Nash Incentive-compatible Online Mechanism Learning via Weakly Differentially Private Online Learning

Joon Suk Huh [1] **Kirthevasan Kandasamy** [1]

Abstract

We study a multi-round mechanism design problem, where we interact with a set of agents over a sequence of rounds. We wish to design an incentive-compatible (IC) online learning scheme to maximize an application-specific objective within a given class of mechanisms, without prior knowledge of the agents' type distributions. Even if each mechanism in this class is IC in a single round, if an algorithm naively chooses from this class on each round, the entire learning process may not be IC against non-myopic buyers who appear over multiple rounds. On each round, our method randomly chooses between the recommendation of a weakly differentially private online learning algorithm (e.g., Hedge), and a commitment mechanism which penalizes non-truthful behavior. Our method is IC and achieves $O(T^{\frac{1+h}{2}})$ regret for the application-specific objective in an adversarial setting, where h quantifies the long-sightedness of the agents. When compared to prior work, our approach is conceptually simpler, it applies to general mechanism design problems (beyond auctions), and its regret scales gracefully with the size of the mechanism class.

1. Introduction

In mechanism design, a social planner designs a protocol, where a set of agents report their *type* to the mechanism, and the mechanism chooses an outcome based on the reported types. It is often desirable, if not necessary, that a mechanism be *incentive-compatible* (IC), i.e. it is in the best interest for each agent to report their types truthfully. When designing the mechanism, we are interested in optimizing some application-specific objectives. This typically requires knowledge of the distribution on agents' types. As examples, consider a seller who wishes to design either a reserve price auction or a posted price mechanism to sell an item. In both cases, to maximize revenue (objective), the seller requires knowledge of the distribution of buyers' (agents) valuations (types) for the item.

However, in practice, this type distribution is often unknown *a priori*. Consequently, learning an optimal mechanism from data has been one of the central themes in the intersection of machine learning and game theory. Several prior works have studied learning optimal mechanisms from an i.i.d. dataset of types in offline settings (Cole & Roughgarden, 2014; Daskalakis et al., 2014; Balcan et al., 2016; 2018). In this work, our focus is on online settings, where there is a stream of types and objectives, and the goal is to design *no-regret* learning procedures which need to trade-off between exploring candidate mechanisms in a mechanism class, while simultaneously exploiting current knowledge to maximize the desired objective (Kleinberg & Leighton, 2003; Blum & Hartline, 2005).

This problem of online mechanism learning has a distinctive challenge: the data sources are themselves strategic agents endowed with their own *long-term* utilities when interacting with the mechanism over multiple rounds. These agents may misreport their type for their own long-term benefit. This is detrimental to learners who will be unable to estimate the true frequency of types and optimize the given objective. In particular, even if a multi-round mechanism is IC in a single round, it does not necessarily translate to IC over the *entire* learning process. Agents may find it beneficial to misreport their type in the early rounds while incurring some short-term loss, so as to manipulate later outcomes in their favor, resulting in long-term gains.

To illustrate this, consider designing a revenue-optimal posted price mechanism (PPM)[1], where a seller has an infinite supply of a single item. The seller chooses a price p without revealing it to the buyers. The buyers report their value, and the seller sells the item to those whose value exceeds p and charges them p. This mechanism is IC on any single round. However, if the seller naively repeats this over a sequence of rounds, and uses reported values from previ-

[1]Department of Computer Science, University of Wisconsin–Madison, Madison, WI, USA. Correspondence to: Joon Suk Huh <joon@cs.wisc.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

[1]This description is different, yet equivalent to another common interpretation of the PPM, where the seller announces the price p and buyers buy if their value is larger than p.

ous rounds to set the current price, buyers may bid lower in the early rounds. While this deceitful bidding strategy may risk not getting the item early on, it would misguide a learning algorithm to think the current price is too high, and lower future prices, benefiting the buyer. Consequently, the seller will not be able to optimize her long-term revenue.

Recent works have recognized this challenge and developed learning algorithms for auctions with various strategy-robustness and IC guarantees (Amin et al., 2013; Mohri & Munoz, 2014; 2015; Drutsa, 2017; 2018; Liu et al., 2018; Drutsa, 2020b; Abernethy et al., 2019; Deng et al., 2020; Mirrokni et al., 2020). However, all methods above only apply to auctions, and not general mechanism design problems. Moreover, in most work, there is only a single parameter to be learned in the mechanism (e.g. reserve price in an auction, selling price in a PPM), and there is no natural extension when there are multiple learnable parameters (e.g. VCG auctions with several reserve prices (Hartline & Roughgarden, 2009; Roughgarden & Wang, 2019)).

Our approach offers the following advantages when compared to prior work. First, it applies to general mechanism design problems with application-specific objectives, including mechanisms without money (Procaccia & Tennenholtz, 2013; Nissim et al., 2012). Second, it can be naturally applied to settings with multiple learnable parameters. In particular, by leveraging a weak notion of differential privacy, we are able to avoid the curse of dimensionality in large parameter spaces. Third, our method is conceptually much simpler than many of the existing work, as it is based on the well-known Hedge algorithm (Hazan, 2016).

1.1. Summary of Contributions and Main Results

Our main contributions in this work are as follows:

1. In §3, we present a framework for IC online learning of mechanisms with full-information feedback. In our framework, an *online mechanism* chooses a single-round mechanism on each round; an environment realizes agents' types and the social planner's objective. Agents report their types, and then the single-round mechanism chooses an outcome (see Protocol 1). Our framework applies to general mechanism design problems where a social planner may have different objectives on each round, and agents may have different types on each round.

2. In §4.1, we show that, under certain conditions on the agents' long-sightedness (Definition 3.3), any *weakly* differentially private (DP) sequence (Definition 2.2) of distributions over mechanisms can be used as a primitive for building an IC online learning mechanism. Our notion of a weak DP sequence is a considerably weaker technical requirement than the usual notion of DP for sequential algorithms used to build IC online mechanisms in prior work (Jain et al., 2012; Abernethy et al., 2019).

3. In §4.2, we use the above result to construct an IC online learning mechanism using the well-known Hedge (a.k.a. exponential weights) algorithm (Hazan, 2016). As we show, Hedge provides a weak DP sequence. Our learning algorithm is *(i)* Nash incentive-compatible (NIC, Definition 3.1) and *(ii)* achieves $O\big((\log|\Pi|)\cdot T^{(1+h)/2}\big)$ regret compared to the best mechanism in any NIC mechanism class Π. Here $h\in[0,1)$ is an exponent that characterizes the long-sightedness of agents (Definition 3.3 and Theorem 4.5).

4. In §5 and Appendix C, we apply our framework to develop no-regret NIC online learning mechanisms for various mechanism design problems. They are *(i)* the k-facility location problem (Lu et al., 2010; Fotakis & Tzamos, 2013) where the goal is to minimize the average distance between citizens and their closest facilities, *(ii)* Social welfare maximization with *ex-post* externality via the VCG mechanism with bidder-specific reserve prices (Hartline & Roughgarden, 2009; Roughgarden & Wang, 2019), and *(iii)* a resource allocation problem (Demers et al., 1989) where we wish to maximize the number of jobs completed.

Key insight. While prior works have used DP to achieve IC for auctions in sequential settings, the standard notion of DP for sequential algorithms (Jain et al., 2012) is stronger than necessary. Our key insight is in recognizing that a weaker notion of DP for sequential algorithms is sufficient, and moreover that the vanilla Hedge algorithm satisfies this property. This helps us reduce the dependence to logarithmic in $|\Pi|$. Conveniently, this approach also provides a straightforward extension to general mechanism design problems, beyond auctions.

Algorithm overview. On each round, our construction randomly chooses a lottery of two options. In the first, we follow a mechanism in Π recommended by Hedge, and in the other, we implement a problem-specific *commitment mechanism* that penalizes any non-truthful behavior (Nissim et al., 2012). Then, the agents report their types after seeing the lottery of mechanisms, which includes the probability of each option that is set based on problem-specific factors.

Proof overview. On each round, the recommendation from Hedge can be viewed as a sequence of randomized maps from past observations to a single-round mechanism. We show that *each* randomized map in the sequence being η-DP is enough to guarantee that truth-telling is a Nash equilibrium in the online setting (Theorem 4.3), for some choice of $\eta\geq 0$. Specifically, while agents can gain by lying in Hedge, the weak DP property ensures that this gain is bounded. Then, the commitment mechanism cancels out this gain by

sufficiently penalizing untruthful reporting. While similar intuitions have been used in prior auction work, the key challenge of applying this idea with our notion of weak DP is that the agents' potential gains are functions of the full sequence of outcomes up to the final round T. Hence, using sequential composition (Dwork et al., 2014), we can only bound the agents' potential gains by an ηT factor of the maximum utility, which grows linearly with T.

To address this, we rely on a backward-induction argument. First, we show that an agent has no incentive to deviate from truth-telling in the last round. Then, we recursively apply this argument to demonstrate that truth-telling is the best strategy in all rounds. The key observation is that our weak DP notion is sufficient for this proof technique to work.

1.2. Discussion of Most Relevant Works

Due to space constraints, we discuss only the most relevant works here, and defer a detailed overview to Appendix D. For discussions on previous regret bounds for settings similar to ours (Abernethy et al., 2019; Liu et al., 2018; Amin et al., 2013), we refer readers to §4.3.

One of the closest lines of research to ours is dynamic mechanism design (Baron & Besanko, 1984; Thomas & Worrall, 1990; Courty & Hao, 2000; Bergemann & Välimäki, 2010; Ashlagi et al., 2016; Mirrokni et al., 2016; 2020; Bergemann & Välimäki, 2019; Deng et al., 2019b). In dynamic mechanism design, the seller possesses precise knowledge of the buyer's value distribution across all rounds. Leveraging this information, the seller designs a revenue-maximizing dynamic mechanism by adjusting the mechanism's rule based on the buyers' past bids. Recently, dynamic mechanism design has been extended to the setting where the exact distributional knowledge requirement is relaxed to approximate ones (Golrezaei et al., 2019; Deng et al., 2019a; 2020). However, even with these extensions, dynamic mechanism design requires a strong assumption of the existence of an underlying distribution on buyers' values, and often necessitates all buyers to share their beliefs. Our approach does not require any of this: We allow the types of agents to be arbitrary sequences of time. Moreover, our Nash incentive-compatibility guarantee does not require agents to share their exact beliefs other than everyone reporting their types truthfully. Furthermore, to the best of our knowledge, dynamic mechanism design has primarily focused on mechanisms involving monetary transactions, whereas our framework operates independently of any payment or quasilinear utility assumptions.

Besides the dynamic mechanism design literature, the closest work to ours is Abernethy et al. (2019), who delve into an online Bayesian optimal auction design problem. They use a strongly DP online learning algorithm to design a version of the Myerson auction that sells a limited supply of identical items to multiple buyers. They show that this approach is η-bid-approximate Bayesian incentive-compatible, meaning that every agent submitting bids within η proximity to their true values is a Nash equilibrium with respect to each agent's Bayesian belief. Our IC guarantee, Nash incentive compatibility, is stronger as it does not depend on any prior type distribution and holds for all true agent types (i.e. in an *ex-post* sense). Moreover, our IC guarantee is valid even when all previous information is *public* to all agents, while previous works (Liu et al., 2018; Abernethy et al., 2019) rely on the private information assumption where agents cannot observe other agents' previous bids.

To realize an incentive-compatible online mechanism, our method uses a lottery between an online learning algorithm and a *commitment mechanism* which penalizes agents for non-truthful behavior. The technique of combining a differentially private algorithm with a commitment mechanism was originally introduced by Nissim et al. (2012) to design approximately optimal singe-round mechanisms. We extend their approach to the online setting without losing efficiency in the sense that our regret bound is logarithmic in $|\Pi|$.

2. Background

In this section, we provide a brief overview of mechanism design and differential privacy.

Notation and assumptions. For $d \in \mathbb{N}$, let $[d] := \{1, \cdots, d\}$. Let $\mathbb{1}[\cdot]$ be the indicator function. For a set S, let $\Delta(S)$ denote the set of probability distributions over S. We use $t \in [T]$ ($i \in [n]$) to denote rounds (agents). For any set of round indexed variable $\{x_t\}$, let $x_{<t} := (x_\tau)_{\tau \in [t-1]}$ and $x_{\leq t} := (x_\tau)_{\tau \in [t]}$. Throughout this paper, we assume types of agents are discrete, that is, we set Θ_i introduced in the next subsection as a countable set.

2.1. Mechanism Design

There are n agents, denoted $[n]$. Agent i has type $\theta_i \in \Theta_i$, where Θ_i is the set of available types. Each agent reports their type to a mechanism, which may not necessarily be the same as their true type θ_i. We use $b_i \in \Theta_i$ to denote the type reported by agent i. Based on the reported types, the mechanism, possibly randomly, chooses an outcome $s \in S$, where S is the set of available outcomes. Then, each agent i experiences utility $u_i(\theta_i, s)$. The social planner's satisfaction is given by an (application-specific) *objective* $G(\theta, s) \in [-1, 1]$, where $\theta := (\theta_i)_{i \in [n]}$ is all agent types.

A mechanism π is a map from reported types to a distribution over outcomes in S, i.e. $\pi : \prod_{i \in [n]} \Theta_i \to \Delta(S)$, where $\pi(b)$ is the distribution induced on the set of outcomes when the agents report $b := (b_i)_{i \in [n]}$. An outcome s is then sampled from $\pi(b)$. The goal of mechanism design is to design π which has large expected objective $\mathbb{E}_{s \sim \pi(\theta)} G(\theta, s)$, when

the agents are reporting truthfully, i.e. $b_i = \theta_i$ for all $i \in [n]$.

However, as agents may misreport their types to maximize their utility, a mechanism should also be IC, i.e. incentivizes agents to report their types truthfully. To state this formally, with slight abuse of notation, let

$$u_i\big(b; \pi, \theta_i\big) := \mathop{\mathbb{E}}_{s \sim \pi(b_i, b_{-i})} \left[u_i(\theta_i, s)\right], \tag{1}$$

denote the expected utility of agent i of type θ_i, given reported types b under mechanism π. Here, u_i on the RHS is the utility of the agent for outcome s as defined above.

One common notion of IC is the dominant-strategy incentive-compatibility (DSIC). We say that π is DSIC if truth-telling is better for all agents regardless of the types reported by the other agents. That is, for all $i \in [n]$, $\theta_i, b_i \in \Theta_i$ and $b_{-i} \in \Theta_{-i} := \prod_{j \in [n] \setminus \{i\}} \Theta_j$, we have

$$u_i\big(\theta_i, b_{-i}; \pi, \theta_i\big) \geq u_i\big(b_i, b_{-i}; \pi, \theta_i\big). \tag{2}$$

While DSIC is a strong IC requirement, it is not always attainable and hence, it is common to relax this requirement. In this work, we consider Nash incentive compatibility (NIC). We say that π is NIC, if for all $i \in [n]$, $\theta_i, b_i \in \Theta_i$ and $\theta_{-i} \in \Theta_{-i}$, we have

$$u_i\big(\theta_i, \theta_{-i}; \pi, \theta_i\big) \geq u_i\big(b_i, \theta_{-i}; \pi, \theta_i\big). \tag{3}$$

That is, given that other agents report their types truthfully ($b_{-i} = \theta_{-i}$), reporting one's type truthfully is an optimal strategy. This means that truth-telling is a Nash equilibrium. Since we aim to learn a mechanism in online, adversarial environments without assuming type distributions, NIC is appropriate for our setting. NIC is also stronger than Bayes-Nash incentive compatibility (BNIC), which only guarantees truth-telling in expectation over type distributions (Nisan et al., 2007).

In this work, we study finding an optimal mechanism in a given class of mechanisms Π. We say that Π is NIC (DSIC) if all $\pi \in \Pi$ is NIC (DSIC).

2.2. Differential Privacy

Next, we briefly review differential privacy. We refer readers to Dwork et al. (2014) for a comprehensive overview. Let X be an input domain and X^m be a Cartesian product of X, representing the space of datasets of size m. Let Y be an output space. We define DP as follows.

Definition 2.1 (Differential privacy)**.** A mapping $q : X^m \to \Delta(Y)$ is η-differentially private (η-DP) if for all $A \subseteq Y$ and all inputs $x = (x_1, \ldots, x_m) \in X^m$ and $x' = (x'_1, \ldots, x'_m) \in \mathcal{X}^m$ that differ in at most one entry ($x_j \neq x'_j$ for at most one $j \in [m]$), the following holds:

$$\Pr_{y \sim q(x)}\big(y \in A\big) \leq e^{\eta} \Pr_{y \sim q(x')}\big(y \in A\big).$$

Intuitively, this means that the output of an η-DP algorithm does not change significantly if we change one input data point. In sequential learning settings, we extend this definition as follows.

Definition 2.2 (Weak η-DP sequence)**.** Let $q_{\leq T} := (q_t)_{t \in [T]}$ be a sequence of mappings where $q_t : X^{t-1} \to \Delta(Y)$. We say that $q_{\leq T}$ is a weak η-DP sequence if q_t is η-DP for each $t \in [T]$.

Readers familiar with the literature will recognize that the above definition is weaker than the more standard definition (e.g. Jain et al. (2012)), which requires that $\prod_{t=1}^{T} q_t(\cdot)$, viewed as a map from X^T to $\Delta(Y^T)$ is η-DP. While the stronger definition is relevant from a data privacy perspective, any compliant algorithm often comes with inevitable performance degradation. However, for our purpose of designing an IC online learning method, the weaker extension in Definition 2.2 is sufficient. In particular, while the vanilla Hedge algorithm is not strongly η-DP (Jain et al., 2012), in §4.2, we show that it is weakly η-DP. In fact, this follows immediately from a well-known DP technique called the exponential mechanism, described below.

Proposition 2.3 (Exponential mechanism (McSherry & Talwar, 2007))**.** *Consider a map $q : X^m \to \Delta(Y)$ such that $q(x)(y) \propto \exp(\eta\, g(x, y))$, for some $g : X^m \times Y \to \mathbb{R}$, for all $x \in X^m$ and $y \in Y$. Let $\Delta g := \max_{x, x', y} |g(x, y) - g(x', y)|$ where the maximum is taken over $x, x' \in X^m$ that differ in at most one entry. Then q is $2\eta\Delta g$-DP.*

3. Nash Incentive-compatible Online Mechanism Learning

In this section, we describe the problem of Nash incentive-compatible online mechanism learning.

3.1. Online Mechanism Learning

A set of n agents repeatedly interact with an online mechanism over a sequence of T rounds, although not all agents need to appear on all rounds. On each round, the environment chooses types $\theta_{i,t}$ one for each participating agent and the social planner's objective G_t. Simultaneously, the online mechanism $\mathcal{M}$ uses past information, to choose a single-round mechanism π_t to collect agent types and choose an outcome $s_t \in S$. Our goal is to design a learning algorithm that is competitive against the single best mechanism in hindsight in some mechanism class Π. We have outlined the protocol for online mechanism learning in Protocol 1. Next, will now describe the individual components.

Agents. There are n agents, indexed by $i \in [n]$. Each agent $i \in [n]$ appears on a subset $\mathcal{T}_i \subset [T]$ of the rounds. Let $\theta_i := (\theta_{i,t})_{t \in \mathcal{T}_i}$ where $\theta_{i,t} \in \Theta_i$ be the sequence of types for agent i in the rounds that she participates. Let $\gamma_i : [T] \to [0, 1]$, be an agent-specific monotonically non-increasing

Protocol 1. Online Mechanism Learning

For each round $t = 1, \ldots, T$:

1. $\mathcal{M}$ chooses a single-round mechanism π_t based on past reported types $b_{<t}$ and observed objective functions $G_{<t}$.
2. Each participating agent observes π_t and reports their types $b_{i,t} \in \Theta_i$.
3. $\mathcal{M}$ executes π_t with input b_t and realizes an outcome $s_t \sim \pi_t(b_t)$.
4. Each participating agent experiences utility $u_i(\theta_{i,t}, s_t)$.
5. $\mathcal{M}$ observes G_t and experiences $G_t\big(\theta_t, s_t\big)$.

function which discounts the agent's future utilities. If the mechanism chooses outcomes $(s_t)_{t\in[T]}$ on each round, the agent's utility is $\sum_{t\in\mathcal{T}_i} \gamma_i(t) u_i(\theta_{i,t}, s_t)$.

If an agent participates in round t, she reports her type $b_{i,t} \in \Theta_i$, not necessarily truthfully to the mechanism. If an agent does not participate, we will set the reported and true types as $b_{i,t} = \theta_{i,t} = \perp$. For what follows, denote $b_t := (b_{i,t})_{i\in[n]}$, $b_{<t} := (b_\tau)_{\tau\in[t-1]}$ and $b_{\leq T} := (b_t)_{t\in[T]}$. Define θ_t, $\theta_{<t}$ and $\theta_{\leq T}$ similarly. Finally, we write the agents as a tuple:

$$A := \big(n, T, (\theta_t)_{t\in[T]}, (\mathcal{T}_i)_{i\in[n]}, (\gamma_i)_{i\in[n]}\big) \tag{4}$$

Objectives. The social planner's objective functions are given by $G_t : \prod_{i\in[n]}(\Theta_i \cup \{\perp\}) \times S \to [-1, 1]$ for each round $t \in [T]$. These objectives could be application-specific, and be chosen by an oblivious adversary.

Online mechanism. An online mechanism $\mathcal{M} = (\mathcal{M}_t)_{t\in\mathbb{N}}$, is a sequence of policies, where $\mathcal{M}_t$ maps past reported types $b_{<t}$ and objectives $G_{<t}$ to a distribution over $\mathcal{U}$, where $\mathcal{U} := \{\pi : \prod_{i\in[n]}(\Theta_i \cup \{\perp\}) \to \Delta(S)\}$ is the class of *all* single-round mechanisms. Some π_t is sampled from this distribution $\mathcal{M}_t(b_{<t}, G_{<t})$ and executed on round t.

Regret. To define the regret, first consider the cumulative objective attained by an online mechanism $\mathcal{M}$, given reported types, $b_{\leq T}$ with respect to agents A.

$$\text{Alg}\big(b_{\leq T}, \mathcal{M}; A\big) := \mathbb{E}\left[\sum_{t=1}^{T} G_t\big(\theta_t, s_t\big)\right].$$

Here, the expectation is over the sequence of outcomes $s_{\leq T}$ generated by running Protocol 1 with $\mathcal{M}$, A and $b_{\leq T}$. We compare the above quantity to the optimal cumulative objective $\text{Opt}(\Pi; A)$ attainable within a class of mechanisms Π with respect to agents A:

$$\text{Opt}(\Pi; A) := \max_{\pi\in\Pi} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{s\sim\pi(\theta_t)} G_t(\theta_t, s)$$

The performance of $\mathcal{M}$ is measured by the gap between the above two quantities, which is called the regret:

$$\text{Reg}\big(b_{\leq T}, \mathcal{M}, \Pi; A\big) := \text{Opt}(\Pi; A) - \text{Alg}\big(b_{\leq T}, \mathcal{M}; A\big).$$

3.2. Nash Incentive Compatibility

We next describe the IC requirement for an online learning mechanism. For this, let $h_{<t}$ denote the history up to round $t-1$. It consists of the previous single-round mechanisms chosen, submitted *and* true types of all agents, outcomes, and objectives, i.e. $h_{<t} := (\pi_{<t}, b_{<t}, \theta_{<t}, s_{<t}, G_{<t})$.

We will assume that throughout the T rounds, agents are aware of their utility function u_i, and their true types $\theta_i = (\theta_{i,t})_{t\in\mathcal{T}_i}$ in past, current, and future rounds. Moreover, on round t, agents may have access to the realized history $h_{<t}$ while not required. Note that while agents may know the true types, the online mechanism does not.

Agent's strategy. An agent i's strategy $\sigma_i := (\sigma_{i,t}(\cdot))_{t\in\mathcal{T}_i}$ is a sequence of policies, where $\sigma_{i,t}$ maps the realized history $h_{<t}$ to a reported type. That is, $b_{i,t} = \sigma_{i,t}(h_{<t}) \in \Theta_i$. If agents follow randomized strategies, our results will hold for all realizations of this strategy. We use $\sigma := (\sigma_i)_{i\in[n]}$ to denote a strategy profile of all agents and σ_{-i} to denote a strategy profile except agent i's.

The *truth-telling strategy* for agent i reports the true type regardless of history. With a slight abuse of notation, we denote this as $\theta_i := (\theta_{i,t}(\cdot))_{t\in\mathcal{T}_i}$ where $\theta_{i,t}(\cdot) := \theta_{i,t}$. We call $\theta := (\theta_i)_{i\in[n]}$ the truth-telling strategy profile.

Nash incentive-compatible learning. To define the NIC requirement, we should define the expected utility for an agent. Given an online mechanism $\mathcal{M}$, agents A, and strategy profile σ, with a slight abuse of notation, we denote the expected utility for agent i as

$$u_i(\sigma; \mathcal{M}, A) := \mathbb{E}\left[\sum_{t\in\mathcal{T}_i} \gamma_i(t) u_i(\theta_{i,t}, s_t)\right]. \tag{5}$$

Here, u_i on the RHS is the utility of the agent with type $\theta_{i,t}$ for outcome s_t, and is similar to the RHS of (1). Moreover, the expectation is over the joint distribution of $s_{\leq T}$, induced by the interaction between the strategies σ, mechanism $\mathcal{M}$, and agents A via Protocol 1. We can now extend the definition of NIC (see (3)) for online mechanisms.

Definition 3.1 (NIC, Online setting)**.** An online mechanism $\mathcal{M}$ is Nash incentive-compatible (NIC) with respect to agents A, if for every agent i,

$$u_i(\theta_i, \theta_{-i}; \mathcal{M}, A) \geq u_i(\sigma_i', \theta_{-i}; \mathcal{M}, A),$$

that is, no agent i has incentive to deviate from the truth-telling strategy θ_i, given that other agents are truth-telling.

In other words, NIC guarantees that the truth-telling strategy profile is a Nash equilibrium of the game induced by $\mathcal{M}$ and E. Finally, we can state our definition of a Nash incentive-compatible online mechanism.

Definition 3.2 (NICOM)**.** Let $\mathcal{A}$ be a class of possible agents (i.e. a set of the tuples in (4)), and let $\mathcal{A}_T \subseteq \mathcal{A}$ be the subset of all $A \in \mathcal{A}$ whose total number of rounds is T. We say that a mechanism $\mathcal{M}$ is a Nash incentive-compatible online mechanism (NICOM) for a single-round mechanism class Π with respect to an agent class $\mathcal{A}$, if,

1. The regret of $\mathcal{M}$ is sub-linear given that agents report their types truthfully. That is,

$$\sup_{A \in \mathcal{A}_T} \mathrm{Reg}(\theta_{\leq T}, \mathcal{M}, \Pi; A) \in o(T).$$

2. For every $A \in \mathcal{A}$, $\mathcal{M}$ is NIC with respect to A.

In the above definition, we assume that the mechanism $\mathcal{M}$ has full knowledge of the agent class $\mathcal{A}$, but not of realized agents $A \in \mathcal{A}$ (e.g. their true types) except for T.

Long-sightedness. Before we proceed, we discuss one more condition necessary to state our results. If agents prioritize future gains and are willing to incur initial losses for a longer period, incentive-compatible online learning gets more challenging. Hence, prior works in this space have assumed that agents either appear a small number of rounds, i.e. $|\mathcal{T}_i| \in o(T)$, and/or that the discounted utilities grow sublinearly, i.e. $\sum_{t=1}^{T} \gamma_i(t) \in o(T)$ to circumvent this issue (Liu et al., 2018; Abernethy et al., 2019). In fact, Amin et al. (2013)'s Theorem 3 implies that if either of these does not hold, it is impossible to design an incentive-compatible posted price mechanism even when there is a single buyer. Below, we use a similar characterization of an agent's long-sightedness, which combines both definitions.

Definition 3.3 (Long-sighted agents)**.** The long-sightedness α of the agents A is,

$$\alpha(A) := \max_{i \in [n],\, t \in [T]} \min\left(\frac{1}{\gamma_i(t)} \sum_{\tau=t}^{T} \gamma_i(\tau),\ |\mathcal{T}_i| \right).$$

Furthermore, for an agent class $\mathcal{A}$, let $\alpha_T(\mathcal{A}) := \sup_{A \in \mathcal{A}_T} \alpha(A)$, where $\mathcal{A}_T$ is the set of all $A \in \mathcal{A}$ whose total rounds are T.

When $\gamma_i(t) = \nu^t$ for all t, for some $\nu \in [0, 1)$, we have $\alpha(A) \leq \frac{1}{1-\nu}$, which is the maximum (discounted) utility an agent can achieve. On the other extreme, when $\gamma_i(\cdot) = 1$, now $\alpha(A) = \max_{i \in [n]} |\mathcal{T}_i|$ is the maximum number of times an agent appears. We will see that our regret bound increases with $\alpha_T(\mathcal{A})$.

4. Method

In this section, we describe our key methodological contributions. In §4.1, we provide a general recipe for building a NICOM from a weak DP sequence. Then, in §4.2, we instantiate this recipe via the Hedge algorithm. In §4.3 and 4.4, we provide a series of discussions on our main result including references to previous bounds for similar settings.

4.1. Designing NIC Online Mechanisms

Commitment mechanism. A key ingredient of our construction is a *commitment mechanism* π^{com}. This refers to any strictly DSIC mechanism–usually far from optimal–that penalizes non-truthful behaviors. Our method will use randomly choose between a mechanism from the target class Π, and the commitment mechanism.

Below, we define the *penalty gap*, which quantifies the power of a commitment mechanism, i.e. its ability to penalize non-truthful behavior. For this, recall that $\mathcal{U}$ is the class of all single-round mechanisms.

Definition 4.1 (Penalty gap)**.** For a single-round mechanism $\pi \in \mathcal{U}$, the penalty gap $\beta(\pi)$ is the minimum difference in any agent's utility between reporting truthfully and non-truthfully, regardless of the others' behavior. That is,

$$\beta(\pi) := \min_{i, \theta_i, b_i \neq \theta_i, b_{-i}} \left[\mathop{\mathbb{E}}_{s \sim \pi(\theta_i, b_{-i})} u_i(\theta_i, s) - \mathop{\mathbb{E}}_{s \sim \pi(b_i, b_{-i})} u_i(\theta_i, s) \right].$$

Our method will not require π^{com} to be in the target mechanism class Π. The purpose of π^{com} is solely to cancel out any potential gains from lying. As such, we require a commitment mechanism with $\beta(\pi^{\mathrm{com}}) > 0$, and our regret bound will scale with how powerful $\beta(\pi^{\mathrm{com}})$ is. As we show in the examples in §5 and Appendix C, it is often possible to find a good commitment mechanism for a problem.

NIC via Weak DP Sequences. Now, we present a general construction of NICOMs, using a weak differentially private sequence (Definition 2.2) and a commitment mechanism.

Definition 4.2 (Weak η-DP mechanism)**.** Consider any $\lambda \in [0, 1]$, commitment mechanism π^{com} and weak η-DP sequence $(q_t^{\eta\text{-wdp}})_t$. Here, $q_t^{\eta\text{-wdp}}$ is a map from $b_{<t}$ and $G_{<t}$ to a distribution over $\mathcal{U}$. Let $\mathcal{M}^{\mathrm{wDP}}$ be an online mechanism that outputs the following mechanism in (6) for each t. Below, $\pi_t^{\mathrm{sr}} \in \mathcal{U}$ is a single-round mechanism sampled from $q_t^{\eta\text{-wdp}}(b_{<t}, G_{<t})$. We have,

$$\pi_t(\cdot) := (1 - \lambda) \cdot \pi_t^{\mathrm{sr}}(\cdot) + \lambda \cdot \pi^{\mathrm{com}}(\cdot). \tag{6}$$

The above definition says that $\mathcal{M}^{\mathrm{wDP}}$ outputs a lottery which realizes the commitment mechanism with probability λ, and

the recommendation of a weak η-DP sequence $q_t^{\eta\text{-wdp}}$ (e.g. an algorithm such as Hedge) with probability $1-\lambda$.

Next, we state our main theoretical result of this subsection.

Theorem 4.3. *Let Π be a* NIC *single-round mechanism class and A be agents. If $\frac{\lambda}{4\eta}\beta(\pi^{\text{com}}) \geq \alpha(A)$, then $\mathcal{M}^{\text{wDP}}$ is NIC with respect to agents A.*

From this theorem, we can conclude that when $\beta(\pi^{\text{com}}) > 0$ and the agents are not too long-sighted (i.e. $\alpha_T(\mathcal{A})$ is sub-linear in T), for a suitable choice of λ and η, the proposed mechanism $\mathcal{M}^{\text{wDP}}$ is Nash incentive-compatible.

Proof sketch: The weak η-DP sequence limits how much the agent can earn by changing a reported type in each round, given that other agents' strategies do not depend on histories. Combining the maximum possible earnings limited by the weak η-DP sequence and the minimum penalty induced by the commitment mechanism, we show that no agent has an incentive to deviate from truth-telling in the last round when other agents are truthful. By recursively applying this argument, we demonstrate that being truthful in all rounds is optimal for any agent whenever others are truthful, making truth-telling a Nash equilibrium. The full proof is in App.A.

4.2. NIC online learning via Hedge

To apply Theorem 4.3 and design a NICOM, we need to find a sequence of maps (online learners) $q_{\leq T}^{\eta\text{-wdp}}$ that satisfies the weak DP constraint, while also achieving sub-linear regret. In this subsection, we show that the Hedge algorithm (Hazan, 2016), one of the most common online learning algorithms for the expert problem, satisfies both properties. For simplicity of exposition, we will assume that the mechanism class Π is finite. If this is not the case, we can discretize the mechanism's learnable parameters, and then separately control the discretization error.

Hedge algorithm. First, we state the Hedge update rule. Here, we are interested in finding a mechanism π in a finite set of mechanisms Π, which maximizes the sum of objectives $\sum_t G_t(\theta, s)$. For any $\eta \geq 0$, past reported types $b_{<t}$, observed objectives $G_{<t}$ and a mechanism class Π, let $q_{\eta,t}^{\text{hedge}}(b_{<t}, G_{<t})$ be the probability distribution over Π whose probability masses are given by

$$q_{\eta,t}^{\text{hedge}}(b_{<t}, G_{<t})(\pi) \propto \exp\left(\eta \sum_{\tau=1}^{t-1} F_\tau(b_\tau, \pi)\right). \quad (7)$$

Here $F_t(\theta_t, \pi) := \mathbb{E}_{s\sim\pi(\theta_t)} G_t(\theta_t, s)$. Consequently, we may view each $q_{\eta,t}^{\text{hedge}}(b_{<t})$ as an exponential mechanism (Proposition 2.3) with $g := \sum_{\tau=1}^{t-1} F_\tau$ and $\Delta g = 2$ as $F_t(\cdot,\cdot) \in [-1, 1]$. Hence, we immediately conclude that:

Lemma 4.4. $\left(q_{\eta,t}^{\text{hedge}}\right)_{t\in[T]}$ *given in Definition 7 is a weak 4η-DP sequence.*

The Hedge NICOM: We can now present our main method, an algorithm for online mechanism learning using Hedge. For any $\lambda \in [0, 1]$, a commitment mechanism π^{com}, and mechanism class Π, let $\mathcal{M}^{\text{Hedge}}$ be an online mechanism that outputs the following mechanism for each $t \in [T]$. Below, $\pi_t^{\text{sr}} \in \Pi$ is a single-round mechansim sampled from $q_{\eta,t}^{\text{hedge}}(b_{<t}, G_{<t})$. That is,

$$\pi_t(\cdot) := (1-\lambda)\cdot \pi_t^{\text{sr}}(\cdot) + \lambda \cdot \pi^{\text{com}}(\cdot). \quad (8)$$

Our theorem below states that $\mathcal{M}^{\text{Hedge}}$ is a NICOM with suitable choices of λ, η and π^{com}.

Theorem 4.5. *Let Π be a* NIC *mechanism class and $\mathcal{A}$ be an agent class. Then, the Hedge learner $\mathcal{M}^{\text{Hedge}}$ (8) is a* NICOM *with regret guarantee*

$$\sup_{A\in\mathcal{A}_T} \text{Reg}(\theta_{\leq T}, \mathcal{M}^{\text{Hedge}}, \Pi; A) \in O\left(\left(\log|\Pi| + \beta(\pi^{\text{com}})^{-1}\right)\cdot\sqrt{\alpha_T(\mathcal{A})\,T}\right)$$

for $\eta := 1/\sqrt{\alpha_T(\mathcal{A})\,T}$ and $\lambda \in \Theta\left(\frac{\sqrt{\alpha_T(\mathcal{A})/T}}{\beta(\pi^{\text{com}})}\right)$. In particular, when $\alpha_T(\mathcal{A}) \in O(T^h)$, the regret is in $O\left(T^{\frac{1+h}{2}}\right)$.

Proof sketch: The proof follows almost directly from the standard regret bound of the Hedge algorithm (Hazan, 2016). By the linearity of expectation, the commitment mechanism causes a small λT offset to the standard regret. After choosing appropriate parameters, we obtain the stated regret. We give the full proof of Theorem 4.5 in Appendix B.

As mentioned in §1.1, thanks to Theorem 4.3, which relies only on our weak notion of sequential DP, we are able to utilize the Hedge algorithm, resulting in a logarithmic dependence on the size of the mechanism class $|\Pi|$.

4.3. Discussion I: Contextualizing the Regret Bound

In this subsection, we provide several pointers to help interpret and contextualize our main regret bound in Theorem 4.5. It is important to beware that the regret bounds cited below, which have been suitably converted to meet our convention, are *not* directly comparable to ours due to differences in problem settings and strategy-proofness guarantees.

The work most closely related to ours is by Abernethy et al. (2019), who present an approximately IC online auction within the Bayesian framework. They establish a regret bound of $O(T/\epsilon + (k\epsilon)^{1/3}T)$ with an IC guarantee, where ϵ is the DP parameter that appears in the approximate IC guarantee and k is the maximum number of rounds an agent can participate in. This bound becomes vacuous when k grows with T, leading to a superlinear regret even when ϵ tends to 0. Specifically, if we formulate this bound in terms of our h parameter so that $k \in O(T^h)$, the bound

becomes $O(T^{1+h/3})$ for fixed ϵ. In contrast, our bound remains sublinear $O(T^{(1+h)/2})$ whenever k is sublinear.

Other relevant works by Amin et al. (2013) and Liu et al. (2018) consider *strategic regret* in the online posted-price auction, which is a regret measured with respect to non-myopic buyers and is defined as $\mathsf{SReg}(A) := \mathsf{Rev}(p^\star, v) - \mathsf{Rev}(A, B)$, where $\mathsf{Rev}(v, p^\star)$ is the revenue by the best-fixed price p^* with respect to the buyer's true values $v := (v_1, ..., v_T)$ and $\mathsf{Rev}(A, B)$ is the revenue of algorithm A with respect to non-myopic buyer B.

For T^h long-sightedness (i.e., effective horizon), Amin et al. (2013) offer strategic regret bounds of $O(T^{1/2+h})$ when the buyer's valuation is fixed in all rounds (Theorem 1 therein), which is slightly worse than our $O(T^{(1+h)/2})$ bound, and $O(cT^{\sqrt{h}})$ for stochastic valuations, where c is an instance-dependent factor (Theorem 2 therein). However, these settings are less general than our adversarial setting.

Liu et al. (2018) demonstrate a regret bound of $\widetilde{O}(T^{(3+h)/4})$ for adversarial settings when there are continuous values, but their analysis can obtain a $\widetilde{O}(T^{(1+h)/2})$ bound for discrete values which match our regret bound. Note that their algorithm is specific to single-item auctions, and conceptually more complicated than ours.

As seen above, the recurring theme of regret bounds that are "robust" against non-myopic agents is their dependency on the agents' long-sightedness. Notably, all of the above bounds become vacuous when the long-sightedness becomes linear, that is, when $h = 1$. This phenomenon aligns with the negative result given in Amin et al. (2013).

4.4. Discussion II: Limitations and Extensions

Our construction relies on the commitment mechanism and our regret bound depends on the power of the commitment mechanism to penalize untruthful agents. Our examples in §5 and Appendix C show that it is often possible to find such a commitment mechanism, quantify its penalty gap, and obtain transparent bounds on the regret. That said, designing a good commitment mechanism may not always be straightforward for a problem.

To simplify the exposition, we only consider *exact* commitment mechanisms that strictly punish any type report that is not *equal* to the true type. The existence of an exact commitment mechanism often relies on the assumption that agents' types are discrete (for continuous types, the gap may be arbitrarily close to zero). Hence, in this work, we only consider mechanisms with discrete agent types. Our construction can be extended to incorporate continuous types via an approximate commitment mechanism, leading to an approximate IC guarantee. However, to simplify the presentation and focus on the main insights, we focus on discrete types and consider exact commitment mechanisms.

In mechanism design, we are often interested in *individual rationality* (IR), which stipulates that the agent's utility when participating truthfully is larger than when not participating. While we have not explicitly considered IR, by choosing an IR mechanism class and commitment mechanism, we straightforwardly obtain an IR online mechanism.

We also reiterate that we assumed $|\Pi|$ was finite in §4.2 when applying the Hedge algorithm. If this is not the case, we can choose a discretization with good coverage, apply Hedge on the discretization, and then control the discretization error to obtain final regret bounds.

Finally, we also wish to mention that our methods can be quite straightforwardly extended to contextual settings, where an online mechanism takes in a piece of information $x \in X$ (such as UserID) as an auxiliary input to choose a mechanism for each round. In this case, we can simply instantiate the Hedge over a subset of policies Π^X.

5. Application Examples

We now apply our framework to develop no-regret NICOMs for three different mechanism design problems: *(i)* Online facility location, *(ii)* VCG auctions with reserve prices under *ex-post* externality, and *(iii)* online optimal resource allocation. Of these, *(i)* and *(iii)* are examples of mechanism design without money. Due to space constraints, we only present the first two problems here, and defer the online optimal resource allocation to Appendix C.

5.1. Online Facility Location.

Consider a government trying to locate k moving facilities (e.g. moving clinics, libraries) every day. The local government wants to locate these facilities to minimize the average distance between each resident and their closest facility, weighted by utilization rate. To simplify the problem, we assume n residents are located on a finite 1d grid $I_m := \left\{0, \frac{1}{m}, \frac{2}{m}, \ldots, 1\right\}$, hence $\Theta_i = I_m$ for all $i \in [n]$. A social outcome $s \in S$ is a tuple $((x_1, o_1), \ldots, (x_k, o_k))$ where $x_l \in I_m$ and $o_l \subseteq [n]$, where x_l indicates the location of facility l and o_l is the set of residents who are allowed to use facility l.

Then, for an outcome s, resident i's utility on round t when their location is $\theta_{i,t}$ is

$$u_i(\theta_{i,t}, s) := 1 - \min_{l \in [k] \text{ s.t. } i \in o_l} |x_l - \theta_{i,t}|,$$

and $u_i(\theta_{i,t}, s) := 0$ if $i \notin o_l$ for any $l \in [k]$. The government's objective on round t given types θ_t is

$$G_t(\theta_t, s) := \sum_{i \in [n]} r_{i,t} \cdot u_i(\theta_{i,t}, s).$$

Here, $r_{i,t} \in [0, 1]$ is the utilization rate of a facility by the resident i during round t.

Posted-location mechanism. We consider the class of "posted-location" mechanisms, which is parametrized by $w \in I_m^k$ representing the locations of facilities. This class $\Pi^{\text{PL}}_{m,k}$ is defined as follows: for each $w \in I_m^k$, $\pi^{\text{pl}}_w \in \Pi^{\text{PL}}_{m,k}$ is a mechanism that deterministically outputs $((w_1, [n]), \ldots, (w_k, [n])) \in S$ regardless of input types. That is, it deterministically sets the locations of the facilities at $w \in I_m^k$, and allows any agents to use any facility so that agents can pick the closest facility to them. As the mechanism's output does not depend on the reported types, it is NIC in a single round; but if we naively use previous reports to set future locations, the online mechanism is not NIC.

Commitment mechanism. We construct a commitment mechanism π^{com} for the online facility locating as follows: Given type reports $b_t := (b_{i,t})_{i\in[n]}$, π^{com} samples $l \in [m]$ uniformly at random and place one facility in $\frac{l-1}{m}$ and other $k-1$ facilities in $\frac{l}{m}$ (i.e. $x_1 = \frac{l-1}{m}$ and $x_l = \frac{l}{m}$ for $2 \le l \le k$). Then, π^{com} sets $o_l = \arg\min_{i\in[n] \text{ s.t. } b_{i,t}\neq\perp} |x_l - b_{i,t}|$ for each $l \in [k]$, that is, facility l is only accessible to residents whose closest facility is l according to their reports.

Then, the penalty gap $\beta(\pi^{\text{com}})$ is at least m^{-2}. To see this, suppose $b_{i,t} > \theta_{i,t}$. If π^{com} chooses an outcome s such that $x_1 = \theta_{i,t}$, the resident i suffers at least m^{-1} utility loss compared to that of reporting $\theta_{i,t}$, as they have to use some facility $l \ge 2$ located at $x_l = \theta_{i,t} + m^{-1}$. As π^{com} chooses such an outcome with probability at least m^{-1}, the expected utility loss is at least m^{-2}. By a similar argument, the same holds when $b_{i,t} < \theta_{i,t}$. Hence $\beta(\pi^{\text{com}}) \ge m^{-2}$.

Regret bound. Finally, as there are $|I_m^k|$ possible placements of facilities, we have that $\left|\Pi^{\text{PL}}_{m,k}\right| = |I_m|^k = (m+1)^k$. Hence, there is a NICOM for $\Pi^{\text{PL}}_{m,k}$ with regret $O\Big((k \log m + m^2) \cdot T^{\frac{1+h}{2}}\Big)$ for any agent class $\mathcal{A}$ such that $\sup_{A\in\mathcal{A}_T} \alpha(A) \in O(T^h)$ for some $h \in [0, 1)$.

5.2. VCG Mechanism under *Ex-post* Externalities

Consider a government selling a permit to n firms via VCG mechanism (Nisan et al., 2007). We use $o \subseteq [n]$ to denote the set of firms who get the permit and $p \in [0, 1]^n$ is the price of the permit charged to each firm. Then, the social outcome space is $S := \{(o, p) : o \subseteq [n], p \in [0, 1]^n\}$. For all $i \in [n]$, let type space Θ_i be $I_m := \left\{0, \frac{1}{m}, \frac{2}{m}, \ldots, 1\right\}$, representing a firm's possible valuations of the permit.

The government's objective function is the social welfare: $G_t(\theta_t, s) := \sum_{i\in o} \theta_{i,t} - c_t(o)$ where $\theta_{i,t}$ is firm i's valuation of the permit at round t and $c_t(\cdot)$ is an *ex-post* externality (such as pollution) revealed after round t. To balance the total value and externalities, the government introduces bidder-specific reserve prices $w \in I_m^n$, where w_i is the minimum price of the permit charged to firm i.

VCG mechanism with bidder-specific reserve prices. We define a class of mechanisms $\Pi^{\text{VCG}}_{n,m}$ as follows: Given input bids $b := (b_i)_{i\in[n]}$, $\pi^{\text{vcg}}_w \in \Pi^{\text{VCG}}$ with bidder-specific reserve prices $w \in I_m^n$ decides the set of winners $o^\star \subseteq \{0, 1\}^n$ as follows:

$$o^\star \in \arg\max_{o\subseteq\mathcal{P}_w(b)} \sum_{j\in\mathcal{P}_w(b)} b_j \cdot \mathbb{1}\big[j \in o\big],$$

where $\mathcal{P}_w(b) := \{i \in [n] : b_i \neq \perp,\, b_i \ge w_i\}$, and charges $p^\star \in [0, 1]^n$ where

$$p_i^\star = \max_{o\subseteq\mathcal{P}_w(b)} \Bigg(\sum_{j\in\mathcal{P}_w(b)\setminus\{i\}} b_j \cdot \mathbb{1}\big[j \in o\big] \Bigg) - \sum_{j\in\mathcal{P}_w(b)\setminus\{i\}} b_j \cdot \mathbb{1}\big[j \in o^\star(b, w)\big].$$

For any reserve price, this mechanism is NIC (actually, DSIC) with respect to the quasi-linear utility $u_i(\theta_i, s) := (\theta_i - p_i) \cdot \mathbb{1}[i \in o]$ for $s \in S$ (Nisan et al., 2007).

Commitment mechanism. Our choice of π^{com} for this problem is as follows: Given bids $b_t := (b_{i,t})_{i\in[n]}$, π^{com} samples $z \in I_{2m}$ uniformly at random and outputs (o, p) where $o = \{i \in [n] : b_{i,t} \neq \perp,\, b_{i,t} \ge z\}$ and $p_i = z$ for all $i \in [n]$. Then, the penalty gap $\beta(\pi^{\text{com}})$ is at least $(2m)^{-2}$ which can be seen as follows: Suppose $b_{i,t} < \theta_{i,t}$. If π^{com} chooses an outcome such that $p_i = \theta_{i,t} - (2m)^{-1}$, the firm i suffers at least $(2m)^{-1}$ utility loss compared to that of reporting $\theta_{i,t}$. Similarly, the firm i suffers at least $(2m)^{-1}$ when $b_{i,t} > \theta_{i,t}$, if π^{com} chooses an outcome such that $p_i = \theta_{i,t} + (2m)^{-1}$. As these events happens with probability at least $(2m)^{-1}$, $\beta(\pi^{\text{com}}) \ge (2m)^{-2}$.

Regret bound. To sum up, we have that $\beta(\pi^{\text{com}}) \ge (2m)^{-2}$ and $|\Pi^{\text{VCG}}_{n,m}| = |I_m^n| = (m+1)^n$. Therefore, by Theorem 4.5, there is a NICOM for this problem with regret $O\Big((n \log m + m^2) \cdot T^{\frac{1+h}{2}}\Big)$ under any agent class $\mathcal{A}$ such that $\alpha_T(\mathcal{A}) \in O(T^h)$ for some $h \in [0, 1)$.

6. Summary

We studied an online mechanism design setting, where we wish to design an online mechanism that is Nash incentive-compatible (NIC) throughout the entire learning process against non-myopic agents, while maximizing an application-specific objective. In contrast with prior work, our method applies to general mechanism design problems in online adversarial settings. We show that a weak notion of DP, which is satisfied by the Hedge algorithm, is sufficient to achieve NIC. This allows us to obtain regret bounds that scale logarithmically in the size of the mechanism class. We apply our framework to design NIC online learning procedures for three different mechanism design problems.

Acknowledgements

We would like to thank the reviewers for taking the time and effort necessary to review the manuscript. We sincerely appreciate their valuable comments and suggestions, which helped improve the quality of the manuscript.

Impact Statement

This paper presents work whose goal is to advance the field of machine learning and mechanism Design. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Abernethy, J. D., Cummings, R., Kumar, B., Taggart, S., and Morgenstern, J. H. Learning auctions with robust incentive guarantees. *Advances in Neural Information Processing Systems*, 32, 2019.

Amin, K., Rostamizadeh, A., and Syed, U. Learning prices for repeated auctions with strategic buyers. *Advances in Neural Information Processing Systems*, 26, 2013.

Amin, K., Rostamizadeh, A., and Syed, U. Repeated contextual auctions with strategic buyers. *Advances in Neural Information Processing Systems*, 27, 2014.

Arora, S., Hazan, E., and Kale, S. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Ashlagi, I., Daskalakis, C., and Haghpanah, N. Sequential mechanisms with ex-post participation guarantees. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 213–214, 2016.

Babaioff, M., Sharma, Y., and Slivkins, A. Characterizing truthful multi-armed bandit mechanisms. In *Proceedings of the 10th ACM Conference on Electronic commerce*, pp. 79–88, 2009.

Babaioff, M., Kleinberg, R. D., and Slivkins, A. Truthful mechanisms with implicit payment computation. *Journal of the ACM (JACM)*, 62(2):1–37, 2015.

Balcan, M.-F., Sandholm, T., and Vitercik, E. A general theory of sample complexity for multi-item profit maximization. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pp. 173–174, 2018.

Balcan, M.-F. F., Sandholm, T., and Vitercik, E. Sample complexity of automated mechanism design. *Advances in Neural Information Processing Systems*, 29, 2016.

Baron, D. P. and Besanko, D. Regulation and information in a continuing relationship. *Information Economics and Policy*, 1(3):267–302, 1984.

Bergemann, D. and Välimäki, J. The dynamic pivot mechanism. *Econometrica*, 78(2):771–789, 2010.

Bergemann, D. and Välimäki, J. Dynamic mechanism design: An introduction. *Journal of Economic Literature*, 57(2):235–274, 2019.

Blum, A. and Hartline, J. D. Near-optimal online auctions. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1156–1163, 2005.

Cole, R. and Roughgarden, T. The sample complexity of revenue maximization. In *Proceedings of the forty-sixth annual ACM Symposium on Theory of Computing*, pp. 243–252, 2014.

Courty, P. and Hao, L. Sequential screening. *The Review of Economic Studies*, 67(4):697–717, 2000.

Daskalakis, C., Deckelbaum, A., and Tzamos, C. The complexity of optimal mechanism design. In *Proceedings of the twenty-fifth annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1302–1318. SIAM, 2014.

Demers, A., Keshav, S., and Shenker, S. Analysis and simulation of a fair queueing algorithm. *ACM SIGCOMM Computer Communication Review*, 19(4):1–12, 1989.

Den Boer, A. V. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in Operations Research and Management Science*, 20(1): 1–18, 2015.

Deng, Y., Lahaie, S., and Mirrokni, V. A robust non-clairvoyant dynamic mechanism for contextual auctions. *Advances in Neural Information Processing Systems*, 32, 2019a.

Deng, Y., Mirrokni, V., and Zuo, S. Non-clairvoyant dynamic mechanism design with budget constraints and beyond. *Available at SSRN 3383231*, 2019b.

Deng, Y., Lahaie, S., and Mirrokni, V. Robust pricing in dynamic mechanism design. In *International Conference on Machine Learning*, pp. 2494–2503. PMLR, 2020.

Devanur, N. R. and Kakade, S. M. The price of truthfulness for pay-per-click auctions. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, pp. 99–106, 2009.

Drutsa, A. Horizon-independent optimal pricing in repeated auctions with truthful and strategic buyers. In *Proceedings of the 26th International Conference on World Wide Web*, pp. 33–42, 2017.

Drutsa, A. Weakly consistent optimal pricing algorithms in repeated posted-price auctions with strategic buyer. In *International Conference on Machine Learning*, pp. 1319–1328. PMLR, 2018.

Drutsa, A. Optimal non-parametric learning in repeated contextual auctions with strategic buyer. In *International Conference on Machine Learning*, pp. 2668–2677. PMLR, 2020a.

Drutsa, A. Reserve pricing in repeated second-price auctions with strategic bidders. In *International Conference on Machine Learning*, pp. 2678–2689. PMLR, 2020b.

Dwork, C., Roth, A., et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Fotakis, D. and Tzamos, C. Winner-imposing strategyproof mechanisms for multiple facility location games. *Theoretical Computer Science*, 472:90–103, 2013.

Golrezaei, N., Javanmard, A., and Mirrokni, V. Dynamic incentive-aware learning: Robust pricing in contextual auctions. *Advances in Neural Information Processing Systems*, 32, 2019.

Hartline, J. D. and Roughgarden, T. Simple versus optimal mechanisms. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, pp. 225–234, 2009.

Hazan, E. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Jain, P., Kothari, P., and Thakurta, A. Differentially private online learning. In *Conference on Learning Theory*, pp. 24–1. JMLR Workshop and Conference Proceedings, 2012.

Kandasamy, K., Sela, G.-E., Gonzalez, J. E., Jordan, M. I., and Stoica, I. Online learning demands in max-min fairness. *arXiv preprint arXiv:2012.08648*, 2020.

Kearns, M., Pai, M., Roth, A., and Ullman, J. Mechanism design in large games: Incentives and privacy. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, pp. 403–410, 2014.

Kleinberg, R. and Leighton, T. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *Proceedings of 44th Annual IEEE Symposium on Foundations of Computer Science*, pp. 594–605. IEEE, 2003.

Liu, J., Huang, Z., and Wang, X. Learning optimal reserve price against non-myopic bidders. *Advances in Neural Information Processing Systems*, 31, 2018.

Lu, P., Sun, X., Wang, Y., and Zhu, Z. A. Asymptotically optimal strategy-proof mechanisms for two-facility games. In *Proceedings of the 11th ACM Conference on Electronic Commerce*, pp. 315–324, 2010.

McSherry, F. and Talwar, K. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science*, pp. 94–103. IEEE, 2007.

Mirrokni, V., Leme, R. P., Tang, P., and Zuo, S. Dynamic auctions with bank accounts. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 387–393, 2016.

Mirrokni, V., Paes Leme, R., Tang, P., and Zuo, S. Non-clairvoyant dynamic mechanism design. *Econometrica*, 88(5):1939–1963, 2020.

Mohri, M. and Munoz, A. Optimal regret minimization in posted-price auctions with strategic buyers. *Advances in Neural Information Processing Systems*, 27, 2014.

Mohri, M. and Munoz, A. Revenue optimization against strategic buyers. *Advances in Neural Information Processing Systems*, 28, 2015.

Moulin, H. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V. V. *Algorithmic game theory*. Cambridge University Press, 2007.

Nissim, K., Smorodinsky, R., and Tennenholtz, M. Approximately optimal mechanism design via differential privacy. In *Proceedings of the 3rd Innovations in Theoretical Computer Science*, pp. 203–213, 2012.

Procaccia, A. D. and Tennenholtz, M. Approximate mechanism design without money. *ACM Transactions on Economics and Computation (TEAC)*, 1(4):1–26, 2013.

Roughgarden, T. and Wang, J. R. Minimizing regret with multiple reserves. *ACM Transactions on Economics and Computation (TEAC)*, 7(3):1–18, 2019.

Thomas, J. and Worrall, T. Income fluctuation and asymmetric information: An example of a repeated principal-agent problem. *Journal of Economic Theory*, 51(2):367–390, 1990.

Zhang, L., Zhu, T., Xiong, P., Zhou, W., and Yu, P. S. More than privacy: Adopting differential privacy in game-theoretic mechanism design. *ACM Computing Surveys (CSUR)*, 54(7):1–37, 2021.

Zhiyanov, A. and Drutsa, A. Bisection-based pricing for repeated contextual auctions against strategic buyer. In *International Conference on Machine Learning*, pp. 11469–11480. PMLR, 2020.

A. Proof of Theorem 4.3

In this appendix, we give a proof of Theorem 4.3 restated in below.

Theorem 4.3. *Let* Π *be a* NIC *single-round mechanism class and* A *be agents. If* $\frac{\lambda}{4\eta}\beta(\pi^{\text{com}}) \geq \alpha(A)$*, then* $\mathcal{M}^{\text{wDP}}$ *is NIC with respect to agents* A*.*

Proof. Fix any $i \in [n]$. For any strategy profile (σ_i, σ_{-i}), we can rewrite agent i's expected utility as follows:

$$\begin{aligned} u_i\big(\sigma_i, \sigma_{-i}; \mathcal{M}^{\text{wDP}}, A\big) &:= \mathbb{E}\left[\sum_{t\in\mathcal{T}_i} \gamma_i(t) u_i(\theta_{i,t}, s_t)\right] \\ &= \sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{b_{\leq t}}\left[\mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, b_{\leq t}\Big]\right] \\ &= \mathop{\mathbb{E}}_{b_{\leq T}}\left[\sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, b_{\leq t}\Big]\right], \end{aligned} \tag{A.1}$$

where the first expectation is over all randomness and $\mathbb{E}_{b_{\leq t}}$ $\big(\mathbb{E}_{b_{\leq T}}\big)$ is the expectation with respect to the distribution of $b_{\leq t}$ $(b_{\leq T})$ induced by the interaction between agents A employing strategy (σ_i, σ_{-i}) and online mechanism $\mathcal{M}^{\text{wDP}}$ via Protocol 1. And $\mathbb{E}_{s_t}[\,\cdot\,|b_{\leq t}]$ is the expectation with respect to the following distribution:

$$\mathcal{D}(b_{\leq t}) := (1-\lambda)\cdot q_t^{\eta\text{–wdp}}(b_{<t})(b_t) + \lambda \cdot \pi^{\text{com}}(b_t) \in \Delta(S),$$

where $q_t^{\eta\text{–wdp}}(b_{<t})(b_t) \in \Delta(S)$ is the distribution defined by process $s \sim \pi(b_t)$, $\pi \sim q_t^{\eta\text{–wdp}}(b_{<t})$.

For each $t \in [T]$, let $b_{i,\leq t} := (b_{i,\tau})_{\tau\in[t]}$ where $b_{i,\tau} :=\perp$, if and only if $\tau \notin \mathcal{T}_i$ and $b_{-i,\leq t} := (b_{j,\leq t})_{j\in[n]\setminus\{i\}}$, define $\theta_{i,\leq t}$, $\theta_{-i,\leq t}$ similarly but with true types. With these notations, (A.1) can be rewritten as

$$u_i\big(\sigma_i, \sigma_{-i}; \mathcal{M}^{\text{wDP}}, A\big) = \mathop{\mathbb{E}}_{b_{i,\leq T}, b_{-i,\leq T}}\left[\sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, b_{i,\leq t},\, b_{-i,\leq t}\Big]\right].$$

Suppose other players are truthful, that is, $\sigma_{-i} = \theta_{-i}$. Substituting $\sigma_{-i} = \theta_{-i}$ to the above, we have that $b_{-i,\leq T} = \theta_{-i,\leq T}$ and hence

$$u_i\big(\sigma_i, \theta_{-i}; \mathcal{M}^{\text{wDP}}, A\big) = \mathop{\mathbb{E}}_{b_{i,\leq T}}\left[\sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, b_{i,\leq t},\, \theta_{-i,\leq t}\Big] \,\middle|\, \theta_{-i,\leq T}\right], \tag{A.2}$$

where $\mathbb{E}_{b_{i,\leq T}}$ is the distribution over agent i's reported types $(b_{i,t})_{t\in\mathcal{T}_i}$ induced by playing σ_i while other agents' strategies are θ_{-i}. Using Hölder's inequality to (A.2), we have that

$$u_i\big(\sigma_i, \theta_{-i}; \mathcal{M}^{\text{wDP}}, A\big) \leq \sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, \hat{b}_{i,\leq t},\, \theta_{-i,\leq t}\Big] \tag{A.3}$$

for some $\hat{b}_{i,\leq T}$.

Now, we show that the above upper bound is attained by $\hat{b}_{i,\leq T}$ such that $\hat{b}_{i,t} = \theta_{i,t}$ for all $t \in \mathcal{T}_i$. To see this, define a strategy modification function $\phi_i : \prod_{t=1}^{T}(\Theta_i \cup \{\perp\}) \to \prod_{t=1}^{T}(\Theta_i \cup \{\perp\})$ as follows:

1. Given input $b_{i,\leq T}$, ϕ_i checks whether there exists $t \in \mathcal{T}_i$ such that $b_{i,t} \neq \theta_{i,t}$.
2. If there is no such $t \in [T]$, return unmodified $(b_{i,t})_{t\in\mathcal{T}_i}$.
3. Otherwise, let $t^\star$ be the largest $t \in \mathcal{T}_i$ such that $b_{i,t} \neq \theta_{i,t}$, and return the modified input $(b_{i,t})_{t\in\mathcal{T}_i}$ where $b_{i,t^\star}$ is set to $\theta_{i,t^\star}$ and other elements are unchanged.

Let $\phi_i(b_{i,\le t})$ denote the restriction of $\phi_i(b_{i,\le T})$ up to round t. Then, we claim that for any $b_{i,\le T}$, the following holds.

$$\sum_{t\in\mathcal{T}_i}\gamma_i(t)\,\mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t},s_t)\,\Big|\,b_{i,\le t},\,\theta_{-i,\le t}\Big]\le\sum_{t\in\mathcal{T}_i}\gamma_i(t)\,\mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t},s_t)\,\Big|\,\phi_i(b_{i,\le t})\,,\,\theta_{-i,\le t}\Big]. \tag{A.4}$$

Proof of (A.4)**.** First, it suffices to consider only $b_{i,\le T}$ such that $b_{i,t}\neq\theta_{i,t}$ for some $t\in\mathcal{T}_i$ as otherwise $\phi_i(b_{i,\le T})=b_{i,\le T}$. Moreover, as ϕ_i only modifies $b_{i,t^\star}$ where $t^\star$ is the largest $t\in\mathcal{T}_i$ such that $b_{i,t}\neq\theta_{i,t}$, it suffices to show

$$\sum_{t\in\mathcal{T}_i\cap[t^\star:T]}\gamma_i(t)\,\mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t},s_t)\,\Big|\,b_{i,\le t},\,\theta_{-i,\le t}\Big]\le\sum_{t\in\mathcal{T}_i\cap[t^\star:T]}\gamma_i(t)\,\mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t},s_t)\,\Big|\,\phi_i(b_{i,\le t}),\,\theta_{-i,\le t}\Big],$$

where $[t^\star:T]:=\{t^\star,t^\star+1,\dots,T\}$. By the linearity of expectation, $\sum_{t\in\mathcal{T}_i}\gamma_i(t)\,\mathbb{E}_{s_t}\big[u_i(\theta_{i,t},s_t)\big|b_{i,\le t},\theta_{-i,\le t}\big]$ can be unfolded as

$$(1-\lambda)\cdot\gamma_i(t)\sum_{t\in\mathcal{T}_i\cap[t^\star:T]}\mathop{\mathbb{E}}_{s_t\sim\mathcal{A}_t(b_{i,\le t})}u_i(\theta_{i,t},s_t)+\lambda\cdot\gamma_i(t)\sum_{t\in\mathcal{T}_i\cap[t^\star:T]}\mathop{\mathbb{E}}_{s_t\sim\mathcal{B}_t(b_{i,t})}u_i(\theta_{i,t},s_t), \tag{A.5}$$

where the former expectations are over $\mathcal{A}_t(b_{i,\le t}):=q_t^{\eta\text{–wdp}}\big(b_{i,<t},\theta_{-i,\le t}\big)\big(b_{i,t},\theta_{-i,t}\big)$ and the later expectations are over $\mathcal{B}_t(b_{i,t}):=\pi^{\text{com}}\big(b_{i,t},\theta_{-i,t}\big)$.

Since $\pi(\cdot,w)$ is NIC and ϕ_i modifies $b_{i,t^\star}$ to $\theta_{i,t^\star}$, the following holds for $t=t^\star$.

$$\mathop{\mathbb{E}}_{s_t\sim\mathcal{A}_{t^\star}(b_{i,\le t^\star})}u_i(\theta_{i,t^\star},\,s_{t^\star})\le\mathop{\mathbb{E}}_{s_t\sim\mathcal{A}_{t^\star}(\phi_i(b_{i,\le t^\star}))}u_i(\theta_{i,t^\star},\,s_{t^\star}). \tag{A.6}$$

On the other hand, for each $t>t^\star$, $\mathcal{A}_t$ is η-DP with respect to $b_{i,<t}$ by the post-processing lemma (Lemma A.1). Hence, for all $t>t^\star$ we obtain

$$\mathop{\mathbb{E}}_{s_t\sim\mathcal{A}_t(b_{i,\le t})}\Big[\,u_i(\theta_{i,t},s_t)+1\,\Big]\le e^\eta\cdot\mathop{\mathbb{E}}_{s_t\sim\mathcal{A}_t(\phi_i(b_{i,\le t}))}\Big[\,u_i(\theta_{i,t},s_t)+1\,\Big],$$

which follows from the definition of the weak η-DP sequence (Definition 2.2) , the identity $\mathbb{E}[X]=\int_0^\infty\Pr[X\ge x]dx$ for any non-negative random variable X and $u_i(\cdot,\cdot)\in[-1,1]$. As $e^\eta\le1+2\eta$ for $\forall\eta\in[0,1]$, the above implies

$$\mathop{\mathbb{E}}_{s_t\sim\mathcal{A}_t(b_{i,\le t})}u_i(\theta_{i,t},s_t)\le\mathop{\mathbb{E}}_{s_t\sim\mathcal{A}_t(\phi_i(b_{i,\le t}))}u_i(\theta_{i,t},s_t)+4\eta, \tag{A.7}$$

for all $t\in\mathcal{T}_i\cap[t^\star+1:T]$.

Next, we consider the later expectations over $\mathcal{B}_t$ in (A.5). Since $\mathcal{B}_t(b_{i,t}):=\pi^{\text{com}}\big(b_{i,t},\theta_{-i,t}\big)$ does not depend on past reported types $b_{i,t<t}$, for all $t\in\mathcal{T}_i\cap[t^\star+1:T]$, we have that

$$\mathop{\mathbb{E}}_{s_t\sim\mathcal{B}_t(b_{i,t})}u_i(\theta_{i,t},s_t)=\mathop{\mathbb{E}}_{s_t\sim\mathcal{B}_t(b_{i,t})}u_i(\theta_{i,t},s_t). \tag{A.8}$$

On the other hand, for $t=t^\star$ term, by the definition of $\beta(\pi^{\text{com}})$ (Definition 4.1), we obtain

$$\mathop{\mathbb{E}}_{s_t\sim\mathcal{B}_{t^\star}(b_{i,t^\star})}u_i(\theta_{i,t^\star},s_{t^\star})\le\mathop{\mathbb{E}}_{s_t\sim\mathcal{B}_{t^\star}(\theta_{i,t^\star})}u_i(\theta_{i,t^\star},s_{t^\star})-\beta(\pi^{\text{com}}), \tag{A.9}$$

as ϕ_i modifies $b_{i,t^\star}$ to $\theta_{i,t^\star}$.

Finally, by combining (A.5) – (A.9), we obtain

$$\begin{aligned}\sum_{t\in\mathcal{T}_i\cap[t^\star:T]}\gamma_i(t)\,\mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t},s_t)\,\Big|\,\phi_i(b_{i,\le t}),\,\theta_{-i,\le t}\Big]-\sum_{t\in\mathcal{T}_i\cap[t^\star:T]}\gamma_i(t)\,\mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t},s_t)\,\Big|\,b_{i,\le t},\,\theta_{-i,\le t}\Big]\\ \ge\gamma_i(t^\star)\cdot\lambda\cdot\beta(\pi^{\text{com}})-4\eta\cdot\sum_{t\in\mathcal{T}_i\cap[t^\star+1:T]}\gamma_i(t)\end{aligned}$$

$$\geq \gamma_i(t^\star) \cdot \lambda \cdot \beta(\pi^{\text{com}}) - 4\eta \cdot \min\left(\sum_{t=t^\star}^{T} \gamma_i(t),\, \gamma_i(t^\star) \cdot |\mathcal{T}_i|\right)$$

(by the monotonicity of $\gamma_i(\cdot)$)

$$\geq 4\eta \cdot \gamma_i(t^\star) \cdot \left(\frac{\lambda \cdot \beta(\pi^{\text{com}})}{4\eta} - \alpha(A)\right)$$

$$\geq 0,$$

as $\lambda\beta(\pi^{\text{com}})/4\eta \geq \alpha(A)$. Hence, we have (A.4). □

Now, we apply (A.4) recursively to itself:

$$\begin{aligned}
\sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, b_{i,\leq t},\, \theta_{-i,\leq t}\Big] &\leq \sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, \phi_i(b_{i,\leq t}),\, \theta_{-i,\leq t}\Big] \\
&\leq \sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, \phi_i \circ \phi_i(b_{i,\leq t}),\, \theta_{-i,\leq t}\Big] \\
&\leq \cdots
\end{aligned}$$

Since there are only finitely many rounds, after a finite number of applications of ϕ_i, the above process reaches to a fixed point $b_{i,\leq T} = \phi_i(b_{i,\leq T})$ such that $b_{i,t} = \theta_{i,t}$ for all $t \in \mathcal{T}_i$. Hence,

$$\begin{aligned}
u_i\big(\sigma_i, \theta_{-i}; \mathcal{M}^{\text{wDP}}, A\big) &\leq \sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, \hat{b}_{i,\leq t},\, \theta_{-i,\leq t}\Big], && \text{(by (A.3))} \\
&\leq \sum_{t\in\mathcal{T}_i} \gamma_i(t) \mathop{\mathbb{E}}_{s_t}\Big[u_i(\theta_{i,t}, s_t) \,\Big|\, \theta_{i,\leq t},\, \theta_{-i,\leq t}\Big] \\
&= u_i\big(\theta_i, \theta_{-i}; \mathcal{M}^{\text{wDP}}, A\big)\,.
\end{aligned}$$

Therefore, the strategy profile $\theta := (\theta_i)_{i\in[n]}$ is a Nash equilibrium. □

Lemma A.1 (Post-processing (Dwork et al., 2014))**.** *Let $f : Y \to Y'$ be an arbitrary function. If q is η-DP then $f \circ q : X^m \to \Delta(Y')$ is also η-DP.*

B. Proof of Theorem 4.5

In this appendix, we give a proof of Theorem 4.5 restated in below.

Theorem 4.5. *Let Π be a* NIC *mechanism class and $\mathcal{A}$ be an agent class. Then, the Hedge learner $\mathcal{M}^{\text{Hedge}}$ (8) is a* NICOM *with regret guarantee*

$$\sup_{A\in\mathcal{A}_T} \text{Reg}(\theta_{\leq T},\, \mathcal{M}^{\text{Hedge}}, \Pi; A) \in O\left(\left(\log|\Pi| + \beta(\pi^{\text{com}})^{-1}\right) \cdot \sqrt{\alpha_T(\mathcal{A})\, T}\right)$$

for $\eta := 1/\sqrt{\alpha_T(\mathcal{A})\, T}$ and $\lambda \in \Theta\left(\frac{\sqrt{\alpha_T(\mathcal{A})/T}}{\beta(\pi^{\text{com}})}\right)$. In particular, when $\alpha_T(\mathcal{A}) \in O(T^h)$, the regret is in $O\big(T^{\frac{1+h}{2}}\big)$.

Proof. Recall that $\mathcal{M}^{\text{Hedge}}$ is given as follows: For given $T \in \mathbb{N}$, $\mathcal{M}^{\text{Hedge}}$ outputs the following mechanism for each $t \in [T]$:

$$\pi_t(\cdot) := (1-\lambda) \cdot q^{\text{hedge}}_{\eta,t}(b_{<t})(\cdot) + \lambda \cdot \pi^{\text{com}}(\cdot),$$

where $\big(q^{\text{hedge}}_{\eta,t}(b_{<t})\big)_{t\in[T]}$ is the exponential weights sequence (Definition 7) with objectives $F_t(\theta_t, \pi) := \mathbb{E}_{s\sim\pi(\theta_t)}\, G_t(\theta_t, s)$.

First, we show that, for every $T \in \mathbb{N}$, there is a choice of (λ, η) such that the truth-telling profile is a Nash equilibrium of any game induced by $\mathcal{M}^{\text{Hedge}}$ and $A \in \mathcal{A}_T$. By Lemma 4.4 and Theorem 4.3, this holds whenever

$$\frac{\lambda\beta(\pi^{\text{com}})}{16\eta} \geq \sup_{A\in\mathcal{A}_T} \alpha(A) := \alpha_T(\mathcal{A}) \;\text{ and }\; \eta \in [0,1]. \tag{B.1}$$

Then, for $\eta := 1/\sqrt{\alpha_T(\mathcal{A})\,T} \in [0,1]$, there is a choice of λ for each $T \in \mathbb{N}$ such that (B.1) holds for all $T \in \mathbb{N}$ and $\lambda \in \Theta\Big(\frac{\sqrt{\alpha_T(\mathcal{A})/T}}{\beta(\pi^{\text{com}})}\Big)$. Moreover, to choose such λs, one only needs the knowledge of the whole class $\mathcal{A}$; hence such choices can be made before starting any round.

Next, we show that $\mathcal{M}^{\text{Hedge}}$ guarantees no-regret: $\sup_{A\in\mathcal{A}_T} \text{Reg}\big(\theta_{\le T}, \mathcal{M}^{\text{Hedge}}, \Pi; A\big) \in o(T)$. By the linearity of expectation and the standard regret bound of the exponential weights algorithm (Arora et al., 2012; Hazan, 2016), we have that

$$\sup_{A\in\mathcal{A}_T} \text{Reg}\big(\theta_{\le T}, \mathcal{M}^{\text{Hedge}}, \Pi; A\big) \le 4\eta\,T + \frac{1}{\eta}\log|\Pi| + \lambda\,T.$$

Since $\eta := 1/\sqrt{\alpha_T(\mathcal{A})\,T} \in [0,1]$ and $\lambda \in \Theta\Big(\frac{\sqrt{\alpha_T(\mathcal{A})/T}}{\beta(\pi^{\text{com}})}\Big)$, the LHS of the above bound is in $O\big((\log|\Pi| + \beta(\pi^{\text{com}})^{-1}) \cdot \sqrt{\alpha_T(\mathcal{A})\,T}\big)$. □

C. Application: Online Optimal Resource Allocation

In this appendix, we provide an additional application of Theorem 4.5: Designing a Nash incentive-compatible online mechanism for optimal resource allocation.

Setting. Consider there are k CPUs in a cluster shared by n users. A user's type $\theta \in [n]$ represents her true demand on CPU. The social outcome s is the CPU allocation vector, that is, $s_i \in \mathbb{N}$ represents the number of CPUs allocated to user i. Hence, the outcome space is $S_{n,k} := \big\{s \in \mathbb{N}^n : \sum_{i\in[n]} s_i = k\big\}$. We assume all users have identical utility functions given by $u(\theta, s) := \mathbb{1}\,[s \ge \theta]$, that is, an agent experiences utility 1 with enough resources and 0 for insufficient resources. Then, the cluster manager's objective function is the number of jobs completed, weighted by job importance. That is, $G_t(\theta_t, s) := \sum_{i\in[n]} r_{i,t} \cdot u(\theta_{i,t}, s)$, where $r_{i,t} \in [0,1]$s are *ex-post* weighting factors (such as *ex-post* resource utilization rates), specified by the cluster manager.

Target mechanisms. The cluster manager chooses any DSIC/NIC off-the-shelf single-round resource allocation mechanism with parameters. An example would be the max-min fair allocation mechanism $\Pi^{\text{MMF}}_{n,k}$ whose parameter is the initial endowment vector $w \in S_{n,k}$ (Demers et al., 1989; Kandasamy et al., 2020). Another example would be the posted-allocation mechanism whose outcome is always its parameter vector $w \in S_{n,k}$. With any choice of parametrized mechanism, the cluster manager's goal is to compete against the best parameter in terms of the cumulative social welfare $\sum_{t\in[T]} G_t(\theta_t, s)$.

Commitment mechanism. Our commitment mechanism π^{com} is as follows: Given input demands $b \in [k]^n$, $\pi^{\text{com}}(b)$ samples $i \in [n]$ and $j \in [2k]$ uniformly at random. If $j > b_i$, set $s_i = b_i$ and $s_j = \big\lfloor \frac{k-b_i}{n-1} \big\rfloor$ for all $j \ne i, \perp$. Otherwise, set $s_i = 0$ and $s_j = \big\lfloor \frac{k}{n-1} \big\rfloor$ for all $j \ne i, \perp$. Then, we show the following.

Lemma C.1. *For the commitment mechanism described above, we have that* $\beta(\pi^{\text{com}}) \ge (2nk)^{-1}$.

Proof. First, recall the definition of $\beta(\pi^{\text{com}})$ here. The penalty gap $\beta(\pi^{\text{com}})$ is the minimum of the following quantity,

$$\mathop{\mathbb{E}}_{s\sim\pi(\theta_i,b_{-i})} u(\theta_i, s) - \mathop{\mathbb{E}}_{s\sim\pi(b_i,b_{-i})} u(\theta_i, s),$$

over all i, b_{-i} and $\theta_i, b_i \in \Theta_i$ such that $b_i \ne \theta_i$. Fix an $i \in [n]$ and b_{-i}. Then, consider two cases $b_i > \theta_i$ and $b_i < \theta_i$.

Case $b_i > \theta_i$. In this case, we have that

$$\begin{aligned}
\mathop{\mathbb{E}}_{s\sim\pi(b_i,b_{-i})} u(\theta_i, s) &= \left(1 - \frac{1}{n}\right) \cdot C + \frac{1}{n} \cdot \frac{2k - b_i}{2k} \\
&= \left(1 - \frac{1}{n}\right) \cdot C + \frac{1}{n}\left(\frac{2k - \theta_i}{2k} - \frac{b_i - \theta_i}{2k}\right) \\
&= \underbrace{\left(1 - \frac{1}{n}\right) \cdot C + \frac{1}{n} \cdot \frac{2k - \theta_i}{2k}}_{\mathbb{E}_{s\sim\pi(\theta_i,b_{-i})} u(\theta_i, s)} - \frac{b_i - \theta_i}{2nk}
\end{aligned}$$

$$\leq \mathop{\mathbb{E}}_{s\sim\pi(\theta_i,b_{-i})} u(\theta_i, s) - \frac{1}{2nk} \qquad \text{(as } b_i - \theta_i \geq 1\text{)}$$

where C is a constant that does not depend on the input demand of user i. Hence, we have that

$$\mathop{\mathbb{E}}_{s\sim\pi(\theta_i,b_{-i})} u(\theta_i, s) - \mathop{\mathbb{E}}_{s\sim\pi(b_i,b_{-i})} u(\theta_i, s) \geq \frac{1}{2nk},$$

when $b_i > \theta_i$.

Case $b_i < \theta_i$. In this case, observe that

$$\begin{aligned}
\mathop{\mathbb{E}}_{s\sim\pi(b_i,b_{-i})} u(\theta_i, s) &= \left(1 - \frac{1}{n}\right) \cdot C \\
&= \left(1 - \frac{1}{n}\right) \cdot C + \frac{1}{n} \cdot \frac{2k - \theta_i}{2k} - \frac{1}{n} \cdot \frac{2k - \theta_i}{2k} \\
&= \mathop{\mathbb{E}}_{s\sim\pi(\theta_i,b_{-i})} u(\theta_i, s) - \frac{1}{n} \cdot \frac{2k - \theta_i}{2k} \\
&\leq \mathop{\mathbb{E}}_{s\sim\pi(\theta_i,b_{-i})} u(\theta_i, s) - \frac{1}{2n} \qquad \text{(as } \theta_i \leq k\text{)}
\end{aligned}$$

Hence, we have that

$$\mathop{\mathbb{E}}_{s\sim\pi(\theta_i,b_{-i})} u(\theta_i, s) - \mathop{\mathbb{E}}_{s\sim\pi(b_i,b_{-i})} u(\theta_i, s) \geq \frac{1}{2n} \geq \frac{1}{2nk},$$

when $b_i < \theta_i$. □

Regret bound. To sum up, we have that $\beta(\pi^{\text{com}}) \geq (2nk)^{-1}$ for this problem. If the cluster manager's goal is to maximize the social welfare with respect to $\Pi^{\text{MMF}}_{n,k}$, our Theorem 4.5 implies there is a NICOM for $\Pi^{\text{MMF}}_{n,k}$ with regret $O\Big((n \log k + nk) \cdot T^{\frac{1+h}{2}}\Big)$ under any agent class $\mathcal{A}$ such that $\alpha_T(\mathcal{A}) \in O(T^h)$ for some $h \in [0, 1)$.

D. Related works

Dynamic pricing. Online learning for auctions traces back to the seminal work of Kleinberg & Leighton (2003). Online learning of posted-price auctions is often referred to as dynamic pricing, and there is a wealth of literature on this subject following the aforementioned seminal works. Therefore, we direct readers to a review of this topic (Den Boer, 2015).

Strategy-robust pricing. One of the first works to investigate non-myopic agents was by Amin et al. (2013), who introduced a notion of regret that remains robust under strategic buyers. They devised a robust pricing algorithm for the repeated single-bidder single-item posted price mechanism, assuming that the buyer's value is drawn independently from an unknown value distribution.

This setting has subsequently been generalized in two ways. The first is the contextual setting, where the seller observes contextual information (such as UserID) while selling an item. Amin et al. (2014) were the first to study the problem of robustly learning prices under the contextual setting. Drutsa (2020a), Zhiyanov & Drutsa (2020), Golrezaei et al. (2019), and Deng et al. (2019a) further optimized regret bounds or extended the problem setting by considering noisy observations.

While previous works relied on distributional assumptions, a generalization of the previous stochastic setting to an online adversarial setup was explored by Liu et al. (2018). Particularly, Liu et al. (2018) extended this to a multi-bidder second-price auction in an adversarial environment, using a strongly differentially private algorithm for the expert problem. In comparison to our work, previous studies focus on scenarios where the objective is to learn a single parameter (price or reserve price).

Motivated by online advertising, a line of work has studied bandit learning in single-item auctions under truthfulness constraints (Babaioff et al., 2009; 2015; Devanur & Kakade, 2009). Here, buyers place a single bid at the beginning of the game representing their value for the bid, and on each round, the seller observes the number of clicks for the buyer she chooses for the round. When compared to our setting, as the buyers place only a single bid at the beginning, they have significantly less opportunity to manipulate outcomes in their favor.

Differential privacy and mechanism design. Since the seminal work of McSherry & Talwar (2007), differential privacy and mechanism design have become closely intertwined fields. Subsequently, Nissim et al. (2012) and Kearns et al. (2014) expanded the scope of mechanism design through differential privacy. Notably, Nissim et al. (2012) introduced the commitment mechanism and constructed general mechanisms from the exponential mechanism, including mechanisms without money (Moulin, 1980; Procaccia & Tennenholtz, 2013). Our work can be seen as a conceptual extension of Nissim et al. (2012)'s work to an online adversarial setting. As discussed earlier, Liu et al. (2018) employed differential privacy to design a strategy-robust online learning algorithm for single-item auctions. Additionally, as discussed in the main text, Abernethy et al. (2019) applied a differentially private primitive to design a strategy-robust version of Myerson's optimal auction in the online setting. For further exploration of the growing trend of using differential privacy to control incentives, we recommend reading a review by Zhang et al. (2021).