Online Resource Allocation with Stochastic Resource Consumption

Jiashuo Jiang   Jiawei Zhang

Department of Technology, Operations & Statistics, Stern School of Business, New York University

Abstract: we consider an online resource allocation problem where multiple resources, each with an individual initial capacity, are available to serve random requests arriving sequentially over multiple discrete time periods. At each time period, one request arrives and its associated reward and size are drawn independently from a known distribution that could be resource-dependent. Upon its arrival and revealing itself, an online decision has to be made on whether or not to accept the request. If accepted, another online decision should also be made to specify on assigning which resource to serve the request, then a certain amount of the resource equal to the size of the request will be consumed and a reward will be collected. The objective of the decision maker is to maximize the total collected reward subject to the capacity constraints. We develop near-optimal policies for our problem under two settings separately. When the reward distribution has a finite support, we assume that given each reward realization, the size distributions for different resources are perfectly positive correlated and we propose an adaptive threshold policy. We show that with certain regularity conditions on the size distribution, our policy enjoys an optimal $O(\log T)$ regret bound, where T denotes the total time periods. When the support of the reward distribution is not necessarily finite, we develop another adaptive threshold policy with a $O(\log T)$ regret bound when both the reward and the size of each request are resource-independent and the size distribution satisfies the same conditions.

1. Introduction

We study an online resource allocation problem where multiple resources, each with an initial individual capacity, are available to serve a stream of requests arriving over a horizon of finite time periods. At each period, one request arrives with an associated reward and size, whose values are assumed to be random, following a given probability distribution that could be resource-dependent. Upon the arrival of a request, its reward and size are revealed. Then a decision-maker has to irrevocably decide to reject or accept it and to assign an available resource to serve it if accepted. If a request is served by a resource, a reward is collected and a certain amount of the resource equal to the size of the request is consumed. The decision maker's objective is to maximize the total collected reward without violating the resource capacity constraints.

The single-resource case of our model reduces to the online knapsack problem, which already enjoys a wide range of applications in logistics, scheduling and pricing, among others, as described in Papastavrou et al. (1996). Recent research on online knapsack problem have also been motivated by applications in online advertising (e.g. Zhou et al. (2008), Balseiro and Gur (2019)). In these models,

the requests could arrive in an adversary order (e.g. Zhou et al. (2008)) or in a stochastic order (e.g. Papastavrou et al. (1996), Balseiro and Gur (2019)). Our model assumes a stationary stochastic arriving order, i.e., the values of the reward and size of every request are drawn independently from a known distribution which is stationary across time.

The objective of this paper is to develop near optimal policies for our model. Throughout the paper, the performance of our policies is compared against that of the prophet, i.e., the optimal offline decision maker. We will use *regret* to measure the performance of our policy, which is concerned about the *additive* expected difference between the prophet and our policy. The formal definition of *regret* are provided in Section 2 after introducing the problem formulation and notations.

1.1. Main Results and Contributions

With additional assumptions, we develop policies with regret bounds of the order $\log T$, where T is the total number of periods. More specifically, we assume that either the reward distribution has a finite support and given each reward realization, the size distributions for different resources are perfectly positive correlated, or both the reward and the size of each request are resource-independent and the size distribution satisfies the same conditions. The same bound was previously known only for the single-resource case with unitary rewards Arlotto and Xie (2020). The authors left it as an open question whether the bound still holds when the reward could take multiple values. We provide an affirmative answer to this question, and even more, we generalize it to the multi-resource case. Notice that the numerical experiments in Arlotto and Xie (2020) indicate that the $O(\log T)$ should be order tight even for the single-resource unitary-reward case, thus it is unlikely to improve our bound.

Our regret bounds are based on deriving a linear programming to serve as the upper bound of the expected reward collected by the prophet. A key step is to establish a closed-form characterization of the optimal LP solution. We observe that the optimal LP solution possesses a threshold structure, which motivates us to consider an adaptive threshold policy. Also, the closed-form characterization allows us to derive certain convexity that is crucial to prove our regret bounds. As discussed in Section 3, our LP-based approach is different from that in Arlotto and Xie (2020). Moreover, our approach also allows us to analyze how the expected reward obtained by the prophet scales over T, which in turn establishes the asymptotic optimality of our policies when $T \to \infty$.

Note that several recent papers assume that both the reward distribution and the size distribution have a finite support and derive uniform regret bounds that are independent of T (e.g. Jasin and Kumar (2012), Bumpensanti and Wang (2020), Arlotto and Gurvich (2019), Vera and Banerjee (2020), Vera et al. (2019), Freund and Banerjee (2019)). However, these bounds depend implicitly

on the cardinality of the support of the reward and size. Moreover, it is shown in Arlotto and Gurvich (2019) that even for the simplest multi-secretary setting where there is only one resource and the size of each request equals 1, the optimal regret will depend linearly on the reciprocal of the minimal mass of the reward and size distribution. Therefore, their methods and results could not be generalized to our setting where the size is assumed to be continuously distributed.

1.2. Other Related Literature

Our model is closely related to the knapsack problem and several streams of literature on online optimization. In addition to the results mentioned in the previous subsection, we briefly review other related results.

The knapsack problem is mainly studied under two different settings, the offline setting where the requests are all available to be served before any decision is made and the online setting where the requests arrive sequentially over time. The early formulation of the offline setting, as described in Dantzig (1957), features deterministic reward and size for each request. Then, an offline and stochastic setting is considered in the literature, where the reward and size of each request are assumed to be random and will be revealed only after being included in the knapsack (e.g. Derman et al. (1975), Bhalgat et al. (2011), Ma (2018)). Notably, Dean et al. (2008) considered the setting where the reward is deterministic and the size is random for each request, and provided constant bound between optimal nonadaptive policy and optimal adaptive policy, which is the first result that addressed the *benefit of adaptivity*. For the online setting, when requests arrive in an adversary order, Marchetti-Spaccamela and Vercellis (1995) show that no online algorithm can achieve a constant competitive ratio. However, Zhou et al. (2008) prove a parametric competitive ratio of $\log(U/L) + 1$ where L and U are the lower and upper bounds, respectively, of the reward/size ratios of the requests. Moreover, Babaioff et al. (2007) obtain constant competitive ratio when the requests arrive in a random order and Dütting et al. (2017) prove a 0.2 competitive ratio when the requests arrive in a non-stationary stochastic order. Besides the constant bounds, Marchetti-Spaccamela and Vercellis (1995) prove a $O(\log^{3/2} T)$ regret bound when both the reward and the size of each requests are independently and uniformly distributed. Lueker (1998) further improve the previous bound to $O(\log T)$ and show that the order of the bound is tight.

Our problem is also related to the Online Linear Programming (OLP) literature. The OLP problem takes a standard linear programming as the underlying form, while the columns of the constraint matrix and the corresponding coefficients of the objective function will arrive sequentially over time. The trade-off between algorithm competitiveness and resource capacity, denoted as c, is studied in the literature (e.g. Buchbinder and Naor (2009), Agrawal et al. (2014), Devanur et al. (2019)) and it is shown that in order to have a $1 - \varepsilon$ competitive ratio, it is essential for c to be

at least $\Omega(1/\varepsilon^2)$, which implies that it is necessary for c to be scaled up simultaneously with T to have any regret bound sublinear in the prophet. Also, a recent work Li and Ye (2019) adopted an asymptotic setting that c scales up linearly in T and proved a $O(\log T \log \log T)$ regret bound. However, a term $\frac{T}{c}$ shows up in their regret bound and is abbreviated in notation $O(\cdot)$ since it is a constant in their framework. As a result, a direct application of their result to our problem would imply a regret at least linear in T since we consider our problem under the setting that c is fixed as $T \to \infty$. The above discussions imply that, although the formulation of OLP seems like a non-parametric generalization of our model, the fixed resource capacity setting considered in this paper essentially differentiate our problem from the OLP problem and would require analysis other than those in the OLP literature. Thus, though closely related, our work is not covered by the OLP literature and serves as a supplement to online decision making literature under an asymptotic setting other than that of the OLP literature.

2. Problem Formulation

In our problem, there are m resources. Each resource j has an initial capacity c_j. The resources are used to serve a stream of requests arriving sequentially over T discrete time periods. At each time period t, one request arrives, denoted as request t, which is associated with a nonnegative reward $\tilde{\mathbf{r}}^{\mathbf{t}} = [\tilde{r}_1^t, \tilde{r}_2^t, \ldots, \tilde{r}_m^t]$ and a positive size $\tilde{\mathbf{d}}^{\mathbf{t}} = [\tilde{d}_1^t, \tilde{d}_2^t, \ldots, \tilde{d}_m^t]$. Here, $(\tilde{\mathbf{r}}^{\mathbf{t}}, \tilde{\mathbf{d}}^{\mathbf{t}})$ is assumed to be stochastic and its value is drawn independently from a known distribution which is stationary across time. After the value of $(\tilde{\mathbf{r}}^{\mathbf{t}}, \tilde{\mathbf{d}}^{\mathbf{t}})$ is revealed, denoted as $(\mathbf{r}^t, \mathbf{d}^t)$, the decision maker has to irrevocably decide to reject or accept request t. If rejected, no resource will be consumed and no reward will be collected. If accepted, the decision maker has to assign one of the resources to serve it. If resource j is assigned to serve request t, then d_j^t units of resource j will be consumed and a reward r_j^t will be collected. The objective of the decision maker is to maximize the total collected reward without violating the capacity constraints.

Any policy π is defined via a sequence of binary variables $\{x_j^\pi(t)\}$, where $x_j^\pi(t)$ denotes whether request t is served by resource j. Here, $x_j^\pi(t)$ is random and its realization is dependent on the arriving sequence of the requests. Policy π is feasible as long as π is non-anticipating, i.e., for any j, $x_j^\pi(t)$ can only depend on $\{(\mathbf{r}^1, \mathbf{d}^1), (\mathbf{r}^2, \mathbf{d}^2), \ldots, (\mathbf{r}^t, \mathbf{d}^t)\}$, and π satisfies the following constraints:

$$\sum_{t=1}^{T} d_j^t \cdot x_j^\pi(t) \leq c_j \quad \forall j \quad \text{and} \quad \sum_{j=1}^{m} x_j^\pi(t) \leq 1 \quad \forall t \tag{1}$$

The total reward collected by policy π is denoted as:

$$V^\pi(\mathbf{I}) = \sum_{t=1}^{T} \sum_{j=1}^{m} r_j^t \cdot x_j^\pi(t) \tag{2}$$

where $\mathbf{I} = \{(\mathbf{r}^1, \mathbf{d}^1), (\mathbf{r}^2, \mathbf{d}^2), \ldots, (\mathbf{r}^T, \mathbf{d}^T)\}$ denotes the sample path of request arrivals over the entire time horizon. Then the expected total reward collected by the policy π is denoted as $\mathbb{E}[V^\pi(\mathbf{I})]$, where $\mathbb{E}[\cdot]$ denotes taking expectation over the request arrivals $\mathbf{I}$.

We compare the performance of any feasible policy π to the prophet, an offline decision maker with full knowledge of the arriving sequence of the requests and always applying the optimal policy in hindsight. Similarly, given the sample path $\mathbf{I}$, we use a binary variable $x_j^t(\mathbf{I})$ to denote whether request t is served by resource j under the optimal policy in hindsight. Here, $x_j^t(\mathbf{I})$ is allowed to depend on the whole sample path $\mathbf{I}$. Indeed, given $\mathbf{I}$, $\{x_j^t(\mathbf{I})\}$ is the optimal solution to the following optimal offline problem:

$$
\begin{aligned}
V^{\text{off}}(\mathbf{I}) = \max \quad & \sum_{t=1}^{T}\sum_{j=1}^{m} r_j^t \cdot x_j^t \\
\text{s.t.} \quad & \sum_{t=1}^{T} d_j^t \cdot x_j^t \le c_j \quad \forall j \\
& \sum_{j=1}^{m} x_j^t \le 1 \quad \forall t \\
& x_j^t \in \{0,1\} \quad \forall j, \forall t
\end{aligned}
\tag{3}
$$

Then, the expected total reward collected by the prophet is denoted as $\mathbb{E}[V^{\text{off}}(\mathbf{I})]$.

We will use *regret* to measure the performance of any feasible policy. The regret of the policy π, denoted as $\text{Regret}(\pi)$, is defined as the difference between the expected total reward collected by the prophet and the expected total reward collected by the policy π:

$$
\text{Regret}(\pi) = \mathbb{E}[V^{\text{off}}(\mathbf{I})] - \mathbb{E}[V^\pi(\mathbf{I})]
\tag{4}
$$

3. Regret Bound for Finite-Support Reward Case

In this section, we develop our policy when the reward distribution has a finite support. Specifically, we make the following assumption:

Assumption 1 *For each t, the value of $\tilde{\mathbf{r}}^{\mathbf{t}}$ is drawn independently from a finite set $\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_n\}$, where $\mathbf{r}_i = [r_{i1}, r_{i2}, \ldots, r_{im}]$. Moreover, given that $\tilde{\mathbf{r}}^{\mathbf{t}}$ is realized as $\mathbf{r}_i$, we have $\tilde{\mathbf{d}}^{\mathbf{t}} = \mathbf{b}_i \cdot \tilde{u}_i^t$, where $\mathbf{b}_i = [b_{i1}, b_{i2}, \ldots, b_{im}]$ denotes the positive deterministic weight and $\tilde{u}_i^t$ is a single-dimensional random variable.*

For each t, we use p_i to denote the probability that $\tilde{\mathbf{r}}^{\mathbf{t}}$ is realized as $\mathbf{r}_i$ and use $F_i(\cdot)$ to denote the distribution of $\tilde{u}_i^t$, where its realization is denoted as u_i^t. Since the hindsight optimum obtained by solving the integer programming (3) often preserves complex structure thus is very hard to analyze, we first derive a tractable upper bound of $\mathbb{E}[V^{\text{off}}(\mathbf{I})]$. Moreover, we show that under Assumption 1, the prophet upper bound we have derived could be reduced equivalently into n sub-LPs and

accordingly, our original online resource allocation problem could be reduced to n separate sub-problems, where in each sub-problem the reward of each request is deterministic and each sub-LP serves as the prophet upper bound for each sub-problem.

Then, we show that the optimal solution of each sub-LP possesses a threshold structure and the thresholds could be given in a closed-form expression. By assuming some additional regularity conditions on the distribution function $F_i(\cdot)$ for each i, we further show a convex structure of the thresholds and then obtain a second-order upper estimates of how the optimal value of each sub-LP would vary over the right hand side (RHS) of the constraints. By applying a adaptive threshold policy in each sub-problem, we show in our regret analysis that the first order terms in the estimate will cancel out and only the second order terms will be left, which will finally lead to a $O(\log T)$ regret bound.

3.1. Prophet Upper Bound

We obtain an upper bound of $\mathbb{E}[V^{\text{off}}(\mathbf{I})]$ by considering the following LP:

$$
\begin{aligned}
\text{LP}^{\text{UB}} := \max \quad & T \cdot \sum_{i=1}^{n} p_i \cdot \mathbb{E}_i[\sum_{j=1}^{m} r_{ij} \cdot x_{ij}(u_i)] \\
\text{s.t.} \quad & T \cdot \sum_{i=1}^{n} p_i \cdot b_{ij} \cdot \mathbb{E}_i[u_i \cdot x_{ij}(u_i)] \le c_j \quad \forall j \\
& \sum_{j=1}^{m} x_{ij}(u_i) \le 1 \quad \forall i, \forall u_i \\
& 0 \le x_{ij}(u_i) \le 1 \quad \forall i, \forall j, \forall u_i,
\end{aligned}
\tag{5}
$$

Here, $\mathbb{E}_i[\cdot]$ denotes taking expectation over realization u_i with distribution function $F_i(\cdot)$ and $x_{ij}(u_i)$ could be interpreted as the probability that request t is served using resource j given the realization $(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)$, regardless of the time period t.

Lemma 1 *It holds that $LP^{UB} \ge \mathbb{E}[V^{off}(\mathbf{I})]$.*

Proof: We use $\{x_j^t(\mathbf{I})\}$ to denote one optimal solution of (3) given the arriving sequence of requests $\mathbf{I}$ and denote $x_{ij}^t(u_i) = \mathbb{E}[x_j^t(\mathbf{I})|(\mathbf{r}^t, \mathbf{d}^t) = (\mathbf{r}_i, \mathbf{b}_i \cdot u_i)]$. Then we have

$$
0 \le x_{ij}^t(u_i) \le 1 \quad \text{and} \quad \sum_{j=1}^{m} x_j^t(\mathbf{I}) \le 1 \Rightarrow \sum_{j=1}^{m} x_{ij}^t(u_i) \le 1
\tag{6}
$$

Also, we have that

$$
\begin{aligned}
\mathbb{E}[V^{\text{off}}(\mathbf{I})] &= \mathbb{E}[\sum_{t=1}^{T} \sum_{j=1}^{m} r_j^t \cdot x_j^t(\mathbf{I})] = \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}[\sum_{j=1}^{m} r_j^t \cdot x_j^t(\mathbf{I})|(\mathbf{r}^t, \mathbf{d}^t) = (\mathbf{r}_i, \mathbf{b}_i \cdot u_i)]\right] \\
&= \sum_{t=1}^{T} \mathbb{E}[\sum_{j=1}^{m} r_j \cdot x_{ij}^t(u_i)]
\end{aligned}
\tag{7}
$$

For each j, we could further obtain the following constraint regarding $x_{ij}^t(u_i)$:

$$\sum_{t=1}^{T}\sum_{i=1}^{n} p_i \cdot \mathbb{E}_i[b_{ij} \cdot u_i \cdot x_{ij}^t(u_i)] = \sum_{t=1}^{T} \mathbb{E}\left[\mathbb{E}[d_j^t \cdot x_j^t(\mathbf{I})|(\mathbf{r}^t, \mathbf{d}^t) = (\mathbf{r}_i, \mathbf{b}_i \cdot u_i)]\right]$$

$$= \mathbb{E}[\sum_{t=1}^{T} d_j^t \cdot x_j^t(\mathbf{I})] \leq c_j \tag{8}$$

Since the request distribution is stationary across different time periods and the offline optimal solution made by the prophet is not influenced by the order of the sequence $\mathbf{I}$, there must exists one optimal solution of (3) such that $x_{ij}^{t_1}(u_i) = x_{ij}^{t_2}(u_i)$ for arbitrary t_1 and t_2. Thus, we could suppress the notation t in $x_{ij}^t(u_i)$. Then given (6), (7) and (8), it holds that $\mathrm{LP}^{\mathrm{UB}} \geq \mathbb{E}[V^{\mathrm{off}}(\mathbf{I})]$. $\square$

We then show how to reduce (5) equivalently into n sub-LPs and characterize the optimal solution of each sub-LP in a closed-form expression. We consider the sub-LP of the following formulation:

$$G_i(\tilde{\mathbf{c}}_i, t) := \max \quad (T - t + 1) \cdot p_i \cdot \mathbb{E}_i[\sum_{j=1}^{m} r_{ij} \cdot x_{ij}(u_i)]$$

$$\text{s.t.} \quad (T - t + 1) \cdot p_i \cdot b_{ij} \cdot \mathbb{E}_i[u_i \cdot x_{ij}(u_i)] \leq \tilde{c}_{ij} \quad \forall j$$

$$\sum_{j=1}^{m} x_{ij}(u_i) \leq 1 \quad \forall u_i \tag{9}$$

$$0 \leq x_{ij}(u_i) \leq 1 \quad \forall j, \forall u_i$$

where $\tilde{\mathbf{c}}_i = [\tilde{c}_{i1}, \tilde{c}_{i2}, \ldots, \tilde{c}_{im}]$.

Lemma 2 *It holds that*

$$\mathrm{LP}^{UB} = \max \quad \sum_{i=1}^{n} G_i(\tilde{\mathbf{c}}_i, 1)$$

$$s.t. \quad \sum_{i=1}^{n} \tilde{c}_{ij} \leq c_j \quad \forall j \tag{10}$$

$$\tilde{c}_{ij} \geq 0 \quad \forall i, \forall j$$

We now show the threshold structure of the optimal solution of the sub-LP (9). Specifically, for each i, denote $\{i_1, i_2, \ldots, i_m\}$ as a permutation of $\{1, 2, \ldots, m\}$ such that $r_{ii_1} \geq r_{ii_2} \geq \cdots \geq r_{ii_m} \geq 0$ and we set $r_{ii_{m+1}} = 0$, then for each i, j, t, we define the threshold $\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)$ as the solution of the following equation:

$$p_i \cdot \int_0^{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} u dF_i(u) = \frac{\sum_{k=1}^{j} \tilde{c}_{ii_k}/b_{ii_k}}{T - t + 1} \tag{11}$$

If there is no $\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)$ satisfying the above equation, we set $\mu_t^{ii_j}(\tilde{\mathbf{c}}_i) = \infty$.

Lemma 3 *For each i, j, u_i, define the solution $\hat{x}_{ii_j}(u_i)$ by*

$$\hat{x}_{ii_j}(u_i) = \begin{cases} 1, & \text{if } \mu_t^{ii_{j-1}}(\tilde{\mathbf{c}}_i) < u_i \leq \mu_t^{ii_j}(\tilde{\mathbf{c}}_i) \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where $\mu_t^{ii_0}(\tilde{\mathbf{c}}_i)$ *is set to be* 0. *Then* $\{\hat{x}_{ii_j}(u_i)\}$ *is an optimal solution to* (9) *and it holds that*

$$G_i(\tilde{\mathbf{c}}_i, t) = (T - t + 1) \cdot p_i \cdot \sum_{j=1}^{m} (r_{ii_j} - r_{ii_{j+1}}) \cdot F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) \tag{13}$$

The proof of Lemma 2 and Lemma 3 is relegated to the appendix. Together, Lemma 2 and Lemma 3 will provide a closed-form characterization of the optimal solution of the linear programming prophet upper bound (5). Indeed, notice that $G_i(\tilde{\mathbf{c}}_i, t)$ is a concave function over $\tilde{\mathbf{c}}_i$, the optimization problem (10) is a convex optimization problem, and we could use the following formula to obtain the supergradient of $G_i(\tilde{\mathbf{c}}_i, t)$ immediately:

$$\frac{\partial G_i(\tilde{\mathbf{c}}_i, t)}{\partial \tilde{c}_{i_j}} = \sum_{k=j}^{m} \frac{r_{ii_k} - r_{ii_{k+1}}}{b_{ii_j} \cdot \mu_t^{ii_k}(\tilde{\mathbf{c}}_i)}, \quad \forall j$$

Thus, the optimization problem (10) is tractable and could be solved efficiently.

Remark: when there is only one resource and the reward is unitary, i.e., $m = n = 1$, our prophet upper bound (5) is identical to the upper bound in Arlotto and Xie (2020). However, their approach to derive the prophet upper bound is based on an observation that when there is only one resource and the reward is unitary, the hindsight optimal policy is to simply serve the requests in an increasing order of the size. Thus, the order statistics of the size of the requests is enough to characterize the hindsight optimum. Their approach could not be directly applied to a more general setting where this observation doesn't hold. Specifically, when there are multiple resources and the reward could take multiple values, the hindsight optimum is an integer programming (3), the solution of which should not be serving the request from the smallest size. Instead, our approach is based on directly dealing with the linear programming prophet upper bound (5) and show its specific threshold structure.

3.2. Regularity Conditions

In this subsection, we will describe the regularity conditions that the distribution functions $F_i(\cdot)$ need to satisfy. A special case of our problem where there is only one resource and the reward is unitary is identical to the dynamic and stochastic knapsack problem with equal reward studied in a stream of literature: Coffman et al. (1987), Bruss and Robertson (1991), Papastavrou et al. (1996), Arlotto and Xie (2020). All these papers assumed some regularity conditions on the distribution functions to obtain their results. One may see the necessity of the regularity conditions by referring to Section 5 in Papastavrou et al. (1996) which shows that one could not expect some structural properties such as the monotonicity of the optimal threshold and the concavity of the optimal value for general distributions. Specifically, Arlotto and Xie (2020) proved a $O(\log T)$ regret bound for this knapsack problem by assuming the distribution function belongs to the so-called *typical class*.

In what follows, we adopt the *typical class* conditions with only one minor change which will be discussed later.

Assumption 2 *For some $\bar{\omega} > 0$ and for all i, the distribution function $F_i(\cdot)$ has a continuous density function $f_i(\cdot)$ and satisfies the following two conditions:*

1. *(Behavior at Zero) There exists a constant $0 < \lambda < 1$ and two constants $1 < \gamma_1 < \gamma_2$ such that*

$$\gamma_1 \cdot F_i(\lambda w) \le F_i(w) \le \gamma_2 \cdot F_i(\lambda w) \quad \text{for all } w \in (0, \bar{\omega}) \tag{14}$$

2. *(Monotonicity) The function $w :\to w^3 f_i(w)$ is a non-decreasing function over $(0, \bar{\omega})$.*

The only difference between our Assumption 2 and the *typical class* conditions in Arlotto and Xie (2020) lies in condition 1. *Typical class* only assumed $F_i(w)$ to be lower bounded by $F_i(\lambda w)$, i.e., they only assumed $\gamma_1 \cdot F_i(\lambda w) \le F_i(w)$, while in our assumption, we assume $F_i(w)$ to be both lower bounded and upper bounded by $F_i(\lambda w)$. The discussions in Section 5 of Arlotto and Xie (2020) showed that their *typical class* conditions could cover a very broad class of continuous distributions. Moreover, it is easy to check that the following distributions they have considered continue to satisfy both conditions in Assumption 2:

1. Uniform distribution, exponential distribution and logit-normal distribution.
2. Truncated normal distribution and truncated logistic distribution.
3. Power distribution with distribution function $F(x) = Ax^\alpha$ for $A, \alpha > 0$ truncated on $(0, b)$, where b is any positive constant.
4. Distributions with convex distribution function $F(\cdot)$ and continuous density function $f(\cdot)$ such that $f(0) > 0$.
5. The mixture distribution of some distributions satisfying both conditions in Assumption 2.

We show in the next lemma that condition 1 in Assumption 2 will enable us to derive the following estimate over the distribution functions $F_i(\cdot)$.

Lemma 4 *If there exists a continuous distribution $F(\cdot)$ such that for some $\bar{\omega} > 0$, condition 1 in Assumption 2 is satisfied, i.e., there exists a constant $0 < \lambda < 1$ and two constants $1 < \gamma_1 < \gamma_2$ such that*

$$\gamma_1 \cdot F(\lambda w) \le F(w) \le \gamma_2 \cdot F(\lambda w) \quad \text{for all } w \in (0, \bar{\omega}) \tag{15}$$

then there exists a constant $M > 1$ such that for any w_1, w_2 satisfying $0 < w_1 < w_2 < \bar{\omega}$, we have

$$\frac{w_2(F(w_2) - F(w_1))}{\int_{w_1}^{w_2} w \, dF(w)} \le M \tag{16}$$

Also, we could obtain the following convex property regarding the threshold $\mu_t^{iij}(\tilde{\mathbf{c}}_i)$ for each i and j under condition 2 in Assumption 2.

Lemma 5 *Suppose Assumption 2 holds, then for each i, j, and any $k = 1, 2, \ldots, j$, the function $\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}$ is a convex function over $\tilde{c}_{ii_k}$.*

The proofs of the above two lemmas are relegated to the appendix. Together, Lemma 4 and Lemma 5 would enable us to obtain the second-order upper estimate of the sensitivity of $G_i(\tilde{\mathbf{c}}_i, t)$ over $\tilde{\mathbf{c}}_i$. Indeed, given the formulation of $G_i(\tilde{\mathbf{c}}_i, t)$ in (13), it is enough to consider the sensitivity of $F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))$ over $\tilde{\mathbf{c}}_i$ for each j. The following second-order estimate generalizes the estimate in Arlotto and Xie (2020) from the single-resource setting to the multi-resource setting and the proof could be found in the appendix.

Lemma 6 *Suppose Assumption 2 holds, then for each i, j and any $k = 1, 2, \ldots, j$, we have that*

$$F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k})) \le \frac{(M-1) \cdot u^2}{(\tilde{c}_{ii_k}/b_{ii_k}) p_i (T - t + 1) \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} + \frac{u}{p_i (T - t + 1) \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} \quad (17)$$

holds for any $u \in [0, \tilde{c}_{ii_k}/b_{ii_k}]$.

3.3. Adaptive Threshold Policy

We now derive our adaptive threshold policy and prove the corresponding regret bound. Since the optimal solution and the sensitivity of each sub-LP (9) could be fully characterized in (12) and in (17), we will reduce the original online resource allocation problem into n sub-problems to utilize this characterization. Specifically, we will divide each resource j into n parts and assign a capacity c_{ij}^* to each ith part of resource j, where $\{c_{ij}^*\}$ is denoted as the optimal solution of (10). Then, for each request with a realized reward $\mathbf{r}_i$, we will only consider using the ith part of each resource to serve it. By doing so, the ith sub-problem is only concerned about serving request with a deterministic reward $\mathbf{r}_i$ using a separate part of each resource. Thus, each sub-problem will be independent from each other and it is enough to focus on each sub-problem separately.

For ith sub-problem, we will construct our control policy based on re-solving (11) to obtain the thresholds adaptively. Specifically, at each time period t, denote the remaining capacity of the ith part of resource j as c_{ij}^t and denote $\mathbf{c}_i^t = [c_{i1}^t, c_{i2}^t, \ldots, c_{im}^t]$, then given $\tilde{\mathbf{r}}^t = \mathbf{r}_i$ and $\tilde{\mathbf{d}}^t = \mathbf{b}_i \cdot u_i^t$, the threshold structure of the optimal solution of $G_i(\mathbf{c}_i^t, t)$ in (12) implies us to assign resource i_j to serve the request as long as $\mu_t^{ii_{j-1}}(\mathbf{c}_i^t) < u_i^t \le \mu_t^{ii_j}(\mathbf{c}_i^t)$. However, the actual threshold will be truncated by the remaining capacity $\mathbf{c}_i^t$. Thus, in our adaptive threshold policy, we will compare u_i^t sequentially to the truncated thresholds $h_t^{ii_j}(\mathbf{c}_i^t) = \min\{\mu_t^{ii_j}(\mathbf{c}_i^t), c_{ii_j}^t/b_{ii_j}\}$ from $j = 1$ to m and assign resource i_j to serve the request once $u_i^t \le h_t^{ii_j}(\mathbf{c}_i^t)$. Our adaptive threshold policy is formally presented in Algorithm 1.

Next, we discuss the performance of our policy. For each i, denote $V_i^{\mathrm{ATP}_1}(\mathbf{I})$ as the total reward

---

Algorithm 1 Adaptive Threshold Policy (ATP$_1$)

1: Solve (10) and denote the optimal solution as $\{\mathbf{c}_i^1\}$.

2: At time period $t = 1, 2, \ldots, T$:

3: After request t arrives and reveals itself as $(\mathbf{r}_i, \mathbf{b}_i \cdot u_i^t)$, re-solve (11) to obtain the thresholds $\mu_t^{ii_j}(\mathbf{c}_i^t)$ for each j.

4: For each j, define the truncated threshold $h_t^{ii_j}(\mathbf{c}_i^t) = \min\{c_{ii_j}^t/b_{ii_j}, \mu_t^{ii_j}(\mathbf{c}_i^t)\}$ and then do the following:

   (i) If for all j, $u_i^t > h_t^{ii_j}(\mathbf{c}_i^t)$, reject request t and then set $c_{i'j'}^{t+1} = c_{i'j'}^t$ for each i' and j'.

   (ii) Otherwise, compare u_i^t to the truncated thresholds $h_t^{ii_1}(\mathbf{c}_i^t), h_t^{ii_2}(\mathbf{c}_i^t), \ldots, h_t^{ii_m}(\mathbf{c}_i^t)$ sequentially and assign the first resource i_j to serve request t such that $u_i^t \leq h_t^{ii_j}(\mathbf{c}_i^t)$. Then set $c_{ii_j}^{t+1} = c_{ii_j}^t - b_{ii_j} \cdot u_i^t$ and set $c_{i'j'}^{t+1} = c_{i'j'}^t$ for all other (i', j') such that $(i', j') \neq (i, i_j)$.

---

collected by Algorithm 1 for the ith sub-problem, i.e., the total reward collected from serving the requests with a realized reward $\mathbf{r}_i$, then we have

$$\mathbb{E}[V^{\mathrm{ATP_1}}(\mathbf{I})] = \sum_{i=1}^n \mathbb{E}[V_i^{\mathrm{ATP_1}}(\mathbf{I})] \tag{18}$$

Thus, from (10), the regret of Algorithm 1 could be expressed as the summation of the regret of each sub-problem:

$$\mathrm{Regret}(\mathrm{ATP_1}) \leq \sum_{i=1}^n G_i(\mathbf{c}_i^*, 1) - \mathbb{E}[V_i^{\mathrm{ATP_1}}(\mathbf{I})] \tag{19}$$

where $\{\mathbf{c}_i^*\}$ denotes the optimal solution of (10). For each i, we have the following upper bound of the term $G_i(\mathbf{c}_i^*, 1) - \mathbb{E}[V_i^{\mathrm{ATP_1}}(\mathbf{I})]$.

Lemma 7 *Suppose Assumption 2 holds and let K_i be a constant defined as follows:*

$$K_i = \left\lceil \frac{\sum_{j=1}^m c_{ii_j}^*/b_{ii_j}}{p_i \cdot \int_0^{\bar{\omega}} w dF_i(w)} \right\rceil \tag{20}$$

Then, we have the following bound:

$$G_i(\mathbf{c}_i^*, 1) - \mathbb{E}[V_i^{ATP_1}(\mathbf{I})] \leq \sum_{j=1}^m r_{ii_j}[(j+1)M - j](\log T + 1) + r_{\max}^i \cdot K_i = O(m^2 \log T) \tag{21}$$

where $r_{\max}^i = \max\{r_{i1}, r_{i2}, \ldots, r_{im}\}$ and M is the constant defined in Lemma 4.

The proof of Lemma 7 is based on first deriving the recursion of the "to-go" expected reward of the policy given the current time period t and the remaining capacity $\mathbf{c}_i^t$, denoted as $V_t(\mathbf{c}_i^t)$, in the following way:

$$V_t(\mathbf{c}_i^t) = (1 - p_i \cdot F(h_t^{ijW}(\mathbf{c}_i^t))) \cdot V_{t+1}(\mathbf{c}_i^t) + p_i \cdot \sum_{w=1}^W \int_{h_t^{ijw-1}(\mathbf{c}_i^t)}^{h_t^{ijw}(\mathbf{c}_i^t)} \{r_{jw} + V_{t+1}(\mathbf{c}_i^t - b_{ijw} u \cdot \mathbf{e}_{jw})\} dF_i(u)$$

where $\{h_t^{ijw}(\mathbf{c}_i^t)\}$ denotes the monotone increasing subsequence of $\{h_t^{ii_j}(\mathbf{c}_i^t)\}$. Note that although the thresholds $\{\mu_t^{ii_j}(\mathbf{c}_i^t)\}$ is monotone increasing in j, the actual thresholds $\{h_t^{ii_j}(\mathbf{c}_i^t)\}$ involved in our policy may not be monotone due to the truncation of the remaining capacity, thus, only the monotone subsequence $\{h_t^{ijw}(\mathbf{c}_i^t)\}$ will show up in the above recursion. Then denoting $\rho_t(\mathbf{c}_i^t) = G_i(\mathbf{c}_i^t, t) - V_t(\mathbf{c}_i^t)$, we could substitute $\rho_t(\mathbf{c}_i^t)$ into the above formula and obtain a recursion over $\rho_t(\mathbf{c}_i^t)$. We will use the second-order upper estimate in Lemma 6 to upper bound each term in the recursion, and together with an analysis on the influence of truncation, i.e., comparing the subsequence $\{h_t^{ijw}(\mathbf{c}_i^t)\}$ with $\{\mu_t^{ii_j}(\mathbf{c}_i^t)\}$, we could show that the first-order term will cancel out in the recursion and only the second-order term will be left. Thus, we show that for any $\mathbf{c}_i^t$, we have $\rho_t(\mathbf{c}_i^t) = O(\sum_{t'=T-t+1}^T \frac{1}{t'})$, which implies that $\rho_1(\mathbf{c}_i^*) = O(\log T)$. The complete proof is relegated to the appendix. Then, from (19), we could establish the following regret bound of Algorithm 1.

Theorem 1 *Suppose Assumption 2 hold, then we have the following regret bound:*

$$Regret(ATP_1) \le \sum_{i=1}^n \sum_{j=1}^m r_{ii_j}[(j+1)M - j](\log T + 1) + \sum_{i=1}^n r_{max}^i \cdot K_i = O(nm^2 \log T) \qquad (22)$$

It is interesting to note that the re-solving procedure for each sub-problem in Algorithm 1 has also been applied to designing near-optimal policy for network revenue management problem with a so-called "non-degeneracy" assumption in Jasin and Kumar (2012) and following-up work Jasin (2015). We will conclude this subsection by making a comparison between our approach and their approach in the following remark.

Remark: a similarity between our approach and the approach in Jasin and Kumar (2012) and Jasin (2015) is that the optimal solution for the LP upper bound could be expressed in closed-form via the remaining capacity of each resource. However, our regret analysis is fundamentally different from theirs. A key step in their analysis is that they show under the "non-degeneracy" assumption, the closed-form expression could be obtained provided that $\frac{c_j^t}{T-t+1}$ for all j, where c_j^t denotes the remaining capacity of resource j at time t, doesn't variate too much from $\frac{c_j}{T}$ so that the optimal bases for the LP upper bound doesn't change. Then they show that this condition could be guaranteed for a sufficiently long time in their policy by constructing a martingale involving $\frac{c_j^t}{T-t+1}$ and applying the Doob's inequality for martingale. As a result, their regret bound is obtained when the ratio $\frac{c_j}{T}$ is fixed and their approach could not be directly applied to the setting where c_j could grow sublinearly in T. In contrast, our closed-form expression is obtained for any c_j^t and our regret analysis is based on considering the curvature of the distribution function of the size instead of constructing a martingale, thus our approach works for the setting where c_j is fixed and $T \to \infty$. We also need to analyze the influence of the fact that the actual threshold is truncated by the remaining capacity, which is not needed in their analysis since c_j grows linearly in T thus

c_j^t will always exceeds the size of the request except for the last few time periods. Moreover, the re-solving procedure in Jasin and Kumar (2012) and Jasin (2015) requires solving a LP, which may be time consuming, while our re-solving procedure in Algorithm 1 could be done quickly with the closed-form characterization (11).

3.4. Asymptotic Optimality

In this subsection we show how does $\mathbb{E}[V^{\text{off}}(\mathbf{I})]$ scale over T. Our next theorem shows that $\mathbb{E}[V^{\text{off}}(\mathbf{I})] = \Omega(T^{\frac{1}{1+\alpha}})$, where α is a positive constant depending on λ, γ_2 in Assumption 2. Thus, we immediately show the asymptotic optimality of our adaptive threshold policy as $T \to +\infty$.

Theorem 2 *Suppose Assumption 1 and Assumption 2 hold, then we have*

$$\mathbb{E}[V^{off}(\mathbf{I})] = \Omega(T^{\frac{1}{1+\alpha}}) \tag{23}$$

where α is a positive constant satisfying $\frac{1}{\gamma_2} = \lambda^{\alpha}$.

4. Regret Bound for Resource-independent Case

The main purpose of this section is to study whether the term n in our previous $O(\log T)$ regret bound could be removed and whether we could directly deal with our online resource allocation problem without reducing the problem into n separate problems. We show that under a special case where both the reward and the size of each request are resource-independent, we can indeed derive a similar adaptive threshold policy to achieve a $O(\log T)$ regret bound without any dependence on n. The main difference of our analysis from the previous analysis is that we directly show the threshold structure of the prophet upper bound without dividing it into n sub-LPs. As a result, we could consider our online resource allocation problem as a whole without any reduction. Although we only analyze the case where the reward distribution has a finite support to keep the notation in consistent with the previous sections, our analysis could be directly applied to the case where the support of the reward distribution is infinite and our regret bound continues to hold. In addition to Assumption 1, we further make the following assumption:

Assumption 3 *For each i, $r_{ij} = r_i$ for every j and $\mathbf{b}_i$ is an all-one vector .*

In what follows, we will first show the threshold structure of our prophet upper bound and then derive our policy and prove the corresponding regret bound.

Theorem 3 *We have the following upper bound of the expected total reward collected by the prophet:*

$$\mathbb{E}[V^{off}(\mathbf{I})] \leq T \cdot \sum_{i=1}^{n} p_i \cdot r_i \cdot F_i(r_i \cdot \mu(\mathbf{c})) \tag{24}$$

where the threshold $\mu(\mathbf{c})$ is obtained by solving the following equation:

$$\sum_{i=1}^{n} p_i \cdot \int_{0}^{r_i \cdot \mu(\mathbf{c})} u_i dF_i(u_i) = \frac{\sum_{j=1}^{m} c_j}{T} \tag{25}$$

If there is no $\mu(\mathbf{c})$ satisfying the above equation, we set $\mu(\mathbf{c}) = +\infty$.

Proof: First, we have that

$$
\begin{aligned}
\mathrm{LP}^{\mathrm{UB}} \leq \max \quad & T \cdot \sum_{i=1}^{n} p_i \cdot \mathbb{E}_i[r_i \cdot y_i(u_i)] \\
\text{s.t.} \quad & T \cdot \sum_{i=1}^{n} p_i \cdot \mathbb{E}_i[d \cdot y_i(u_i)] \leq \sum_{j=1}^{m} c_j \\
& 0 \leq y_i(u_i) \leq 1 \quad \forall i, \forall u_i
\end{aligned}
\tag{26}
$$

The inequality holds by noting that the solution $\{\hat{y}_i(u_i)\}$ defined by $\hat{y}_i(u_i) = \sum_{j=1}^{m} x_j^*(r_i, u_i)$, where $\{x_j^*(r_i, u_i)\}$ is the optimal solution of (5), is feasible to the RHS of (26), and we have

$$\mathrm{LP}^{\mathrm{UB}} = T \cdot \sum_{i=1}^{n} p_i \cdot \mathbb{E}_i[\sum_{j=1}^{m} r_i \cdot x_j^*(r_i, u_i)] = T \cdot \sum_{i=1}^{n} p_i \cdot \mathbb{E}_i[r_i \cdot \hat{y}_i(u_i)]$$

Suppose the optimal solution of the RHS of (26) is denoted as $\{y_i^*(u_i)\}$, then define $\hat{z}_i = \mathbb{E}_i[y_i^*(u_i)] \in [0,1]$ for every i and we have the following constraint regarding $\hat{z}_i$:

$$\sum_{j=1}^{m} c_i \geq T \cdot \sum_{i=1}^{n} p_i \cdot \mathbb{E}_i[u_i \cdot y_i^*(u_i)] \geq T \cdot \sum_{i=1}^{n} p_i \cdot \left[\int_{0}^{F_i^{-1}(\hat{z}_i)} u_i dF_i(u_i) \right] = T \cdot \sum_{i=1}^{n} p_i \cdot \left[\int_{0}^{\hat{z}_i} F_i^{-1}(v_i) dv_i \right] \tag{27}$$

where the second inequality holds by noting that $y_i^*(u_i) \in [0,1]$ thus $\mathbb{E}_i[u_i \cdot y_i^*(u_i)]$ must be lower bounded by integrating u_i below the $\mathbb{E}_i[y_i^*(u_i)]$ quantile with distribution $F_i(\cdot)$, and the last equality holds by replacing variable u_i using $F_i^{-1}(v_i)$. Thus, we have

$$
\begin{aligned}
\mathrm{LP}^{\mathrm{UB}} \leq \max \quad & T \cdot \sum_{i=1}^{n} p_i \cdot r_i \cdot z_i \\
\text{s.t.} \quad & T \cdot \sum_{i=1}^{n} p_i \cdot \int_{0}^{z_i} F_i^{-1}(v_i) dv_i \leq \sum_{j=1}^{m} c_j \\
& 0 \leq z_i \leq 1 \quad \forall i
\end{aligned}
\tag{28}
$$

The RHS of (28) is a convex optimization problem over $\{z_i\}$ by noting that for each i, if we define $g_i(z) = \int_{0}^{z} F_i^{-1}(v) dv$, then $g_i'(z) = F_i^{-1}(z)$ is a non-decreasing function over z, which implies that $g_i(z)$ is a convex function over z. Then strong duality implies that

$$\mathrm{LP}^{\mathrm{UB}} \leq \min_{\mu \geq 0} \max_{0 \leq z_i \leq 1} L(\mathbf{z}, \mu) \tag{29}$$

where $L(\mathbf{z}, \mu)$ is the lagrangian function defined as follows:

$$L(\mathbf{z}, \mu) = T \cdot \sum_{i=1}^{n} r_i \cdot p_i \cdot z_i + \mu \cdot [\sum_{j=1}^{m} c_i - T \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{z_i} F_i^{-1}(v_i) dv_i] \tag{30}$$

Denote $\{z_i^*, \mu^*\}$ as the optimal solution of the dual problem, then from the KKT condition, we have that

$$\frac{\partial L(\mathbf{z}^*, \mu^*)}{\partial z_i} = 0 \Rightarrow r_i \cdot p_i = \mu^* \cdot p_i \cdot F_i^{-1}(z_i^*) \Rightarrow z_i^* = F_i(\frac{r_i}{\mu^*}) \tag{31}$$

and

$$\mu^* \cdot [\sum_{j=1}^{m} c_i - T \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{z_i^*} F_i^{-1}(v_i) dv_i] = 0 \tag{32}$$

Then, $\mu^* > 0$ implies that

$$\sum_{j=1}^{m} c_i - T \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{z_i^*} F_i^{-1}(v_i) dv_i = 0 \Rightarrow \sum_{j=1}^{m} c_i - T \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{F_i^{-1}(\frac{r_i}{\mu^*})} F_i^{-1}(v_i) dv_i = 0$$

Thus, the optimal dual variable μ^* is obtained by solving the following equation:

$$T \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{\frac{r_i}{\mu^*}} w_i dF_i(w_i) = T \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{F_i^{-1}(\frac{r_i}{\mu^*})} F_i^{-1}(v_i) dv_i = \sum_{j=1}^{m} c_i$$

and $\mu^* = 0$ if no μ^* makes the above equation hold. Denoting $\mu(\mathbf{c}) = 1/\mu^*$, we finally have $z_i^* = r_i \cdot \mu(\mathbf{c})$ for each i and

$$\text{LP}^{\text{UB}} \leq \min_{\mu \geq 0} \max_{0 \leq z_i \leq 1} L(\mathbf{z}, \mu) = T \cdot \sum_{i=1}^{n} p_i \cdot r_i \cdot F_i(r_i \cdot \mu(\mathbf{c}))$$

□

When the request arrives and reveals itself as (r_i, u_i), based on the threshold structure of the prophet upper bound (24), we will make the accept/reject decision by comparing size u_i to the threshold $r_i \cdot \mu$, where we will compute the threshold μ adaptively by (25). If the request is accepted, since the reward is resource-independent, it is natural for our policy to assign the resource with the largest capacity to serve it. Our policy is formally presented in Algorithm 2. Following a similar way of proving Lemma 7, we could obtain the following regret bound of Algorithm 2.

Theorem 4 *Suppose Assumption 2 holds and $\hat{K}$ is the constant defined as follows:*

$$\hat{K} = \left\lceil \frac{\sum_{j=1}^{m} c_j}{\sum_{i=1}^{n} p_i \cdot \int_0^{r_i \cdot \frac{\bar{\omega}}{r_{\max}}} w_i dF_i(w_i)} \right\rceil \tag{34}$$

where $r_{\max} = \max\{r_1, r_2, \ldots, r_n\}$. Then we have the following regret bound

$$Regret(ATP_2) \leq (\log T + 1) \cdot (2mM - m) \cdot r_{\max} + r_{\max} \cdot \hat{K} = O(m \log T) \tag{35}$$

Algorithm 2 Adaptive Threshold Policy (ATP$_2$)

1: At time period $t = 1, 2, \ldots, T$

2: Denote the remaining capacity as $\mathbf{c}^t$, and compute the threshold $\mu_t(\mathbf{c}^t)$ by solving the equation

$$\sum_{i=1}^{n} p_i \cdot \int_0^{r_i \cdot \mu_t(\mathbf{c}^t)} w_i dF_i(w_i) = \frac{\sum_{j=1}^{m} c_j^t}{T - t + 1} \tag{33}$$

If there is no threshold $\mu_t(\mathbf{c}^t)$ making the equation holds, we set $\mu_t(\mathbf{c}^t) = +\infty$.

3: Denote j_t as the resource with the largest remaining capacity, i.e., $c_{j_t}^t = \max\{c_1^t, c_2^t, \ldots, c_m^t\}$.

4: For each i, define the truncated threshold $h_t^i(\mathbf{c}^t) = \min\{c_{j_t}^t, r_i \cdot \mu_t(\mathbf{c}^t)\}$.

5: After request t arrives and reveals itself as (r_i, u_i), do the following:

   (i) Assign resource j_t to serve request t if $u_i \leq h_t^i(\mathbf{c}^t)$.

   (ii) Reject request t if $u_i > h_t^i(\mathbf{c}^t)$.

It is also worth-noting that when there is only one resource, then the *typical class* conditions from Arlotto and Xie (2020) is enough to establish the above regret bound. The reason is that in the proof, when there is only one resource, we only need to bound the term $\frac{wF(w)}{\int_0^w udF(u)}$, which could be upper bounded by a constant M under the *typical class* conditions due to Lemma 4 in Arlotto and Xie (2020). Also, as shown in the next theorem, the $\Omega(T^{\frac{1}{1+\alpha}})$ lower bound of $\mathbb{E}[V^{\text{off}}(T)]$ continues to hold. Thus, given the $O(\log T)$ regret bound, Algorithm 2 is asymptotically optimal as $T \to \infty$.

Theorem 5 *Under Assumption 1, Assumption 2 and Assumption 3, we have that*

$$\mathbb{E}[V^{off}(\mathbf{I})] = \Omega(T^{\frac{1}{1+\alpha}}) \tag{36}$$

where α is a positive constant satisfying $\frac{1}{\gamma_2} = \lambda^\alpha$. Moreover, the constant terms in $\Omega(\cdot)$ is independent of n.

Note that both the regret bound in Theorem 4 and the lower bound of the expected reward obtained by the prophet in Theorem 5 depend only on $\hat{K}$ instead of n. Thus, Algorithm 2 is asymptotically optimal as $T \to \infty$ even when the reward distribution has an infinite support.

5. Concluding Remarks

In this paper, we consider a general model of the online resource allocation problem and we derive near-optimal policies for our problem under two conditions separately: the reward could only take n values and given each reward realization, the size distribution equals a deterministic weight multiplied by a single-dimensional random variable, or both the reward and the size are independent of which resource assigned. Under each condition, we propose an adaptive threshold policy and

prove a $O(\log T)$ regret bound. Our approaches for both conditions are based on showing that the optimal solution to the corresponding prophet upper bound possesses a threshold structure and we could mimick the prophet by following an adaptive threshold policy. Specifically, under the first condition, our policy first divide each resource into n parts and only use one specific part of each resource to serve the requests with the corresponding reward. As a result, the regret bound depends on n. We then explore if there exists a policy without dividing the resources and if the regret bound will not depend on the cardinality of the support set of the reward. We show that under the second condition, we could propose another adaptive threshold policy which utilizes each resource as a whole and we show that the corresponding regret bound will be independent of n. Our analysis could be directly applied to the case where the cardinality of the support set of the reward is infinitely large and our regret bound continue to hold. Also, we show that under both conditions, the expected reward collected by the prophet is at least $\Omega(T^{\frac{1}{1+\alpha}})$, where α is a positive constant depending on the distribution functions, thus, together with the $O(\log T)$ regret bound, we establish the asymptotic optimality of our policies.

Our future efforts will be devoted to deriving near optimal policies and proving regret bounds for more general settings. For example, the setting where the size distribution could not be expressed explicitly by a single-dimensional random variable, or the setting where the reward distribution has an infinite support and the reward and the size could be resource-dependent.

References

Agrawal, Shipra, Zizhuo Wang, Yinyu Ye. 2014. A dynamic near-optimal algorithm for online linear programming. *Operations Research* **62**(4) 876–890.

Arlotto, Alessandro, Itai Gurvich. 2019. Uniformly bounded regret in the multisecretary problem. *Stochastic Systems* **9**(3) 231–260.

Arlotto, Alessandro, Xinchang Xie. 2020. Logarithmic regret in the dynamic and stochastic knapsack problem with equal rewards. *Stochastic Systems* .

Babaioff, Moshe, Nicole Immorlica, David Kempe, Robert Kleinberg. 2007. A knapsack secretary problem with applications. *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*. Springer, 16–28.

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

Bhalgat, Anand, Ashish Goel, Sanjeev Khanna. 2011. Improved approximation results for stochastic knapsack problems. *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 1647–1665.

Bruss, F Thomas, James B Robertson. 1991. wald's lemma'for sums of order statistics of iid random variables. *Advances in applied probability* **23**(3) 612–623.

Buchbinder, Niv, Joseph Naor. 2009. Online primal-dual algorithms for covering and packing. *Mathematics of Operations Research* **34**(2) 270–286.

Bumpensanti, Pornpawee, He Wang. 2020. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science* .

Coffman, EG, L Flatto, RR Weber. 1987. Optimal selection of stochastic intervals under a sum constraint. *Advances in applied probability* **19**(2) 454–473.

Dantzig, George B. 1957. Discrete-variable extremum problems. *Operations research* **5**(2) 266–288.

Dean, Brian C, Michel X Goemans, Jan Vondrák. 2008. Approximating the stochastic knapsack problem: The benefit of adaptivity. *Mathematics of Operations Research* **33**(4) 945–964.

Derman, Cyrus, Gerald J Lieberman, Sheldon M Ross. 1975. A stochastic sequential allocation model. *Operations Research* **23**(6) 1120–1130.

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

Dütting, Paul, Michal Feldman, Thomas Kesselheim, Brendan Lucier. 2017. Prophet inequalities made easy: Stochastic optimization by pricing non-stochastic inputs. *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 540–551.

Freund, Daniel, Siddhartha Banerjee. 2019. Uniform loss algorithms for online stochastic decision-making with applications to bin packing. *Available at SSRN 3479189* .

Jasin, Stefanus. 2015. Performance of an lp-based control for revenue management with unknown demand parameters. *Operations Research* **63**(4) 909–915.

Jasin, Stefanus, Sunil Kumar. 2012. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research* **37**(2) 313–345.

Li, Xiaocheng, Yinyu Ye. 2019. Online linear programming: Dual convergence, new algorithms, and regret bounds. *arXiv preprint arXiv:1909.05499* .

Lueker, George S. 1998. Average-case analysis of off-line and on-line knapsack problems. *Journal of Algorithms* **29**(2) 277–305.

Ma, Will. 2018. Improvements and generalizations of stochastic knapsack and markovian bandits approximation algorithms. *Mathematics of Operations Research* **43**(3) 789–812.

Marchetti-Spaccamela, Alberto, Carlo Vercellis. 1995. Stochastic on-line knapsack problems. *Mathematical Programming* **68**(1-3) 73–104.

Papastavrou, Jason D, Srikanth Rajagopalan, Anton J Kleywegt. 1996. The dynamic and stochastic knapsack problem with deadlines. *Management Science* **42**(12) 1706–1718.

Vera, Alberto, Siddhartha Banerjee. 2020. The bayesian prophet: A low-regret framework for online decision making. *Management Science* .

Vera, Alberto, Siddhartha Banerjee, Itai Gurvich. 2019. Online allocation and pricing: Constant regret via bellman inequalities. *arXiv preprint arXiv:1906.06361* .

Zhou, Yunhong, Deeparnab Chakrabarty, Rajan Lukose. 2008. Budget constrained bidding in keyword auctions and online knapsack problems. *International Workshop on Internet and Network Economics*. Springer, 566–576.

Appendix. Proofs of Lemmas, Propositions and Theorems

A. Proof of Lemma 2

We will prove the lemma by showing that both relationships "$\leq$", "$\geq$" hold between $\mathrm{LP}^{\mathrm{UB}}$ and the right hand side of (10). We first prove "$\leq$". Denote $\{x_j^*(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)\}$ as the optimal solution of (5) and define $\hat{c}_{ij} = T \cdot p_i \cdot b_{ij} \cdot \mathbb{E}_i[u_i \cdot x_j^*(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)]$. Obviously, the capacity constraints of (5) imply that $\{\hat{c}_{ij}\}$ forms a feasible solution of the RHS of (10). Then we focus on $G_i(\hat{\mathbf{c}}_i, 1)$ and denote $\hat{x}_{ij}(u_i) = x_j^*(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)$. It is obvious that $\{\hat{x}_{ij}(u_i)\}$ forms a feasible solution of (9), thus we have that

$$G_i(\hat{\mathbf{c}}_i, 1) \geq T \cdot p_i \cdot \mathbb{E}_i[\sum_{j=1}^m r_{ij} \cdot \hat{x}_{ij}(u_i)] = T \cdot p_i \cdot \mathbb{E}_i[\sum_{j=1}^m r_{ij} \cdot x_j^*(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)]$$

which implies that

$$\mathrm{LP}^{\mathrm{UB}} = \sum_{i=1}^n T \cdot p_i \cdot \mathbb{E}_i[\sum_{j=1}^m r_{ij} \cdot x_j^*(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)] \leq \sum_{i=1}^n G_i(\hat{\mathbf{c}}_i, 1) \leq \text{the RHS of (10)}$$

Next, we prove "$\geq$". Denote $\{c_{ij}^*\}$ as the optimal solution of the RHS of (10) and denote $\{x_{ij}^*(u_i)\}$ as the optimal solution of (9) for $G_i(\mathbf{c}_i^*, 1)$. Then we define $\hat{x}_j(\mathbf{r}_i, \mathbf{b}_i \cdot u_i) = x_{ij}^*(u_i)$ and we have that

$$0 \leq \hat{x}_j(\mathbf{r}_i, \mathbf{b}_i \cdot u_i) \leq 1 \quad \text{and} \quad \sum_{j=1}^m \hat{x}_j(\mathbf{r}_i, \mathbf{b}_i \cdot u_i) = \sum_{j=1}^m x_{ij}^*(u_i) \leq 1$$

Also, we have that

$$\sum_{i=1}^n T \cdot p_i \cdot b_{ij} \cdot \mathbb{E}_i[u_i \cdot \hat{x}_j(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)] = \sum_{i=1}^n T \cdot p_i \cdot b_{ij} \cdot \mathbb{E}_i[u_i \cdot x_{ij}^*(u_i)] = \sum_{i=1}^n c_{ij}^* \leq c_j$$

Thus, $\{\hat{x}_j(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)\}$ forms a feasible solution of (5) and we have

$$\mathrm{LP}^{\mathrm{UB}} \geq \sum_{i=1}^n T \cdot p_i \cdot \mathbb{E}_i[\sum_{j=1}^m r_{ij} \cdot \hat{x}_j(\mathbf{r}_i, \mathbf{b}_i \cdot u_i)] = \sum_{i=1}^n T \cdot p_i \cdot \mathbb{E}_i[\sum_{j=1}^m r_{ij} \cdot x_{ij}^*(u_i)] = \sum_{i=1}^n G_i(\mathbf{c}_i^*, 1) = \text{the RHS of (10)}$$

which completes our proof.

B. Proof of Lemma 3

We first show that $G_i(\tilde{\mathbf{c}}_i, t) \leq (T - t + 1) \cdot p_i \cdot \sum_{j=1}^{m} (r_{ii_j} - r_{ii_{j+1}}) \cdot F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))$. Suppose $\{x_{ii_j}^*(u_i)\}$ is the optimal solution of (9) and we define that $\hat{y}_{ii_j}(u_i) = \sum_{k=1}^{j} x_{ii_k}^*(u_i)$ for each j, then it follows that

$$0 \leq \hat{y}_{ii_j}(u_i) \leq 1 \quad \text{and} \quad (T - t + 1) \cdot p_i \cdot \mathbb{E}_i[u_i \cdot \hat{y}_{ii_j}(u_i)] \leq \sum_{k=1}^{j} \tilde{c}_{ii_k} / b_{ii_k}$$

Then we define $\hat{z}_{i_j} = \mathbb{E}_i[\hat{y}_{ii_j}(u_i)]$ for each j. It is obvious that $0 \leq \hat{z}_{i_j} \leq 1$ and we also have the following constraint regarding $\hat{z}_{i_j}$

$$(T - t + 1) \cdot p_i \cdot \int_0^{F_i^{-1}(\hat{z}_{i_j})} u_i dF_i(u_i) \leq (T - t + 1) \cdot p_i \cdot \mathbb{E}_i[u_i \cdot \hat{y}_{ii_j}(u_i)] \leq \sum_{k=1}^{j} \tilde{c}_{ii_k} / b_{ii_k} \qquad (37)$$

The first inequality holds by noting that $\hat{y}_{ii_j}(u_i) \in [0, 1]$, then the expectation $\mathbb{E}_i[u_i \cdot \hat{y}_{ii_j}(u_i)]$ must be lower bounded by integrating u_i from the smallest realization 0 to a realization a such that $P(\{u_i \in [0, a]\}) = \mathbb{E}_i[\hat{y}_{ii_j}(u_i)] = \hat{z}_{i_j}$. Also note that

$$G_i(\tilde{\mathbf{c}}_i, t) = (T - t + 1) \cdot p_i \cdot \sum_{j=1}^{m} r_{ii_j} \cdot \mathbb{E}_i[x_{ii_j}^*(u_i)] = (T - t + 1) \cdot p_i \cdot \sum_{j=1}^{m} (r_{ii_j} - r_{ii_{j+1}}) \cdot \mathbb{E}_i[\hat{y}_{ii_j}(u_i)]$$

$$= (T - t + 1) \cdot p_i \cdot \sum_{j=1}^{m} (r_{ii_j} - r_{ii_{j+1}}) \cdot \hat{z}_{i_j}$$

Thus, we have

$$
\begin{aligned}
G_i(\tilde{\mathbf{c}}_i, t) \leq \max \quad & (T - t + 1) \cdot p_i \cdot \sum_{j=1}^{m} (r_{ii_j} - r_{ii_{j+1}}) \cdot z_{i_j} \\
\text{s.t.} \quad & (T - t + 1) \cdot p_i \cdot \int_0^{F_i^{-1}(z_{i_j})} w dF_i(w) \leq \sum_{k=1}^{j} \tilde{c}_{ii_k} / b_{ii_k} \quad \forall j \\
& 0 \leq z_{i_j} \leq 1 \quad \forall j
\end{aligned}
\qquad (38)
$$

Note that each constraint in the RHS of (38) only involves a single z_{i_j} and the objective function is linear in z_{i_j}, thus, the optimization problem is separable over $\{z_{i_j}\}$ and for each j, the optimal solution $z_{i_j}^*$ is given by solving the following single-variable optimization problem:

$$
\begin{aligned}
\max \quad & (T - t + 1) \cdot p_i \cdot (r_{ii_j} - r_{ii_{j+1}}) \cdot z_{i_j} \\
\text{s.t.} \quad & (T - t + 1) \cdot p_i \cdot \int_0^{F_i^{-1}(z_{i_j})} w dF_i(w) \leq \sum_{k=1}^{j} \tilde{c}_{ii_k} / b_{ii_k} \\
& 0 \leq z_{i_j} \leq 1
\end{aligned}
$$

Since $r_{ii_j} - r_{ii_{j+1}} \geq 0$, obviously the optimal solution $z_{i_j}^*$ satisfies

$$(T - t + 1) \cdot p_i \cdot \int_0^{F_i^{-1}(z_{i_j}^*)} w dF_i(w) = \sum_{k=1}^{j} c_{ii_k} / b_{ii_k}, \quad \text{if } (T - t + 1) \cdot p_i \cdot \int_0^{\infty} w dF_i(w) \geq \sum_{k=1}^{j} \tilde{c}_{ii_k} / b_{ii_k}$$

and

$$z_{i_j}^* = 1, \quad \text{if } (T - t + 1) \cdot p_i \cdot \int_0^\infty w dF_i(w) < \sum_{k=1}^j \tilde{c}_{ii_k}/b_{ii_k}$$

Thus, we have $\mu_t^{ii_j}(\tilde{\mathbf{c}}_i) = F_i^{-1}(z_{i_j}^*)$ and it follows that $G_i(\tilde{\mathbf{c}}_i, t) \leq (T - t + 1) \cdot p_i \cdot \sum_{j=1}^m (r_{ii_j} - r_{ii_{j+1}}) \cdot F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))$. Note that

$$(T - t + 1) \cdot p_i \cdot \mathbb{E}_i[\sum_{j=1}^m r_{ij} \cdot \hat{x}_{ij}(u)] = (T - t + 1) \cdot p_i \cdot \sum_{j=1}^m (r_{ii_j} - r_{ii_{j+1}}) \cdot F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))$$

It remains to show that $\{\hat{x}_{ii_j}(u_i)\}$ is feasible to (9). Obviously for each u, we have that $\sum_{j=1}^m \hat{x}_{ii_j}(u_i) \leq 1$ and for each j, we have that

$$(T - t + 1) \cdot p_i \cdot \mathbb{E}_i[u_i \cdot \hat{x}_{ii_j}(u_i)] = (T - t + 1) \cdot p_i \cdot \int_{\mu_t^{ii_{j-1}}(\tilde{\mathbf{c}}_i)}^{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} u_i dF_i(u_i) \leq \tilde{c}_{ii_j}/b_{ii_j}$$

which completes our proof.

C. Proof of Lemma 4

We will first show that under the condition, there exists two constants $M_1, M_2 > 0$ such that

$$M_1 \leq \frac{wF(w)}{\int_0^w u dF(u)} \leq M_2 \quad \text{for all } w \in (0, \bar{\omega}) \tag{39}$$

The second inequality has already been established in Lemma 4 in Arlotto and Xie (2020) and the proof is replicated as follows for completeness:

$$0 < 1 - \gamma_1 \leq 1 - \frac{F(\lambda w)}{F(w)} = \int_{\lambda w}^w \frac{dF(u)}{F(w)} \quad \text{for all } w \in (0, \bar{\omega})$$

$$\Rightarrow \quad \lambda w(1 - \gamma_1) \leq \int_{\lambda w}^w \frac{\lambda w}{F(w)} dF(u) \leq \int_0^w \frac{u}{F(w)} dF(u)$$

$$\Rightarrow \quad \frac{wF(w)}{\int_0^w u dF(u)} \leq \frac{1}{\lambda(1 - \gamma_1)} = M_2 \quad \text{for all } w \in (0, \bar{\omega})$$

Thus, we only focus on proving the first inequality. We have that for any $w \in (0, \bar{\omega})$ and any positive integer k,

$$1 - \frac{F(\lambda^k w)}{F(\lambda^{k-1} w)} \leq 1 - \frac{1}{\gamma_2} \Rightarrow \int_{\lambda^k w}^{\lambda^{k-1} w} \frac{dF(u)}{F(\lambda^{k-1} w)} \leq 1 - \frac{1}{\gamma_2}$$

then we have

$$\int_{\lambda^k w}^{\lambda^{k-1} w} \frac{u dF(u)}{F(\lambda^{k-1} w)} \leq \int_{\lambda^k w}^{\lambda^{k-1} w} \frac{\lambda^{k-1} w dF(u)}{F(\lambda^{k-1} w)} \leq (1 - \frac{1}{\gamma_2}) \cdot \lambda^{k-1} w$$

also, notice that $\frac{1}{F(\lambda^{k-1} w)} \geq \gamma_1^{k-1} \cdot \frac{1}{F(w)}$, we have

$$\int_{\lambda^k w}^{\lambda^{k-1} w} \frac{u dF(u)}{F(w)} \leq (1 - \frac{1}{\gamma_2}) \cdot (\frac{\lambda}{\gamma_1})^{k-1} w$$

Thus, it holds that

$$\int_0^w \frac{u dF(u)}{F(w)} = \sum_{k=1}^{+\infty} \int_{\lambda^k w}^{\lambda^{k-1} w} \frac{u dF(u)}{F(w)} \le \sum_{k=1}^{+\infty} (1 - \frac{1}{\gamma_2}) \cdot (\frac{\lambda}{\gamma_1})^{k-1} w = \frac{w\gamma_1(\gamma_2 - 1)}{\gamma_1 \gamma_2 - \lambda \gamma_2}$$

which implies

$$\frac{wF(w)}{\int_0^w u dF(u)} \ge \frac{\gamma_1 \gamma_2 - \lambda \gamma_2}{\gamma_1(\gamma_2 - 1)}$$

and completes our proof of (39). Moreover, We could also find a positive integer l such that $\frac{M_1}{\lambda^{l-1}} \le M_2 \le \frac{M_1}{\lambda^l}$, then for any $0 < w_1 < w_2 < \bar{\omega}$, we consider two situations:

(i). If $\lambda^l \cdot w_2 \le w_1 < w_2$, we have

$$\frac{w_2(F(w_2) - F(w_1))}{\int_{w_1}^{w_2} u dF(u)} \le \frac{w_2(F(w_2) - F(w_1))}{\int_{w_1}^{w_2} w_1 dF(u)} = \frac{w_2}{w_1} \le 1/\lambda^l$$

(ii). If $w_1 < \lambda^l \cdot w_2$, we have

$$w_2 \cdot F(w_1) \ge \frac{1}{\lambda^l} \cdot w_1 \cdot F(w_1) \ge \frac{M_1}{\lambda^l} \cdot \int_0^{w_1} u dF(u)$$

Thus, we have that

$$w_2(F(w_2) - F(w_1)) \le M_2 \cdot \int_0^{w_2} u dF(u) - \frac{M_1}{\lambda^l} \cdot \int_0^{w_1} u dF(u) = M_2 \cdot \int_{w_1}^{w_2} uF(u) + (M_2 - \frac{M_1}{\lambda^l}) \cdot \int_0^{w_1} u dF(u)$$

$$\le M_2 \cdot \int_{w_1}^{w_2} u dF(u)$$

Our proof is completed by setting $M = \max\{\frac{1}{\lambda^l}, M_2\}$.

D. Proof of Lemma 5

Since $F_i(\cdot)$ has a continuous density function $f_i(\cdot)$, implicit function theorem implies that the threshold $\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)$ is differentiable over $\tilde{c}_{ii_k}$ and the derivatives can be obtained by taking derivative over both sides of (11). We have that

$$\frac{\partial \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}{\partial \tilde{c}_{ii_k}} = \frac{1}{p_i \cdot b_{ii_k} \cdot (T - t + 1) \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i) \cdot f_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))} \tag{40}$$

Thus, we have that

$$\frac{\partial}{\partial \tilde{c}_{ii_k}}\Big(\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}\Big) = -\frac{1}{(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))^2} \cdot \frac{\partial \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}{\partial \tilde{c}_{ii_k}} = -\frac{1}{p_i \cdot b_{ii_k} \cdot (T - t + 1) \cdot (\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))^3 \cdot f_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))}$$

From Condition 2 in Assumption 2, $(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))^3 \cdot f_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))$ is a non-decreasing function over $\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)$. Also, noting that $\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)$ is non-decreasing over $\tilde{c}_{ii_k}$, we have $(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))^3 \cdot f_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))$ is a non-decreasing function over $\tilde{c}_{ii_k}$. Thus, we have that $\frac{\partial}{\partial \tilde{c}_{ii_k}}\Big(\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}\Big)$ is a non-decreasing function over $\tilde{c}_{ii_k}$, which implies the function $\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}$ is a convex function over $\tilde{c}_{ii_k}$.

E. Proof of Lemma 6

First, note that

$$\frac{\partial F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i))}{\partial \tilde{c}_{ii_k}} = f_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) \cdot \frac{\partial \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}{\partial \tilde{c}_{ii_k}} = \frac{1}{p_i \cdot b_{ii_k} \cdot (T-t+1) \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} \tag{41}$$

which implies

$$p_i \cdot (T-t+1) \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} \cdot u \cdot \mathbf{e}_{i_k}))) = \int_{\tilde{c}_{ii_k}/b_{ii_k}-u}^{\tilde{c}_{ii_k}/b_{ii_k}} \frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - (\tilde{c}_{ii_k} - b_{ii_k} \cdot v) \cdot \mathbf{e}_{i_k})} dv \tag{42}$$

From Lemma 5, the function $\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - (\tilde{c}_{ii_k} - b_{ii_k} \cdot v) \cdot \mathbf{e}_{i_k})}$ is a convex function over v, which implies that

$$\int_{\tilde{c}_{ii_k}/b_{ii_k}-u}^{\tilde{c}_{ii_k}/b_{ii_k}} \frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - (\tilde{c}_{ii_k} - b_{ii_k} \cdot v) \cdot \mathbf{e}_{i_k})} dv \leq \frac{u}{2} \cdot \left[\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} + \frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} \cdot u \cdot \mathbf{e}_{i_k})} \right] \tag{43}$$

Thus, we have that

$$p_i \cdot (T-t+1) \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - u \cdot \mathbf{e}_{i_k}))) \leq \frac{u}{2} \cdot \left[\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} + \frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} \cdot u \cdot \mathbf{e}_{i_k})} \right] \tag{44}$$

We then define a function $g(u)$ as follows:

$$g(u) := \frac{p_i \cdot (T-t+1)}{u^2} \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}})) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}} - b_{ii_k} \cdot u \cdot \mathbf{e}_{i_k}))) - \frac{1}{u \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} \tag{45}$$

Then the derivative of $g(u)$ over u could be obtained as follows:

$$g'(u) = -\frac{2p_i(T-t+1)}{u^3} \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k}))) + \frac{1}{u^2} \cdot \left[\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} + \frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k})} \right]$$

$$= -\frac{2}{u^3} \left\{ p_i(T-t+1)(F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k}))) - \frac{u}{2} \cdot \left[\frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} + \frac{1}{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k})} \right] \right\} \tag{46}$$

Thus, we have that $g'(u) \geq 0$, which implies that $g(u) \leq g(\tilde{c}_{ii_k}/b_{ii_k})$. Then we have

$$g(u) = \frac{p_i(T-t+1)}{u^2} \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k}))) - \frac{1}{u \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}$$

$$\leq g(\tilde{c}_{ii_k}/b_{ii_k}) = \frac{p_i(T-t+1)}{(\tilde{c}_{ii_k}/b_{ii_k})^2} \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - \tilde{c}_{ii_k} \cdot \mathbf{e}_{i_k}))) - \frac{1}{\tilde{c}_{ii_k}/b_{ii_k} \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}$$

Thus, we have that

$$F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k})) \leq \frac{u^2}{(\tilde{c}_{ii_k}/b_{ii_k})^2} \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - \tilde{c}_{ii_k} \cdot \mathbf{e}_{i_k})))$$

$$+ \frac{u}{p_i \cdot (T-t+1) \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} - \frac{u^2}{\tilde{c}_{ii_k}/b_{ii_k} \cdot p_i \cdot (T-t+1) \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}$$

Moreover, due to Lemma 4, we have that

$$\frac{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i) \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - \tilde{c}_{ii_k} \cdot \mathbf{e}_{i_k})))}{\tilde{c}_{ii_k}} = \frac{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i) \cdot (F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - \tilde{c}_{ii_k} \cdot \mathbf{e}_{i_k})))}{p_i b_{ii_k}(T - t + 1) \cdot \int_{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - \tilde{c}_{ii_k} \cdot \mathbf{e}_{i_k})}^{\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} w dF_i(w)}$$

$$\leq \frac{M}{p_i b_{ii_k}(T - t + 1)}$$

Thus, we have that

$$F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i)) - F_i(\mu_t^{ii_j}(\tilde{\mathbf{c}}_i - b_{ii_k} u \cdot \mathbf{e}_{i_k})) \leq \frac{(M - 1) \cdot u^2}{\tilde{c}_{ii_k}/b_{ii_k} \cdot p_i \cdot (T - t + 1) \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)} + \frac{u}{p_i \cdot (T - t + 1) \cdot \mu_t^{ii_j}(\tilde{\mathbf{c}}_i)}$$

which completes our proof.

F. Proof of Lemma 7

Since we focus on a fixed i during the proof, we will omit the index i by using index j to denote index i_j, using $\mu_t^j(\mathbf{c}^t)$ to denote $\mu_t^{ii_j}(\mathbf{c}_i^t)$ and using b_j to denote b_{ii_j}.

We only focus on the time period t such that $t < T - K$. The reason is that for any $t < T - K$, we must have for each j and every possible remaining capacity $\mathbf{c}^t$, the threshold $\mu_t^j(\mathbf{c}^t)$ will satisfy $\mu_t^j(\mathbf{c}^t) < \bar{\omega}$. Thus, by only focusing on $t < T - K$, the value of $\mu_t^j(\mathbf{c}^t)$ is always finite and we could fully exploit the two conditions over the distribution function proposed in Assumption 2. Since K is a constant independent of T, the cost of ignoring the last K time periods is at most a constant regret, which must be upper bounded by $r_{\max} \cdot K$. In what follows, we assume $t < T - K$. Define

$$\rho_t(\mathbf{c}^t) = (T - t + 1) \cdot p \cdot \sum_{j=1}^{m} (r_j - r_{j+1}) \cdot F(\mu_t^j(\mathbf{c}^t)) - V_t(\mathbf{c}^t) \qquad (47)$$

where $V_t(\mathbf{c}^t)$ is the "to-go" expected total reward collected by the policy, i.e., the expected cumulative reward collected by the policy from serving requests with reward $\mathbf{r}_i$ from time period t to the last time period T given the remaining capacity at time period t is $\mathbf{c}^t$. Then our final goal is to bound $\rho_1(\mathbf{c}^*)$. Given the remaining capacity $\mathbf{c}^t$ and the current time period t, denote $\{j_1, j_2, \ldots, j_W\}$ as a subset of $\{1, 2, \ldots, m\}$ where $j_1 = 1$ and $\{j_1, j_2, \ldots, j_W\}$ satisfies that

$$h_t^1(\mathbf{c}^t) = h_t^{j_1}(\mathbf{c}^t) \leq h_t^{j_2}(\mathbf{c}^t) \leq \cdots \leq h_t^{j_W}(\mathbf{c}^t) = \max\{h_t^1(\mathbf{c}^t), h_t^2(\mathbf{c}^t), \ldots, h_t^m(\mathbf{c}^t)\}$$

and for each $1 \leq w \leq W$ and every j such that $j_w < j \leq j_{w+1} - 1$, we must have $h_t^j(\mathbf{c}^t) < h_t^{j_w}(\mathbf{c}^t)$, where we set $j_{W+1} = m + 1$. Note that the definition of $\{j_1, j_2, \ldots, j_W\}$ will enable us to obtain the following relationship.

Claim 1 *For any $1 \leq w \leq W$, we must have*

$$\mu_t^{j_w}(\mathbf{c}^t) \geq c_j^t / b_j \quad \text{for every } j \text{ such that } j_w < j \leq j_{w+1} - 1 \qquad (48)$$

Proof: we have that $\mu_t^{j_w}(\mathbf{c}^t) \geq h_t^{j_w}(\mathbf{c}^t) > h_t^j(\mathbf{c}^t)$ for any j such that $j_w < j \leq j_{w+1} - 1$. Also, it is obvious that $\mu_t^{j_w}(\mathbf{c}^t) \leq \mu_t^j(\mathbf{c}^t)$ since $j_w < j$. Note that $h_t^j(\mathbf{c}^t) = \min\{c_j^t/b_j, \mu_t^j(\mathbf{c}^t)\}$, it must hold that $\mu_t^{j_w}(\mathbf{c}^t) \geq c_j^t/b_j$. $\square$

Moreover, according to our policy in Algorithm 1, for each $1 \leq w \leq W$, request t with size $\mathbf{b} \cdot u$ will be served by resource j_w if and only if $h_t^{j_{w-1}}(\mathbf{c}^t) < u \leq h_t^{j_w}(\mathbf{c}^t)$, where we set $j_0 = 0$ and $h_t^{j_0}(\mathbf{c}^t) = 0$, and request t will be discarded if and only if $u > h_t^{j_W}(\mathbf{c}^t)$. Then, we can write down recursion of $V_t(\mathbf{c}^t)$ as follows:

$$V_t(\mathbf{c}^t) = (1-p)V_{t+1}(\mathbf{c}^t) + p \cdot \sum_{w=1}^{W} \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \{r_{j_w} + V_{t+1}(\mathbf{c}^t - b_{j_w}u \cdot \mathbf{e}_{j_w})\}dF(u) + p \cdot (1 - F(h_t^{j_W}(\mathbf{c}^t))) \cdot V_{t+1}(\mathbf{c}^t)$$

$$= (1 - p \cdot F(h_t^{j_W}(\mathbf{c}^t))) \cdot V_{t+1}(\mathbf{c}^t) + p \cdot \sum_{w=1}^{W} \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \{r_{j_w} + V_{t+1}(\mathbf{c}^t - b_{j_w}u \cdot \mathbf{e}_{j_w})\}dF(u)$$

(49)

where $\mathbf{e}_{j_w}$ denotes a m dimensional vector with 1 at j_wth component and 0 at other components. We could substitute $V_t(\cdot)$ by $(T-t+1)p \cdot \sum_{j=1}^{m}(r_j - r_{j+1}) \cdot F(\mu_t^j(\cdot)) - \rho_t(\cdot)$ and substitute $V_{t+1}(\cdot)$ by $(T-t)p \cdot \sum_{j=1}^{m}(r_j - r_{j+1}) \cdot F(\mu_{t+1}^j(\cdot)) - \rho_{t+1}(\cdot)$ in the above recursion. Then we will get the recursion of $\rho_t(\mathbf{c}^t)$ as follows:

$$(T-t+1)p\sum_{j=1}^{m}(r_j - r_{j+1})F(\mu_t^j(\mathbf{c}^t)) - \rho_t(\mathbf{c}^t) = (1 - pF(h_t^{j_W}(\mathbf{c}^t)))\{(T-t)p\sum_{j=1}^{m}(r_j - r_{j+1})F(\mu_{t+1}^j(\mathbf{c}^t)) - \rho_{t+1}(\mathbf{c}^t)\}$$

$$+ p\sum_{w=1}^{W} \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \{r_{j_w} + (T-t)p\sum_{k=1}^{m}(r_k - r_{k+1})F(\mu_{t+1}^k(\mathbf{c}^t - b_{j_w}u \cdot \mathbf{e}_{j_w})) - \rho_{t+1}(\mathbf{c}^t - b_{j_w}u \cdot \mathbf{e}_{j_w})\}dF(u)$$

(50)

which implies that

$$\rho_t(\mathbf{c}^t) = (1 - pF(h_t^{j_W}(\mathbf{c}^t))) \cdot \rho_{t+1}(\mathbf{c}^t) + p\sum_{w=1}^{W} \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \rho_{t+1}(\mathbf{c}^t - b_{j_w}u \cdot \mathbf{e}_{j_w})dF(u)$$

$$\underbrace{+ p \cdot \sum_{j=1}^{m}(r_j - r_{j+1})F(\mu_t^j(\mathbf{c}^t)) - p \cdot \sum_{w=1}^{W} r_{j_w} \cdot (F(h_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_{w-1}}(\mathbf{c}^t)))}_{\text{I}}$$

$$\underbrace{+ (T-t)p \cdot \sum_{j=1}^{m}(r_j - r_{j+1}) \cdot (F(\mu_t^j(\mathbf{c}^t)) - F(\mu_{t+1}^j(\mathbf{c}^t)))}_{\text{II}}$$

(51)

$$\underbrace{+ p \cdot \sum_{w=1}^{W} \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \{(T-t)p\sum_{k=1}^{m}(r_k - r_{k+1}) \cdot (F(\mu_{t+1}^k(\mathbf{c}^t)) - F(\mu_{t+1}^k(\mathbf{c}^t - b_{j_w}u \cdot \mathbf{e}_{j_w})))\}dF(u)}_{\text{III}}$$

Denote $\bar{\rho}_{t+1}$ as the maximal value of $\rho_{t+1}(\mathbf{c}^{t+1})$ for all the possible remaining capacity $\mathbf{c}^{t+1}$. Then we have

$$\rho_t(\mathbf{c}^t) - \bar{\rho}_{t+1} \leq \text{I} + \text{II} + \text{III}$$

In the following, we will bound the three terms I, II, III separately.

Bound I: By rearranging terms, we get

$$
\begin{aligned}
\mathrm{I} &= p \cdot \sum_{j=1}^{m} (r_j - r_{j+1}) F(\mu_t^j(\mathbf{c}^t)) - p \cdot \sum_{w=1}^{W} r_{j_w} \cdot (F(h_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_{w-1}}(\mathbf{c}^t))) \\
&= p \cdot \sum_{w=1}^{W} \{ \sum_{j=j_w}^{j_{w+1}-1} (r_j - r_{j+1}) \cdot [F(\mu_t^j(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))] \}
\end{aligned}
\tag{52}
$$

Then, for each fixed j such that $j_w \le j \le j_{w+1} - 1$, we will bound the term $F(\mu_t^j(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))$. We will first prove the following bound.

Claim 2 *For each $1 \le w \le W$ and each j such that $j_w \le j \le j_{w+1} - 1$, we must have*

$$
F(\mu_t^j(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t)) \le \frac{M \cdot (j - j_w)}{p \cdot (T - t + 1)} + [F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t / b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}}
\tag{53}
$$

Proof: note that from Lemma 4, we have that

$$
\mu_t^j(\mathbf{c}^t) \cdot (F(\mu_t^j(\mathbf{c}^t)) - F(\mu_t^{j_w}(\mathbf{c}^t))) \le M \cdot \int_{\mu_t^{j_w}(\mathbf{c}^t)}^{\mu_t^j(\mathbf{c}^t)} u \, dF(u) = \frac{M \cdot (c_{j_w+1}^t / b_{j_w+1} + \cdots + c_j^t / b_j)}{p \cdot (T - t + 1)}
\tag{54}
$$

Thus, we have

$$
\begin{aligned}
F(\mu_t^j(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t)) &= F(\mu_t^j(\mathbf{c}^t)) - F(\mu_t^{j_w}(\mathbf{c}^t)) + F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t)) \\
&\le \frac{M \cdot (c_{j_w+1}^t / b_{j_w+1} + \cdots + c_j^t / b_j)}{\mu_t^j(\mathbf{c}^t) \cdot p \cdot (T - t + 1)} + F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t)) \\
&\le \frac{M \cdot (j - j_w)}{p \cdot (T - t + 1)} + F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))
\end{aligned}
\tag{55}
$$

where the first inequality holds due to (54) and the second inequality holds by noting that according to (48), we must have

$$
\mu_t^j(\mathbf{c}^t) \ge \mu_t^{j_w}(\mathbf{c}^t) \ge \max\{c_{j_w+1}^t / b_{j_w+1}, \dots, c_j^t / b_j\}
$$

Note that $h_t^{j_w}(\mathbf{c}^t) = \min\{c_{j_w}^t / b_{j_w}, \mu_t^{j_w}(\mathbf{c}^t)\}$, then $h_t^{j_w}(\mathbf{c}^t)$ differentiates from $\mu_t^{j_w}(\mathbf{c}^t)$ if and only if $c_{j_w}^t / b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)$, thus we have

$$
F(\mu_t^j(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t)) \le \frac{M \cdot (j - j_w)}{p \cdot (T - t + 1)} + [F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t / b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}}
$$

□

We then bound the term $[F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t / b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}}$.

Claim 3 *For any $1 \le w \le W$, it holds that*

$$
[F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t / b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}} \le \frac{M \cdot j_w}{p \cdot (T - t + 1)}
\tag{56}
$$

Proof: we will prove (56) by induction. For $j_1 = 1$, we have that

$$[F(\mu_t^{j_1}(\mathbf{c}^t)) - F(h_t^{j_1}(\mathbf{c}^t))] \cdot 1_{\{c_{j_1}^t/b_{jw} < \mu_t^{j_1}(\mathbf{c}^t)\}} \leq F(\mu_t^{j_1}(\mathbf{c}^t)) \cdot 1_{\{c_{j_1}^t/b_{j_1} < \mu_t^{j_1}(\mathbf{c}^t)\}}$$

$$\leq \frac{M \cdot c_{j_1}^t/b_{j_1}}{\mu_t^{j_1}(\mathbf{c}^t) \cdot p \cdot (T-t+1)} \cdot 1_{\{c_{j_1}^t/b_{j_1} < \mu_t^{j_1}(\mathbf{c}^t)\}}$$

$$\leq \frac{M}{p \cdot (T-t+1)}$$

where the second inequality holds due to Lemma 4. Now suppose that for $w-1$, we have

$$[F(\mu_t^{j_{w-1}}(\mathbf{c}^t)) - F(h_t^{j_{w-1}}(\mathbf{c}^t))] \cdot 1_{\{c_{j_{w-1}}^t/b_{j_{w-1}} < \mu_t^{j_{w-1}}(\mathbf{c}^t)\}} \leq \frac{M \cdot j_{w-1}}{p \cdot (T-t+1)}$$

Then note that $h_t^{j_{w-1}}(\mathbf{c}^t) \leq h_t^{j_w}(\mathbf{c}^t)$, we have

$$[F(\mu_t^{j_w}(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}}$$

$$\leq [F(\mu_t^{j_w}(\mathbf{c}^t)) - F(\mu_t^{j_{w-1}}(\mathbf{c}^t)) + F(\mu_t^{j_{w-1}}(\mathbf{c}^t)) - F(h_t^{j_{w-1}}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}}$$

$$\leq [F(\mu_t^{j_w}(\mathbf{c}^t)) - F(\mu_t^{j_{w-1}}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}} + F(\mu_t^{j_{w-1}}(\mathbf{c}^t)) - F(h_t^{j_{w-1}}(\mathbf{c}^t))$$

$$= [F(\mu_t^{j_w}(\mathbf{c}^t)) - F(\mu_t^{j_{w-1}}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}} + [F(\mu_t^{j_{w-1}}(\mathbf{c}^t)) - F(h_t^{j_{w-1}}(\mathbf{c}^t))] \cdot 1_{\{c_{j_{w-1}}^t/b_{j_{w-1}} < \mu_t^{j_{w-1}}(\mathbf{c}^t)\}}$$

$$\leq [F(\mu_t^{j_w}(\mathbf{c}^t)) - F(\mu_t^{j_{w-1}}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}} + \frac{M \cdot j_{w-1}}{p \cdot (T-t+1)}$$

Note that from Lemma 4, it holds that

$$[F(\mu_t^{j_w}(\mathbf{c}^t)) - F(\mu_t^{j_{w-1}}(\mathbf{c}^t))] \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}} \leq \frac{M \cdot (c_{j_{w-1}+1}^t/b_{j_{w-1}+1} + \cdots + c_{j_w}^t/b_{j_w})}{\mu_t^{j_w}(\mathbf{c}^t) \cdot p \cdot (T-t+1)} \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}}$$

$$\leq \frac{M \cdot (c_{j_{w-1}+1}^t/b_{j_{w-1}+1} + \cdots + c_{j_w-1}^t/b_{j_w-1})}{\mu_t^{j_w}(\mathbf{c}^t) \cdot p \cdot (T-t+1)} + \frac{M \cdot c_{j_w}^t/b_{j_w}}{\mu_t^{j_w}(\mathbf{c}^t) \cdot p \cdot (T-t+1)} \cdot 1_{\{c_{j_w}^t/b_{j_w} < \mu_t^{j_w}(\mathbf{c}^t)\}}$$

$$\leq \frac{M \cdot (c_{j_{w-1}+1}^t/b_{j_{w-1}+1} + \cdots + c_{j_w-1}^t/b_{j_w-1})}{\mu_t^{j_w}(\mathbf{c}^t) \cdot p \cdot (T-t+1)} + \frac{M}{p \cdot (T-t+1)}$$

$$\leq \frac{M \cdot (j_w - 1 - j_{w-1})}{p \cdot (T-t+1)} + \frac{M}{p \cdot (T-t+1)} = \frac{M \cdot (j_w - j_{w-1})}{p \cdot (T-t+1)}$$

where the last inequality holds by noting that according to (48), we must have

$$\mu_t^{j_w}(\mathbf{c}^t) \geq \mu_t^{j_{w-1}}(\mathbf{c}^t) \geq \max\{c_{j_{w-1}+1}^t/b_{j_{w-1}+1}, \ldots, c_{j_w-1}^t/b_{j_w-1}\}$$

Thus, by induction, for any $1 \leq w \leq W$, (56) holds. $\square$

Thus from (53) and (56), we have that

$$F(\mu_t^j(\mathbf{c}^t)) - F(h_t^{j_w}(\mathbf{c}^t)) \leq \frac{M \cdot j}{p \cdot (T-t+1)} \tag{57}$$

Then by substituting (57) into (52), we can finally upper bound the term I by:

$$\text{I} \leq \frac{M}{T-t+1} \cdot \sum_{j=1}^{m}(r_j - r_{j+1}) \cdot j = \frac{M}{T-t+1} \cdot \sum_{j=1}^{m} r_j \tag{58}$$

We then proceed to bound the term II.

Bound II: From the definition of the threshold $\mu_t^j(\mathbf{c}^t)$ in (11), for each $j = 1, 2, \ldots, m$, we have that

$$\frac{c_1^t/b_1 + \cdots + c_j^t/b_j}{T-t} - \frac{c_1^t/b_1 + \cdots + c_j^t/b_j}{T-t+1} = p \cdot \int_{\mu_t^j(\mathbf{c}^t)}^{\mu_{t+1}^j(\mathbf{c}^t)} u\, dF(u) \le p \cdot \mu_{t+1}^j(\mathbf{c}^t) \cdot [F(\mu_{t+1}^j(\mathbf{c}^t)) - F(\mu_t^j(\mathbf{c}^t))]$$

which implies

$$(T-t) \cdot p \cdot [F(\mu_t^j(\mathbf{c}^t)) - F(\mu_{t+1}^j(\mathbf{c}^t))] \le -\frac{c_1^t/b_1 + \cdots + c_j^t/b_j}{\mu_{t+1}^j(\mathbf{c}^t) \cdot (T-t+1)}$$

Thus, we have the following upper bound regarding the term II:

$$\text{II} \le -\sum_{j=1}^m (r_j - r_{j+1}) \cdot \frac{c_1^t/b_1 + \cdots + c_j^t/b_j}{\mu_{t+1}^j(\mathbf{c}^t) \cdot (T-t+1)} \tag{59}$$

At last, it remains to bound the term III.

Bound III: We begin with an observation that from the definition of the threshold $\mu_t^j(\mathbf{c}^t)$ in (11), if $k < j_w$, then for any u, we must have $\mu_{t+1}^k(\mathbf{c}^t) = \mu_{t+1}^k(\mathbf{c}^t - b_{j_w} u \cdot \mathbf{e}_{j_w})$. Thus, we could rewrite the term III as:

$$\text{III} = p \cdot \sum_{w=1}^W \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \left\{ (T-t) \cdot p \cdot \sum_{k=j_w}^m (r_k - r_{k+1}) \cdot (F(\mu_{t+1}^k(\mathbf{c}^t)) - F(\mu_{t+1}^k(\mathbf{c}^t - b_{j_w} u \cdot \mathbf{e}_{j_w}))) \right\} dF(u) \tag{60}$$

Note that from Lemma 6, for any $k \ge j_w$, we have that

$$p(T-t) \cdot (F(\mu_{t+1}^k(\mathbf{c}^t)) - F(\mu_{t+1}^k(\mathbf{c}^t - b_{j_w} u \cdot \mathbf{e}_{j_w}))) \le \frac{(M-1)u^2}{(c_{j_w}^t/b_{j_w}) \cdot \mu_{t+1}^k(\mathbf{c}^t)} + \frac{u}{\mu_{t+1}^k(\mathbf{c}^t)}$$

Then, we have that

$$\text{III} \le \underbrace{p(M-1) \cdot \sum_{w=1}^W \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \left\{ \sum_{k=j_w}^m (r_k - r_{k+1}) \cdot \frac{u^2}{(c_{j_w}^t/b_{j_w}) \cdot \mu_{t+1}^k(\mathbf{c}^t)} \right\} dF(u)}_{\text{IV}}$$
$$+ \underbrace{p \cdot \sum_{w=1}^W \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \sum_{k=j_w}^m (r_k - r_{k+1}) \cdot \frac{u}{\mu_{t+1}^k(\mathbf{c}^t)} dF(u)}_{\text{V}} \tag{61}$$

We next bound the term IV and V separately. We have that

$$\text{IV} = p(M-1) \sum_{w=1}^W \sum_{k=j_w}^m \frac{r_k - r_{k+1}}{(c_{j_w}^t/b_{j_w}) \cdot \mu_{t+1}^k(\mathbf{c}^t)} \cdot \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u^2 dF(u)$$
$$\le p(M-1) \sum_{w=1}^W \sum_{k=j_w}^m \frac{r_k - r_{k+1}}{(c_{j_w}^t/b_{j_w}) \cdot \mu_{t+1}^k(\mathbf{c}^t)} \cdot h_t^{j_w}(\mathbf{c}^t) \cdot \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u\, dF(u) \tag{62}$$
$$\le p(M-1) \sum_{w=1}^W \sum_{k=j_w}^m \frac{r_k - r_{k+1}}{(c_{j_w}^t/b_{j_w})} \cdot \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u\, dF(u)$$

where the second inequality holds by noting that when $k \ge j_w$, we must have $h_t^{j_w}(\mathbf{c}^t) \le \mu_t^{j_w}(\mathbf{c}^t) \le \mu_t^k(\mathbf{c}^t) \le \mu_{t+1}^k(\mathbf{c}^t)$. We then bound the term $\frac{1}{c_{j_w}^t/b_{j_w}} \cdot \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u\, dF(u)$ in (62).

Claim 4 *For any $1 \le w \le W$, it holds that*

$$\frac{1}{c_{j_w}^t / b_{j_w}} \cdot \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u\, dF(u) \le \frac{j_w}{p(T-t+1)} \tag{63}$$

Proof: For each j_w, we define the index $j(w)$:

$$j(w) := \max\{j : 0 \le j < j_w \text{ and } c_j / b_j \ge \mu_t^j(\mathbf{c}^t)\} \tag{64}$$

where we will denote $c_0 = \mu_t^0(\mathbf{c}^t) = 0$ and such a $j(w)$ always exist. Since $j(w) < j_w$, from the definition of $\{j_1, j_2, \ldots, j_W\}$, it must hold that

$$\mu_t^{j(w)}(\mathbf{c}^t) = h_t^{j(w)}(\mathbf{c}^t) \le h_t^{j_{w-1}}(\mathbf{c}^t)$$

then we have

$$\frac{1}{c_{j_w}^t / b_{j_w}} \cdot \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u\, dF(u) \le \frac{1}{c_{j_w}^t / b_{j_w}} \cdot \int_{h_t^{j(w)}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u\, dF(u) \le \frac{1}{c_{j_w}^t / b_{j_w}} \cdot \int_{\mu_t^{j(w)}(\mathbf{c}^t)}^{\mu_t^{j_w}(\mathbf{c}^t)} u\, dF(u)$$

$$= \frac{1}{c_{j_w}^t / b_{j_w}} \cdot \frac{c_{j(w)+1}^t / b_{j(w)+1} + \cdots + c_{j_w}^t / b_{j_w}}{p(T-t+1)}$$

where the last equality holds due to the definition of $\mu_t^{j(w)}(\mathbf{c}^t)$ and $\mu_t^{j_w}(\mathbf{c}^t)$ in (9). Moreover, note that from the definition of the index $j(w)$, for any index q such that $j(w) < q < j_w$, it must hold $c_q^t / b_q < \mu_t^q(\mathbf{c}^t)$. Thus, we have that $c_q^t / b_q = h_t^q(\mathbf{c}^t) \le h_t^{j_w}(\mathbf{c}^t) \le c_{j_w}^t / b_{j_w}$, which implies that

$$\frac{c_{j(w)+1}^t / b_{j(w)+1} + \cdots + c_{j_w}^t / b_{j_w}}{c_{j_w}^t / b_{j_w}} \le j_w - j(w) \le j_w$$

Thus, the proof is completed. $\square$

By substituting (63) into (62), we have that

$$\text{IV} \le \frac{M-1}{T-t+1} \cdot \sum_{w=1}^{W} \sum_{k=j_w}^{m} (r_k - r_{k+1}) j_w = \frac{M-1}{T-t+1} \cdot \sum_{w=1}^{W} r_{j_w} \cdot j_w \le \frac{M-1}{T-t+1} \cdot \sum_{j=1}^{m} j \cdot r_j \tag{65}$$

At last, we bound the term V as follows:

$$\text{V} = p \cdot \sum_{w=1}^{W} \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} \sum_{k=j_w}^{m} (r_k - r_{k+1}) \cdot \frac{u}{\mu_{t+1}^k(\mathbf{c}^t)} dF(u) = p \cdot \sum_{k=1}^{m} \frac{r_k - r_{k+1}}{\mu_{t+1}^k(\mathbf{c}^t)} \cdot \sum_{j_w : j_w \le k} \int_{h_t^{j_{w-1}}(\mathbf{c}^t)}^{h_t^{j_w}(\mathbf{c}^t)} u\, dF(u)$$

$$\le p \cdot \sum_{k=1}^{m} \frac{r_k - r_{k+1}}{\mu_{t+1}^k(\mathbf{c}^t)} \cdot \int_0^{\mu_t^k(\mathbf{c}^t)} u\, dF(u) = \sum_{k=1}^{m} (r_k - r_{k+1}) \cdot \frac{c_1^t / b_1 + \cdots + c_k^t / b_k}{\mu_{t+1}^k(\mathbf{c}^t) \cdot (T-t+1)}$$

$$\tag{66}$$

where the first inequality holds by noting that for all $j_w \le k$, we have $h_t^{j_w}(\mathbf{c}^t) \le \mu_t^{j_w}(\mathbf{c}^t) \le \mu_t^k(\mathbf{c}^t)$.
Finally, we can bound term III by:

$$\text{III} \le \text{IV} + \text{V} \le \frac{M-1}{T-t+1} \cdot \sum_{j=1}^{m} j \cdot r_j + \sum_{k=1}^{m} (r_k - r_{k+1}) \cdot \frac{c_1^t / b_1 + \cdots + c_k^t / b_k}{\mu_{t+1}^k(\mathbf{c}^t) \cdot (T-t+1)} \tag{67}$$

By combining (58), (59) and (67), we have

$$
\begin{aligned}
\rho_t(\mathbf{c}^t) - \bar{\rho}_{t+1} &\leq \frac{M}{T-t+1} \cdot \sum_{j=1}^{m} r_j - \sum_{j=1}^{m}(r_j - r_{j+1}) \cdot \frac{c_1^t/b_1 + \cdots + c_j^t/b_j}{\mu_{t+1}^j(\mathbf{c}^t) \cdot (T-t+1)} \\
&\quad + \frac{M-1}{T-t+1} \cdot \sum_{j=1}^{m} j \cdot r_j + \sum_{k=1}^{m}(r_k - r_{k+1}) \cdot \frac{c_1^t/b_1 + \cdots + c_k^t/b_k}{\mu_{t+1}^k(\mathbf{c}^t) \cdot (T-t+1)} \\
&= \frac{1}{T-t+1} \cdot \sum_{j=1}^{m} r_j \cdot (M(j+1) - j)
\end{aligned}
\tag{68}
$$

holds for any $\mathbf{c}^t$, which implies that

$$
\bar{\rho}_t - \bar{\rho}_{t+1} \leq \frac{1}{T-t+1} \cdot \sum_{j=1}^{m} r_j \cdot (M(j+1) - j)
$$

Thus, by noting that $\bar{\rho}_{T-K+1} \leq r_{\max} \cdot K$, we have

$$
\begin{aligned}
\rho_1(\mathbf{c}^*) \leq \bar{\rho}_1 &\leq \sum_{t=1}^{T-K}(\bar{\rho}_t - \bar{\rho}_{t+1}) + \bar{\rho}_{T-K+1} \leq \sum_{t=1}^{T-K}\left(\frac{1}{T-t+1}\right) \cdot \sum_{j=1}^{m}(r_j \cdot (M(j+1) - j)) + r_{\max} \cdot K \\
&\leq (\log T + 1) \cdot \sum_{j=1}^{m}(r_j \cdot (M(j+1) - j)) + r_{\max} \cdot K
\end{aligned}
$$

which completes our proof.

G. Proof of Theorem 2

Since we have shown

$$
\mathbb{E}[V^{\mathrm{off}}(\mathbf{I})] \geq \mathbb{E}[V^{\mathrm{ATP}_1}(\mathbf{I})] \geq \sum_{i=1}^{n} G_i(\mathbf{c}_i^*, 1) - O(\log T)
$$

it is enough to show that $\sum_{i=1}^{n} G_i(\mathbf{c}_i^*, 1) = \Omega(T^{\frac{1}{1+\alpha}})$. Also, there must exists a i and i_j such that $p_i > 0$, $r_{ii_j} - r_{ii_{j+1}} > 0$ and $\sum_{k=1}^{j} c_{ii_k}^* \geq \frac{\sum_{j=1}^{m} c_j}{n}$. If we can show that

$$
T \cdot p_i \cdot (r_{ii_j} - r_{ii_{j+1}}) \cdot F_i(\mu_1^{ii_j}(\mathbf{c}_i^*)) = \Omega(T^{\frac{1}{1+\alpha}})
$$

then the proof is completed immediately. In what follows, we will use μ_T to denote the threshold $\mu_1^{ii_j}(\mathbf{c}_i^*)$ and denote l as the positive integer such that $\lambda^l \cdot \bar{\omega} \leq \mu_T \leq \lambda^{l-1} \cdot \bar{\omega}$. Then we have

$$
\frac{\sum_{k=1}^{j} c_{ii_k}^*}{\bar{b} \cdot p_i \cdot T} \geq \frac{\sum_{k=1}^{j} c_{ii_k}^*/b_{ii_k}}{p_i \cdot T} = \int_0^{\mu_T} u \, dF_i(u) \geq \sum_{k=l}^{+\infty} \int_{\lambda^{k+1}\bar{\omega}}^{\lambda^k\bar{\omega}} u \, dF_i(u)
$$

where $\bar{b} = \min_{1 \leq k \leq j}\{b_{ii_k}\}$. We also have that

$$
\int_{\lambda^{k+1}\bar{\omega}}^{\lambda^k\bar{\omega}} u \, dF_i(u) \geq \lambda^{k+1}\bar{\omega} \cdot [F_i(\lambda^k\bar{\omega}) - F_i(\lambda^{k+1}\bar{\omega})] \geq \lambda^{k+1}\bar{\omega}(\gamma_1 - 1) \cdot F_i(\lambda^{k+1}\bar{\omega}) \geq (\frac{\lambda}{\gamma_2})^{k+1}\bar{\omega}(\gamma_1 - 1)F_i(\bar{\omega})
$$

The last inequality holds by noting $F_i(\bar{\omega}) \le \gamma_2^{k+1} \cdot F_i(\lambda^{k+1}\bar{\omega})$. Thus, we have that

$$\frac{\sum_{k=1}^{j} c_{ii_k}^*}{\bar{b} \cdot p_i \cdot T} \ge \bar{\omega}(\gamma_1 - 1) F_i(\bar{\omega}) \cdot \sum_{k=l}^{+\infty} (\frac{\lambda}{\gamma_2})^{k+1} = \bar{\omega}(\gamma_1 - 1) F_i(\bar{\omega})(\frac{\lambda}{\gamma_2})^{l+1} \cdot \frac{\gamma_2}{\gamma_2 - \lambda}$$

Define a constant $Q = \frac{(\sum_{k=1}^{j} c_{ii_k}^*) \cdot (\gamma_2 - \lambda)}{\bar{b}\bar{\omega}(\gamma_1 - 1)\lambda p_i F_i(\bar{\omega})}$, we have that

$$(\frac{\lambda}{\gamma_2})^l \le \frac{Q}{T}$$

Also, since $0 < \lambda < 1$ and $0 < \frac{1}{\gamma_2} < 1$, there exists a constant $\alpha > 0$ such that $\frac{1}{\gamma_2} = \lambda^\alpha$. Thus, we have

$$\lambda^{(1+\alpha)l} \le \frac{Q}{T} \Rightarrow \lambda^l \le (\frac{Q}{T})^{\frac{1}{1+\alpha}}$$

Note that from (39), we have that

$$F_i(\mu_T) \ge \frac{M_1}{\mu_T} \cdot \int_0^{\mu_T} u dF_i(u) = \frac{M_1 \cdot (\sum_{k=1}^{j} c_{ii_k}^*/b_{ii_k})}{\mu_T \cdot p_i \cdot T} \ge \frac{M_1 \lambda (\sum_{k=1}^{j} c_{ii_k}^*)}{\hat{b} \cdot \bar{\omega} \cdot p_i \cdot T} \cdot \frac{1}{\lambda^l}$$

where $\hat{b} = \max_{1 \le k \le j}\{b_{ii_k}\}$. Thus, we have that

$$
\begin{aligned}
T \cdot p_i \cdot (r_{ii_j} - r_{ii_{j+1}}) \cdot F_i(\mu_T) &\ge \frac{M_1 \lambda (r_{ii_j} - r_{ii_{j+1}})(\sum_{k=1}^{j} c_{ii_k}^*)}{\hat{b} \cdot \bar{\omega}} \cdot \frac{1}{\lambda^l} \\
&\ge \frac{M_1 \lambda^{\frac{2+\alpha}{1+\alpha}} (r_{ii_j} - r_{ii_{j+1}})((\gamma_1 - 1)\bar{b}p_i F_i(\bar{\omega}))^{\frac{1}{1+\alpha}}}{\hat{b}\bar{\omega}^{\frac{\alpha}{1+\alpha}}(\gamma_2 - \lambda)^{\frac{1}{1+\alpha}}} \cdot (\sum_{k=1}^{j} c_{ii_k}^*)^{\frac{\alpha}{1+\alpha}} \cdot T^{\frac{1}{1+\alpha}} \\
&\ge \frac{M_1 \lambda^{\frac{2+\alpha}{1+\alpha}} (r_{ii_j} - r_{ii_{j+1}})((\gamma_1 - 1)\bar{b}p_i F_i(\bar{\omega}))^{\frac{1}{1+\alpha}} (\sum_{j=1}^{m} c_j)^{\frac{\alpha}{1+\alpha}}}{\hat{b}\bar{\omega}^{\frac{\alpha}{1+\alpha}}(\gamma_2 - \lambda)^{\frac{1}{1+\alpha}} n^{\frac{\alpha}{1+\alpha}}} \cdot T^{\frac{1}{1+\alpha}} \\
&= \Omega(T^{\frac{1}{1+\alpha}})
\end{aligned}
$$

which completes our proof.

H. Proof of Theorem 4

We first prove the following lemma regarding the threshold $\mu_t(\mathbf{c}^t)$.

Lemma 8 *For any $j = 1, 2, \ldots, m$, the function $\frac{1}{\mu_t(\mathbf{c}^t)}$ is a convex function over c_j^t.*

Proof: Since for each i, $F_i(\cdot)$ has a continuous density function $f_i(\cdot)$, then implicit function theorem implies that the threshold $\mu_t(\mathbf{c}^t)$ is differentiable over c_j^t for each j and the derivatives can be obtained by taking derivative over both sides of (33). We have that

$$\frac{\partial \mu_t(\mathbf{c}^t)}{\partial c_j^t} = \frac{1}{\sum_{i=1}^{n} p_i \cdot (T - t + 1) \cdot r_i^2 \cdot \mu_t^j(\mathbf{c}^t) \cdot f_i(r_i \mu_t^j(\mathbf{c}^t))}$$

Thus, we have the function $\frac{1}{\mu_t(\mathbf{c}^t)}$ is differentiable over c_j^t. Note that

$$\frac{\partial}{\partial c_j^t}(\frac{1}{\mu_t(\mathbf{c}^t)}) = -\frac{1}{(\mu_t(\mathbf{c}^t))^2} \cdot \frac{\partial \mu_t(\mathbf{c}^t)}{\partial c_j^t} = -\frac{1}{\sum_{i=1}^{n} \frac{p_i(T-t+1)}{r_i} \cdot (r_i \mu_t(\mathbf{c}^t))^3 \cdot f_i(r_i \mu_t(\mathbf{c}^t))}$$

From Condition 2 in Assumption 2, for each i, we have that $(r_i\mu_t(\mathbf{c}^t))^3 \cdot f_i(r_i\mu_t(\mathbf{c}^t))$ is a non-decreasing function over $r_i\mu_t(\mathbf{c}^t)$. Also, obviously $r_i\mu_t(\mathbf{c}^t)$ is non-decreasing over c_j^t, thus we have $(r_i\mu_t(\mathbf{c}^t))^3 \cdot f_i(r_i\mu_t(\mathbf{c}^t))$ is a non-decreasing function over c_j^t. Finally, we have that $\frac{\partial}{\partial c_j^t}(\frac{1}{\mu_t(\mathbf{c}^t)})$ is a non-decreasing function over c_j^t, which implies the function $\frac{1}{\mu_t(\mathbf{c}^t)}$ is a convex function over c_j^t. $\square$

To derive the regret bound, we study the gap between the expected total reward collected by our policy and the prophet upper bound in Theorem 3. Specifically, define

$$\rho_t(\mathbf{c}^t) = (T-t+1) \cdot \sum_{i=1}^{n} p_i \cdot r_i \cdot F_i(r_i\mu_t(\mathbf{c}^t)) - V_t(\mathbf{c}^t) \tag{69}$$

where $V_t(\mathbf{c}^t)$ is the "to-go" expected total reward collected by the our policy, i.e., the expected cumulative reward collected by our policy from time period t to the last time period T given the remaining capacity at time period t is denoted as $\mathbf{c}^t$. Then we have that

$$\text{Regret}(\text{ATP}_2) \leq \rho_1(\mathbf{c})$$

Since j_t is denoted as the resource with largest remaining capacity given the remaining capacity is $\mathbf{c}^t$, i.e., $c_{j_t}^t = \max\{c_1^t, c_2^t, \ldots, c_m^t\}$, and our policy will assign resource j_t to serve request t, revealed as (r_i, u_i), as long as $u_i \leq h_t^i(\mathbf{c}^t) = \min\{c_{j_t}^t, r_i\mu_t(\mathbf{c}^t)\}$, we have the following recursion of $V_t(\mathbf{c}^t)$:

$$V_t(\mathbf{c}^t) = (1 - \sum_{i=1}^{n} p_i)V_{t+1}(\mathbf{c}^t) + \sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} \{r_i + V_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})\}dF_i(w_i) + \sum_{i=1}^{n} p_i(1 - F_i(h_t^i(\mathbf{c}^t)))V_{t+1}(\mathbf{c}^t)$$

$$\Rightarrow$$

$$V_t(\mathbf{c}^t) = (1 - \sum_{i=1}^{n} p_i F_i(h_t^i(\mathbf{c}^t)))V_{t+1}(\mathbf{c}^t) + \sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} \{r_i + V_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})\}dF_i(w_i)$$

where $\mathbf{e}_{j_t}$ denotes a m dimensional vector with 1 at j_tth component and 0 at other components. Then, by substituting (69) into the above equation, we get the recursion of $\rho_t(\mathbf{c}^t)$.

$$(T-t+1) \cdot \sum_{i=1}^{n} p_i r_i F_i(r_i\mu_t(\mathbf{c}^t)) - \rho_t(\mathbf{c}^t) = (1 - \sum_{i=1}^{n} p_i F_i(h_t^i(\mathbf{c}^t))) \cdot [(T-t) \sum_{i=1}^{n} p_i r_i F_i(r_i\mu_{t+1}(\mathbf{c}^t)) - \rho_{t+1}(\mathbf{c}^t)]$$

$$+ \sum_{i=1}^{m} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} \{r_i + (T-t) \sum_{k=1}^{n} p_k r_k F_k(r_k\mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})) - \rho_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})\}dF_i(w_i)$$

which implies that

$$\rho_t(\mathbf{c}^t) = (1 - \sum_{i=1}^{n} p_i F_i(h_t^i(\mathbf{c}^t))) \cdot \rho_{t+1}(\mathbf{c}^t) + \sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} \rho_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})dF_i(w_i)$$

$$+ \underbrace{\sum_{i=1}^{n} p_i r_i \cdot (F_i(r_i\mu_t(\mathbf{c}^t)) - F_i(h_t^i(\mathbf{c}^t)))}_{\text{I}} + \underbrace{(T-t) \cdot \sum_{i=1}^{n} p_i r_i \cdot (F_i(r_i\mu_t(\mathbf{c}^t)) - F_i(r_i\mu_{t+1}(\mathbf{c}^t)))}_{\text{II}}$$

$$+ \underbrace{\sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} \{(T-t) \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k\mu_{t+1}(\mathbf{c}^t)) - F_k(r_k\mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})))\}dF_i(w_i)}_{\text{III}}$$

By denoting $\bar{\rho}_{t+1}$ as the maximal value of $\rho_{t+1}(\mathbf{c}^{t+1})$ for all the possible remaining capacity $\mathbf{c}^{t+1}$, we have that

$$\rho_t(\mathbf{c}^t) - \bar{\rho}_{t+1} \leq \text{I} + \text{II} + \text{III}$$

In the following, we will bound the term I, II, III separately.

Bound I: From Lemma 4, we have that

$$\frac{r_i \mu_t(\mathbf{c}^t) F_i(r_i \mu_t(\mathbf{c}^t))}{\int_0^{r_i \mu_t(\mathbf{c}^t)} w_i dF_i(w_i)} \leq M \quad \forall i \quad \Rightarrow \quad \frac{\sum_{i=1}^n p_i r_i \mu_t(\mathbf{c}^t) F_i(r_i \mu_t(\mathbf{c}^t))}{\sum_{i=1}^n p_i \cdot \int_0^{r_i \mu_t(\mathbf{c}^t)} w_i dF_i(w_i)} \leq M$$

Also from the definition of $\mu_t(\mathbf{c}^t)$ in (33), we have that

$$\sum_{i=1}^n p_i r_i \mu_t(\mathbf{c}^t) F_i(r_i \mu_t(\mathbf{c}^t)) \leq \frac{M \cdot (c_1^t + \cdots + c_m^t)}{T - t + 1}$$

Thus, note that $h_t^i(\mathbf{c}^t) = \min\{c_{j_t}^t, r_i \cdot \mu_t(\mathbf{c}^t)\}$, we have the following bound over the term I:

$$\begin{aligned}
\text{I} &= \sum_{i=1}^n p_i r_i \cdot (F_i(r_i \mu_t(\mathbf{c}^t)) - F_i(h_t^i(\mathbf{c}^t))) \leq \sum_{i=1}^n p_i r_i F_i(r_i \mu_t(\mathbf{c}^t)) \cdot 1_{\{r_i \cdot \mu_t(\mathbf{c}^t) > c_{j_t}^t\}} \\
&\leq \sum_{i=1}^n p_i r_i F_i(r_i \mu_t(\mathbf{c}^t)) \cdot 1_{\{r_{\max} \cdot \mu_t(\mathbf{c}^t) > c_{j_t}^t\}} = 1_{\{r_{\max} \cdot \mu_t(\mathbf{c}^t) > c_{j_t}^t\}} \cdot \sum_{i=1}^n p_i r_i F_i(r_i \mu_t(\mathbf{c}^t)) \\
&\leq 1_{\{r_{\max} \cdot \mu_t(\mathbf{c}^t) > c_{j_t}^t\}} \cdot \frac{M \cdot (c_1^t + \cdots + c_m^t)}{\mu_t(\mathbf{c}^t) \cdot (T - t + 1)} \leq \frac{M \cdot (c_1^t + \cdots + c_m^t) \cdot r_{\max}}{c_{j_t}^t \cdot (T - t + 1)} \\
&\leq \frac{M \cdot r_{\max} \cdot m}{T - t + 1}
\end{aligned}$$

where $r_{\max} = \{r_1, r_2, \ldots, r_n\}$. In the next, we bound the term II.

Bound II: From the definition of $\mu_t(\mathbf{c}^t)$ in (33), we have

$$\frac{c_1^t + \cdots + c_m^t}{T - t} - \frac{c_1^t + \cdots + c_m^t}{T - t + 1} = \sum_{i=1}^n p_i \cdot \int_{r_i \mu_t(\mathbf{c}^t)}^{r_i \mu_{t+1}(\mathbf{c}^t)} w_i dF_i(w_i) \leq \sum_{i=1}^n p_i r_i \mu_{t+1}(\mathbf{c}^t) \cdot (F_i(r_i \mu_{t+1}(\mathbf{c}^t)) - F_i(r_i \mu_t(\mathbf{c}^t)))$$

Thus, we have that

$$\text{II} = (T - t) \cdot \sum_{i=1}^n p_i r_i \cdot (F_i(r_i \mu_t(\mathbf{c}^t)) - F_i(r_i \mu_{t+1}(\mathbf{c}^t))) \leq -\frac{c_1^t + \cdots + c_m^t}{\mu_{t+1}(\mathbf{c}^t) \cdot (T - t + 1)}$$

At last, it remains to bound III.

Bound III: Note that

$$\frac{\partial \mu_{t+1}(\mathbf{c}^t)}{\partial c_{j_t}^t} = \frac{1}{(T - t) \cdot \sum_{i=1}^n p_i r_i^2 f_i(r_i \mu_{t+1}(\mathbf{c}^t)) \cdot \mu_{t+1}(\mathbf{c}^t)}$$

Thus, we have that

$$\frac{\partial}{\partial c_{j_t}^t}\left((T - t) \cdot \sum_{k=1}^n p_k r_k F_k(r_k \mu_{t+1}(\mathbf{c}^t))\right) = (T - t) \cdot \sum_{k=1}^n p_k r_k^2 f_k(r_k \mu_{t+1}(\mathbf{c}^t)) \cdot \frac{\partial \mu_{t+1}(\mathbf{c}^t)}{c_{j_t}^t} = \frac{1}{\mu_{t+1}(\mathbf{c}^t)}$$

which implies that

$$(T-t) \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t}))) = \int_{c_{j_t}^t - w_i}^{c_{j_t}^t} \frac{1}{\mu_{t+1}(\mathbf{c}^t(u))} du$$

where $\mathbf{c}^t(u) = \mathbf{c}^t + (u - c_{j_t}^t) \cdot \mathbf{e}_{j_t}$. Also from Lemma 8, $1/\mu_{t+1}(\mathbf{c}^t)$ is convex over $c_{j_t}^t$ for any $\mathbf{c}^t$, then we have that:

$$(T-t) \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t}))) \le \frac{w_i}{2} \cdot \left[\frac{1}{\mu_{t+1}(\mathbf{c}^t)} + \frac{1}{\mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})} \right]$$

Define a function $g(w_i)$ as

$$g(w_i) := \frac{T-t}{w_i^2} \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t}))) - \frac{1}{w_i \cdot \mu_{t+1}(\mathbf{c}^t)}$$

then we have that

$$g'(w_i) = -\frac{2(T-t)}{w_i^3} \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t}))) + \frac{1}{w_i^2 \cdot \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})} + \frac{1}{w_i^2 \cdot \mu_{t+1}(\mathbf{c}^t)}$$

$$= -\frac{2}{w_i^3} \cdot \left[(T-t) \sum_{k=1}^{n} p_k r_k (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t}))) - \frac{w_i}{2} \left(\frac{1}{\mu_{t+1}(\mathbf{c}^t)} + \frac{1}{\mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})} \right) \right]$$

$$\ge 0$$

Thus, $g(w_i)$ is non-decreasing over w_i, which implies that, for any $w_i \le c_{j_t}^t$, we have

$$g(w_i) = \frac{T-t}{w_i^2} \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t}))) - \frac{1}{w_i \cdot \mu_{t+1}(\mathbf{c}^t)}$$

$$\le g(c_{j_t}^t) = \frac{T-t}{(c_{j_t}^t)^2} \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t}))) - \frac{1}{c_{j_t}^t \cdot \mu_{t+1}(\mathbf{c}^t)}$$

which implies

$$(T-t) \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})))$$

$$\le \frac{w_i^2 (T-t)}{(c_{j_t}^t)^2} \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t}))) + \frac{1}{\mu_{t+1}(\mathbf{c}^t)} \cdot [w_i - \frac{w_i^2}{c_{j_t}^t}] \tag{70}$$

Also note that from Lemma 4, for each i, we have that

$$\frac{r_i \mu_{t+1}(\mathbf{c}^t) \cdot (F_i(r_i \mu_{t+1}(\mathbf{c}^t)) - F_i(r_i \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t})))}{\int_{r_i \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t})}^{r_i \mu_{t+1}(\mathbf{c}^t)} u_i dF_i(u_i)} \le M$$

which implies

$$\frac{\sum_{i=1}^{n} p_i r_i \mu_{t+1}(\mathbf{c}^t) \cdot (F_i(r_i \mu_{t+1}(\mathbf{c}^t)) - F_i(r_i \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t})))}{\sum_{i=1}^{n} p_i \cdot \int_{r_i \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t})}^{r_i \mu_{t+1}(\mathbf{c}^t)} u_i dF_i(u_i)} \le M$$

And from the definition of the threshold $\mu_t(\mathbf{c}^t)$ in (33), we have

$$\sum_{i=1}^{n} p_i \cdot \int_{r_i \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t})}^{r_i \mu_{t+1}(\mathbf{c}^t)} u_i dF_i(u_i) = \frac{c_{j_t}^t}{T-t}$$

Thus, we have that

$$\sum_{i=1}^{n} p_i r_i \cdot (F_i(r_i \mu_{t+1}(\mathbf{c}^t)) - F_i(r_i \mu_{t+1}(\mathbf{c}^t - c_{j_t}^t \cdot \mathbf{e}_{j_t}))) \leq \frac{M \cdot c_{j_t}^t}{\mu_{t+1}(\mathbf{c}^t) \cdot (T-t)} \tag{71}$$

Substituting (71) into (70), we have that

$$(T-t) \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})))$$

$$\leq \frac{(M-1) \cdot w_i^2}{\mu_{t+1}(\mathbf{c}^t) \cdot c_{j_t}^t} + \frac{w_i}{\mu_{t+1}(\mathbf{c}^t)}$$

As a result, we could bound the term III by:

$$\text{III} = \sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} \{(T-t) \cdot \sum_{k=1}^{n} p_k r_k \cdot (F_k(r_k \mu_{t+1}(\mathbf{c}^t)) - F_k(r_k \mu_{t+1}(\mathbf{c}^t - w_i \cdot \mathbf{e}_{j_t})))\} dF_i(w_i)$$

$$\leq \underbrace{\frac{M-1}{\mu_{t+1}(\mathbf{c}^t) \cdot c_{j_t}^t} \cdot \left[\sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} w_i^2 dF_i(w_i)\right]}_{\text{IV}} + \underbrace{\frac{1}{\mu_{t+1}(\mathbf{c}^t)} \cdot \left[\sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} w_i dF_i(w_i)\right]}_{\text{V}}$$

In what follows, we upper bound the term IV and V separately and we then get the upper bound of III. Note that

$$\sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} w_i^2 dF_i(w_i) \leq \sum_{i=1}^{n} p_i h_t^i(\mathbf{c}^t) \cdot \int_0^{h_t^i(\mathbf{c}^t)} w_i dF_i(w_i) \leq r_{\max} \cdot \mu_t(\mathbf{c}^t) \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} w_i dF_i(w_i)$$

$$\leq r_{\max} \cdot \mu_t(\mathbf{c}^t) \cdot \sum_{i=1}^{n} p_i \cdot \int_0^{r_i \mu_t(\mathbf{c}^t)} w_i dF_i(w_i) = r_{\max} \cdot \mu_t(\mathbf{c}^t) \cdot \frac{c_1^t + \cdots + c_m^t}{T-t+1}$$

$$\leq r_{\max} \cdot \mu_{t+1}(\mathbf{c}^t) \cdot \frac{c_1^t + \cdots + c_m^t}{T-t+1}$$

Thus, we have

$$\text{IV} = \frac{M-1}{\mu_{t+1}(\mathbf{c}^t) \cdot c_{j_t}^t} \cdot \left[\sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} w_i^2 dF_i(w_i)\right] \leq \frac{(M-1) \cdot r_{\max} \cdot (c_1^t + \cdots + c_m^t)}{c_{j_t}^t \cdot (T-t+1)} \leq \frac{(M-1) r_{\max} m}{T-t+1}$$

It also holds that

$$\text{V} = \frac{1}{\mu_{t+1}(\mathbf{c}^t)} \cdot \left[\sum_{i=1}^{n} p_i \cdot \int_0^{h_t^i(\mathbf{c}^t)} w_i dF_i(w_i)\right] \leq \frac{1}{\mu_{t+1}(\mathbf{c}^t)} \cdot \left[\sum_{i=1}^{n} p_i \cdot \int_0^{r_i \mu_t(\mathbf{c}^t)} w_i dF_i(w_i)\right]$$

$$= \frac{c_1^t + \cdots + c_m^t}{\mu_{t+1}(\mathbf{c}^t) \cdot (T-t+1)}$$

Then, we obtain the following upper bound of the term III:

$$\text{III} \leq \frac{(M-1)r_{\max}m}{T-t+1} + \frac{c_1^t + \cdots + c_m^t}{\mu_{t+1}(\mathbf{c}^t) \cdot (T-t+1)}$$

Thus, we have

$$\rho_t(\mathbf{c}^t) - \bar{\rho}_{t+1} \leq \text{I} + \text{II} + \text{III} \leq \frac{Mr_{\max}m}{T-t+1} - \frac{c_1^t + \cdots + c_m^t}{\mu_{t+1}(\mathbf{c}^t) \cdot (T-t+1)} + \frac{(M-1)r_{\max}m}{T-t+1} + \frac{c_1^t + \cdots + c_m^t}{\mu_{t+1}(\mathbf{c}^t) \cdot (T-t+1)}$$

$$= \frac{(2mM - m) \cdot r_{\max}}{T-t+1}$$

holds for all possible $\mathbf{c}^t$, which implies that

$$\bar{\rho}_t - \bar{\rho}_{t+1} \leq \frac{(2mM-m) \cdot r_{\max}}{T-t+1}$$

Moreover, note that $\bar{\rho}_{T-\hat{K}+1} \leq r_{\max} \cdot \hat{K}$, we have that

$$\rho_1(\mathbf{c}) \leq \bar{\rho}_1 \leq \sum_{t=1}^{T-\hat{K}} (\bar{\rho}_t - \bar{\rho}_{t+1}) + \bar{\rho}_{T-\hat{K}+1} \leq \sum_{t=1}^{T-\hat{K}} \frac{(2mM-m) \cdot r_{\max}}{T-t+1} + r_{\max} \cdot \hat{K}$$

$$\leq (\log T + 1) \cdot (2mM - m) \cdot r_{\max} + r_{\max} \cdot \hat{K}$$

which completes our proof.

I. Proof of Theorem 5

Without lose of generality, we normalize $r_{\max}$ to be 1 and we assume that T is large enough such that for each i, $r_i \cdot \mu(\mathbf{c}) \leq \bar{\omega}$. Since we have shown in Theorem 1 that

$$\mathbb{E}[V^{\text{off}}(\mathbf{I})] \geq \mathbb{E}[V^{\text{ATP}_1}(\mathbf{I})] \geq T \cdot \sum_{i=1}^n p_i \cdot r_i \cdot F_i(r_i \cdot \mu(\mathbf{c})) - O(\log T)$$

it is enough to show that $T \cdot \sum_{i=1}^n p_i r_i F_i(r_i \mu(\mathbf{c})) = \Omega(T^{\frac{1}{1+\alpha}})$. In what follows, we will use μ_T to denote the threshold $\mu(\mathbf{c})$ in (25) and denote l as the positive integer such that $\lambda^l \cdot \bar{\omega} \leq \mu_T \leq \lambda^{l-1} \cdot \bar{\omega}$. We have from (25) that

$$\frac{\sum_{j=1}^n c_j}{T} \geq \sum_{i=1}^n p_i \cdot \int_0^{r_i \cdot \mu_T} u_i dF_i(u_i) \geq \sum_{i=1}^n p_i \cdot \sum_{k=l}^{+\infty} \int_{r_i \lambda^{k+1}\bar{\omega}}^{r_i \lambda^k \bar{\omega}} u_i dF_i(u_i)$$

We also have that

$$\int_{r_i \lambda^{k+1}\bar{\omega}}^{r_i \lambda^k \bar{\omega}} u_i dF_i(u_i) \geq r_i \lambda^{k+1}\bar{\omega} \cdot [F_i(r_i \lambda^k \bar{\omega}) - F_i(r_i \lambda^{k+1}\bar{\omega})] \geq r_i \lambda^{k+1}\bar{\omega}(\gamma_1 - 1) \cdot F_i(r_i \lambda^{k+1}\bar{\omega})$$

$$\geq (\frac{\lambda}{\gamma_2})^{k+1}\bar{\omega}(\gamma_1 - 1)r_i F_i(r_i \bar{\omega})$$

The last inequality holds by noting $F_i(r_i \bar{\omega}) \leq \gamma_2^{k+1} \cdot F_i(r_i \lambda^{k+1}\bar{\omega})$. Thus, we have that

$$\frac{\sum_{j=1}^m c_j}{T} \geq \bar{\omega}(\gamma_1 - 1)\sum_{i=1}^n r_i p_i F_i(r_i \bar{\omega}) \cdot \sum_{k=l}^{+\infty}(\frac{\lambda}{\gamma_2})^{k+1} = \bar{\omega}(\gamma_1 - 1)\sum_{i=1}^n r_i p_i F_i(r_i \bar{\omega})(\frac{\lambda}{\gamma_2})^{l+1} \cdot \frac{\gamma_2}{\gamma_2 - \lambda}$$

Define a constant $Q = \frac{(\sum_{j=1}^{m} c_j) \cdot (\gamma_2 - \lambda)}{\bar{\omega}(\gamma_1 - 1)\lambda \sum_{i=1}^{n} r_i p_i F_i(r_i \bar{\omega})}$, we have that

$$(\frac{\lambda}{\gamma_2})^l \leq \frac{Q}{T}$$

Also, since $0 < \lambda < 1$ and $0 < \frac{1}{\gamma_2} < 1$, there exists a constant $\alpha > 0$ such that $\frac{1}{\gamma_2} = \lambda^\alpha$. Thus, we have

$$\lambda^{(1+\alpha)l} \leq \frac{Q}{T} \Rightarrow \lambda^l \leq (\frac{Q}{T})^{\frac{1}{1+\alpha}}$$

Notice that from (25), we have

$$\sum_{i=1}^{n} p_i r_i F_i(r_i \mu_T) \geq \frac{M_1}{\mu_T} \cdot \sum_{i=1}^{n} p_i \int_0^{r_i \mu_T} u_i dF_i(u_i) = \frac{M_1 \cdot (\sum_{j=1}^{m} c_j)}{\mu_T \cdot T} \geq \frac{M_1 \lambda (\sum_{j=1}^{m} c_j)}{\bar{\omega} \cdot T} \cdot \frac{1}{\lambda^l}$$

Then, we have that

$$T \cdot \sum_{i=1}^{n} p_i r_i F_i(r_i \mu_T) \geq \frac{M_1 \lambda (\sum_{j=1}^{m} c_j)}{\bar{\omega}} \cdot \frac{1}{\lambda^l} \geq \frac{M_1 \lambda (\sum_{j=1}^{m} c_j)}{Q^{\frac{1}{1+\alpha}} \bar{\omega}} \cdot T^{\frac{1}{1+\alpha}}$$

Further note that from (34), we have

$$\hat{K} + 1 \geq \frac{\sum_{j=1}^{m} c_j}{\sum_{i=1}^{n} p_i \cdot \int_0^{r_i \cdot \frac{\bar{\omega}}{r_{\max}}} w_i dF_i(w_i)} \geq \frac{\sum_{j=1}^{m} c_j}{\sum_{i=1}^{n} p_i r_i \bar{\omega} F_i(r_i \bar{\omega})}$$

which implies that $Q \leq \frac{(\gamma_2 - \lambda)(\hat{K}+1)}{(\gamma_1 - 1)\lambda}$. Thus we have that

$$T \cdot \sum_{i=1}^{n} p_i r_i F_i(r_i \mu_T) \geq \frac{M_1 \lambda (\sum_{j=1}^{m} c_j) \cdot (\lambda(\gamma_1 - 1))^{\frac{1}{1+\alpha}}}{\bar{\omega} \cdot ((\gamma_2 - \lambda)(\hat{K}+1))^{\frac{1}{1+\alpha}}} \cdot T^{\frac{1}{1+\alpha}} = \Omega(T^{\frac{1}{1+\alpha}})$$

which completes our proof.