DATAFED: BID-AGNOSTIC AUCTIONS FOR HIERARCHICAL CLIENT SELECTION IN ONLINE FEDERATED EDGE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Dynamic client availability and resource variations motivate online federated edge learning (OFEL), where real-time client selection requires truthful auctions to align clients' bids with their actual costs. However, existing auctions face three key challenges: (i) two-sided incomplete information in bid-agnostic settings, (ii) hierarchical objective misalignment between cloud and edge servers, and (iii) lack of long-term optimization in dynamic environments. To address these challenges, we propose DataFed, a bid-agnostic auction mechanism that integrates dynamic Bayesian bid distribution estimation with Stackelberg-based client selection. Specifically, we formulate hierarchical client selection as a dynamic Bayesian Stackelberg game with a multi-armed bandit objective to minimize cumulative expected regret. With a known bid distribution, we establish a Stackelberg equilibrium for the normalized proxy model. When the distribution is unknown but fixed, we derive estimation error bounds under independent proxy feedback. With sufficient exploration, we prove sublinear cumulative expected regret relative to the proxy benchmark. By adaptively selecting clients with favorable utility-cost trade-offs, DataFed improves learning efficiency and stabilizes online training in dynamic environments. Experiments show that DataFed reduces mean cumulative regret by 23.0% on average across four real-world datasets and improves mean cloud utility by 7.9% through ES response anticipation.

1 INTRODUCTION

Federated Edge Learning (FEL) has emerged as a promising paradigm for distributed intelligence in large-scale edge networks, enabling collaborative model training with reduced communication latency and improved scalability (Luo et al., 2020; Ren et al., 2020). In such an cloud-edge framework, hierarchical client selection becomes necessary, where edge servers (ESs) first pre-select candidate clients, and the cloud server (CS) subsequently selects a subset as final training participants. When client participation is dynamic due to time-varying communication and computation resources, limited server-side capacity prevents involving all clients simultaneously. Moreover, clients' instantaneous participation states are not fully observed at decision time, forcing real-time client selection under incomplete information. This setting naturally motivates the adoption of online federated edge learning (OFEL) (Xin et al., 2025; Chen et al., 2020), where hierarchical client selection becomes an online decision-making problem. Such OFEL settings commonly arise in vehicular networks and IoT edge systems with dynamic client availability and resource fluctuations.

To coordinate client participation in such online settings, auction-based client selection has emerged as an effective approach for allocating limited training opportunities among self-interested clients while aligning bids with participation costs. However, most existing auctions rely on bid-aware assumptions, where servers can observe clients' bids before making selection decisions (Mittelmann et al., 2023; Jiao et al., 2021; Deng et al., 2024). Such assumptions are often impractical in OFEL, where client availability and resource conditions vary dynamically, forcing servers to make decisions under bid-agnostic incomplete information. Existing auctions remain limited in OFEL for two main reasons. On the one hand, *mismatch between bid-aware auction design and online decision-making.* Most existing mechanisms determine winner allocation based on observable bids, which is incompatible with the decision-before-observation nature of OFEL (Le et al., 2021; Lu et al., 2023; Hu et al., 2024).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

On the other hand, *imbalanced client competition caused by hierarchical pre-selection.* Most existing studies focus only on client selection at the CS while overlooking candidate pre-selection at ESs, where inappropriate pre-selection may lead to excessive competition (over-selection) or insufficient incentives (under-selection) (Jin et al., 2024; Liu et al., 2024; Weng et al., 2021). These limitations prevent existing auctions from being directly applied to OFEL, motivating the design of a bid-agnostic hierarchical auction mechanism.

To address the aforementioned limitations, it is necessary to design a bid-agnostic auction for client selection in OFEL. However, such a design must confront the following challenges: *(i) Two-sided incomplete constraints under bid-agnostic settings:* On the server side, online constraints prevent the server from observing clients' bids before making client selection decisions in each round. On the client side, bids are sealed and each client only observes its own bid. Then, ensuring truthful bidding and estimating clients' dynamic bid distribution from partial history bidding information becomes a challenge for the server (Yuan et al., 2022; Deng et al., 2021). *(ii) Hierarchical objective misalignment between ES and CS:* The CS aims to optimize the global model performance, whereas ESs aim to maximize regional utility. The misalignment of objectives across hierarchical servers poses significant coordination challenges, particularly under incomplete bid information, where bids are key parameters in their strategy determination (Qu et al., 2022; Tian et al., 2024). *(iii) Lack of long-term optimization in dynamic online environments.* Most existing auctions optimize per-round utility in static settings, whereas OFEL requires sequential client selection that maximizes long-term cumulative utility under dynamic environments (Baudry et al., 2024; Xin et al., 2025). Consequently, effective auctions must adapt to real-time decision-making to support appropriate hierarchical client selection under incomplete bid information.

To overcome the abovementioned challenges, we propose a bid-agnostic auction for client selection in OFEL, which integrates a Dynamic Bayesian bid distribution estimation rule and a Stackelberg-based auction rule (DataFed). The former rule estimates the bid distribution from historical auction payment thresholds. The latter rule addresses hierarchical client selection by determining candidate clients at the ES and final training clients at the CS. Due to incomplete information in online settings, minimizing cumulative expected regret is a natural objective for evaluate long-term utility. By integrating these two rules, DataFed enables servers to adopt adaptive strategies based on learned bids distribution over time, thereby facilitating regret minimization. Thus, DataFed enables an effective bid-agnostic auction for hierarchical client selection, thereby facilitating practical deployment in real-world online scenario for FL.

We summarize our main contributions as follows: (i) We formulate bid-agnostic hierarchical client selection as an online decision-making problem under incomplete information. Since servers must make sequential decisions before observing clients' bids, we design DataFed, an online auction mechanism that estimates bid distributions from partially observable historical bids and adapts its strategies over time to minimize cumulative expected regret. (ii) To characterize the benchmark under a known bid distribution, we model the interaction between the CS and ESs as a dynamic Bayesian Stackelberg game and establish the existence of an equilibrium for the normalized proxy model. Furthermore, we derive estimation error bounds for normalized proxy feedback given bids independently drawn from a fixed distribution. With sufficient exploration, we prove sublinear cumulative expected regret relative to the known distribution proxy benchmark. (iii) Across four real-world datasets, DataFed reduces mean cumulative expected regret by 23.0% on average relative to the best baseline in each setting. Further experiments demonstrate the effectiveness of DataFed's hierarchical coordination, with an 11.0% reduction in mean cumulative regret and a 7.9% increase in mean cloud utility.

2 Related Work

In this section, we review recent studies on auctions and client selection, and analyze their limitations in OFEL, as summarized in Table 1.

Auctions for OFEL. Properly designed truthful auctions incentivize clients to report their costs while ensuring individual rationality. However, applying existing auctions to online FL faces two limitations. *(i) Incomplete-information constraints:* Servers choose the candidate and winner counts before observing current bids. The auction then selects winners from the submitted bids (Hu et al.,

Table 1: Comparison with existing works for OFEL.

Category	Bid-Agnostic	Online Bid Learning	Hierarchical Selection	Multi-Client Selection	Main Limitation in OFEL
Auction-based selection (Hu et al., 2024; Damaskinos et al., 2022; Han et al., 2020)	×	×	×	✓	Needs bid access
Auction-based allocation (Mittelmann et al., 2023; Jiao et al., 2021; Deng et al., 2024)	×	×	×	✓	Stationary bids
Hierarchical selection (Qu et al., 2022; Tian et al., 2024)	×	×	✓	✓	Edge imbalance
Online selection (Xin et al., 2025; Agrawal et al., 2023)	✓	✓	×	×	Single-client only
DataFed (Ours)	✓	✓	✓	✓	Supports all four

2024; Damaskinos et al., 2022; Han et al., 2020). *(ii) Non-stationary bid distributions.* Time-varying computation and communication resources can cause bid distributions to drift in practice. Our regret analysis considers bids independently drawn from a fixed distribution, while the additional drift experiments assess behavior outside that guarantee. Existing auctions rely on full bid observability or a fixed bid model, limiting their applicability when decisions precede bids or distributions drift (Mittelmann et al., 2023; Jiao et al., 2021; Deng et al., 2024). To bridge this gap, we propose a bid-agnostic k-reverse Vickrey auction that learns from partial feedback. Given a fixed bid distribution and sufficient exploration, the normalized proxy analysis yields sublinear cumulative expected regret.

Client Selection for OFEL. Client selection can efficiently improve the performance of OFEL by balancing social utility and individual client utility, as this balance promotes both system efficiency and sustained client participation, which are essential for long-term system performance. However, when extending existing works to hierarchical structures, multi-level client selection in online settings prevents these mechanisms from being directly applicable, for the following reasons. *(i) Misalignment in cloud-edge client selection.* Existing works performed client selection in a centralized manner at the CS, treating ESs as a communication intermediaries between the CS and clients. This design ignores the impact of improper client pre-selection at the ES level, which may lead to over-selection, causing excessive competition, or under-selection, resulting in insufficient client participation (Qu et al., 2022; Tian et al., 2024). *(ii) Oversimplification in single-client selection under online settings.* To satisfy online constraints without observing bids, some existing studied reduce the client selection problem to a single-client decision, which constitutes an overly simplistic strategy and is impractical in real-world OHFL systems (Xin et al., 2025; Agrawal et al., 2023). To bridge this gap, we propose a bid-agnostic hierarchical client selection framework that selects multiple clients per round through a two-level decision structure, thereby enabling balanced client participating across hierarchy levels and avoiding both overly aggressive and overly conservative selection behaviors, while improving overall system performance.

3 Problem Formulation

In this section, we introduce the client selection process in OFEL, present the k-reverse Vickrey auction, and formulate the bid-agnostic online client selection problem.

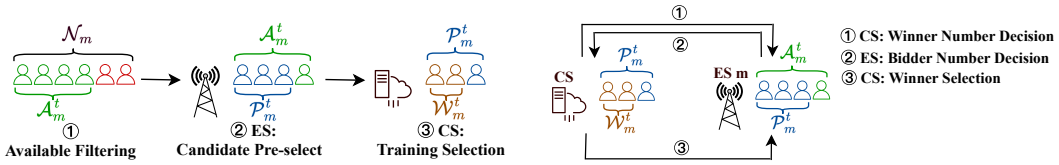
3.1 System Model

OFEL consists of a CS, a set of ESs $\mathcal{M} = \{1, \cdots, M\}$, and a set of clients $\mathcal{N} = \{1, \cdots, N\}$. Let $\mathcal{N}_m$ denote the clients that can communicate with ES $m \in \mathcal{M}$. Due to backhaul transmission constraints, each client is assigned to one ES, so $\mathcal{N}_m \cap \mathcal{N}_{m'} = \emptyset$ for distinct $m, m' \in \mathcal{M}$ (Qu et al., 2022). Client availability varies across time slots $\mathcal{T} = \{1, \cdots, T\}$. Figure 1(a) illustrates hierarchical client selection. At each $t \in \mathcal{T}$, let $\mathcal{A}_m^t \subseteq \mathcal{N}_m$ denote the available clients at ES m (①). Each ES m preselects candidates $\mathcal{P}_m^t \subseteq \mathcal{A}_m^t$ (②), from which the CS selects participants $\mathcal{W}_m^t \subseteq \mathcal{P}_m^t$ (③).

We define x_n^t and ω_n^t as the local training data and local model parameters of client $n \in \mathcal{W}_m^t$, respectively. Client n performs its local update process as follows:

$$\omega_n^{t+1} = \omega_n^t - \eta \nabla l(\omega^t, x_n^t), \tag{1}$$

where ω^t denotes the global model parameters, η is the learning rate, and $\nabla l(\cdot)$ is the gradient of local average loss $l(\cdot)$ (Han et al., 2021). After client n completes its local update, ES m performs an



(a) Hierarchical Client Selection (b) Bid-Agnostic k-reverse Vickrey Auction

Figure 1: Bid-Agnostic Auction for Hierarchical Client Selection in OFEL.

edge aggregation to update the edge model ω_m^t:

$$\omega_m^{t+1} = \omega_m^t + \frac{1}{|\mathcal{W}_m^t|} \sum_{n \in \mathcal{W}_m^t} \omega_n^t, \tag{2}$$

where $|\cdot|$ denotes the cardinality of a set (Fang et al., 2024). The server then integrates the received edge models into the global model as $\omega^t = \frac{1}{|\mathcal{M}|} \sum_{m \in \mathcal{M}} \omega_m^t$.

3.2 k-REVERSE VICKREY AUCTION

Let c_n^t denote client n's training cost at time slot t. The CS selects clients with relatively low costs (Yu et al., 2020). When c_n^t is private, a reverse Vickrey auction incentivizes truthful cost reporting (Zhao et al., 2023; Bubeck et al., 2019). OFEL requires selecting multiple clients with low bids and paying each selected client. To this end, we propose a k-reverse Vickrey auction, where the CS designates the $k-1$ lowest bidders as winners and pays each the k-th lowest bid. All candidate clients in $\bigcup_{m \in \mathcal{M}} \mathcal{P}_m^t$ submit bids and are ranked in ascending order of their bids. The proposed k-reverse Vickrey auction is defined as follows.

Definition 1 *(k-reverse Vickrey auction) At each time slot t, each client n submits a bid b_n^t, the utility function for her is given by:*

$$u_n^t = (b_k^t - c_n^t)\mathbf{1}(b_n^t < b_k^t), \tag{3}$$

Here c_n^t is client n's true participation cost, b_n^t is its bid, and b_k^t is the payment to each winner. The indicator equals one when client n wins and zero otherwise. The k-reverse Vickrey auction is truthful in each round. Given other clients' bids, reporting $b_n^t = c_n^t$ weakly dominates other bids for client n.

Proposition 1 *The k-reverse Vickrey auction is truthful.*

The k-reverse Vickrey auction is truthful within each round, with other clients' bids held fixed. Per-round truthfulness does not guarantee dynamic incentive compatibility when bids affect future learning and allocations. The auction also satisfies individual rationality (IR), as truthful bidding yields non-negative utility. We therefore omit a detailed analysis of IR and focus on truthfulness.

3.3 BID-AGNOSTIC ONLINE CLIENT SELECTION PROBLEM

Figure 1(b) illustrates the bid-agnostic k-reverse Vickrey auction. At time slot t, the CS sets the training participant count $\zeta^t = k^t - 1$ (①). ES m then sets its bidder count $\alpha_m^t = |\mathcal{P}_m^t|$, while $\alpha_{-m}^t = \sum_{j \neq m} \alpha_j^t$ is the total bidder count of the other ESs (②). For a fixed round, we omit the time index t for simplicity. The CS selects ζ winners and pays each $b_{\zeta+1}(\alpha_m, \alpha_{-m})$, the $(\zeta+1)$-th lowest bid among all bidders (③). Since the ES serves as an intermediary between the CS and its associated clients, its utility depends on the clients connected to that ES, i.e.,

$$v_m(\alpha_m, \alpha_{-m}, \zeta) = \sum_{n=1}^{\alpha_m} [b_{\zeta+1}(\alpha_m, \alpha_{-m}) - b_n] \cdot \mathbf{1}[b_n < b_{\zeta+1}(\alpha_m, \alpha_{-m})]. \tag{4}$$

The CS's utility is its valuation minus the payments to training clients. The CS's valuation is defined as the improvement in global model accuracy, which is characterized as a function of the aggregate

batch size across all training clients (Hu et al., 2022; Huang et al., 2023; Li et al., 2024). Without access to clients' local data, the CS models the expected global accuracy improvement as $q(\zeta)$, a function of the selected client count. The CS utility is therefore given by

$$s(\alpha_m, \alpha_{-m}, \zeta) = q(\zeta) - \zeta b_{\zeta+1}(\alpha_m, \alpha_{-m}). \tag{5}$$

In the online setting, the CS and ESs optimize their utilities before observing current bids. After each auction, the learning module receives the k-th lowest bid, which determines the uniform payment to the winners. For analysis, clients' bids are drawn independently from a fixed unknown distribution on $[0, 1]$ with CDF $\phi(\cdot)$. This distributional model supports decisions before bids are observed (Li et al., 2024; Baudry et al., 2024). The OFEL objective is to maximize the CS utility in each round, balancing model quality improvement against client payments, as follows.

$$\begin{aligned} \mathbb{P}_1 : \arg\max \mathbb{E}_{b\sim\phi}[s(\alpha_m, \alpha_{-m}, \zeta, \phi)], \\ \text{s.t. Eqs. } (4) \text{ and } (5). \end{aligned} \tag{6}$$

Problem $\mathbb{P}_1$ considers bid-agnostic hierarchical client selection under a k-reverse Vickrey auction, with joint CS and ES strategies $(\zeta, \alpha_m, \alpha_{-m})$ and an unknown bid distribution $\phi(\cdot)$. In each round, the objective is to maximize the CS's expected utility over these strategies, averaging over bids.

4 Proposed Method: DataFed

In this section, we formulate the cumulative expected regret based on a dynamic Bayesian Stackelberg game model and propose a novel DataFed mechanism to address it.

4.1 Dynamic Bayesian Stackelberg Game

The Bayesian Stackelberg Game models leader-follower interactions under incomplete information, where the leader maximizes its expected utility by anticipating the followers' responses. In repeated auctions, servers choose their strategies before observing current bids. We formulate this interaction as a Dynamic Bayesian Stackelberg Game (DBSG). As the equilibrium cannot be computed without prior knowledge of follower types, the leader adopts online learning to adapt its strategy, with the objective of minimizing cumulative expected regret relative to the equilibrium under known follower types (Xu et al., 2016; Abdel Wahab et al., 2021; Shen et al., 2025). The dynamic aspect lies in the learner's information updates after each auction. The regret analysis considers a fixed underlying bid distribution. The model does not posit an exogenous state transition process.

Here, we model the interaction between the CS and the ESs in an online setting as a DBSG. In each round, the CS first announces the winner count ζ as the leader. Each ES then chooses its bidder count α_m from the available clients as a follower. The CS subsequently selects ζ winners from all bidders. We define the DBSG as follows:

Definition 2 *The DBSG can be characterized by a* 4*-tuple* $< \mathcal{R}, \mathcal{S}, \mathcal{F}, \mathcal{U} >$ *, where*

- *Player set:* $\mathcal{R} = \{\mathcal{L}, \mathcal{F}\}$*, where* $\mathcal{L} = \{CS\}$*,* $\mathcal{F} = \mathcal{M}$*.*
- *Strategy set:* $\mathcal{S} = \mathcal{L}_S \times \prod_{m=1}^{M} \mathcal{F}_S^m$*, where* $\mathcal{L}_S = \{\zeta | \zeta \in [1, \sum_{m=1}^{M} \alpha_m] \cap \mathbb{Z}^+\}$ *and* $\mathcal{F}_S^m = \{\alpha_m | \alpha_m \in [1, |\mathcal{A}_m|] \cap \mathbb{Z}^+\}$*.*
- *Distribution set:* $\mathcal{G} = \{\phi : [0, 1] \to [0, 1] | \phi(0) = 0, \phi(1) = 1, \phi$ *is non-decreasing*$\}$*.*
- *Payoff set:* $\mathcal{U} = \{s(\cdot, \cdot), v(\cdot, \cdot)\}$*.*

To minimize cumulative expected regret in the DBSG, we formulate the interaction between the CS and ESs as a Multi-Armed Bandit (MAB) problem, enabling online learning from repeated interactions and feedback (Wu & Li, 2024; Wang et al., 2024). The ES m sequentially explores and exploits candidate counts $\alpha_m \in \mathcal{F}_S^m$, while the CS anticipates the ES responses when selecting the winner count ζ. After the auction, the learner observes $b^t_{\zeta_t+1}$, the $(\zeta_t + 1)$-st lowest bid and the uniform payment to winners. Individual bids and counterfactual thresholds for unselected actions

are unavailable. We measure cumulative expected regret against the bid-aware benchmark under the same true bid distribution:

$$\mathbb{P}_2 : \min_{\{\mathcal{S}_t\}_{t=1}^T} R(T) = \sum_{t=1}^{T} \mathbb{E}[s(\mathcal{S}^*, \phi) - s(\mathcal{S}_t, \phi)]\,, \quad \text{s.t. Eqs. (4) and (5).} \tag{7}$$

Here $\phi \in \mathcal{G}$ is the true but unknown bid CDF, $\mathcal{S}^* \in \mathcal{S}$ is the bid-aware benchmark strategy, and $\mathcal{S}_t \in \mathcal{S}$ is selected using auction feedback available before round t. Bid distribution estimates guide $\mathcal{S}_t$, while both utilities are evaluated under the same ϕ. We measure regret in terms of the CS utility. Problem $\mathbb{P}_2$ requires online hierarchical decisions under an unknown bid distribution and partial auction feedback. Client selection thus jointly updates bid estimates and cloud and edge decisions.

4.2 Formulation of DataFed

To minimize cumulative expected regret in the DBSG, we propose DataFed, which jointly optimizes CS and ES strategies while estimating the bid distribution. Given α_{-m}, we write the hierarchical selection strategy as (α_m, ζ) and abbreviate α_m as α. However, the exact expected utilities depend on the $(\zeta+1)$-th order statistic among $M\alpha$ bids, complicating the analysis. To address this, we exploit symmetry and introduce an expectation-level normalization that maps the multi-winner setting to a scaled single-winner formulation (see Appendix A.2 for details). Based on this mapping, we construct a tractable benchmark that captures the marginal competition effect induced by the lower tail of the bid distribution.

Lemma 1 *The benchmark expected utilities of ES m and the CS are given by:*

$$\begin{cases} \widetilde{v}(\alpha,\zeta) = \alpha \int_0^1 \varphi^{\theta-1}(x) - \varphi^{\theta}(x)\,dx, \\ \widetilde{s}(\alpha,\zeta) = q(\zeta) - \zeta \int_0^1 (1-\theta)\varphi^{\theta}(x) + \theta\varphi^{\theta-1}(x)\,dx, \end{cases} \tag{8}$$

where $\theta = \frac{M\alpha}{\zeta}$, and $\varphi(x) = 1 - \phi(x)$.

Derived from a normalized single-winner formulation, these expressions approximate the expected utilities of the original multi-winner mechanism. They retain the dependence on lower-tail competition and the scaling with α and ζ used in the subsequent analysis. For the ES, increasing α improves the winning probability of its nominated bidders while intensifies competition. For the CS, ζ induces a trade-off between competition intensity and payment, balancing utility and participation incentives. We define DataFed for bid-agnostic hierarchical client selection as follows:

Definition 3 *DataFed is formally defined as a structured pair $(\mathbb{G}, \mathbb{A})$, i.e., a dynamic Bayesian bid distribution estimation rule $\mathbb{G}$, and a Stackelberg-based auction rule $\mathbb{A}$.*

- *$\mathbb{G}$: $\prod[0,1] \to \mathcal{G}$ denotes a distribution estimation rule that infers the client bid distribution from historical observations of the k-th lowest bids, which are publicly revealed as payment to winners, i.e.,*
$$\phi \leftarrow \{b_k^t | t \in \mathcal{T}\}. \tag{9}$$
- *$\mathbb{A}$: $\mathcal{G} \to \mathcal{S}$ denotes a decision mapping that, for a given bid distribution $\phi \in \mathcal{G}$, returns the strategy pair $(\alpha, \zeta) \in \mathcal{S}$ selected by the ES and the CS, respectively, i.e.,*
$$(\alpha, \zeta) \leftarrow \arg\max_{(\alpha,\zeta)} s(\alpha, \zeta | \phi). \tag{10}$$

Based on Definition 3, we use the normalized proxy model to formulate an approximate DataFed design for $\mathbb{P}_2$:

$$\mathbb{P}_3 : \arg\min_{(\mathbb{G},\mathbb{A})} R(T), \quad \text{s.t. Eqs. (7), (8), (9) and (10).} \tag{11}$$

This problem minimizes cumulative expected regret by jointly designing bid distribution estimation and hierarchical client selection under a k-reverse Vickrey auction. Problem $\mathbb{P}_3$ remains challenging because the estimation rule $\mathbb{G}$ and auction rule $\mathbb{A}$ are coupled through online auction outcomes. Consequently, bid distribution estimates and hierarchical decisions must be updated jointly online.

5 APPROXIMATE DESIGN OF DATAFED

The unknown bid distribution and partial auction feedback make the original multi-winner auction analytically challenging. Therefore, we adopt an expectation-level normalized approximation to construct the benchmark and guide online decision-making. In this section, we characterize the equilibrium of the normalized proxy game and analyze estimation and regret under independent proxy feedback and sufficient exploration. The algorithm observes the actual auction payment threshold.

5.1 REGRET BENCHMARK UNDER THE BID-AWARE SETTING

Given a known client bid distribution $\phi(\cdot)$, we derive the dynamic Bayesian Stackelberg equilibrium (DBSE) of the normalized proxy game via backward induction. Given the CS's winner count, each ES optimizes its bidder selection strategy. The CS then anticipates the ESs' equilibrium responses to determine the optimal winner number, leading to the following result.

Theorem 1 *There exists a DBSE (α, ζ) for the DBSG $< \mathcal{R}, \mathcal{S}, \mathcal{F}, \mathcal{U} >$ under bid distribution ϕ.*

5.2 BID DISTRIBUTION ESTIMATION UNDER THE BID-AGNOSTIC SETTING

The CS and ESs do not observe the current bids before choosing their actions. Fix a decision pair (α, ζ), and consider the λ auctions in which this pair was used. Let p_i be the observed $(\zeta+1)$-st lowest bid in the i-th such auction. For $x \in [0, 1]$, the empirical CDF of these payment thresholds is

$$D(x) = \frac{1}{\lambda} \sum_{i=1}^{\lambda} \mathbf{1}[p_i \leq x]. \tag{12}$$

Here D is the empirical CDF of payment thresholds observed under the fixed pair, rather than the bid CDF ϕ. The bid CDF can be estimated from D by inverting the CDF of the $(\zeta+1)$-st order statistic. For analysis, we use a separate normalized proxy model with $\theta = M\alpha/\zeta$ and $\varphi = 1 - \phi$, as follows:

Lemma 2 *Under the normalized proxy model, the threshold has survival function $G(x) = \varphi(x)^\theta$.*

The proxy survival function G generally differs from that of the observed payment threshold, whose empirical CDF is D. Given λ independent proxy threshold samples, let $\widehat{G}$ denote their empirical survival function and set $\widehat{\varphi}(x) = \widehat{G}(x)^{1/\theta}$. The following bounds apply to these proxy samples, not the realized auction thresholds in Eq. (12).

Theorem 2 *Given λ independent normalized proxy threshold samples from a fixed bid distribution, for any fixed x and $\theta > 1$, the following estimation error bounds hold:*

$$\mathbb{E}\left[\left|\hat{\varphi}^l(x) - \varphi^l(x)\right|\right] \leq \begin{cases} \dfrac{1}{2\sqrt{\lambda}}, & \text{if } l = \theta, \\ \left(\dfrac{1}{4\lambda}\right)^{\frac{\theta-1}{2\theta}}, & \text{else if } l = \theta - 1. \end{cases} \tag{13}$$

Remark. The bound depends on the sample size λ and exponent θ. Increasing λ tightens the error bounds. For the $(\theta-1)$-th power term, the decay exponent $\frac{\theta-1}{2\theta}$ increases toward $\frac{1}{2}$ as θ grows. Within the normalized proxy, larger θ places more weight on lower-tail bids. These bounds do not directly quantify estimation error under realized payment-threshold feedback.

5.3 THEORETICAL ANALYSIS OF DATAFED

In the bid-aware setting, we characterize a DBSE of the normalized proxy game. In the bid-agnostic setting, Algorithm 1 estimates the bid distribution from historical payment thresholds and uses it to adapt hierarchical client selection. Algorithm 1 takes the ES and client sets as input and returns the candidate and winner counts (α, ζ) (lines 1–2). It initializes $(\alpha_1, \zeta_1) = (1, 1)$ and records the first payment threshold (line 3). In each subsequent round, it retains the previous action pair if its threshold observation count $n_t \leq \delta \log t$ (lines 4–7). Otherwise, it estimates the bid CDF from the

Table 2: Cumulative expected regret performance.

		M=5				M=10			
		MNIST	FEMNIST	CIFAR-10	SVHN	MNIST	FEMNIST	CIFAR-10	SVHN
$\mathcal{N}_m = 25$	DaringFed	448.6±162.2	207.2±153.6	503.1±181.9	357.0±120.3	537.8±110.9	**104.5±108.2**	493.1±167.0	210.6±60.1
	RLB-AFL	194.4±82.5	349.4±72.3	386.7±144.7	417.0±128.2	299.1±54.0	379.0±104.6	351.1±130.6	254.7±73.7
	DualGFL	324.2±73.6	380.9±70.5	401.6±146.9	302.3±100.3	349.9±45.0	421.1±111.8	370.9±137.7	**181.7±50.2**
	TBCIM	243.4±43.9	307.2±53.5	256.9±80.3	360.4±129.3	243.4±19.6	322.2±74.0	245.0±82.5	204.5±61.5
	DataFed	**165.9±47.7**	**189.6±68.0**	**215.2±109.8**	**154.5±38.3**	**152.1±33.2**	146.0±40.3	**192.2±57.4**	190.8±84.4
$\mathcal{N}_m = 50$	DaringFed	645.0±132.6	425.0±108.2	566.7±201.1	329.9±99.3	412.1±49.3	341.7±62.5	549.8±197.4	329.9±99.3
	RLB-AFL	299.1±54.0	567.3±132.6	578.7±216.1	295.6±81.9	721.4±105.4	616.2±146.8	524.9±198.9	295.6±81.9
	DualGFL	349.9±45.0	421.1±111.8	370.9±137.7	204.4±58.1	342.8±46.6	442.1±100.5	336.3±128.6	204.4±58.1
	TBCIM	278.7±26.5	357.7±77.8	268.8±84.6	243.7±75.0	259.8±33.7	369.9±75.0	**255.6±89.3**	243.7±75.0
	DataFed	**182.7±43.9**	**98.3±35.6**	**207.2±52.6**	**165.1±75.3**	**132.5±17.5**	**104.1±13.4**	309.6±267.4	**165.1±75.3**

Algorithm 1: Bid-Agnostic Learning in DataFed

1: **Input:** $\mathcal{M}, \mathcal{N}, \{\mathcal{N}_m | m \in \mathcal{M}\}, \{\mathcal{A}_m | m \in \mathcal{M}\}$
2: **Output:** (α, ζ)
3: $(\alpha_1, \zeta_1) = (1, 1)$; observe $b^1_{\zeta_1+1}$
4: **for** $t \geq 2$ **do**
5: **if** $n_t \leq \delta \log t$ **then**
6: $\alpha_t \leftarrow \alpha_{t-1}$
7: $\zeta_t \leftarrow \zeta_{t-1}$
8: **else**
9: Estimate the bid CDF from D
10: $\alpha^*(\zeta) \leftarrow \arg\max_\alpha \widetilde{v}(\alpha, \zeta)$
11: $\zeta_t \leftarrow \arg\max_\zeta \widetilde{s}(\alpha^*(\zeta), \zeta)$
12: $\alpha_t \leftarrow \alpha^*(\zeta_t)$
13: **end if**
14: Observe $b^t_{\zeta_t+1}$
15: Update the current pair's D with $b^t_{\zeta_t+1}$
16: **end for**

pair's empirical payment-threshold CDF D using the order-statistic relation (lines 8–9). Using this estimate, it computes the ES best response $\alpha^*(\zeta)$ for each feasible ζ (line 10). The CS anticipates these responses to select ζ_t, then sets $\alpha_t = \alpha^*(\zeta_t)$ (lines 11–12). Finally, it observes $b^t_{\zeta_t+1}$ and updates D for the selected pair (lines 14–16). Estimating the empirical bid distribution from λ observations costs $O(\lambda)$. Exhaustive strategy evaluation over $1 \leq \alpha \leq |\mathcal{A}^t_m|$ and $1 \leq \zeta \leq M\alpha - 1$ costs $O(M|\mathcal{A}^t_m|^2)$. Thus, the per-round complexity is $O(\lambda + M|\mathcal{A}^t_m|^2)$, upper bounded by $O(\lambda + M|\mathcal{N}_m|^2)$.

Let $\mathcal{A}$ be a finite set of candidate-count actions, each paired with a feasible winner count ζ_α. Given bids independently drawn from a fixed distribution, the proxy learner receives independent normalized proxy feedback. After sampling each action $m_T = \lceil \delta \log T \rceil$ times for $\delta > 0$, it selects the action with the highest estimated proxy utility. Here, $R(T)$ denotes the learner's cumulative expected regret against the normalized proxy benchmark with a known bid distribution.

Theorem 3 *The cumulative expected regret is* $R(T) = O\big(T(\log T)^{-\beta}\big)$, *where* $\beta = \min_{\alpha \in \mathcal{A}}(M\alpha - \zeta_\alpha)/(2M\alpha)$.

Remark. Each feasible action satisfies $1 \leq \zeta_\alpha \leq M\alpha - 1$. Since $\mathcal{A}$ is finite, $\beta > 0$. Thus, average proxy regret vanishes as T grows, implying sublinear cumulative proxy regret. This bound applies under independent normalized proxy feedback and the stated exploration schedule. It does not directly bound Algorithm 1's regret under actual auction feedback.

6 Experiments

In this section, we evaluate DataFed on four real-world and one synthetic datasets to demonstrate its effectiveness for bid-agnostic auction-based online client selection in OFEL.

Performance on cumulative expected regret. We evaluate DataFed on MNIST, FEMNIST, SVHN, and CIFAR-10 in terms of cumulative regret, and compare it against DaringFed (Xin et al., 2025), RLB-AFL (Tang et al., 2025), DualGFL (Chen et al., 2025), and TBCIM (Ma et al., 2025). Detailed real dataset setups and baseline descriptions are provided in Appendix A.7.1 and Appendix A.7.2. As shown in Table 2, DataFed reduces mean cumulative regret by 23.0% on average across all tested configurations. We compute the percentage reduction against the baseline with the lowest mean regret in each configuration, then average the reductions with equal weights. Its plotted regret curves usually grow more slowly than those of the baselines over the reported horizons. In particular, DataFed typically undergoes a short exploration phase before entering a stable regime with slow

Table 3: Accuracy, payment, and cumulative expected regret.

	MNIST			CIFAR-10		
Method	Final accuracy	Total payment	Cumulative regret	Final accuracy	Total payment	Cumulative regret
DaringFed	$93.00\% \pm 1.88\%$	111.99 ± 46.60	84.97 ± 57.41	$78.02\% \pm 0.20\%$	215.80 ± 62.00	219.74 ± 53.33
RLB-AFL	$91.54\% \pm 3.76\%$	117.49 ± 43.99	167.82 ± 81.40	$77.71\% \pm 0.65\%$	277.78 ± 117.71	404.67 ± 67.37
DualGFL	$90.84\% \pm 3.36\%$	104.19 ± 38.08	124.78 ± 68.39	$77.88\% \pm 0.18\%$	253.27 ± 96.26	322.67 ± 53.60
TBCIM	$93.01\% \pm 1.50\%$	157.51 ± 37.94	139.60 ± 53.58	$78.01\% \pm 0.24\%$	271.39 ± 44.24	322.67 ± 50.38
DataFed	$92.82\% \pm 0.88\%$	133.26 ± 43.47	$\mathbf{73.14 \pm 15.46}$	$78.23\% \pm 0.38\%$	267.83 ± 36.69	$\mathbf{121.88 \pm 30.63}$

Table 4: CDF Error estimation under different $\frac{M\alpha}{\zeta}$ and λ

$\frac{M\alpha}{\zeta}$	$\lambda = 50$				$\lambda = 100$				$\lambda = 300$			
	0.2	0.3	0.4	0.5	0.2	0.3	0.4	0.5	0.2	0.3	0.4	0.5
5	0.136	0.136	0.136	0.145	0.125	0.125	0.125	0.137	**0.128**	**0.128**	**0.128**	**0.139**
15	0.057	0.057	0.057	0.065	0.073	0.073	0.073	0.083	**0.064**	**0.064**	**0.064**	**0.065**
25	0.800	0.800	0.800	0.780	0.800	0.800	0.800	0.780	**0.004**	**0.004**	**0.004**	**0.016**

regret growth, indicating that the learned bid distribution progressively improves hierarchical client selection over time. DataFed combines bid distribution estimation from auction payment thresholds with hierarchical decisions that balance utility and payment under incomplete information. Detailed regret curves are provided in Appendix A.6.1. Appendix A.7.5 shows the benefit of anticipating ES responses. Over 800 rounds, DataFed reduces mean cumulative regret by 11.0% and increases mean per-round cloud utility by 7.9% compared with sequential selection without anticipation.

Accuracy and payment. Table 3 reports final accuracy, total payment, and cumulative expected regret on MNIST and CIFAR-10. Results are means and 95% confidence intervals over five random seeds. Bold values indicate the lowest mean cumulative regret for each dataset. DataFed achieves the lowest mean regret on both datasets, with final accuracy close to the best baseline on MNIST and the highest mean accuracy on CIFAR-10. Its total payment is not the lowest, highlighting the importance of evaluating accuracy and payment jointly. Detailed discussion is provided in Appendix A.7.6.

The effects of λ and θ on bid distribution estimation. We conduct synthetic experiments on bid distribution estimation and hierarchical selection. Detailed synthetic dataset setups are provided in Appendix A.7.1. As shown in Table 4, the estimation accuracy is jointly affected by the sampling frequency λ and competition parameter θ. Moderate θ yields relatively stable estimation across different λ, whereas large θ produces high errors with limited samples and much lower errors with sufficient samples. This pattern reflects stronger lower-tail concentration at large θ, making estimation more sensitive to sample size. Appendix A.7.3 further presents synthetic experiments on utility-payment trade-offs, hierarchical selection behaviors, and bid distribution estimation through detailed curves. In the tested synthetic settings, the proxy-selected actions retain 90.18% and 92.32% of the exact optimal utility for $M = 5$ and $M = 10$, respectively. Both the proxy-selected and exact optimal actions are evaluated using the original multi-winner utility. Detailed comparisons are provided in Appendix A.7.7. Under gradual bid drift, DataFed's final cumulative regret remains close to the stationary result, while shifts every 40 rounds prevent recovery within each regime (Appendix A.7.4).

7 Conclusion

In this paper, we have proposed DataFed for bid-agnostic hierarchical client selection in online federated edge learning. DataFed estimates bid distributions from historical auction payment thresholds and uses a Stackelberg-based auction rule to coordinate candidate and winner counts before observing current bids. For the normalized proxy model, we establish equilibrium existence and estimation error bounds. Given independent bids from a fixed distribution, the proxy learner achieves sublinear cumulative expected regret with independent proxy feedback and sufficient exploration. Experiments on four real-world datasets show a 23.0% average reduction in mean cumulative regret relative to the best baseline in each setting. Coordination experiments show a 7.9% increase in mean cloud utility, while synthetic results illustrate utility and payment trade-offs.

AI USE STATEMENT

Generative AI helped us improve the language and presentation of the manuscript by refining wording, correcting grammar, and removing repetition. We reviewed and checked all suggested changes and made the final decisions on the paper's content.

ETHICS STATEMENT

This study uses benchmark datasets and simulated incentives. It does not collect new data from human participants or involve real payments. Any practical deployment should protect client data, ensure informed participation, and follow applicable regulations.

REPRODUCIBILITY STATEMENT

The main paper describes the assumptions and algorithm. Appendices A.1–A.6 contain the proofs, while Appendix A.7 provides the experimental settings and additional results. The supplementary material provides the implementation and experimental records needed to reproduce the reported results.

REFERENCES

Siqi Luo, Xu Chen, Qiong Wu, Zhi Zhou, and Shuai Yu. HFEL: Joint Edge Association and Resource Allocation for Cost-efficient Hierarchical Federated Edge Learning. *IEEE Transactions on Wireless Communications*, 19(10):6535–6548, 2020.

Jinke Ren, Yinghui He, Dingzhu Wen, Guanding Yu, Kaibin Huang, and Dongning Guo. Scheduling for Cellular Federated Edge Learning with Importance and Channel Awareness. *IEEE Transactions on Wireless Communications*, 19(11):7690–7703, 2020.

Yun Xin, Jianfeng Lu, Shuqin Cao, Gang Li, Haozhao Wang, and Guanghui Wen. DaringFed: A Dynamic Bayesian Persuasion Pricing for Online Federated Learning under Two-Sided Incomplete Information. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 6687–6695. Montreal, Quebec: IJCAI Organization, 2025.

Yujing Chen, Yue Ning, Martin Slawski, and Huzefa Rangwala. Asynchronous Online Federated Learning for Edge Devices with Non-IID Data. In *Proceedings of the IEEE International Conference on Big Data*, pp. 15–24. Atlanta, GA, USA: IEEE, 2020.

Munyque Mittelmann, Bastien Maubert, Aniello Murano, and Laurent Perrussel. Formal Verification of Bayesian Mechanisms. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37(10), pp. 11621–11629. USA: AAAI Press, 2023.

Yutao Jiao, Ping Wang, Dusit Niyato, Bin Lin, and Dong In Kim. Toward an Automated Auction Framework for Wireless Federated Learning Services Market. *IEEE Transactions on Mobile Computing*, 20(10):3034–3048, 2021.

Yuan Deng, Jieming Mao, Vahab Mirrokni, Hanrui Zhang, and Song Zuo. Efficiency of the First-price Auction in the Autobidding World. In *Advances in Neural Information Processing Systems 37*, pp. 139270–139293. Vancouver, Canada: Curran Associates, Inc, 2024.

Tra Huong Thi Le, Nguyen H. Tran, Yan Kyaw Tun, Minh N. H. Nguyen, Shashi Raj Pandey, Zhu Han, and Choong Seon Hong. An Incentive Mechanism for Federated Learning in Wireless Cellular Networks: An Auction Approach. *IEEE Transactions on Wireless Communications*, 20(8): 4874–4887, 2021.

Renhao Lu, Weizhe Zhang, Yan Wang, Qiong Li, Xiaoxiong Zhong, Hongwei Yang, and Desheng Wang. Auction-Based Cluster Federated Learning in Mobile Edge Computing Systems. *IEEE Transactions on Parallel and Distributed Systems*, 34(4):1145–1158, 2023.

Miao Hu, Wenzhuo Yang, Zhenxiao Luo, Xuezheng Liu, Yipeng Zhou, Xu Chen, and Di Wu. AutoFL: A Bayesian Game Approach for Autonomous Client Participation in Federated Edge Learning. *IEEE Transactions on Mobile Computing*, 23(1):194–208, 2024.

Billy Jin, Thomas Kesselheim, Will Ma, and Sahil Singla. Sample Complexity of Posted Pricing for a Single Item. In *Advances in Neural Information Processing Systems 37*, pp. 82296–82317. Vancouver, Canada: Curran Associates, Inc, 2024.

Linfeng Liu, Zhipeng Zhang, Jia Xu, and Ping Wang. A Bayesian Game based Bidding Scheme for Mobile Charging Services in IoEV. *IEEE Transactions on Services Computing*, 17(2):349–363, 2024.

Jiasi Weng, Jian Weng, Hongwei Huang, Chengjun Cai, and Cong Wang. FedServing: A Federated Prediction Serving Framework based on Incentive Mechanism. In *Proceedings of the IEEE Conference on Computer Communications*, pp. 1–10. Vancouver, BC, Canada: IEEE, 2021.

Dorian Baudry, Hugo Richard, Maria Cherifa, Clément Calauzènes, and Vianney Perchet. Optimizing the Coalition Gain in Online Auctions with Greedy Structured Bandits. In *Advances in Neural Information Processing Systems 37*, pp. 19591–19635. Vancouver Convention Centre, Vancouver, BC, Canada: Curran Associates, Inc, 2024.

Bowen Zhao, Ximeng Liu, Wei-Neng Chen, and Robert H. Deng. CrowdFL: Privacy-Preserving Mobile Crowdsensing System via Federated Learning. *IEEE Transactions on Mobile Computing*, 22(8):4607–4619, 2023.

Yulan Yuan, Lei Jiao, Konglin Zhu, and Lin Zhang. Incentivizing Federated Learning under Long-term Energy Constraint via Online Randomized Auctions. *IEEE Transactions on Wireless Communications*, 21(7):5129–5144, 2022.

Yongheng Deng, Feng Lyu, Ju Ren, Yi-Chao Chen, Peng Yang, Yuezhi Zhou, and Yaoxue Zhang. FAIR: Quality-aware Federated Learning with Precise User Incentive and Model Aggregation. In *Proceedings of the IEEE Conference on Computer Communications*, pp. 1–10. Vancouver, BC, Canada: IEEE, 2021.

Georgios Damaskinos, Rachid Guerraoui, Anne-Marie Kermarrec, Vlad Nitu, Rhicheek Patra, and Francois Taiani. FLeet: Online Federated Learning via Staleness Awareness and Performance Prediction. *ACM Transactions on Intelligent Systems and Technology*, 13(5):1–30, 2022.

Pengchao Han, Shiqiang Wang, and Kin K. Leung. Adaptive Gradient Sparsification for Efficient Federated Learning: An Online Learning Approach. In *Proceedings of the IEEE International Conference on Distributed Computing Systems*, pp. 300–310. Singapore, Singapore. IEEE, 2020.

Zhe Qu, Rui Duan, Lixing Chen, Jie Xu, Zhuo Lu, and Yao Liu. Context-aware Online Client Selection for Hierarchical Federated Learning. *IEEE Transactions on Parallel and Distributed Systems*, 33(12):4353–4367, 2022.

Dong-Jun Han, Minseok Choi, Jungwuk Park, and Jaekyun Moon. FedMes: Speeding up Federated Learning with Multiple Edge Servers. *IEEE Journal on Selected Areas in Communications*, 39(12): 3870–3885, 2021.

Yann LeCun, Corinna Cortes, and Christopher J. C. Burges. MNIST Hand-written Digit Database. 2021.

Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading Digits in Natural Images with Unsupervised Feature Learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.

Alex Krizhevsky. Learning Multiple Layers of Features From Tiny Images. *Technical Report*, 2009.

Han Yu, Zelei Liu, Yang Liu, Tianjian Chen, Mingshu Cong, Xi Weng, Dusit Niyato, and Qiang Yang. A Fairness-aware Incentive Scheme for Federated Learning. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pp. 393–399. ACM, 2020.

Haifeng Xu, Rupert Freeman, Vincent Conitzer, Shaddin Dughmi, and Milind Tambe. Signaling in Bayesian Stackelberg Games. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pp. 150–158. Singapore: IFAAMAS, 2016.

Fengsen Tian, Xinglin Zhang, Xiumin Wang, and Yue-Jiao Gong. Two-layer Optimization with Utility Game and Resource Control for Federated Learning in Edge Networks. *IEEE Transactions on Mobile Computing*, 23(12):13834–13850, 2024.

Miao Hu, Di Wu, Yipeng Zhou, Xu Chen, and Min Chen. Incentive-aware Autonomous Client Participation in Federated Learning. *IEEE Transactions on Parallel and Distributed Systems*, 33(10):2612–2627, 2022.

Omar Abdel Wahab, Jamal Bentahar, Hadi Otrok, and Azzam Mourad. Resource-aware Detection and Defense System Against Multi-type Attacks in the Cloud: Repeated Bayesian Stackelberg Game. *IEEE Transactions on Dependable and Secure Computing*, 18(2):605–622, 2021.

Herbert A. David and Haikady N. Nagaraja. *Order Statistics*. 3rd edition. Wiley, 2004.

Guangjing Huang, Xu Chen, Tao Ouyang, Qian Ma, Lin Chen, and Junshan Zhang. Collaboration in Participant-centric Federated Learning: A Game-theoretical Perspective. *IEEE Transactions on Mobile Computing*, 22(11):6311–6326, 2023.

Youqi Li, Fan Li, Song Yang, Chuan Zhang, Liehuang Zhu, and Yu Wang. A Cooperative Analysis to Incentivize Communication-efficient Federated Learning. *IEEE Transactions on Mobile Computing*, 23(10):10175–10190, 2024.

Junfan Li, Zheshun Wu, Zenglin Xu, and Irwin King. On the Necessity of Collaboration for Online Model Selection with Decentralized Data. In *Advances in Neural Information Processing Systems 37*, pp. 85583–85629. Vancouver, Canada: Curran Associates, Inc, 2024.

Sébastien Bubeck, Nikhil R. Devanur, Zhiyi Huang, and Rad Niazadeh. Multi-scale Online Learning: Theory and Applications to Online Auctions and Pricing. *Journal of Machine Learning Research*, 20(62):1–37, 2019.

Wenzhi Fang, Dong-Jun Han, Evan Chen, Shiqiang Wang, and Christopher G. Brinton. Hierarchical Federated Learning with Multi-timescale Gradient Correction. In *Advances in Neural Information Processing Systems 37*, pp. 78863–78904. Vancouver, Canada: Curran Associates, Inc, 2024.

Yizhou Shen, Carlton Shepherd, Chuadhry Mujeeb Ahmed, Shigen Shen, and Shui Yu. Privacy Preservation Strategies for Malware-infected Edge Intelligence Systems: A Bayesian Stochastic Game-based Approach. *IEEE Transactions on Mobile Computing*, 24(8):7121–7135, 2025.

Xiaoyi Wu and Bin Li. Achieving Regular and Fair Learning in Combinatorial Multi-armed Bandit. In *Proceedings of the IEEE Conference on Computer Communications*, pp. 361–370. Vancouver, BC, Canada: IEEE, 2024.

Shufan Wang, Guojun Xiong, and Jian Li. Online Restless Multi-armed Bandits with Long-term Fairness Constraints. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38(14), pp. 15616–15624. Vancouver, BC, Canada: AAAI Press, 2024.

Shipra Agrawal, Yiding Feng, and Wei Tang. Dynamic Pricing and Learning with Bayesian Persuasion. In *Advances in Neural Information Processing Systems 36*, pp. 59273–59285. New Orleans, Louisiana, United States: Curran Associates, Inc, 2023.

Sebastian Caldas, Sai Meher Karthik Duddu, Peter Wu, Tian Li, Jakub Konečný, H. Brendan McMahan, Virginia Smith, and Ameet Talwalkar. LEAF: A Benchmark for Federated Settings. *arXiv preprint arXiv:1812.01097*, 2018.

Lianbo Ma, Ying Qian, Guo Yu, Zhetao Li, Liang Wang, Qing Li, Xingwei Wang, and Guangjie Han. TBCIM: Two-Level Blockchain-Aided Edge Resource Allocation Mechanism for Federated Learning Service Market. *IEEE Transactions on Networking*, 33(6):3408–3423, 2025.

Xiaoli Tang, Han Yu, and Xiaoxiao Li. A Reinforcement Learning-based Bidding Strategy for Data Consumers in Auction-based Federated Learning. In *Advances in Neural Information Processing Systems*, volume 38, pp. 82666–82684. Curran Associates, Inc., 2025.

Xiaobing Chen, Xiangwei Zhou, Songyang Zhang, and Mingxuan Sun. DualGFL: Federated Learning with a Dual-Level Coalition-Auction Game. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39(15), pp. 15904–15912. Philadelphia, PA, USA: AAAI Press, 2025.

A Appendix

A.1 Proof of Proposition 1

Fix a round and all bids except that of client n. Let p_{-n} be the $(k-1)$-st lowest bid among the other clients. This critical value does not change when client n changes its own report. If $b_n < p_{-n}$, client n is selected and receives p_{-n}, so its utility is $p_{-n} - c_n$. Otherwise, its utility is zero. Under the continuous bid model, ties occur with probability zero.

If $c_n < p_{-n}$, truthful bidding selects the client and gives nonnegative utility. Any other winning bid gives the same utility, while a losing bid gives zero. If $c_n \geq p_{-n}$, truthful bidding gives zero utility. A deviation that wins gives utility $p_{-n} - c_n \leq 0$, while a deviation that loses also gives zero. Thus truthful bidding weakly dominates every alternative report in this round. Because a report can affect later learning, this argument does not establish full dynamic incentive compatibility for the history-dependent online mechanism.

A.2 Proof of Lemma 1

Since the clients' bids are i.i.d. and the utility structures are identical across ESs, the game is symmetric. We therefore focus on symmetric equilibria, where all ESs adopt identical strategies, i.e., $\alpha_m = \alpha$ for all m. Hence, there are $M\alpha$ bidders in total across all ESs, and the CS selects ζ clients with the lowest bids as winners under the ζ-reverse Vickrey auction.

Selecting ζ winners from $M\alpha$ bidders involves a combinatorial decision space of size $\binom{M\alpha}{\zeta}$, which is difficult to analyze directly. To address this challenge, we focus on the expected number of selected bidders and exploit the symmetry induced by i.i.d. bids. Based on this, we establish an expectation-level normalization relation that connects the ζ-winner setting to a corresponding single-winner formulation, which significantly simplifies the analysis of utilities.

A.2.1 Normalization Mapping for Hierarchical Client Selection.

Since the clients' bids are i.i.d. and the utility structures are identical across ESs, the game is symmetric. We therefore focus on symmetric equilibria, where all ESs adopt identical strategies, i.e., $\alpha = \alpha_m$ for all m. Hence, there are $M\alpha$ bidders in total across all ESs, and the CS selects ζ clients with the lowest bids as winners under the ζ-reverse Vickrey auction.

Let $z_{\alpha,\zeta}$ denote the number of winners from ES m. It can be written as

$$z_{\alpha,\zeta} = \sum_{n=1}^{\alpha} \mathbf{1}[b_n < b_{\zeta+1}], \tag{14}$$

where $b_{\zeta+1}$ is the $\zeta+1$-th order statistic among all $M\alpha$ bids.

Since all bids are i.i.d., all bidders are statistically symmetric. Therefore, each bidder has the same probability of being selected among the ζ lowest bids, namely

$$\mathbb{P}(b_n < b_{\zeta+1}) = \frac{\zeta}{M\alpha}. \tag{15}$$

Taking expectation on both sides yields

$$\mathbb{E}[z_{\alpha,\zeta}] = \mathbb{E}[\sum_{n=1}^{\alpha} \mathbf{1}[b_n < b_{\zeta+1}]] = \sum_{n=1}^{\alpha} \mathbb{E}[\mathbf{1}[b_n < b_{\zeta+1}]] = \alpha \cdot \frac{\zeta}{M\alpha} = \frac{\zeta}{M}. \tag{16}$$

To simplify the analysis and reveal a scaling structure, we relate the ζ-winner setting to a corresponding single-winner normalized at the expectation level. Here $(\alpha/\zeta, 1)$ is an expectation-level scaling and need not describe an integer number of bidders. The normalized expression is

$$\mathbb{E}[z_{\alpha/\zeta,1}] = \frac{\alpha/\zeta}{M\alpha/\zeta} = \frac{1}{M}. \tag{17}$$

Multiplying both sides by ζ, we obtain

$$\zeta\mathbb{E}[z_{\alpha/\zeta,1}] = \mathbb{E}[z_{\alpha,\zeta}], \tag{18}$$

which establishes the expectation-level normalization relation.

A.2.2 BENCHMARK UTILITY OF THE ES.

We next introduce a tractable benchmark for the expected utility of ES m based on the expectation-level normalization established above.

Under symmetric equilibria, there are $M\alpha$ bidders in total, among which α belong to ES m. The CS selects the ζ lowest bids as winners under the ζ-reverse Vickrey auction. Although the normalization relation is exact for the expected number of winners, it does not directly imply an exact utility equivalence, since the payment in the original multi-winner mechanism is determined by the global $(\zeta+1)$-th lowest bid. To maintain analytical tractability, we therefore use a normalized single-winner formulation as a benchmark for the marginal surplus generated by competition at the bottom of the bid distribution.

Specifically, let $\theta = M\alpha/\zeta > 1$ be the effective competition parameter of the normalized proxy. We define $b_{(1)}$ and $b_{(2)}$ as analytical proxy variables through the survival functions below. When θ is an integer of at least two, these coincide with the lowest and second-lowest bids in a θ-bidder market. By symmetry across ESs, we assign ES m a $1/M$ share of the proxy surplus. Hence, we define the benchmark utility of ES m as

$$\widetilde{v}(\alpha,\zeta) := \frac{\zeta}{M}\,\mathbb{E}[b_{(2)} - b_{(1)}], \tag{19}$$

which captures the expected marginal surplus induced by ζ winners through an expectation-level scaled single-winner representation.

To compute the proxy surplus, we use the following survival functions. Let $\phi(x)$ denote the CDF of bids. Then,

$$\begin{cases} \mathbb{P}(b_{(1)} > x) = [1-\phi(x)]^{\theta}, \\ \mathbb{P}(b_{(2)} > x) = [1-\phi(x)]^{\theta} + \theta\phi(x)[1-\phi(x)]^{\theta-1}. \end{cases} \tag{20}$$

For real $\theta > 1$, the second expression in Eq. (20) defines a valid survival function. Writing $u = 1-\phi(x)$, its derivative with respect to u is $\theta(\theta-1)(1-u)u^{\theta-2} \geq 0$ for $u \in (0,1)$, and it is no smaller than the first expression. Using the identity $\mathbb{E}[X] = \int_0^1 \mathbb{P}(X > x)dx$, we obtain

$$\mathbb{E}[b_{(2)} - b_{(1)}] = \int_0^1 \left(\mathbb{P}(b_{(2)} > x) - \mathbb{P}(b_{(1)} > x)\right) dx. \tag{21}$$

Substituting (23) yields

$$\mathbb{E}[b_{(2)} - b_{(1)}] = \int_0^1 \theta\phi(x)[1-\phi(x)]^{\theta-1}dx. \tag{22}$$

Therefore, the benchmark utility of ES m becomes

$$\widetilde{v}(\alpha,\zeta) = \frac{\zeta}{M}\int_0^1 \theta\phi(x)[1-\phi(x)]^{\theta-1}dx. \tag{23}$$

Substituting $\theta = \frac{M\alpha}{\zeta}$, we further obtain

$$\widetilde{v}(\alpha,\zeta) = \alpha\int_0^1 \phi(x)[1-\phi(x)]^{\theta-1}dx. \tag{24}$$

By introducing the survival function $\varphi(x) = 1-\phi(x)$, we can alternatively express this as:

$$\widetilde{v}(\alpha,\zeta) = \alpha\int_0^1 \varphi^{\theta-1}(x) - \varphi^{\theta}(x)dx. \tag{25}$$

This benchmark will be used in the following analysis as a tractable proxy for the expected utility of ES m.

A.2.3 BENCHMARK UTILITY OF THE CS.

We next construct a tractable benchmark for the expected utility of the CS, following the same expectation-level normalization used in the ES benchmark.

Suppose that under the ζ-reverse Vickrey auction, there are $M\alpha$ bidders in total. In the original multi-winner mechanism, the CS selects the ζ lowest bids as winners, and the payment to each winner is determined by the global $(\zeta+1)$-th lowest bid. Since this payment threshold is analytically cumbersome, we use the normalized proxy with effective competition parameter $\theta = M\alpha/\zeta > 1$ and take $b_{(2)}$ as an analytical proxy for the payment threshold.

We define the benchmark utility of the CS as

$$\widetilde{s}(\zeta, \alpha) := q(\zeta) - \zeta\, \mathbb{E}[b_{(2)}]. \tag{26}$$

To compute $\mathbb{E}[b_{(2)}]$, let $\phi(x)$ denote the CDF of bids on $[0,1]$. By the proxy survival function,

$$\mathbb{P}(b_{(2)} > x) = [1-\phi(x)]^{\theta} + \theta\phi(x)[1-\phi(x)]^{\theta-1}. \tag{27}$$

Using the identity $\mathbb{E}[X] = \int_0^1 \mathbb{P}(X > x)\, dx$ for nonnegative $X \in [0,1]$, we obtain

$$\mathbb{E}[b_{(2)}] = \int_0^1 \left([1-\phi(x)]^{\theta} + \theta\phi(x)[1-\phi(x)]^{\theta-1}\right)\, dx. \tag{28}$$

Therefore, the benchmark utility of the CS becomes

$$\widetilde{s}(\zeta, \alpha) = q(\zeta) - \zeta \int_0^1 \left([1-\phi(x)]^{\theta} + \theta\phi(x)[1-\phi(x)]^{\theta-1}\right)\, dx, \tag{29}$$

where $\theta = \frac{M\alpha}{\zeta}$. Equivalently, this can be written as

$$\widetilde{s}(\zeta, \alpha) = q(\zeta) - \zeta \int_0^1 [1-\phi(x)]^{\theta-1}\big(1 + (\theta-1)\phi(x)\big)\, dx. \tag{30}$$

By introducing the survival function $\varphi(x) = 1 - \phi(x)$, we can alternatively express this as:

$$\widetilde{s}(\zeta, \alpha) = q(\zeta) - \zeta \int_0^1 (1-\theta)\varphi^{\theta}(x) + \theta\varphi^{\theta-1}(x)\, dx. \tag{31}$$

This benchmark will be used in the following analysis as a tractable proxy for the expected utility of the CS.

A.3 Proof of Theorem 1

For each feasible winner count $\zeta \in \mathcal{Z}$, let $\mathcal{A}(\zeta)$ be the nonempty finite set of feasible ES candidate counts, including the constraint $M\alpha \geq \zeta + 1$. The ES best-response set is

$$BR(\zeta) = \arg\max_{\alpha \in \mathcal{A}(\zeta)} \widetilde{v}(\alpha, \zeta). \tag{32}$$

When several ES responses maximize $\widetilde{v}$, we use the response that maximizes the CS proxy utility. A DBSE (α^*, ζ^*) therefore satisfies

$$\alpha^* \in BR(\zeta^*), \qquad \zeta^* \in \arg\max_{\zeta \in \mathcal{Z}} \max_{\alpha \in BR(\zeta)} \widetilde{s}(\alpha, \zeta), \qquad \alpha^* \in \arg\max_{\alpha \in BR(\zeta^*)} \widetilde{s}(\alpha, \zeta^*). \tag{33}$$

This definition follows the CS commitment and the subsequent ES response in the model.

For every $\zeta \in \mathcal{Z}$, the set $\mathcal{A}(\zeta)$ is nonempty and finite. Hence $\widetilde{v}(\alpha, \zeta)$ attains a maximum over $\mathcal{A}(\zeta)$, and $BR(\zeta)$ is nonempty.

For each $\zeta \in \mathcal{Z}$, the maximum of $\widetilde{s}(\alpha, \zeta)$ over the nonempty finite set $BR(\zeta)$ exists. Since $\mathcal{Z}$ is nonempty and finite, this induced CS utility also attains a maximum at some $\zeta^* \in \mathcal{Z}$.

Choose α^* from the CS-favorable best responses in $BR(\zeta^*)$. The resulting pair (α^*, ζ^*) satisfies the preceding DBSE definition. Thus a DBSE exists for the finite feasible proxy game.

A.4 Proof of Lemma 2

A.4.1 Preliminaries on Order Statistics

Let $X_1, \ldots, X_n$ be i.i.d. random variables drawn from a continuous distribution with cumulative distribution function (CDF) $\phi(x)$. Denote the associated survival function by

$$\varphi(x) = 1 - \phi(x).$$

Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ denote the order statistics (David & Nagaraja, 2004). The CDF of the minimum order statistic $X_{(1)}$ is

$$\mathbb{P}(X_{(1)} \le x) = 1 - \mathbb{P}(X_1 > x, \ldots, X_n > x). \tag{34}$$

By independence,

$$\mathbb{P}(X_1 > x, \ldots, X_n > x) = \prod_{i=1}^{n} \mathbb{P}(X_i > x) = \varphi^n(x). \tag{35}$$

Hence,

$$\mathbb{P}(X_{(1)} \le x) = 1 - \varphi^n(x). \tag{36}$$

More generally, the CDF of the k-th order statistic $X_{(k)}$ is

$$\mathbb{P}(X_{(k)} \le x) = \sum_{j=k}^{n} \binom{n}{j} \phi(x)^j \varphi(x)^{n-j}. \tag{37}$$

A.4.2 Application to the ζ-Reverse Vickrey Auction

In the ζ-reverse Vickrey auction, there are in total $M\alpha$ bidders, whose bids are drawn i.i.d. from $\phi(x)$. The mechanism selects the ζ lowest bids as winners, and each winner is paid according to the $(\zeta+1)$-th lowest bid. Therefore, the payment threshold is determined by the order statistic $X_{(\zeta+1)}$. By (37), the exact CDF of the payment threshold is

$$\mathbb{P}\big(X_{(\zeta+1)} \le x\big) = \sum_{j=\zeta+1}^{M\alpha} \binom{M\alpha}{j} \phi(x)^j \varphi(x)^{M\alpha-j}. \tag{38}$$

A.4.3 Normalized Benchmark Representation

The exact CDF in Eq. (38) is analytically cumbersome. We therefore define an expectation-level normalized proxy with the real-valued competition parameter $\theta = M\alpha/\zeta$. The proxy is an analytical distribution; its definition does not require an integer number of synthetic bidders. Let

$$D^{\text{proxy}}_{\alpha,\zeta}(x) := 1 - \varphi(x)^{\theta}. \tag{39}$$

Equivalently,

$$D^{\text{proxy}}_{\alpha,\zeta}(x) = 1 - \varphi(x)^{M\alpha/\zeta}. \tag{40}$$

Define its survival function by

$$G^{\text{proxy}}_{\alpha,\zeta}(x) := 1 - D^{\text{proxy}}_{\alpha,\zeta}(x). \tag{41}$$

Then

$$G^{\text{proxy}}_{\alpha,\zeta}(x) = \varphi(x)^{M\alpha/\zeta}. \tag{42}$$

The proxy CDF in Eq. (40) differs from the exact payment-threshold CDF in Eq. (38). It retains the lower-tail competition dependence used by the proxy utilities in Eq. (8), while the gap to the original multi-winner objective remains an approximation error. For the fixed pair in Lemma 2, G abbreviates $G^{\text{proxy}}_{\alpha,\zeta}$, whereas D in Eq. (12) is the empirical CDF of realized payment thresholds.

A.5 Proof of Theorem 2

Fix an action (α, ζ) and let $\theta = M\alpha/\zeta > 1$. This proof concerns λ independent observations $Y_1, \ldots, Y_\lambda$ from the normalized proxy distribution $D^{\text{proxy}}_{\alpha,\zeta}$, rather than the realized multi-winner payment thresholds in Eq. (12). Define

$$\widehat{G}^{\text{proxy}}_{\alpha,\zeta}(x) = \frac{1}{\lambda}\sum_{i=1}^{\lambda} \mathbf{1}\{Y_i > x\}, \qquad \widehat{\varphi}(x) = \left(\widehat{G}^{\text{proxy}}_{\alpha,\zeta}(x)\right)^{1/\theta}. \tag{43}$$

With the action pair fixed, the main text abbreviates $\widehat{G}^{\text{proxy}}_{\alpha,\zeta}$ as $\widehat{G}$. For fixed x, each indicator is Bernoulli with mean $u = \varphi(x)^\theta$ by Eq. (42). Consequently,

$$\mathbb{E}[\widehat{G}^{\text{proxy}}_{\alpha,\zeta}(x)] = u, \qquad \text{Var}(\widehat{G}^{\text{proxy}}_{\alpha,\zeta}(x)) = \frac{u(1-u)}{\lambda} \le \frac{1}{4\lambda}. \tag{44}$$

Since $\widehat{\varphi}^\theta = \widehat{G}^{\text{proxy}}_{\alpha,\zeta}$, Cauchy–Schwarz gives

$$\mathbb{E}\big[|\widehat{\varphi}^\theta(x) - \varphi^\theta(x)|\big] \le \sqrt{\text{Var}(\widehat{G}^{\text{proxy}}_{\alpha,\zeta}(x))} \le \frac{1}{2\sqrt{\lambda}}. \tag{45}$$

For $\rho = (\theta-1)/\theta \in (0,1)$ and $a, b \in [0,1]$, the inequality $|a^\rho - b^\rho| \le |a-b|^\rho$ applies. Monotonicity of L^p norms therefore yields

$$\begin{aligned}\mathbb{E}\big[|\widehat{\varphi}^{\theta-1}(x) - \varphi^{\theta-1}(x)|\big] &\le \mathbb{E}\Big[|\widehat{G}^{\text{proxy}}_{\alpha,\zeta}(x) - u|^\rho\Big] \\ &\le \Big(\text{Var}(\widehat{G}^{\text{proxy}}_{\alpha,\zeta}(x))\Big)^{\rho/2} \\ &\le \left(\frac{1}{4\lambda}\right)^{\frac{\theta-1}{2\theta}}.\end{aligned} \tag{46}$$

These are the two bounds in Eq. (13).

A.6 Proof of Theorem 3

Let $\mathcal{A}$ be the finite action set. For each action α, fix its associated feasible winner count ζ_α and write $\theta_\alpha = M\alpha/\zeta_\alpha > 1$. The bid distribution ϕ and these action parameters are fixed over the horizon. Define the proxy utility $s_\alpha = \widetilde{s}(\alpha, \zeta_\alpha; \phi)$ and let $\alpha^* \in \arg\max_{\alpha\in\mathcal{A}} s_\alpha$. For every action, the learner receives independent normalized proxy threshold samples and forms an estimated utility $\widehat{s}_\alpha$ from the empirical proxy survival function as in Theorem 2. These samples are distinct from the realized payment thresholds used by Algorithm 1.

For λ proxy samples of action α, the two error bounds in Eq. (13), the triangle inequality, and integration over $x \in [0,1]$ give

$$\mathbb{E}|\widehat{s}_\alpha - s_\alpha| \le \eta_\lambda(\alpha) := \zeta_\alpha \left[\frac{\theta_\alpha - 1}{2\sqrt{\lambda}} + \theta_\alpha \left(\frac{1}{4\lambda}\right)^{\frac{\theta_\alpha - 1}{2\theta_\alpha}}\right]. \tag{47}$$

The $q(\zeta_\alpha)$ terms cancel because the true and estimated utilities use the same action. The bound is nonincreasing in λ.

Consider the policy that samples each action $m_T = \lceil \delta \log T \rceil$ times and then selects an action maximizing its estimated proxy utility. For sufficiently large T, the finite exploration phase fits within the horizon. Its regret is at most

$$R^{\text{explore}}(T) \le m_T \sum_{\alpha\in\mathcal{A}} (s_{\alpha^*} - s_\alpha) = O(\log T). \tag{48}$$

In any exploitation round, the selected action α_t satisfies $\widehat{s}_{\alpha_t} \ge \widehat{s}_{\alpha^*}$. Hence

$$\begin{aligned}s_{\alpha^*} - s_{\alpha_t} &\le |s_{\alpha^*} - \widehat{s}_{\alpha^*}| + |\widehat{s}_{\alpha_t} - s_{\alpha_t}| \\ &\le |s_{\alpha^*} - \widehat{s}_{\alpha^*}| + \sum_{\alpha\in\mathcal{A}} |\widehat{s}_\alpha - s_\alpha|.\end{aligned} \tag{49}$$

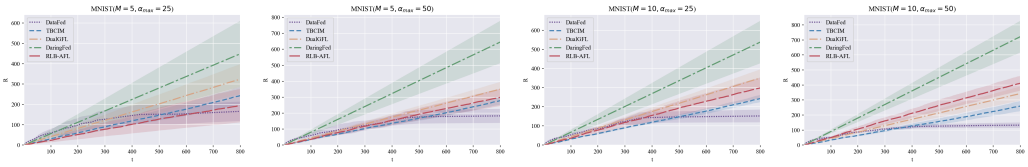


Figure 2: Cumulative expected regret on MNIST under different settings: (a) $M = 5, \mathcal{N}_m = 25$; (b) $M = 5, \mathcal{N}_m = 50$; (c) $M = 10, \mathcal{N}_m = 25$; (d) $M = 10, \mathcal{N}_m = 50$.

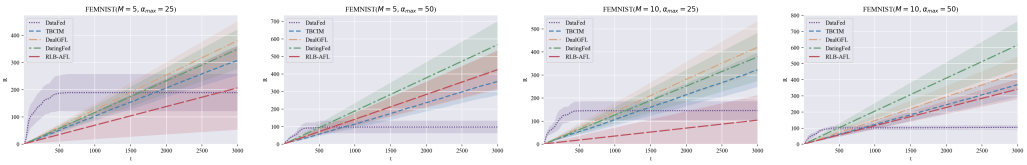


Figure 3: Cumulative expected regret on FEMNIST under different settings: (a) $M = 5, \mathcal{N}_m = 25$; (b) $M = 5, \mathcal{N}_m = 50$; (c) $M = 10, \mathcal{N}_m = 25$; (d) $M = 10, \mathcal{N}_m = 50$.

Taking expectations and applying Eq. (47) with at least m_T samples per action yields

$$R^{\text{exploit}}(T) \leq T \left[\eta_{m_T}(\alpha^*) + \sum_{\alpha \in \mathcal{A}} \eta_{m_T}(\alpha) \right]. \tag{50}$$

Since $\mathcal{A}$ is finite, $\theta_\alpha > 1$, and $\beta = \min_{\alpha \in \mathcal{A}}(\theta_\alpha - 1)/(2\theta_\alpha) = \min_{\alpha \in \mathcal{A}}(M\alpha - \zeta_\alpha)/(2M\alpha) > 0$, the expression in brackets is $O((\log T)^{-\beta})$. Combining exploration and exploitation gives

$$R(T) = O(\log T) + O\big(T(\log T)^{-\beta}\big) = O\big(T(\log T)^{-\beta}\big). \tag{51}$$

Thus the cumulative expected regret is sublinear relative to the known-distribution proxy benchmark under the theorem's stated feedback and exploration conditions.

A.6.1 Results on Real Datasets

A.7 Experimental Details and Additional Results

A.7.1 Experimental Setup

Setup on real datasets. We evaluate cumulative regret on four datasets, including MNIST (LeCun et al., 2021), FEMNIST (Caldas et al., 2018), SVHN (Netzer et al., 2011), and CIFAR-10 (Krizhevsky, 2009). Each experiment is repeated five times with different random seeds to account for randomness in sampling and initialization, and we report the mean cumulative regret with 95% confidence intervals. The settings are $|\mathcal{N}_m| \in \{25, 50\}$, $M \in \{5, 10\}$, and $T \in \{800, 2000, 2500, 3000\}$. In the real-data experiments, we use candidate counts $\alpha \in \{1, \ldots, |\mathcal{N}_m|\}$ and winner counts $\zeta \in \{1, \ldots, \alpha\}$. The cloud valuation is defined as $q(\zeta) = 10 \times Acc(\zeta)$, where $Acc(\zeta) \in [0, 1]$ is the test accuracy achieved with ζ selected clients. Regret is defined as the cumulative utility gap between the optimal and selected actions, where the optimal action is obtained by exhaustive search over all feasible (α, ζ) pairs.

Setup on synthetic datasets. Unless otherwise specified, synthetic experiments use the following two-component truncated Gaussian mixture as the baseline bid distribution. Each ES contains 50 clients, including 10 low-bid clients sampled from $\mathcal{N}(0.15, 0.005^2)$ and 40 main-component clients sampled from $\mathcal{N}(0.65, 0.09^2)$. All bids are truncated to $(0, 1)$ and randomly shuffled within each ES. For ES utility and server utility, we generate $M \in \{5, 10\}$ ES groups. For bid distribution estimation, we use an ES bidding of 50 bids, repeatedly sample 25 bidders for $\lambda = 300$ rounds, and record the second-smallest order statistic as the observable signal for bid-distribution estimation. For ES utility, we consider $\zeta \in \{25, 35, 45\}$ and vary α from 1 to 50, using only valid configurations satisfying $M\alpha \geq \zeta + 1$. For server utility, we set $\alpha \in \{15, 20, 25\}$, vary ζ from 1 to 100, and repeat each configuration 50 times. Note that all experiments are implemented in PyTorch 2.0.1 on an Apple M2 GPU with 8 cores and Metal support.

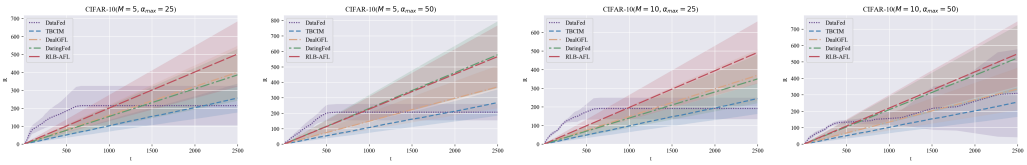


Figure 4: Cumulative expected regret on CIFAR-10 under different settings: (a) $M = 5, \mathcal{N}_m = 25$; (b) $M = 5, \mathcal{N}_m = 50$; (c) $M = 10, \mathcal{N}_m = 25$; (d) $M = 10, \mathcal{N}_m = 50$.

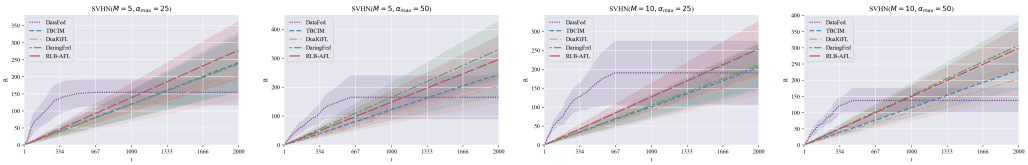


Figure 5: Cumulative expected regret on SVHN under different settings: (a) $M = 5, \mathcal{N}_m = 25$; (b) $M = 5, \mathcal{N}_m = 50$; (c) $M = 10, \mathcal{N}_m = 25$; (d) $M = 10, \mathcal{N}_m = 50$.

A.7.2 BASELINES

The baselines cover diverse incentive and selection paradigms. DaringFed (Xin et al., 2025) models sequential participation via Bayesian persuasion and dynamic pricing but is limited to single-client arrivals. RLB-AFL (Tang et al., 2025) learns bidding strategies through reinforcement learning yet only captures competition without collaboration. DualGFL (Chen et al., 2025) employs a coalition and auction framework but lacks effective intra-coalition regulation, potentially causing imbalanced competition. TBCIM (Ma et al., 2025) integrates reverse auction and Stackelberg game for resource allocation but does not constrain participant scale, which may lead to excessive competition and reduced individual utility.

Performance on cumulative expected regret. The results show that DataFed consistently achieves lower cumulative expected regret in most datasets and parameter settings. Numerically, DataFed usually incurs a short exploration phase at the beginning, after which its regret growth quickly slows down and converges to a stable level. For example, DataFed generally stabilizes at a relatively low regret level after the initial exploration phase across MNIST (Fig. 2), FEMNIST (Fig. 3), CIFAR-10 (Fig. 4), and SVHN (Fig. 5), while the regret becomes slightly higher only in more challenging settings with larger M and N_m. In contrast, most baselines exhibit nearly linear regret accumulation over time. This gap is particularly clear under larger action spaces, such as $N_m = 50$, where DataFed remains almost flat after convergence while the baselines continue to accumulate regret. In a few finite iterates cases, such as SVHN with $M = 10, N_m = 25$ and CIFAR-10 with $M = 10, N_m = 50$, DataFed does not always obtain the lowest final regret, mainly because the number of iterations is not sufficiently large. The improvement comes from DataFed's joint design of bid-distribution learning and hierarchical Stackelberg decision-making: by estimating the bid distribution from the observed $(\zeta + 1)$-th order statistic and jointly optimizing the ES-side bidder number α and the CS-side winner number ζ, DataFed balances competition intensity, payment cost, and model utility. In the plotted horizons, its regret growth generally slows after exploration. These empirical curves do not establish the proxy regret guarantee of Theorem 3 for the realized multi-winner feedback used by DataFed.

A.7.3 RESULTS ON SYNTHETIC DATASETS

The effects of (α, ζ) and M on ES and CS utility As shown in Fig. 6(a), (b) and Fig. 7(a), (b), the ES and CS utilities exhibit non-monotonic behaviors with respect to (α, ζ), and are further influenced by the system scale M. For the ES, under a fixed ζ, increasing α first improves the utility and then leads to a sharp decline, while for the CS, under a fixed α, increasing ζ shows a similar unimodal pattern. This reflects a trade-off between participation scale and payment cost: a larger α increases the chance of winning but intensifies competition and reduces the payment threshold $b_{\zeta+1}$, whereas a larger ζ improves the valuation $q(\zeta)$ but also increases the total payment $\zeta b_{\zeta+1}$.

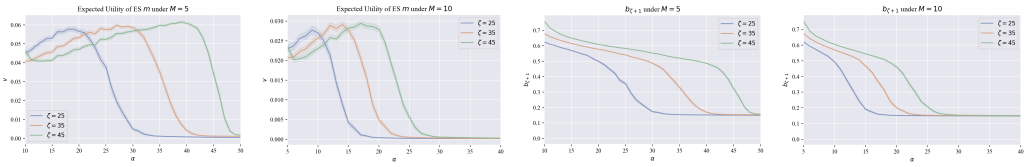


Figure 6: Expected utility of ES m under (a) $M = 5$ and (b) $M = 10$; $b_{\zeta+1}$ under (c) $M = 5$ and (d) $M = 10$.

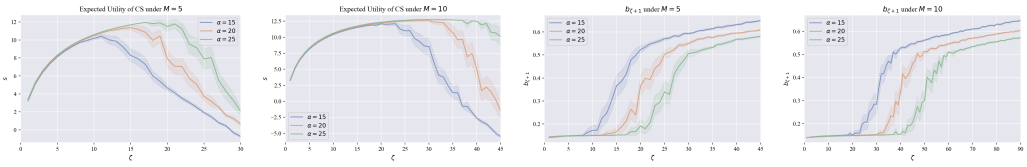


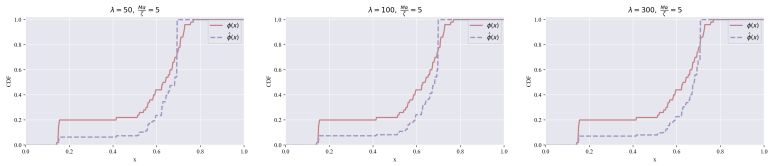
Figure 7: Expected utility of CS under (a) $M = 5$ and (b) $M = 10$; $b_{\zeta+1}$ under (c) $M = 5$ and (d) $M = 10$.

The impact of M further amplifies this trade-off. When M increases, the total number of bidders $M\alpha$ grows, strengthening competition and pushing $b_{\zeta+1}$ toward the lower tail. Consequently, the ES utility decreases and its optimal α shifts to a smaller range, while the CS benefits from lower payments and obtains a wider stable utility region. However, when ζ becomes too large relative to $M\alpha$, the marginal gain from $q(\zeta)$ diminishes and the payment term dominates, leading to a decline in CS utility. Overall, the optimal utility is achieved when (α, ζ) and M jointly balance participation scale, competition intensity, and payment cost, consistent with the dependence on $\theta = \frac{M\alpha}{\zeta}$.

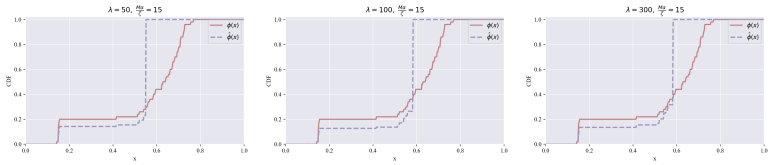
The effects of (α, ζ) and M on payment $b_{\zeta+1}$ As shown in Fig. 6(c), (d) and Fig. 7(c), (d), the payment $b_{\zeta+1}$ decreases with α and increases with ζ, reflecting the order-statistic nature of the k-reverse Vickrey auction. Specifically, increasing α enlarges the total number of bidders $M\alpha$, shifting the $(\zeta + 1)$-th order statistic toward a lower quantile of the bid distribution, while increasing ζ raises the rank of the selected order statistic, leading to a higher payment threshold. This behavior can be interpreted through $\theta = \frac{M\alpha}{\zeta}$, where larger α or smaller ζ increases competition intensity and concentrates the payment on the lower tail, whereas larger ζ weakens competition and raises the payment level.

The system scale M further amplifies this effect. Increasing M enlarges the total bidder pool and reduces $b_{\zeta+1}$ under the same (α, ζ) by strengthening lower-tail competition. As a result, in the $b_{\zeta+1}$ with α plots, the payment drops earlier as α increases, indicating that fewer bidders per ES are sufficient to induce strong competition. In contrast, in the $b_{\zeta+1}$ with ζ plots, the increasing trend shifts to a larger ζ, since a larger bidder pool allows more winners to be selected before the payment rises significantly. Overall, (α, ζ) and M jointly determine the competition intensity and hence the payment level, with larger M accelerating the transition toward a low-payment regime while expanding the feasible range of ζ.

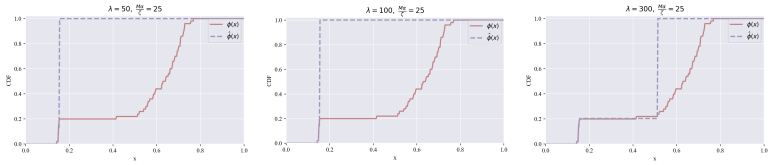
The effects of λ and θ on bid distribution estimation. Figure 8 illustrates the estimated CDFs under different (λ, θ) settings. The true CDF exhibits a mixture structure, with a sharp increase in the low-bid region followed by a gradual rise in the mid-range, consistent with the synthetic setting. When θ is small, the estimated CDF fails to capture the lower-tail behavior accurately. When θ is large but λ is limited, the estimated CDF saturates prematurely in the low-bid region. With sufficiently large λ, however, larger θ better captures lower-tail concentration and produces a more accurate CDF estimate. This reflects the mechanism of DataFed, where λ controls sampling frequency and θ governs lower-tail concentration.



CDF estimation with $\frac{M\alpha}{\zeta} = 5$: (a) $\lambda = 50$, (b) $\lambda = 100$, (c) $\lambda = 300$.



CDF estimation with $\frac{M\alpha}{\zeta} = 15$: (d) $\lambda = 50$, (e) $\lambda = 100$, (f) $\lambda = 300$.



CDF estimation with $\frac{M\alpha}{\zeta} = 25$: (g) $\lambda = 50$, (h) $\lambda = 100$, (i) $\lambda = 300$.

Figure 8: CDF estimation under different λ and $\frac{M\alpha}{\zeta}$.

Table 5: DataFed under stationary and changing bid distributions.

Bid process	Final CDF error	Post-change regret	Recovery rounds	Final cumulative regret
Stationary	0.0224 ± 0.0011	N/A	N/A	357.029 ± 18.377
Gradual drift	0.0213 ± 0.0014	69.643 ± 8.182	81	366.479 ± 19.523
Abrupt shifts	0.0308 ± 0.0014	94.319 ± 8.889	Not within 40 rounds	391.155 ± 19.069

A.7.4 Robustness to Changing Bid Distributions

We compare the original stationary Gaussian mixture with gradual drift toward a higher-cost regime and abrupt distribution changes every 40 rounds. Within each local regime, bids are sampled independently. Mixture weights, variances, the action space, and the feedback format remain as in the original synthetic setting. The results in Table 5 are empirical stress tests; Theorem 3 does not cover a changing underlying distribution.

With gradual drift, the final CDF error stays close to the stationary value and recovery takes 81 rounds. Under repeated abrupt shifts, the 20-round moving-average CDF error peaks at 0.0867 and does not recover before the next change. These observations support empirical adaptation to moderate drift but do not imply nonstationary sublinear regret.

A.7.5 Effect of Hierarchical Response Anticipation

The coordination comparison uses the same feasible actions, realized bids, estimator, feedback, five random seeds, and 800 rounds. Its cumulative regret is measured against centralized joint selection with full information. This reference differs from the normalized proxy benchmark in Theorem 3. The centralized rule is an information and control upper bound rather than a deployable baseline. Sequential selection without anticipation preserves the cloud-first order but selects the winner count using the previous candidate count instead of the current ES best-response mapping.

The independent and sequential baselines have nearly identical mean performance. Relative to sequential selection without anticipation, DataFed reduces mean cumulative regret by 11.0% and

Table 6: Coordination models over 800 rounds. Results are means and 95% confidence intervals.

Coordination model	Cumulative regret	Mean per-round cloud utility
Centralized joint selection	N/A	10.50 ± 0.02
Uncoordinated independent selection	3271.68 ± 366.81	5.57 ± 0.44
Sequential selection without anticipation	3271.80 ± 366.67	5.57 ± 0.44
DataFed with response anticipation	2912.80 ± 201.53	6.01 ± 0.23

Table 7: Exact and proxy optimizer comparison under the synthetic bid distribution.

M	Exact action (α, ζ)	Proxy action (α, ζ)	Exact utility of each action	Exact utility retained
5	$(25, 17)$	$(25, 10)$	11.6403 versus 10.4972	90.18%
10	$(25, 30)$	$(25, 15)$	12.6183 versus 11.6491	92.32%

increases mean per-round cloud utility by 7.9%. This comparison isolates the value of anticipating the ES response beyond merely committing first.

A.7.6 Accuracy, Payment, and Regret

Table 3 in the main paper reports final accuracy, total payment, and cumulative expected regret on MNIST and CIFAR-10 across five random seeds, with 95% confidence intervals. These metrics characterize predictive performance, payment, and regret.

DataFed has the lowest mean cumulative expected regret among the listed methods on both datasets. It does not have the lowest total payment in either dataset, and its MNIST accuracy point estimate is not the highest. The result therefore describes a trade-off across accuracy and payment rather than dominance on every metric.

A.7.7 Proxy and Exact Optimizer Comparison

We compare the proxy-selected action with the exact multi-winner optimizer under the truncated Gaussian mixture used in the synthetic experiments. The candidate counts are drawn from $\alpha \in \{15, 20, 25\}$ and all feasible winner counts up to 100 are considered. The exact utility of the proxy-selected action is evaluated using the original objective, not the proxy utility.

The proxy selects a different winner count in both cases while retaining more than 90% of the exact optimal utility. This comparison quantifies an approximation gap for the tested distribution; it does not make the proxy equilibrium an exact solution to the original multi-winner game.

A.8 Limitations

Our regret guarantee applies to a fixed bid distribution with independent normalized proxy feedback and sufficient exploration. The proxy approximation makes the analysis tractable, while experiments assess its decision quality and robustness to bid drift. Extending the guarantee to realized multi-winner feedback and broader bidding dynamics remains future work.