Optimal Mechanism in a Dynamic Stochastic Knapsack Environment

Jihyeok Jung[1], Chan-Oi Song[2], Deok-Joo Lee[1*], Kiho Yoon[2]

[1]Department of Industrial Engineering, Seoul National University
[2]Department of Economics, Korea University
firedaeman@snu.ac.kr, gnos@korea.ac.kr, leedj@snu.ac.kr, kiho@korea.ac.kr

Abstract

This study introduces an optimal mechanism in a dynamic stochastic knapsack environment. The model features a single seller who has a fixed quantity of a perfectly divisible item. Impatient buyers with a piece-wise linear utility function arrive randomly and they report the two-dimensional private information: marginal value and demanded quantity. We derive a revenue-maximizing dynamic mechanism in a finite discrete time framework that satisfies incentive compatibility, individual rationality, and feasibility conditions. It is achieved by characterizing buyers' utility and deriving the Bellman equation. Moreover, we propose the essential penalty scheme for incentive compatibility, as well as the allocation and payment policies. Lastly, we propose algorithms to approximate the optimal policy, based on the Monte Carlo simulation-based regression method and reinforcement learning.

Introduction

Dynamic resource allocation refers to the distribution of a limited amount of resources in a dynamic environment. This problem arises in various fields where the system has a fixed capacity to serve time-varying demands, such as cloud computing and software-as-a-service (Wu, Garg, and Buyya 2011; Wang, Liang, and Li 2013; Zhang et al. 2015). Therefore, it is crucial for system designers to find an optimal policy to achieve their desired objectives while considering the resource provision for future demands. Although the structure of dynamic allocation problems can vary depending on the objectives and assumptions, this study focuses on a specific category of dynamic resource allocation, the *Dynamic Stochastic Knapsack Problem* (DSKP).

In the original DSKP, the seller tries to sell a fixed amount of an item over a finite time horizon. During each time period, customers with information about the item's value and desired quantity enter the system. The seller, with stochastic knowledge of customer arrivals, aims to determine the optimal allocation strategy to maximize the total expected value (Papastavrou, Rajagopalan, and Kleywegt 1996; Kleywegt and Papastavrou 1998). Moreover, the original DSKP assumes that arriving demands are non-strategic. As each buyer is assumed to behave non-strategically, the problem of maximizing expected value naturally aligns with the goal of maximizing expected revenue, since the value can be interpreted as the maximum willingness to pay.

However, this assumption about non-strategic buyers does not hold true in many real-life scenarios. In reality, customers might strategically misreport their information to achieve favorable outcomes. They could request quantities exceeding their actual desire or overbid the resource's value in an attempt to improve their chances of securing an allocation priority. Given this potential, the seller needs to devise an optimal selling mechanism that maximizes expected revenue when buyers are behaving strategically.

If all participants act strategically, the aforementioned problem can be effectively tackled through mechanism design. By extending the work of Myerson (1981), which proposed the optimal mechanism in one-dimensional and static environments, we can derive the optimal dynamic mechanism in two-dimensional and dynamic environments. However, it's important to note that dynamic mechanisms often involve intricate mathematical formulations, which can pose challenges when directly applied to real-world service systems. To overcome these complexities, the development of approximation algorithms for the derived mathematical solutions becomes essential. This approach aims to render these mechanisms practically implementable, mitigating the challenges associated with their mathematical complexity.

To this end, we consider a revenue-maximizing dynamic mechanism under the dynamic stochastic knapsack environment. Buyers arrive randomly throughout the finite time horizon with private two-dimensional information about their desired quantities and values, which follows a continuous joint probability distribution. In this setting, the optimal allocation and payment rules that satisfy incentive compatibility and individual rationality are derived by modifying the characterization approach and dynamic program. Furthermore, we propose two algorithms to approximate the proposed mathematical solution by implementing Monte Carlo (MC) simulation-based regression method and reinforcement learning. We compare and analyze the performance of these algorithms to evaluate their effectiveness.

The suggested model and algorithms in this study contribute in the following ways:

- We propose a general structure for an optimal dynamic mechanism by relaxing assumptions presented in exist-

*Deok-Joo Lee is the corresponding author.

ing two-dimensional dynamic mechanisms. Firstly, we assume that buyers have a piece-wise linear function instead of the conventional take-it or leave-it form. Secondly, we extend the buyers' type space and the seller's decision set as continuous domains. Lastly, we address an environment where the number of arriving buyers at each time period remains unknown.

- We determine allocation and payment policies contingent on the time period, remaining item quantity, and submitted bids from incoming buyers. Moreover, we devise a penalty scheme to ensure incentive compatibility, which is necessary to prevent buyers from overbidding for the demanded quantity.
- We introduce two algorithms to numerically approximate the proposed mechanism. The first algorithm employs a Monte Carlo (MC) simulation-based regression method, which approximates the state value function with a polynomial function. The second algorithm relies on the deep deterministic policy gradient (DDPG) technique, focusing on learning the allocation policy.

Related Works

Dynamic Stochastic Knapsack Problem

DSKP emerged as a component of the dynamic allocation problem and was initially analyzed in a finite discrete-time framework by Papastavrou, Rajagopalan, and Kleywegt (1996). Subsequently, Kleywegt and Papastavrou (1998) extended this framework to encompass continuous-time scenarios, while Kleywegt and Papastavrou (2001) investigated with random-sized demands. They aimed to find optimal allocation and pricing rules within a non-strategic demand context. Notably, they suggested that the optimal allocation and pricing rules are threshold policies, which offer a uniform price to incoming buyers and the buyers either accept or reject the offer.

Furthermore, various computation algorithms for DSKP have been developed to implement allocation policies in practical scenarios (Zhou, Chakrabarty, and Lukose 2008; Han, Kawase, and Makino 2015; Im et al. 2021; Sun et al. 2022). They focused on online algorithms to allocate resources in response to changing demands and constraints. They offered theoretical bounds to guarantee the efficiency of the proposed algorithms. However, the presence of strategic behaviors among buyers necessitates a mechanism design approach to maximize the seller's expected revenue.

Optimal Mechanism

In the field of mechanism design, the solution to the presented problem—an optimal mechanism in a dynamic stochastic knapsack environment—can be viewed as a two-dimensional type space (value and quantity), multi-unit (perfectly divisible item) dynamic mechanism. Our approach effectively adapts the findings of mechanisms in a static environment to span multiple periods. The cornerstone of the optimal mechanism in a static environment is the work of Myerson (1981), which revolved around a revenue-maximizing auction for a single item using the characterization approach.

Subsequently, much of the literature has expanded to encompass various settings. Che (1993) and Asker and Cantillon (2010) explored the multi-dimensional bidding environment, while Maskin and Riley (1989) extended the model to a multi-unit setting. Furthermore, Iyengar and Kumar (2008) formulated a two-dimensional and multi-unit procurement auction mechanism with perfectly divisible items, laying the groundwork for constructing a periodic allocation model in our study. Additionally, to address the computational challenges of the optimal mechanism, Bhat et al. (2019) suggested an algorithm that transforms integration into a summation operation, and Duetting et al. (2019) used a neural network to approximate the optimal mechanism in various auction environments.

Dynamic Mechanism

In the context of the dynamic mechanism based on the Myersonian approach, Bergemann and Välimäki (2019) noted that the optimal dynamic mechanism has been explored in two main strands. The first strand involves fixed participating agents with evolving values over time (Kakade, Lobel, and Nazerzadeh 2013; Pavan, Segal, and Toikka 2014; Pavan 2017). However, our approach diverges from these works as we consider a scenario where the buyers are impatient, meaning they stay in the system only temporarily.

Our study aligns with the second strand, where the population of bidders changes over time while retaining fixed private values. This line of research primarily focuses on revenue management for sellers dealing with the requests of impatient buyers. Vulcano, Van Ryzin, and Maglaras (2002) tackled the challenge of designing an optimal mechanism for selling indivisible, limited items in a discrete-time space, involving randomly arriving unit-demand buyers. Building upon this, Gershkov and Moldovanu (2009) extended the problem to a scenario with heterogeneous goods and continuous-time arrivals of buyers and Pai and Vohra (2013) considered bids in three dimensions (value, arrival time, deadline) from unit-demand buyers.

The two-dimensional optimal dynamic mechanism where the buyers bid their value and quantity was studied in Dizdar, Gershkov, and Moldovanu (2011) and Wang, Liang, and Li (2013). The former considered an environment where a single buyer arrives every period, while the latter assumed the items are indivisible discrete units. Additionally, both studies simplified the analysis by considering the buyer's utility function as a take-it or leave-it form, where the utility becomes zero if the desired quantity is not obtained. In our model, we aim to improve upon these aspects.

Framework

A monopolistic seller possesses $\bar{Q}$ units of a perfectly divisible item. The seller plans to sell the item during a finite discrete time horizon $\mathcal{T} = \{1, \cdots, T\}$ and aims to maximize the ex-ante expected revenue at the beginning. In every period $t \in \mathcal{T}$, n^t buyers arrive in the market, where n^t follows a probability mass function $g(n)$ with the support $\{1, \cdots, N\}$. Every buyer is assumed to be *impatient*: the buyer arriving at t leaves the market before $t + 1$.

The arriving buyer at time t is denoted as (i,t), where $1 \leq i \leq n^t$. The buyer (i,t) has private information $\theta_i^t = (v_i^t, q_i^t)$, with v_i^t representing the marginal value of the item and q_i^t representing the desired quantity. θ_i^t is independently distributed among the n^t bidders and follows a joint probability density function $f(v,q)$ and cumulative distribution function $F(v,q)$ with the support (or type space) $\Theta = [\underline{v}, \bar{v}] \times [\underline{q}, \bar{q}]$. The lowest values are normalized to 0, i.e., $\underline{v} = 0$ and $\underline{q} = 0$.

We define $\theta^t = (\theta_1^t, \cdots, \theta_{n^t}^t, \varnothing, \cdots, \varnothing) \in \Theta^N$ as the type profile of bidders at period t. The former part describes the type vector of arriving buyers, and the latter describes the dummy types of non-arriving buyers. Here, we let $\varnothing = (0,0)$ to ensure that the dummies do not influence the allocation results. By defining θ^t with dummies, we can express the allocation and payment rules as N-dimensional functions. Additionally, we define θ_{-i}^t as the type profile excluding buyer (i,t).

Then, by the revelation principle, we can focus on the incentive-compatible direct mechanism where buyers truthfully report their types.

Definition 1 (Direct Mechanism) *$\Gamma = (a^t, p^t)_{t \in \mathcal{T}}$ is a direct mechanism where*

$$a^t : \Theta^N \times [0, \bar{Q}] \to \mathbb{R}_+^N \text{ and } p^t : \Theta^N \times [0, \bar{Q}] \to \mathbb{R}_+^N$$

are the allocation rule and payment rule at period t.

The direct mechanism comprises a sequence of time-dependent allocation and payment rules. They are determined based on the reported type profile and the remaining units of the item. Let q^t be the remaining units at the beginning of period t, which is public information. Then, the allocation and payment for buyer (i,t), who reports their type as $\hat{\theta}_i^t = (\hat{v}_i^t, \hat{q}_i^t)$ when others bid truthfully, are denoted as $a_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t)$ and $p_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t)$, respectively. Based on this, we define the utility function for the buyers as follows.

Definition 2 (Ex-Post Utility Function) *Given q^t units remaining in the market at time t, the ex-post utility of buyer (i,t), who reports $\hat{\theta}_i^t$ with the true type θ_i^t, is defined as:*

$$u_i^t(\hat{\theta}_i^t, \theta_{-i}^t | \theta_i^t, q^t) = v_i^t \min\{q_i^t, a_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t)\} - p_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t). \quad (1)$$

The utility function for the buyers is piece-wise linear. The marginal utility remains constant at v_i^t until the allocation reaches the demanded quantity q_i^t, and once it exceeds, the utility no longer increases. This utility function is a natural extension of the linear utility introduced by Myerson (1981). Subsequently, we define the following expectations for the Bayesian equilibrium.

Definition 3 (Expected Allocation and Payment) *Given q^t units remaining in the market at time t, the expected allocation and payment of bidder (i,t) who reports $\hat{\theta}_i^t$ with the true type θ_i^t is defined by*

$$A_i^t(\hat{\theta}_i^t | \theta_i^t, q^t) = \mathbb{E}_{\theta_{-i}^t}[a_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t)] \text{ and}$$

$$P_i^t(\hat{\theta}_i^t | \theta_i^t, q^t) = \mathbb{E}_{\theta_{-i}^t}[p_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t)],$$

provided that the other buyers report their types truthfully.

Definition 4 (Interim Utility Function) *Given q^t units remaining in the market at time t, the expected (interim) utility of bidder (i,t) who reports $\hat{\theta}_i^t$ with the true type θ_i^t is defined by*

$$U_i^t(\hat{\theta}_i^t | \theta_i^t, q^t) = \mathbb{E}_{\theta_{-i}^t}[u_i^t(\hat{\theta}_i^t, \theta_{-i}^t | \theta_i^t, q^t)],$$

provided that the other buyers report their types truthfully.

For the sake of simplicity, denote the expected allocation, payment, and utility function when the buyer truthfully reports its type as $A_i^t(\theta_i^t | q^t) = A_i^t(\theta_i^t | \theta_i^t, q^t)$, $P_i^t(\theta_i^t | q^t) = P_i^t(\theta_i^t | \theta_i^t, q^t)$ and $U_i^t(\theta_i^t | q^t) = U_i^t(\theta_i^t | \theta_i^t, q^t)$.

Then, the objective of the seller is to maximize the ex-ante expected revenue which can be written as

$$\max_{a,p} \quad \sum_{t=1}^{T} \delta^{t-1} \mathbb{E}_{n^t, \theta^t} \left[\sum_{i=1}^{n^t} p_i^t(\theta^t, q^t) \right], \quad (2)$$

with the following constraints:

- **Bayesian incentive compatibility**: For any $t \in \mathcal{T}$ and $0 \leq q^t \leq \bar{Q}$, every arriving buyer (i,t) has no incentive to misreport its true type:

$$U_i^t(\theta_i^t | q^t) \geq U_i^t(\hat{\theta}_i^t | \theta_i^t, q^t), \quad \forall \theta_i^t, \hat{\theta}_i^t \in \Theta. \quad (3)$$

- **Individual rationality**: For any $t \in \mathcal{T}$ and $0 \leq q^t \leq \bar{Q}$, every arriving buyer (i,t) should not be worse off by participating in the mechanism:

$$U_i^t(\theta_i^t | q^t) \geq 0 \quad \forall \theta_i^t \in \Theta. \quad (4)$$

- **Feasibility**: For any $t \in \mathcal{T}$, if the buyers report $\hat{\theta}^t$ and the seller has q^t remaining units,

$$\sum_{i=1}^{N} a_i^t(\hat{\theta}^t, q^t) \leq q^t, \; q^{t+1} = q^t - \sum_{i=1}^{N} a_i^t(\hat{\theta}^t, q^t) \quad (5)$$

$$0 \leq a_i^t(\hat{\theta}^t, q^t) \leq \hat{q}_i^t. \quad (6)$$

The feasibility condition (5) implies that the periodic allocation cannot exceed the current remaining units and there are no newly added items or returns. Also, $q^1 = \bar{Q}$. Meanwhile, the condition (6) implies the individual allocation cannot exceed the reported quantity.

Optimal Mechanism

In this section, we derive the optimal solution to the proposed problem. To achieve this, we initially conducted the characterization of the incentive-compatible mechanism. The proofs of the propositions marked with ♣ are omitted here and can be found in the full version of the paper.

Lemma 1 (♣) *Suppose $\Gamma = (a^t, p^t)_{t \in \mathcal{T}}$ is incentive compatible and feasible. Then at any period $t \in \mathcal{T}$,*

(a) *$U_i^t(v_i^t, q_i^t | q^t)$ is convex with respect to v_i^t.*
(b) *$\forall \varepsilon > 0, U_i^t(v_i^t, q_i^t | q^t) - U_i^t(v_i^t - \varepsilon, q_i^t | q^t) \leq \varepsilon A_i^t(v_i^t, q_i^t | q^t) \leq U_i^t(v_i^t + \varepsilon, q_i^t | q^t) - U_i^t(v_i^t, q_i^t | q^t)$.*

Given (a) of Lemma 1, $U_i^t(v_i^t, q_i^t | q^t)$ is absolutely continuous, implying it is differentiable almost everywhere with respect to v_i^t. Moreover, using (b) of Lemma 1, we can establish the following theorem.

Theorem 1 *Suppose $\Gamma = (a^t, p^t)_{t\in\mathcal{T}}$ is incentive compatible and feasible. Then at any period $t \in \mathcal{T}$,*

(a) $\forall(i,t)$, $A_i^t(v_i^t, q_i^t|q^t)$ *is non-decreasing in v_i^t for fixed q_i^t.*

(b) $U_i^t(v_i^t, q_i^t|q^t) = U_i^t(\underline{v}, q_i^t|q^t) + \int_{\underline{v}}^{v_i^t} A_i^t(\tau, q_i^t|q^t)d\tau.$

Proof.
(a) From (b) of Lemma 1, for any $\varepsilon > 0$, we have

$$\frac{U_i^t(v_i^t, q_i^t|q^t) - U_i^t(v_i^t - \varepsilon, q_i^t|q^t)}{\varepsilon} \le A_i^t(v_i^t, q_i^t|q^t) \le \frac{U_i^t(v_i^t + \varepsilon, q_i^t|q^t) - U_i^t(v_i^t, q_i^t|q^t)}{\varepsilon}$$

If $\varepsilon \to 0$, we get $\partial U_i^t(v_i^t, q_i^t|q^t)/\partial v_i^t = A_i^t(v_i^t, q_i^t|q^t)$ almost every everywhere and since $U_i^t(v_i^t, q_i^t|q^t)$ is convex in v_i^t, $A_i^t(v_i^t, q_i^t|q^t)$ is non-decreasing in v_i^t.
(b) Since $\partial U_i^t(v_i^t, q_i^t|q^t)/\partial v_i^t = A_i^t(v_i^t, q_i^t|q^t)$ almost everywhere, by the fundamental theorem of calculus,

$$\int_{\underline{v}}^{v_i^t} A_i^t(\tau, q_i^t|q^t)d\tau = \int_{\underline{v}}^{v_i^t} \frac{\partial U_i^t(\tau, q_i^t|q^t)}{\partial v_i^t}d\tau = U_i^t(v_i^t, q_i^t|q^t) - U_i^t(\underline{v}, q_i^t|q^t),$$

which completes the proof. ■

Theorem 1 outlines the essential properties that any incentive compatible mechanism should satisfy. Particularly, using (b) of Theorem 1, the original revenue maximization problem is transformed into a problem analogous to Myerson's virtual value maximization problem. The virtual valuation of the two-dimensional bids is defined as follows

Definition 5 (Virtual Valuation) *For any realized buyer (i,t) with $\theta_i^t = (v_i^t, q_i^t)$, the virtual valuation of bidder (i,t) is defined as $\phi_i^t(\theta_i^t) := v_i^t - \frac{1-F(v_i^t|q_i^t)}{f(v_i^t|q_i^t)}$.*

Theorem 2 (♣) *For an incentive compatible, individually rational, and feasible mechanism $\Gamma = (a, p)$, the seller's problem is reduced to the following dynamic stochastic knapsack problem:*

$$\max_{a} \quad \sum_{t=1}^{T} \delta^{t-1}\mathbb{E}_{n^t,\theta^t}\left[\sum_{i=1}^{n^t} \phi_i^t(\theta_i^t)a_i^t(\theta^t, q^t)\right] \tag{7}$$
$$\text{s.t.} \quad (5),\ (6)$$

Subsequently, the dynamic programming approach can be employed to solve the proposed stochastic program. For every possible state (θ^t, q^t), define $V^t(\theta^t, q^t)$ as the state value function, which represents the discounted sum of values that can be obtained from state (θ^t, q^t) until the end of the study period by following the optimal policy. Thus, the optimal solution must satisfy the following Bellman equation for every $t \in \mathcal{T}$ almost everywhere with respect to the underlying probability space.

$$V^t(\theta^t, q^t) = \sup_{0\le x\le q^t} \{R^t(\theta^t, q^t, x) + \delta\mathbb{E}_{n^{t+1},\theta^{t+1}}\left[V^{t+1}(\theta^{t+1}, q^t - x)\right]\}, \tag{8}$$

with the boundary condition $V^{T+1}(\theta^{T+1}, q^{T+1}) = 0$ for every $\theta^{T+1} \in \Theta^N$ and $q^{T+1} \in [0, \bar{Q}]$. Also, $R^t(\theta^t, q^t, x)$ represents the maximum periodic revenue attainable at state (θ^t, q^t) when the seller opts to sell x units, which has a form of

$$R^t(\theta^t, q^t, x) = \max_{a^t} \quad \sum_{i=1}^{n^t} \phi_i^t a_i^t(\theta^t, q^t)$$
$$\text{s.t.} \quad 0 \le a_i^t(\theta^t, q^t) \le q_i^t, \quad \forall(i,t)$$
$$\sum_{i=1}^{n^t} a_i^t(\theta^t, q^t) = x.$$

Since the suggested problem is a linear knapsack problem, the optimal solution is to allocate in descending order of virtual valuation. Consequently, for a given state (θ^t, q^t), the buyers can be reordered based on their virtual valuations, denoted as $\phi_{[1]}^t \ge \cdots \ge \phi_{[n^t]}^t$. Let the order of buyer (i,t) be represented as $[i]$. Then, we have the following theorem.

Theorem 3 *If the seller decides to sell x units, the optimal allocation rule is*

$$(a_i^t)^*(\theta^t, q^t) = \begin{cases} q_i^t & \text{if } [i] \le i^*(\theta^t, x) \\ x - \sum_{j=1}^{i^*(\theta^t,x)} q_{[j]}^t & \text{if } [i] = i^*(\theta^t, x) + 1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

where $i^*(\theta^t, x)$ is the integer that satisfies $\sum_{j=1}^{i^*(\theta^t,x)} q_{[j]}^t \le x < \sum_{j=1}^{i^*(\theta^t,x)+1} q_{[j]}^t$.

Proof. Without loss of generality, assume $\phi_i^t \ge 0$ for every bidder (i,t). The objective function value of the suggested problem is $\sum_{j=1}^{i^*(\theta^t,x)} \phi_{[j]}^t q_{[j]}^t + \phi_{[i^*(\theta^t,x)+1]}(x - \sum_{j=1}^{i^*(\theta^t,x)} q_{[j]}^t)$. Then, consider the dual problem:

$$\min_{\lambda\ge 0,\ \mu} \quad \sum_{i=1}^{n^t} q_i^t\lambda_i + x\mu$$
$$\text{s.t.} \quad \lambda_i^t + \mu \ge \phi_i^t, \quad \forall(i,t).$$

Since $\lambda_i^t = \max\{0, \phi_i^t - \mu\}$, the dual problem is reduced to

$$\min_{\mu} \quad \sum_{i=1}^{n^t} q_i^t \max\{0, \phi_i^t - \mu\} + x\mu$$

If we set $\mu^* = \phi_{[i^*(\theta^t,x)]}^t$, then the objective value of the dual problem is the same as that of the primal problem, which satisfies the strong duality. ■

As depicted in (b) of Theorem 1 and Theorem 3, the optimal allocation rule and payment rule are determined for a given x. Therefore, we are left with the problem of determining the appropriate amount to sell in each period, denoted as $x^*(\theta^t, q^t)$. In order to achieve this, the following is essential.

Lemma 2 (♣) *For a given state (θ^t, q^t),*

(a) $R^t(\theta^t, q^t, x)$ *is non-decreasing with respect to x.*
(b) $R^t(\theta^t, q^t, x)$ *is concave with respect to x.*

(c) $R^t(\theta^t, q^t, x)$ *is concave with respect to* (q^t, x).

Lemma 3 (♣) *For a given state* (θ^t, q^t),

(a) $V^t(\theta^t, q^t)$ *is non-decreasing in* q^t.
(b) $V^t(\theta^t, q^t)$ *is concave with respect to* q^t
(c) $\mathbb{E}_{n^t,\theta^t}[V^t(\theta^t, q^t)]$ *is non-decreasing and concave with respect to* q^t

From Lemma 2 and Lemma 3, we can notice that the functions $R^t(\theta^t, q^t, x)$ and $\mathbb{E}_{n^{t+1},\theta^{t+1}}[V^{t+1}(\theta^{t+1}, q^t - x)]$ are monotone with respect to x. As a result, they are differentiable almost everywhere. Therefore, the differentiation of the state value function is well-defined.

Definition 6 (Marginal Value) *For any state* (θ^t, q^t)*, the marginal value, denoted as* $MV^t(\theta^t, q^t, x)$*, is defined as*

$$MV^t(\theta^t, q^t, x) = \frac{\partial}{\partial x}\{R^t(\theta^t, q^t, x) + \delta\mathbb{E}_{n^{t+1},\theta^{t+1}}[V^{t+1}(\theta^{t+1}, q^t - x)]\}. \quad (10)$$

Note that $MV^t(\theta^t, q^t, x)$ is non-increasing with respect to x since both $R^t(\theta^t, q^t, x)$ and $\mathbb{E}_{n^{t+1},\theta^{t+1}}[V^{t+1}(\theta^{t+1}, q^t - x)]$ are concave. Also, they are continuous almost everywhere, which makes them Riemann integrable. Given that $R^t(\theta^t, q^t, 0) = 0$, the Bellman equation can be rewritten as

$$V^t(\theta^t, q^t) = \sup_{0\le x\le q}\left\{\int_0^x MV^t(\theta^t, q^t, \tau)d\tau\right\} + \delta\mathbb{E}_{n^{t+1},\theta^{t+1}}[V^{t+1}(\theta^{t+1}, q^t)], \quad (11)$$

which implies $x^*(\theta^t, q^t) \in \operatorname{argsup}_{0\le x\le q^t} \int_0^x MV^t(\theta^t, q^t, \tau)d\tau$.

Then, define the set $X^-(\theta^t, q^t)$ that includes x with negative marginal values in a given state (θ^t, q^t), i.e.,

$$X^-(\theta^t, q^t) = \{x \in [0, q^t] | MV^t(\theta^t, q^t, x) < 0\}. \quad (12)$$

If the set $X^-(\theta^t, q^t)$ is non-empty, the infimum is well-defined because 0 is a lower bound. The following theorem proposes the optimal $x^*(\theta^t, q^t)$.

Theorem 4 *For a given state* (θ^t, q^t)*,* $x^*(\theta^t, q^t)$ *is*

$$x^*(\theta^t, q^t) = \begin{cases} q^t & \text{if } X^-(\theta^t, q^t) = \varnothing \\ \inf X^-(\theta^t, q^t) & \text{otherwise.} \end{cases} \quad (13)$$

Proof. Let a state (θ^t, q^t) be given. First, consider the case where $X^-(\theta^t, q^t) = \varnothing$. Then, for all $x \in [0, q^t]$, $MV^t(\theta^t, q^t, x) \ge 0$. So, for all $x \in [0, q^t]$,

$$\int_0^x MV^t(\theta^t, q^t, \tau)d\tau \le \int_0^{q^t} MV^t(\theta^t, q^t, \tau)d\tau.$$

Therefore, $x^*(\theta^t, q^t) = q^t$ in this case. Then, consider the case where $X^-(\theta^t, q^t) \neq \varnothing$. Denote $x^\circ = \inf X^-(\theta^t, q^t)$. If $x^\circ = 0$, it is obvious that x° is optimal because $MV^t(\theta^t, q^t, x) < 0$ for every $x \in (0, q^t]$. If $x^\circ > 0$, we have $MV^t(\theta^t, q^t, x) \ge 0$ for $0 \le x < x^\circ$ and $MV^t(\theta^t, q^t, x) < 0$ for $x^\circ < x \le q^t$ since $MV^t(\theta^t, q^t, x)$ is non-increasing with respect to x. Assume that $x^\circ \notin \operatorname{argsup} \int_0^x MV^t(\theta^t, q^t, \tau)d\tau$. Then, there exists x' that attains the supremum and $x' \neq x^\circ$. If $0 \le x' < x^\circ$,

$$\int_0^{x'} MV^t(\theta^t, q^t, \tau)d\tau \le \int_0^{x^\circ} MV^t(\theta^t, q^t, \tau)d\tau,$$

which is a contradiction that x° does not attain the supremum. Similarly, if $x^\circ < x' \le q^t$,

$$\int_0^{x'} MV^t(\theta^t, q^t, \tau)d\tau < \int_0^{x^\circ} MV^t(\theta^t, q^t, \tau)d\tau,$$

which is a contradiction that x' attains the supremum. ■

Theorem 4 provides us with the optimal solution to the Bellman equation outlined in (8). However, it's crucial to remember that we have not taken into account the non-decreasing property of an incentive compatible mechanism, as presented in (a) of Lemma 1. To ensure the property of incentive compatibility, we define the following assumption.

Assumption 1 (Regularity condition) *The virtual valuation* $\phi_i^t(\theta_i^t) = v_i^t - \frac{1-F(v_i^t|q_i^t)}{f(v_i^t|q_i^t)}$ *is non-decreasing with respect to* v_i^t *and* q_i^t.

Under the regularity condition, buyers with a high value and a greater demanded quantity will have a higher priority if the seller determines to allocate the item as suggested in (9). Therefore, we can obtain the following theorem.

Theorem 5 (♣) *Under the regularity condition,*

(a) $(a_i^t)^*(\theta_i^t, \theta_{-i}^t, q^t)$ *and* $(A_i^t)^*(\theta_i^t|q^t)$ *are non-decreasing with respect to* v_i^t.
(b) $(a_i^t)^*(\theta_i^t, \theta_{-i}^t, q^t)$ *and* $(A_i^t)^*(\theta_i^t|q^t)$ *are non-decreasing with respect to* q_i^t.

Nevertheless, despite deriving the optimal allocation and payment rules based on the necessity condition of incentive compatibility and the regularity condition, it remains to be proven that the proposed mechanism is genuinely incentive compatible. Specifically, it is found that if the mechanism solely consists of the derived allocation and payment rules, it is found that buyers can overbid on their demanded quantities to increase their purchase probability. To prevent this, we develop a penalty scheme that can punish buyers when such quantity overbidding occurs.

Assumption 2 *After the allocation, the seller can observe whether the allocated quantity* $(a_i^t)^*(\hat{\theta}_i^t, \theta_{-i}^t, q^t)$ *is greater than the true demanded quantity* q_i^t*, for every buyer at no cost.*

This assumption indicates that although the seller might not know the buyers' desired quantities initially, they can ascertain them after the bidding and allocation are finished. This assumption is reasonable in the field of rental or service businesses, which is the main motivation of this study, where the system can observe whether allocated resources are actually being utilized after the allocation has taken place. We define the penalty scheme as follows.

Definition 7 (Penalty Scheme) *If a bidder* (i, t) *who reports* $\hat{\theta}_i^t = (v_i^t, \hat{q}_i^t)$ *overbids the quantity, i.e.,* $\hat{q}_i^t > q_i^t$*, and*

subsequently receives an allocation greater than their true quantity, i.e., $a_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t) > q_i^t$*, the bidder must pay a certain amount of penalty defined as*

$$\rho_i^t(\hat{\theta}_i^t, \theta_{-i}^t, q^t) = \frac{\bar{v}\hat{q}_i^t}{\mathcal{P}_{\theta_{-i}^t}((a_i^t)^{*,-1}(\{\hat{q}_i^t\}))}, \tag{14}$$

where $\mathcal{P}_{\theta_{-i}^t}((a_i^t)^{*,-1}(\{\hat{q}_i^t\}))$ *is the probability that the allocated quantity is equal to* $\hat{q}_i^t$*, i.e.,* $(a_i^t)^*(\hat{\theta}_i^t, \theta_{-i}^t, q^t) = \hat{q}_i^t$*.*

With the defined penalty scheme, we prove that the proposed mechanism is incentive compatible, individually rational, and achieves ex-ante revenue maximization.

Theorem 6 (♣) *Under the regularity condition, the following mechanism* $\Gamma^* = (a^*, p^*)$ *with the penalty scheme* ρ *satisfies the incentive compatibility, individual rationality, and the optimality:*

$$(a_i^t)^*(\theta^t, q^t) = \begin{cases} q_i^t & \text{if } [i] \le i^* \\ x - \sum_{j=1}^{i^*} q_{[j]}^t & \text{if } [i] = i^* + 1 \\ 0 & \text{otherwise} \end{cases},$$

$$(p_i^t)^*(\theta^t, q^t) = v_i^t(a_i^t)^*(\theta^t, q^t) - \int_{\underline{v}}^{v_i^t} (a_i^t)^*((\tau, q_i^t), \theta_{-i}^t, q^t)d\tau$$

where $i^* = i^*(\theta^t, x^*(\theta^t, q^t))$ *is the integer that satisfies* $\sum_{j=1}^{i^*(\theta^t, x^*(\theta^t, q^t))} q_{[j]}^t \le x^*(\theta^t, q^t) < \sum_{j=1}^{i^*(\theta^t, x^*(\theta^t, q^t))+1} q_{[j]}^t$ *and*

$$x^*(\theta^t, q^t) = \begin{cases} q^t & \text{if } X^-(\theta^t, q^t) = \varnothing \\ \inf X^-(\theta^t, q^t) & \text{otherwise.} \end{cases}$$

Approximation Algorithms

The optimal mechanism that we have derived comprises the allocation rule, payment rule, and $x^*(\theta^t, q^t)$, which serves to determine the allocation amount for each period. However, the practical implementation of this mechanism necessitates a computational approach. The allocation rule and payment rule entail a computational complexity of $O(N \log N)$ due to the need to sort bidders by their virtual valuations at each time period. Further details of these algorithms are provided in the supplementary material. The complexity arises in computing $x^*(\theta^t, q^t)$ due to the requirement to calculate the expected value of the state value function in a continuous space. To tackle this challenge, we propose two algorithms designed to approximate $x^*(\theta^t, q^t)$.

Approximation Methods

We then present two algorithms to approximate $x^*(\theta^t, q^t)$. To do this, we first consider approximating the expectations of the state value function using the MC simulation-based regression. Assume that there are some basis functions $f_1, \cdots, f_n$. Then, the state value function is approximated as a linear combination of the basis functions: $V^t(\theta^t, q^t) = V^t(q^t) \approx \sum_{j=0}^n c_j^t f_j(q^t)$. The coefficients c_j^t are determined through regression at some fixed states $s_1, \cdots, s_m$. In this study, we use the Chebyshev polynomials and nodes as basis functions and states:

$$f_0(s) = 1,\ f_1(s) = s,\ f_{n+1}(s) = 2sf_n(s) - f_{n-1}(s), \tag{15}$$

Algorithm 1: Monte Carlo Simulation

Input: n (# of basis); m (# of nodes); N^e (# of episodes)

Initialize $V^t(s_k) = 0, t = 1, \cdots, T, k = 1, \cdots, m$
Initialize coefficient matrix $C \in \mathbb{R}^{T \times (n+1)} = \mathbf{0}$
$i = 1$
while $i \le N^e$ **do**
  for $t = T, \cdots, 1$ **do**
    Generate buyers' profiles θ^t
    for $s = s_1, \cdots, s_m$ **do**
      $x^* = \text{argmax}_{x \le s}\{R^t(\theta^t, s, x) + \delta V^{t+1}(s - x)\}$
      $V^{new} = R^t(\theta^t, s, x^*) + \delta V^{t+1}(s - x^*)$
      $V^t(s) \leftarrow \frac{i-1}{i} V^t(s) + \frac{1}{i} V^{new}$
    end for
    $i \leftarrow i + 1$
  end for
end while
for $t = 1, \cdots, T$ **do**
  Fit $\begin{bmatrix} f_0(s_1) & \cdots & f_n(s_1) \\ \vdots & \ddots & \vdots \\ f_0(s_m) & \cdots & f_n(s_m) \end{bmatrix} \begin{bmatrix} c_1^t \\ \vdots \\ c_n^t \end{bmatrix} \approx \begin{bmatrix} V^t(s_1) \\ \vdots \\ V^t(s_m) \end{bmatrix}$
end for
return C

$$s_k = \frac{\bar{Q}}{2}\left\{1 + \cos\left(\frac{2k-1}{m}\pi\right)\right\},\ \text{for } k = 1, \cdots, m. \tag{16}$$

Note that the Chebyshev polynomials have the characteristics that they are orthogonal functions, i.e., $\int_{-1}^{1} f_i(s)f_j(s) = 0$ for any $i \ne j$. Also, the original Chebyshev nodes are $\cos\left(\frac{2k-1}{m}\pi\right)$ which ranges from -1 to 1, so they are resized from 0 to $\bar{Q}$. Then, the original Bellman equation is reduced to the following equation:

$$\sum_{j=0}^{n} c_j^t f_j(s_k) \approx \max_{x \le s_k} R^t(\theta^t, s_k, x) + \delta \sum_{j=1}^{n} c_j^{t+1} f_j(s_k - x). \tag{17}$$

After obtaining the coefficients, the contingent decisions can be made by the approximated state value functions. The simulation algorithm is presented in Algorithm 1.

Meanwhile, an alternative approach is to approximate the optimal policy through policy learning. Specifically, reinforcement learning serves as a technique to learn the optimal policy within a dynamic environment. In our case, as the model encompasses a continuous space, we propose the DDPG method. This actor-critic method is well-suited for handling continuous action spaces. A concise overview of the DDPG algorithm pertaining to dynamic allocation is provided in Algorithm 2. For an in-depth exploration of this algorithm, please refer to Lillicrap et al. (2015).

Numerical Experiment

To compare the effectiveness of the two proposed approximation algorithms, we conduct numerical experiments under various scenarios, especially the length of the study period. We consider three different market environment where

Algorithm 2: Deep Deterministic Policy Gradient method

Input: τ (learning rate), N^e, r (# of random episodes)

Initialize critic $Q(s, a|\theta^Q)$ and actor $\mu(s|\theta^\mu)$
Initialize target critic (Q') and actor (μ') with θ^Q and θ^μ
Initialize Replay buffer R
for $i = 1, \cdots, N^e$ **do**
  if $i \leq r$ **then**
    for $t = 1, \cdots, T$ **do**
      At state $s_t = q^t$, generate buyers' profile θ^t
      Agent chooses random action $a_t = x$
      Compute reward $r_t = R^t(\theta^t, q^t, x)$
      Compute the next state $s_{t+1} = s_t - x$
      Store (s_t, a_t, r_t, s_{t+1}) in replay buffer R
    end for
  else
    Minibatch (s_t, a_t, r_t, s_{t+1}) from replay buffer
    Train critic and actor network, θ^Q and θ^μ
    Update target networks:
    $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$, $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
    for $t = 1, \cdots, T$ **do**
      Choose action $a_t = \mu(s_t|\theta^\mu) + \epsilon_t$
      Store (s_t, a_t, r_t, s_{t+1}) in replay buffer R
    end for
  end if
end for
return μ, μ', Q, Q'

$(T.\bar{Q}) = (10, 10)$, $(30, 30)$, and $(100, 100)$. The distribution of the buyers and the seller's plans are determined as shown in Table 1. For the simulation, we employ 5 basis functions and vary the number of nodes (m) within the range of 5 to 50. In the context of DDPG, we set the learning rates for the actor, critic, and soft update to 0.0001, 0.001, and 0.0001 respectively. A minibatch size of 64 is utilized, and the neural network structure encompasses 3 layers, each with 64 nodes. Also, the initial random choices are set to be 10% of the total training episodes. All the methods are trained by 10,000 episodes and the performances are compared by averaging 20 test episodes. The simulations were performed on a computer with an Intel Core i7-6700 CPU, 16GB RAM, and an NVIDIA GeForce GTX 1060 GPU, using Python numpy and Pytorch packages, with a fixed seed number of 1.

The results, including the average discounted rewards from test episodes and their corresponding training times, are presented in Table 2. The full information case that maximizes the virtual value is presented together to suggest theoretical upper bounds. The MC method outperforms the DDPG approach in terms of average test rewards, even when the number of nodes in the MC method is kept small. However, While increasing the number of nodes in the MC method enhances performance, it considerably slows down training speed. Conversely, DDPG exhibits notably faster computation speed in comparison to the MC method.

Parameters	Values
$g(n)$	Poisson(10) ($g(n) = \frac{10^n e^{-10}}{n!}$)
$f(q)$	Uniform(0, 2) ($f(q) = \frac{1}{2}$ for $0 \leq q \leq 2$)
$f(v\|q)$	Exponential(q) ($f(v\|q) = qe^{-qv}$)
$(T, \bar{Q})$	(10, 10), (30, 30), (100, 100)
δ	0.99

Table 1: Experimental parameters

Methods	Average test rewards (training time: min)		
	$(10, 10)$	$(30, 30)$	$(100, 100)$
MC ($m = 5$)	21.77 (0.9)	56.71 (2.7)	115.88 (9.1)
MC ($m = 10$)	21.80 (3.0)	62.84 (9.3)	132.62 (34.5)
MC ($m = 20$)	21.88 (11.0)	63.48 (34.5)	146.75 (119.4)
MC ($m = 50$)	21.88 (65.4)	63.74 (208.9)	163.40 (714.4)
DDPG	13.40 (1.0)	32.16 (1.9)	91.24 (5.1)
Full information	22.78	65.23	170.62

Table 2: Average test rewards and training time of the approximation methods

To investigate the reasons behind DDPG's lower performance compared to the MC method, we analyze the cumulative average allocation depicted in Figure 1. Notably, the MC method effectively adjusts the distribution of item quantities across different time periods, while DDPG exhibits a tendency to allocate items to buyers arriving early in the study period. This phenomenon can be attributed to two main factors. Firstly, due to the inherent characteristics of DDPG's model, there might be an overestimation of the value function, resulting in an early allocation of resources. Secondly, this premature allocation results from a lack of sufficient state generation for learning during the later stages.

Conclusion

We designed the optimal dynamic mechanism and suggested corresponding approximation algorithms under a more generalized dynamic stochastic knapsack environment. We figured out that the penalty scheme is necessary to preserve incentive compatibility under the regularity condition on a two-dimensional type. Meanwhile, the presented approximation algorithms showed that the DDPG-based method has a lower performance than the regression method. To address these issues, it is evident that improvements in exploration during the policy learning process of DDPG are necessary to train the model using allocation data from the later stages.

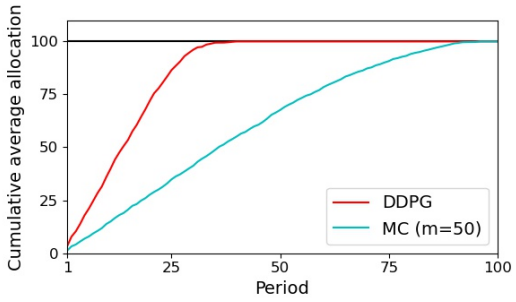

Figure 1: Cumulative allocation of $(T, \bar{Q}) = (100, 100)$

Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1|1A4A01059254).

References

Asker, J.; and Cantillon, E. 2010. Procurement when price and quality matter. *The RAND Journal of Economics*, 41(1): 1–34.

Bergemann, D.; and Välimäki, J. 2019. Dynamic Mechanism Design: An Introduction. *Journal of Economic Literature*, 57(2): 235–74.

Bhat, S.; Jain, S.; Gujar, S.; and Narahari, Y. 2019. An optimal bidimensional multi-armed bandit auction for multi-unit procurement. *Annals of Mathematics and Artificial Intelligence*, 85(1): 1–19.

Che, Y.-K. 1993. Design Competition Through Multidimensional Auctions. *The RAND Journal of Economics*, 24(4): 680.

Dizdar, D.; Gershkov, A.; and Moldovanu, B. 2011. Revenue maximization in the dynamic knapsack problem. *Theoretical Economics*, 6(2): 157–184.

Duetting, P.; Feng, Z.; Narasimhan, H.; Parkes, D.; and Ravindranath, S. S. 2019. Optimal Auctions through Deep Learning. In Chaudhuri, K.; and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 1706–1715. PMLR.

Gershkov, A.; and Moldovanu, B. 2009. Dynamic revenue maximization with heterogeneous objects: A mechanism design approach. *American Economic Journal: Microeconomics*, 1(2): 168–198.

Han, X.; Kawase, Y.; and Makino, K. 2015. Randomized algorithms for online knapsack problems. *Theoretical Computer Science*, 562: 395–405.

Im, S.; Kumar, R.; Montazer Qaem, M.; and Purohit, M. 2021. Online Knapsack with Frequency Predictions. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 2733–2743. Curran Associates, Inc.

Iyengar, G.; and Kumar, A. 2008. Optimal procurement mechanisms for divisible goods with capacitated suppliers. *Review of Economic Design 2008 12:2*, 12(2): 129–154.

Kakade, S. M.; Lobel, I.; and Nazerzadeh, H. 2013. Optimal dynamic mechanism design and the virtual-pivot mechanism. *Operations Research*, 61(4): 837–854.

Kleywegt, A. J.; and Papastavrou, J. D. 1998. The Dynamic and Stochastic Knapsack Problem. *Operations Research*, 46(1): 17–35.

Kleywegt, A. J.; and Papastavrou, J. D. 2001. The dynamic and stochastic knapsack problem with random sized items. *Operations Research*, 49(1): 26–41.

Lillicrap, T. P.; Hunt, J. J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; and Wierstra, D. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Maskin, E.; and Riley, J. 1989. Optimal Multi-unit Auctions. In *The Economics of Missing Markets, Information, and Games*, 312–335. Oxford University Press.

Myerson, R. B. 1981. Optimal Auction Design. *Mathematics of operations research*, 6(1): 58–73.

Pai, M. M.; and Vohra, R. 2013. Optimal dynamic auctions and simple index rules. *Mathematics of Operations Research*, 38(4): 682–697.

Papastavrou, J. D.; Rajagopalan, S.; and Kleywegt, A. J. 1996. The Dynamic and Stochastic Knapsack Problem with Deadlines. *Management Science*, 42(12): 1706–1718.

Pavan, A. 2017. *Dynamic Mechanism Design: Robustness and Endogenous Types*, volume 1 of *Econometric Society Monographs*, 1–62. Cambridge University Press.

Pavan, A.; Segal, I.; and Toikka, J. 2014. Dynamic mechanism design: A myersonian approach. *Econometrica*, 82(2): 601–653.

Sun, B.; Yang, L.; Hajiesmaili, M.; Wierman, A.; Lui, J. C. S.; Towsley, D.; and Tsang, D. H. 2022. The Online Knapsack Problem with Departures. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 6(3).

Vulcano, G.; Van Ryzin, G.; and Maglaras, C. 2002. Optimal dynamic auctions for revenue management. *Management Science*, 48(11): 1388–1407.

Wang, W.; Liang, B.; and Li, B. 2013. Revenue maximization with dynamic auctions in IaaS cloud markets. *IEEE International Workshop on Quality of Service, IWQoS*, 57–62.

Wu, L.; Garg, S. K.; and Buyya, R. 2011. SLA-Based Resource Allocation for Software as a Service Provider (SaaS) in Cloud Computing Environments. In *2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, 195–204.

Zhang, H.; Jiang, H.; Li, B.; Liu, F.; Vasilakos, A. V.; and Liu, J. 2015. A framework for truthful online auctions in cloud computing with heterogeneous user demands. *IEEE Transactions on Computers*, 65(3): 805–818.

Zhou, Y.; Chakrabarty, D.; and Lukose, R. 2008. Budget Constrained Bidding in Keyword Auctions and Online Knapsack Problems. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, 1243–1244. New York, NY, USA: Association for Computing Machinery. ISBN 9781605580852.