and history $\xi_t = \{(y_s, v_s) : 1 \le s \le t\}$.
$\varsigma(\xi_t)$	Sigma-algebra generated by the history through round t, including any initial randomness.
τ_A	First $s < T$ such that $\tilde{B}_{s+1,j} - B_{\mathrm{gap},s+1} \le \bar{b}_j$ for some resource j.
$\mathcal{E}$	Event on which both the budget-estimation bound and the dual-regret bound hold.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [N/A].
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for [N/A]).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will also be asked to include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While [Yes] is generally preferable to [No], it is perfectly acceptable to answer [No] provided a proper justification is given (e.g., error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering [No] or [N/A] is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract and introduction are clearly justified by the theoretical analysis in Sections 4.2, 5, 6, and numerical results in Section 7.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: Limitations of the work are clearly discussed in Section 8

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The full set of assumptions are specified at the start of Section 4. The proofs for all the theoretical results are provided in Sections A, C, D, E, and F.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper in Section G.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Codes for the numerical results will be opensourced once accepted.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.

- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The dataset processing, hyperparameter choices and other setups are detailed in Section G.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer:[Yes]

   Justification: The statistical variance of the algorithms is mainly from the injected privacy noise. The error bars for the main results are provided in Table 1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The comupter resource information is proved in Section G.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper discuss the societal impacts of the work performed in Section 8.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper poses no such risks.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All original owners of assets including the datasets and models are properly credited and cited in Section G.2.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: New assets introduced in the paper are well documented and the documentation is provided alongside the assets.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer:[N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: This research does not involve LLMs as any important, original, or non-standard components

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.