Optimal Fresh Data Sampling and Trading

Junyi He, Qian Ma, Meng Zhang, Jianwei Huang

***Abstract*—Data freshness, measured by Age of information (AoI), is becoming an increasingly significant metric for data valuation. However, most existing data trading markets ignore the impact of such a metric. In this paper, we study a fresh data market, where users with heterogeneous valuations for AoI stochastically arrive over time. The platform decides data sampling (which affects the AoI) and pricing policies (to the users), to maximize its profit. We consider three types of pricing policies with increasing flexibility, i.e., a uniform pricing policy, a dual pricing policy, and a dynamic pricing policy. The joint data sampling and pricing optimization is a non-smooth mixed integer programming problem, which is challenging to solve. Despite the difficulty, we derive the closed-form solutions of the optimal data sampling policies and pricing policies for all three cases. Our analysis yields several interesting practical insights. First, the optimal data prices decrease in the unit sampling cost and increase in the users' arrival rate. Second, for all three pricing policies, the equal-spacing data sampling policy is optimal. Third, numerical results show that the optimal dual pricing policy significantly outperforms the optimal uniform pricing policy. Specifically, the optimal dual pricing policy produces up to 280% of the profit that is achieved by the optimal uniform pricing policy.**

***Index Terms*—Fresh data market, age of information, data sampling and pricing.**

I. Introduction

A. Motivation and Key Questions

As data-driven technologies are becoming essential for numerous applications, data become a new type of valuable digital asset for trading. To facilitate data exchange, the data trading market has recently emerged a new business paradigm (e.g., crowdsensing data trading markets [1]–[3] and IoT data trading markets [4], [5]). With the rapid proliferation of some real-time applications, data freshness, measured by Age of information (AoI), is becoming an increasingly significant metric of data valuation [6]–[8]. Examples of real-time data include traffic conditions, news, sales promotion, and air quality index.

Junyi He is with Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS), School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Guangdong, China (e-mail: junyihe@link.cuhk.edu.cn).

Qian Ma is with School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, Guangdong 518107, China (e-mail: maqian25@mail.sysu.edu.cn).

Meng Zhang is with the Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL 60208 (e-mail: meng.zhang@northwestern.edu).

Jianwei Huang is with School of Science and Engineering, Shenzhen Institute of Artificial Intelligence and Robotics for Society, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China (corresponding author, e-mail: jianweihuang@cuhk.edu.cn).

This work is supported by the National Natural Science Foundation of China under Grant 62002399, the Shenzhen Institute of Artificial Intelligence and Robotics for Society, and the Presidential Fund from the Chinese University of Hong Kong, Shenzhen.

Existing traditional data markets [1]–[5] often ignore the freshness dimension of the data, and the design of fresh data market requires several unique considerations. First, the platform can acquire fresh data via random or periodic data sampling. However, data sampling often incurs cost due to the need of obtaining and processing data. Second, traditional markets assume that users' valuations of data are fixed over time. However, in the fresh data market, users' valuations change with data freshness. The above considerations motivate us to ask the following key question.

Key Question 1: *How does the platform decide the data sampling and pricing policy to maximize its profit, considering the impact of AoI?*

More specifically, we will consider three types of pricing policies.

- *Uniform pricing policy*. The platform adopts a uniform price over the whole time horizon. Many existing data trading markets use such a pricing policy [1]–[3].
- *Dual pricing policy*. The platform decides a full price, a discounted price, and an AoI threshold. The platform charges the full price for data that are fresh enough (i.e., if the AoI is lower than the threshold); otherwise, the platform charges the discounted price. The dual pricing policy is the simplest AoI-aware pricing scheme.
- *Dynamic pricing policy*. The platform determines the price as a real-time function of the ata AoI.

Considering the above three types of pricing policies, it is natural to ask the following question.

Key Question 2: *How do different pricing policies affect the platform's profit in fresh data markets?*

B. Challenges and Key Contributions

In this paper, we study a fresh data market where a platform provides data with different freshness to dynamically arriving users. We propose a two-stage game model to study the interactions between the platform and the users. In Stage I, the platform decides the data sampling and pricing policy to maximize its long-term profit, achieving a balance between the revenue and the incurred sampling cost. In Stage II, each user decides whether to purchase the data upon arrival, based on current data price and AoI.

It is challenging to maximize the platform's profit in such a fresh data market. Since the platform needs to jointly optimize the number of updates, the corresponding sampling time, and the prices, the problem is a non-smooth mixed integer programming problem, which is hard to solve. Despite these challenges, we are able to derive the

ISBN 978-3-903176-37-9 © 2021 IFIP

analytical solutions for the optimal data sampling policy and the optimal three pricing policies. We tackle the challenges by analyzing an appropriate relaxed problem and proving that the optimal solution of the relaxed problem is also optimal for our original problem.

Our key results and contributions are summarized as follows:

- *Fresh data market model*. We propose a new analytical model for the fresh data market, considering users' heterogeneous valuations for AoI. To the best of our knowledge, this is the first work that jointly studies platform's data sampling and pricing policy in the AoI literature. Our model captures the practical scenario where the platform can repeatedly sell the same data to dynamically arriving users.
- *Optimal data sampling policies.* We derive the closed-form optimal sampling policies for all three pricing policies. An interesting result is that for all three pricing policies, the optimal sampling policy is equal-spacing, which is attractive for practical implementation. The specific optimal numbers of updates are different under different pricing policies.
- *Optimal three data pricing policies.* We derive the closed-form solutions of three pricing policies. An important insight is that the optimal uniform price and optimal dual prices decrease in the unit sampling cost and increases in the users' arrival rate. The reason behind this counter-intuitive result is the data freshness. As the unit sampling cost increases or the users' arrival rate decreases, the platform chooses to update data less frequently and users are less willing to pay for the stale data. Hence, the prices decrease. For the dynamic pricing policy, the optimal data prices are the real-time function of the AoI and change periodically over time.
- *Technical contributions.* To get the optimal uniform and dual pricing policies, we first define a proper relaxed problem. Then we optimize the platform's profit of the relaxed problem, which serves as an upper bound of the maximum profit of the original problem. After that, we derive the optimal solution of the relaxed problem and prove that it is also optimal for our original problem. For the dynamic pricing policy, by utilizing the special structures of the problem, we sequentially optimize the data sampling and pricing policy to get the optimal solution.
- *Profit comparison among different pricing polices.* We show that the optimal dual pricing policy significantly outperforms the optimal uniform pricing policy. Specifically, the optimal dual pricing policy produces up to 280% of the profit that is achieved by the optimal uniform pricing policy.

The rest of the paper is organized as follows. In Section II, we review the literature. Section III introduces the system model and problem formulation. Section IV develops the optimal data sampling and uniform pricing policy. Section V develops the optimal data sampling and dual pricing policy. Section VI investigates the optimal data sampling and dynamic pricing policy. Section VII gives the numerical results. Finally, Section VIII concludes the paper.

II. Related Work

Related works can be classified into two categories, AoI minimization and economic issues in AoI.

AoI Minimization: Existing works along this line mainly focus on minimizing time-average AoI under a variety of system settings (e.g., [9]–[20]). Kaul *et. al.* [9] used queuing theory to optimize the average AoI in the queuing systems. Sun *et. al.* [10] proposed optimal sampling policies to minimize the average AoI for a single-source system, which was extended to a multiple-source system in [11]. There are some works studying AoI-aware scheduling for information freshness in the networks [12]–[17]. Hsu *et. al.* [12] and Kadota *et. al.* [13] proposed scheduling policies to minimize the average AoI in wireless broadcast networks. Lu *et. al.* [14] considered the problem of scheduling real-time data traffic with hard deadlines. Bastopcu *et. al.* [15] studied information freshness in a cache updating system. Considering the energy harvesting, Bacinoglu *et. al.* [18], [19] proposed optimal status sampling schemes to minimize the average AoI. Arafa *et. al.* [20] proposed optimal status update policies for energy harvesting sensors in an online setting. However, this line of work ignores the economic issues of controlling AoI.

Economic Issues in AoI: Most existing works study how content platforms incentivize data sources to generate fresh data updates (e.g., [21]–[25]). Li *et. al.* [21] designed efficient reward mechanisms to incentivize mobile users to report fresh data in time, with the goal of keeping the AoI low. Wang *et. al.* [22] studied how the platform minimizes the expected AoI and total payment to the data sources. On the other hand, in work [23], the data sources designs the pricing schemes, and the destination decides the sampling policies. Zhang *et. al.* [24] and Wang *et. al.* [25] studied the information asymmetry in the fresh data acquisition. However, this line of work focused on how the platform acquires fresh data from sources. It remains an open problem regarding how platform should jointly sample fresh data and sell them to users. We will further consider the possibility that the same copy of data can be sold to different users over time (with different AoI values). This further couples the decisions of the data sampling, data pricing, and AoI in the temporal dimension.

III. System Model

In this section, we first introduce the fresh data market, including the data sampling policy and pricing policy. Then, we formulate a two-stage Stackelberg game to study the fresh data trading between the platform and the users.

A. Fresh Data Market

We consider a fresh data market as shown in Fig. 1, where a platform sells data (e.g., traffic data of a particular region or the noise data of a community) to interested users over a time

ISBN 978-3-903176-37-9 © 2021 IFIP

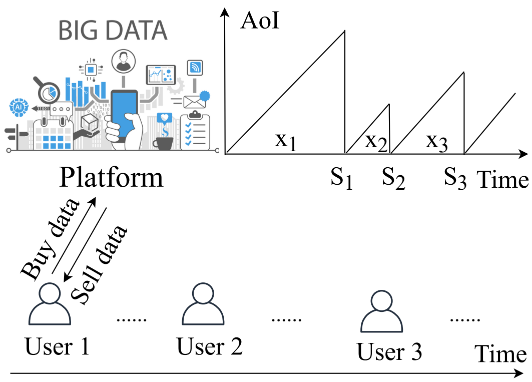


Fig. 1. System model

horizon $\mathcal{T} = [0, T]$. Users dynamically arrive at the fresh data market according to a Poisson Process with an arrival rate λ. Each of the users has unit demand of the data. The platform decides how to sample fresh data and sell them to users.

1) Data sampling policy: We study a generate-at-will updating model [7], [10]. Once the platform decides to sample a data packet, it will arrive at the platform immediately without any delay. Let $\mathbb{N}$ denote the set of all natural numbers. The platform's sampling policy S includes the number of updates and a sequence of sampling time, i.e., $S = \{s_1 \le s_2 \le ... \le s_K\}$, where $K \in \mathbb{N}$ is the total number of updates and $s_k \in [0, T]$ is the time instance when the platform receives the k-th data update.

To facilitate the analysis, we use an equivalent way to express the sampling policy S. We first define $s_0 = 0$ and $s_{K+1} = T$, such that $0 = s_0 \le s_1 \le s_2 \le ... \le s_K \le s_{K+1} = T$. Given a sampling policy S with K updates, we can divide the total time horizon $[0, T]$ into $K+1$ inter-update periods, with the k-th period length satisfying $x_k = s_k - s_{k-1}$. Denote $\boldsymbol{x} = \{x_1, x_2, ..., x_{K+1}\}$, and a sampling policy can be expressed as $S = (K, \boldsymbol{x})$.

***Definition** 1 (Age of information [7], [10], [23]):* The age of information $\Delta(S, t)$ at time t is

$$\Delta(S, t) = t - H_t(S), \tag{1}$$

where $H_t(S)$ denotes the time stamp of the latest update before time t, so that

$$H_t(S) = \max_{k \in \mathbb{N} \setminus \{0\}} \sum_{j=1}^{k} x_j, \quad \text{s.t.} \quad \sum_{j=1}^{k} x_j \le t.$$

2) Data pricing policy: After the platform samples data, the same data could be sold repeatedly to different dynamically arriving users without additional cost. However, the data freshness decreases as time goes by. The platform can choose to price data with different freshness differently. Let $p(t)$ denote the data price at time t and a data pricing policy is denoted as $\mathcal{P} \triangleq \{p(t) \ge 0, \forall t \in \mathcal{T}\}$.

We formulate the interactions between the platform and the users as a two-stage Stackelberg Game, as shown in Fig. 2. In Stage I, the platform decides data sampling and pricing

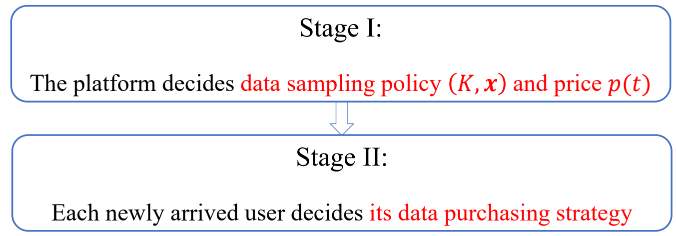


Fig. 2. Two-stage Stackelberg game

policy to maximize its long-term profit. In Stage II, each user decides whether to purchase the data upon arrival. Next, we introduce the details of the two-stage Stackelberg game through backward induction.

B. Stage II: Users' Data Purchasing

Users wish to get the data upon arrival instantaneously and their valuations about data are freshness-sensitive. Users value the freshest data most and their valuations decrease in AoI [26]. To make the analysis tractable, we model the users' data valuations as a decreasing function with AoI,

$$v_i(t) = \frac{\vartheta_i}{\Delta(S, t) + 1},$$

where ϑ_i is the coefficient. We assume that ϑ_i follows a uniform distribution $[0, \vartheta_{\max}]$, where $\vartheta_{\max} \le 1$ [27].

Given the data price $p(t)$ and the AoI $\Delta(S, t)$, user i arriving at time t decides its purchasing strategy $b_i(t) \in \{0, 1\}$ to maximize the payoff:

$$b_i^*(t) = \underset{b_i(t) \in \{0,1\}}{\text{argmax}} \; b_i(t) \left(\frac{\vartheta_i}{\Delta(S, t) + 1} - p(t) \right). \tag{2}$$

It is easy to see that user i's optimal data purchasing strategy has a threshold structure. Due to the limited space, all the proofs are provided in the online appendix [28].

***Lemma** 1 (Threshold-based Purchasing Strategy):* Given a data pricing policy $\mathcal{P}$ and current AoI $\Delta(S, t)$, user i's optimal data purchasing strategy is

$$b_i^*(t) = \begin{cases} 1, & \text{if } \vartheta_i \ge p(t) \left(\Delta(S, t) + 1 \right), \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

C. Stage I: Platform's Data Sampling and Pricing

In Stage I, the platform needs to decide the data sampling and pricing to maximize its long-time profit, considering the impact on users' data purchasing decisions in Stage II. We consider the incomplete information scenario, where the platform only knows the distribution of users' valuations but not the specific valuation of each user. By using Lemma 1, we can show that from the platform's perspective, the probability that a user i arriving at time t buys the data is

$$\Pr_{\{b_i^*(t)=1\}} = 1 - \min \left\{ 1, \frac{p(t)(\Delta(S, t) + 1)}{\vartheta_{\max}} \right\}.$$

The purchasing probability is non-increasing in $p(t)$ and AoI $\Delta(S, t)$, as users become less willing to pay due to high

ISBN 978-3-903176-37-9 © 2021 IFIP

prices and lower valuations of the stale data. When users' arrival rate is λ, the expected revenue of the platform is

$$R(S, \mathcal{P}) = \int_0^T \lambda p(t) \Pr_{\{b_i^*(t)=1\}} dt. \quad (4)$$

The integral term on the right-hand-side of (4) can be written as $\sum_{i=1}^{K+1} \Phi(x_k, \mathcal{P})$, where $\Phi(x_k, \mathcal{P})$ is the k-th period revenue given by

$$\Phi(x_k, \mathcal{P}) = \int_{s_{k-1}}^{s_k} \lambda p(t) \left(1 - \min\left\{1, \frac{p(t)(\Delta(S,t)+1)}{\vartheta_{\max}}\right\}\right) dt. \quad (5)$$

Let $\mathcal{K}_{K+1}$ denote the set $\{1, 2, ..., K+1\}$ and c denote the unit sampling cost. The joint data sampling and pricing problem is formulated as

$$\max_{(K,\boldsymbol{x}),\mathcal{P}} \quad \sum_{i=1}^{K+1} \Phi(x_k, \mathcal{P}) - cK \quad (6a)$$

$$\text{s.t.} \quad \sum_{k=1}^{K+1} x_k = T, \quad (6b)$$

$$\text{var.} \quad x_k \geq 0, \quad \forall k \in \mathcal{K}_{K+1}, \quad (6c)$$

$$K \in \mathbb{N}, \quad (6d)$$

$$p(t) \geq 0, \quad \forall t \in \mathcal{T}. \quad (6e)$$

To study how different pricing policies affect the platform's profit in fresh data markets, we will solve Problem (6) under three types of pricing policies. Different types of pricing policies result in users' different data purchasing strategies by adjusting data prices, thereby affecting the platform's profit. Formally, we define three types of pricing policies as follows,

- A uniform pricing policy $\mathcal{P}_{\text{uniform}}$. The platform charges the uniform price p, i.e., $p(t) = p, \forall t \in \mathcal{T}$.
- A dual pricing policy $\mathcal{P}_{\text{dual}}$. The platform decides a full price p_h, a discounted price p_l, and an AoI threshold $\bar{\Delta}$. The platform charges the full price if the AoI is lower than the threshold, and charges the discounted price otherwise. In others words, in a dual pricing policy, the price at time t is

$$p(t) = \begin{cases} p_h, & \text{if } 0 \leq \Delta(S,t) \leq \bar{\Delta}, \\ p_l, & \text{otherwise.} \end{cases}$$

- A dynamic pricing policy $\mathcal{P}_{\text{dynamic}}$. The platform can adjust the data price $p(t)$ at any time t.

IV. Optimal Data Sampling and Uniform Pricing Policy

This section aims to study the simplest pricing model: the uniform pricing policy, which uses a single price for the whole horizon $\mathcal{T}$. We first define the profit maximization problem under the uniform pricing policy, which is difficult to solve directly. Then we study a general relaxed uniform pricing policy optimization problem, and show that its optimal solution corresponds to the optimal uniform pricing policy of our original problem.

A. Profit Maximization Under Uniform Pricing Policy

The platform will decide the data sampling policy $(K, \boldsymbol{x})$ and the uniform pricing policy $\mathcal{P}_{\text{uniform}}$ to maximize its total profit as follows,

$$\max_{S=(K,\boldsymbol{x}),\mathcal{P}_{\text{uniform}}} \quad \sum_{k=1}^{K+1} \Phi(x_k, \mathcal{P}_{\text{uniform}}) - cK \quad (7a)$$

$$\text{s.t.} \quad (6b), (6c), (6d), (6e), \quad (7b)$$

where the k-th period revenue is given by

$$\Phi(x_k, \mathcal{P}_{\text{uniform}}) = \int_0^{x_k} \lambda p \left(1 - \min\left\{1, \frac{p \cdot (t+1)}{\vartheta_{\max}}\right\}\right) dt.$$

Since the platform seeks to jointly optimize the number of updates, the sampling time, and the price, Problem (7) is a mixed integer program. Furthermore, the objective function is non-smooth due to the $\min$ operation. Under a fixed sampling policy, it is difficult to derive the closed-form expression of the optimal uniform pricing policy to Problem (7). However, by exploiting the special structures and features of Problem (7), we can characterize the optimal solutions analytically.

B. Profit Maximization Under General Uniform Pricing Policy

In this subsection, we will study a relaxed version of Problem (7). We first generalize the uniform pricing policy by considering a multi-period uniform pricing policy that sets different prices for different periods. Then, we derive the maximum profit under the proposed general policy, which provides an upper bound of the maximum profit under the uniform pricing policy. Finally, we will show the equivalence relation between the derived optimal multi-period uniform pricing policy and the optimal uniform pricing policy.

Formally, given an arbitrary sampling policy S, a multi-period uniform pricing policy $\mathcal{P}_{\text{uniRelax}}(S)$ sets the data price p_k for the k-th period, i.e.,

$$\mathcal{P}_{\text{uniRelax}}(S) \triangleq \{p(t) = p_k, \text{if } t \in [s_{k-1}, s_k], \forall t \in \mathcal{T}\}.$$

Instead of directly solving problem (7), we will study the profit maximization problem for the multi-period uniform pricing policy as follows,

$$\max_{S=(K,\boldsymbol{x}),\mathcal{P}_{\text{uniRelax}}(S)} \quad \sum_{k=1}^{K+1} \Phi(x_k, \mathcal{P}_{\text{uniRelax}}(S)) - cK \quad (8a)$$

$$\text{s.t.} \quad (6b), (6c), (6d), (6e), \quad (8b)$$

where the k-th period revenue is given by

$$\Phi(x_k, \mathcal{P}_{\text{uniRelax}}(S)) = \int_0^{x_k} \lambda p_k \left(1 - \min\left\{1, \frac{p_k \cdot (t+1)}{\vartheta_{\max}}\right\}\right) dt.$$

The maximum profit in Problem (8) provides an upper bound over the maximum profit under a uniform pricing policy. The advantage of considering the multi-period uniform pricing policy is as follows. When fixing the sampling

ISBN 978-3-903176-37-9 © 2021 IFIP

policy, we can derive the closed-form expression of the optimal multi-period uniform pricing policy in terms of the sampling policy. Then, we can optimize the sampling policy to get the analytical form of the optimal solution.

***Proposition** 1:* The platform's optimal sampling policy $S^*_{\text{uniRelax}} = (K^*_{\text{uniRelax}}, \boldsymbol{x}^*_{\text{uniRelax}})$ and optimal multi-period uniform pricing policy $\mathcal{P}^*_{\text{uniRelax}}$ satisfy

$$K^*_{\text{uniRelax}} = \underset{K\in\mathbb{N}}{\text{argmax}}\, \frac{\lambda\vartheta_{\max}}{2}\frac{(K+1)T}{T+2(K+1)} - cK, \quad (9)$$

$$x^*_{k,\text{uniRelax}} = \frac{T}{1+K^*_{\text{uniRelax}}},\ \forall k \in \mathcal{K}_{K^*_{\text{uniRelax}}+1}, \quad (10)$$

$$p^*_k = \frac{\vartheta_{\max}}{\frac{T}{K^*_{\text{uniRelax}}+1}+2},\ \forall k \in \mathcal{K}_{K^*_{\text{uniRelax}}+1}. \quad (11)$$

An interesting insight from Proposition 1 is that the optimal prices p^*_k is the same across different periods, and hence corresponds to a uniform pricing policy. Further, since Problem (8) is the relaxed version of Problem (7), while (9)-(11) is a feasible solution to Problem (7), it is optimal to Problem (7) as well.

C. Optimal Uniform Pricing Policy

***Theorem** 1:* For Problem (7), the platform's optimal sampling policy is $S^*_{\text{unifrom}} = S^*_{\text{uniRelax}}$ in (9)-(10) and the optimal uniform price is $p^*_{\text{uniform}} = p^*_k$ in (11).

Theorem 1 leads to some important insights. *First*, the optimal uniform data price is non-increasing in the unit sampling cost c. This is counter-intuitive, as when the marginal cost c of data acquisition increases, we would expect the platform to increase the data prices in order to recover the increasing costs. However, such an intuition is derived from the traditional data market. In our fresh data market, when the sampling cost is high, the platform chooses to update data less frequently, causing lower data freshness. Due to users' freshness-sensitive data valuations, users are less willing to pay for the stale data, hence the platform needs to reduce the price accordingly. *Second*, as users' arrival rate λ increases, the optimal uniform data price increases. This suggests that when facing a larger market, the platform can raise the data price to gain more profit, which seems to go against pricing rules for data goods. Intuitively, the supply of data goods is unlimited due to low costs of duplicate. Hence, the data price should have nothing to do with the market size. However, in the fresh data market, when facing a larger number of users, the platform can get larger revenue and afford a larger number of updates. As a result, the average AoI decreases, and users are willing to pay more for the fresher data. Therefore, the platform can raise the data price. *Third*, the optimal sampling policy is an equal-spacing one, which is simple and attractive for practical implementation.

Nevertheless, the maximum profit under the uniform pricing policy may be relatively low due to its simplicity. This motivates us to consider a dual pricing policy, which is the simplest AoI-aware pricing scheme.

V. Optimal Data Sampling and Dual Pricing Policy

In this section, we study a more sophisticated pricing policy: the dual pricing policy. To derive the optimal data sampling and dual pricing policy, we study a general relaxed dual pricing policy. Then we show that the optimal general dual pricing policy is equivalent to the optimal dual pricing policy.

A. Profit Maximization Under Dual Pricing Policy

The platform will optimize the sampling policy $(K, \boldsymbol{x})$ and dual pricing policy $\mathcal{P}_{\text{dual}}$ to maximize its total profit as follows,

$$\max_{S=(K,\boldsymbol{x}),\mathcal{P}_{\text{dual}}} \quad \sum_{k=1}^{K+1} \Phi(x_k, \mathcal{P}_{\text{dual}}) - cK \quad (12a)$$

$$\text{s.t.} \quad (6b), (6c), (6d), (6e), \quad (12b)$$

where

$$\Phi(x_k, \mathcal{P}_{\text{dual}}) = \int_0^{\bar{\Delta}} \lambda p_h \left(1 - \min\left\{1, \frac{p_h\cdot(t+1)}{\vartheta_{\max}}\right\}\right) dt + \int_{\bar{\Delta}}^{x_k} \lambda p_l \left(1 - \min\left\{1, \frac{p_l\cdot(t+1)}{\vartheta_{\max}}\right\}\right) dt.$$

Similar to the profit maximization problem under the uniform pricing policy, we cannot solve Problem (12) explicitly. Furthermore, unlike the uniform pricing policy, the joint optimization of the full price p_h, the discounted price p_l, and the AoI threshold $\bar{\Delta}$ makes it more hard to analyze the non-smooth objective function.

B. Profit Maximization Under General Dual Pricing Policy

In this subsection, we first define a general dual pricing policy, which is a multi-period dual pricing policy. The platform can use different dual pricing policies during different periods. Then we will study the profit maximization under the general dual pricing policy.

Given an arbitrary sampling policy S, a multi-period dual pricing policy is denoted by $\mathcal{P}_{\text{dualRelax}}(S)$, where the price

$$p(t) = \begin{cases} p_{h,k}, & \text{if } 0 \le \Delta(S,t) \le \bar{\Delta}_k, t \in [s_{k-1}, s_k], \\ p_{l,k}, & \text{if } \bar{\Delta}_k < \Delta(S,t) \le x_k, t \in [s_{k-1}, s_k]. \end{cases}$$

Instead of directly solving Problem (12), we will study profit maximization problem under the multi-period dual pricing policy as follows,

$$\max_{S=(K,\boldsymbol{x}),\mathcal{P}_{\text{dualRelax}}(S)} \quad \sum_{k=1}^{K+1} \Phi(x_k, \mathcal{P}_{\text{dualRelax}}(S)) - cK \quad (13a)$$

$$\text{s.t.} \quad (6b), (6c), (6d), (6e). \quad (13b)$$

The optimal solution of Problem (12) is a special case of the multi-period dual pricing policy. Hence, the maximum profit in Problem (13) provides an upper bound over the maximum profit in Problem (12). Next, we will optimize Problem (13) to get the upper bound. Similarly, we find that the optimal multi-period dual pricing policy also establishes a feasible dual pricing policy. Since Problem (13) is the relaxed version of Problem (12), we can get the optimal solution for Problem (12).

ISBN 978-3-903176-37-9 © 2021 IFIP

C. Optimal Dual Pricing Policy

Before presenting the following theorem, we first define a function

$$G(x_k) = \frac{\lambda \vartheta_{\max} x_k \sqrt{x_k+1}}{(\sqrt{x_k+1}+1)(\sqrt{x_k+1}+1+x_k)}.$$

***Theorem** 2:* The platform's optimal sampling policy $S^*_{\text{dual}} = (K^*_{\text{dual}}, \boldsymbol{x}^*_{\text{dual}})$ and optimal dual pricing policy $\mathcal{P}^*_{\text{dual}}$ satisfy

$$K^*_{\text{dual}} = \underset{K \in \mathbb{N}}{\operatorname{argmax}} \sum_{k=1}^{K+1} G\left(\frac{T}{K+1}\right) - cK, \tag{14}$$

$$x^*_{\text{dual},k} = x^*_{\text{dual}} = \frac{T}{K^*_{\text{dual}}+1}, \forall k \in \mathcal{K}_{K^*_{\text{dual}}+1}, \tag{15}$$

$$p^*_h = \frac{\vartheta_{\max}}{\bar{\Delta}+2},\ p^*_l = \frac{\vartheta_{\max}}{x^*_{\text{dual}}+\bar{\Delta}+2}, \tag{16}$$

$$\bar{\Delta} = \sqrt{x^*_{\text{dual}}+1} - 1. \tag{17}$$

Theorem 2 reveals some important insights. The optimal full price p^*_h and the optimal discounted price p^*_l decrease in the unit sampling cost and increase in the users' arrival rate. This insight is similar as that from Theorem 1. A new observation here is that the ratio $\bar{\Delta}/x_{\text{dual}} = (\sqrt{x^*_{\text{dual}}+1}-1)/x_{\text{dual}}$ also decreases in the unit sampling cost and increase in the users' arrival rate. This means that the platform is more likely to sell fresh data at the discounted price, when facing a higher unit sampling cost or a smaller users' arrival rate.

Although a dual pricing policy can provide price discounts for users, the data prices are limited to two scalars. This motivates us to consider a dynamic pricing policy, where the platform can arbitrarily change data prices at each time.

VI. Optimal Data Sampling and Dynamic Pricing Policy

In this section, we aim to study the optimal data sampling and the most flexible dynamic pricing policy.

The platform optimizes the sampling policy $S = (K, \boldsymbol{x})$ and dynamic pricing policy $\mathcal{P}_{\text{dynamic}}$ to maximize its total profit as follows,

$$\max_{S=(K,\boldsymbol{x}),\mathcal{P}_{\text{dynamic}}} \quad \sum_{k=1}^{K+1} \Phi(x_k, \mathcal{P}_{\text{dynamic}}) - cK \tag{18a}$$

$$\text{s.t.} \quad (6b), (6c), (6d), (6e). \tag{18b}$$

Problem (18) is a non-smooth non-convex problem due to the $\min$ operation in the objective function and the integer decision variable K. We will use the following three steps to solve it. First, we fix the sampling policy $(K, \boldsymbol{x})$ and compute the optimal dynamic pricing policy $\mathcal{P}^*_{\text{dynamic}}(K, \boldsymbol{x})$. Second, under the optimal dynamic pricing policy $\mathcal{P}^*_{\text{dynamic}}(K, \boldsymbol{x})$, we fix K and compute $\boldsymbol{x}^*(K)$. Finally, we derive the optimal number of updates K^*.

A. Step 1: Optimizing the Dynamic Pricing Policy

Given the sampling policy $(K, \boldsymbol{x})$, the optimal dynamic pricing policy is to decide price at time t so that the integrand in (5) is maximized.

***Lemma** 2:* Given fixed sampling policy $S = (K, \boldsymbol{x})$, the optimal dynamic price at time t is

$$p^*_{\text{dynamic}}(t) = \frac{\vartheta_{\max}}{2(1+\Delta(S,t))}. \tag{19}$$

From Lemma 2, we can see that the optimal dynamic price $p^*_{\text{dynamic}}(t)$ increases with $\vartheta_{\max}$, and decreases with the AoI $\Delta(S,t)$.

B. Step 2: Optimizing the Inter-update Periods

Based on Lemma 2, we can rewrite Problem (18) as

$$\max_{S=(K,\boldsymbol{x})} \quad \sum_{k=1}^{K+1} \frac{\lambda \vartheta_{\max}}{4} \ln(1+x_k) - cK \tag{20a}$$

$$\text{s.t.} \quad (6b), (6c), (6d). \tag{20b}$$

Given the total number of updates K, Problem (20) is a convex problem with respect to $\boldsymbol{x} = \{x_1, ..., x_{K+1}\}$. Hence, we can derive the following result.

***Lemma** 3:* Given the number of updates K, the platform's optimal inter-update periods $\boldsymbol{x}^*$ in Problem (20) satisfies

$$x^*_{k,\text{dynamic}}(K) = \frac{T}{K+1}, \forall k \in \mathcal{K}_{K+1}. \tag{21}$$

We can see that the inter-update period $x^*_{k,\text{dynamic}}(K)$ increases in T and decreases in K.

C. Step 3: Optimizing the Number of Updates

By substituting $x^*_{k,\text{dynamic}}(K)$ into the objective function of Problem (20), we can write Problem (20) as the following problem:

$$\max_{K\in\mathbb{N}} \frac{\lambda \vartheta_{\max}}{4}(K+1)\ln\left(\frac{T}{K+1}+1\right) - cK.$$

***Theorem** 3:* The platform's optimal sampling policy $S^*_{\text{dynamic}} = (K^*_{\text{dynamic}}, \boldsymbol{x}^*_{\text{dynamic}})$ and optimal dynamic pricing policy $\mathcal{P}^*_{\text{dynamic}}$ satisfy

$$K^*_{\text{dynamic}} = \underset{K\in\mathbb{N}}{\operatorname{argmax}} \frac{\lambda \vartheta_{\max}}{4}(K+1)\ln\left(\frac{T}{K+1}+1\right) - cK,$$

$$x^*_{\text{dynamic},k} = \frac{T}{K^*_{\text{dynamic}}+1}, \forall k \in \mathcal{K}_{K^*_{\text{dynamic}}+1},$$

$$p^*_{\text{dynamic}}(t) = \frac{\vartheta_{\max}}{2(1+\Delta(S^*_{\text{dynamic}}, t))}.$$

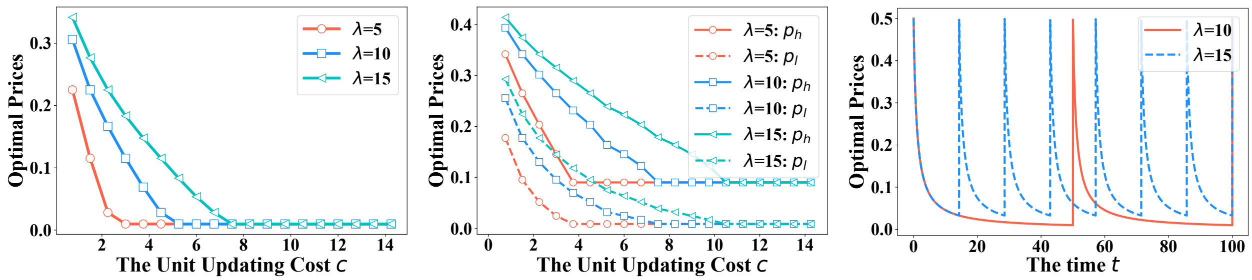
Theorem 3 implies the following results. Although data prices vary over time, the optimal sampling policy is still equal-spacing. Hence, the AoI changes periodically and the optimal dynamic price also changes periodically.

VII. Numerical Results

In this section, we perform experiments to study the proposed pricing polices. We consider a time interval of $T = 100$ (days) and normalize the maximum valuation to $\vartheta_{\max} = 1$.

In Fig. 3, we plot the price evolution under three types of optimal pricing policies with different parameters. Fig. 3(a) shows that under the optimal uniform pricing policy, as the unit sampling cost c increases, the optimal uniform price decreases

ISBN 978-3-903176-37-9 © 2021 IFIP



(a) Optimal uniform pricing policy.

(b) Optimal dual pricing policy.

(c) Optimal dynamic pricing policy.

Fig. 3. Price change under three types of pricing policies.

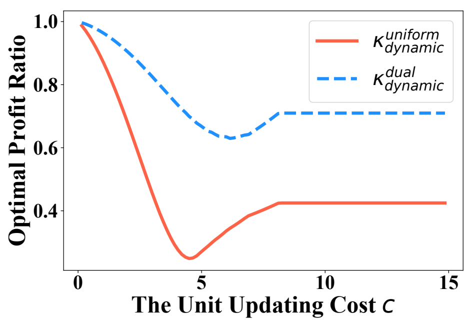


Fig. 4. Profit ratio under unit sampling cost c.

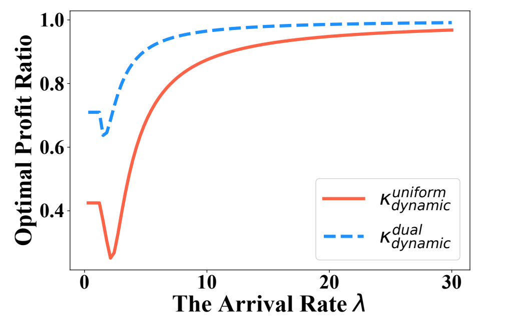


Fig. 5. Profit ratio under user arrival rate λ.

to a fixed value. This is because when the unit sampling cost is too high, the platform will choose the zero-update policy. In Fig. 3(b), as the unit sampling cost c, the optimal dual price decrease to two different values. In Fig. 3(c), as the arrival rate λ increases, the lowest price in the optimal pricing policy and the number of price change cycles increase. This is because the platform will choose more updates as users arrive more often.

Next, we compare the maximum profits under the three optimal pricing policies. The platform's profit is affected by both the sampling policy and the price flexibility of the pricing policies. Let Π^*_{uniform}, Π^*_{dual}, and Π^*_{dynamic} denote the maximum profit under the uniform pricing policy, the dual pricing policy, and the dynamic pricing policy, respectively. We first define the profit ratios as follows:

$$\kappa^{\text{uniform}}_{\text{dynamic}} = \frac{\Pi^*_{\text{uniform}}}{\Pi^*_{\text{dynamic}}} \quad \text{and} \quad \kappa^{\text{dual}}_{\text{dynamic}} = \frac{\Pi^*_{\text{dual}}}{\Pi^*_{\text{dynamic}}}.$$

Fig. 4 shows how the profit ratios change with the varying value of unit sampling cost c. The results are as follows.

- When c approaches 0, both $\kappa^{\text{uniform}}_{\text{dynamic}}$ and $\kappa^{\text{dual}}_{\text{dynamic}}$ will converge to 1. This is because the optimal number of updates goes to infinity, and the data will keep the freshest for all the time. This suggests that when facing a near-zero sampling cost, the platform can get almost the maximum profit by just using the simple uniform pricing policy.
- When c increases from 0, both $\kappa^{\text{uniform}}_{\text{dynamic}}$ and $\kappa^{\text{dual}}_{\text{dynamic}}$ initially decrease. In this case, the platform updates the data less frequently due to a higher sampling cost. The optimal dynamic pricing policy with the highest price flexibility can better exploit the AoI-sensitivity of users' valuations. Hence, it can produce more profit compared to the uniform policy and the dual policy. After $\kappa^{\text{uniform}}_{\text{dynamic}}$ and $\kappa^{\text{dual}}_{\text{dynamic}}$ reach the lowest points, they will rise as c increases. This happens when the data freshness is relatively low and the difference of data prices under the three types of optimal pricing policies narrows. Hence, the profit gaps of three optimal pricing polices also narrow.
- When c is large (more than 10 in Fig. 4), both $\kappa^{\text{uniform}}_{\text{dynamic}}$ and $\kappa^{\text{dual}}_{\text{dynamic}}$ no longer change in c. This is because the platform will choose a zero-update sampling policy under all three pricing policies due to the high sampling cost. Since the data is not always the freshest, different pricing policies produce different profits and the ratios are not equal to one.

Fig. 5 shows how the profit ratios change with the arrival rate λ, with results shown as follows.

- When λ is close to 0, both $\kappa^{\text{uniform}}_{\text{dynamic}}$ and $\kappa^{\text{dual}}_{\text{dynamic}}$ do not change with λ. This is because the platform will choose a zero-update sampling policy under all three pricing policies.
- When λ increases from 0, both $\kappa^{\text{uniform}}_{\text{dynamic}}$ and $\kappa^{\text{dual}}_{\text{dynamic}}$ decrease. This is because the platform can make more profit by selling the same fresh data to more users. Hence, the platform can afford to have more updates, and the data

ISBN 978-3-903176-37-9 © 2021 IFIP

freshness increases. Since the optimal dynamic pricing can better exploit users' time sensitivity and make more profit, two ratios keep decreasing.

- When λ further increases to infinity, both $\kappa_{\text{dynamic}}^{\text{uniform}}$ and $\kappa_{\text{dynamic}}^{\text{dual}}$ will increase and approach 1. As the optimal number of updates goes infinity, the data will almost always keep the freshest. The maximum profit under the three types of pricing policies will converge to the same limit.

Moreover, Fig. 4 and Fig. 5 show that the optimal dual pricing policy significantly outperforms the optimal uniform pricing policy. Specifically, the optimal dual pricing policy produces up to 280% of the profit that is achieved by the optimal uniform pricing policy.

VIII. Conclusion

In this paper, we proposed a fresh data market, where a platform sells data with different AoI values to dynamically arriving users. We proposed three types of pricing policies, i.e., the uniform pricing policy, the dual pricing policy, and the dynamic pricing policy. The joint data sampling and pricing problem is a challenging non-smooth mixed integer programming problem. For the uniform pricing and dual pricing, we tackle the challenge by analyzing a relaxed problem, and we prove that the optimal solution of the relaxed problem is also optimal for our original problem. Several interesting practical insights emerge. First, the optimal data prices decrease in the unit sampling cost and increase in the users' arrival rate. Second, for all three pricing policies, the equal-spacing data sampling policy is optimal. Third, numerical results show that the optimal dual pricing policy significantly outperforms the optimal uniform pricing policy.

In the future work, we will study how multiple data sources affect the platform's data sampling and pricing polices, as well as the users' data purchasing behaviors. Multiple data sources decide how much the platform should pay for each update. The platform decides which sources to choose, the data sampling policy, and the data pricing policy, considering the unit price decided by the data sources and the users purchasing behaviors. Furthermore, we can consider the scenario where the platform does not know the sources' update generating costs and use the contract theory to deal with such information asymmetry.

References

[1] S. He, D.-H. Shin, J. Zhang, J. Chen, and P. Lin, "An exchange market approach to mobile crowdsensing: pricing, task allocation, and walrasian equilibrium," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 921–934, 2017.

[2] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "Trading data in the crowd: Profit-driven data acquisition for mobile crowdsensing," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 2, pp. 486–501, 2017.

[3] T. Jung, X. Li, W. Huang, J. Qian, L. Chen, J. Han, J. Hou, and C. Su, "Accounttrade: Accountable protocols for big data trading against dishonest consumers," *IEEE INFOCOM*, pp. 1–9, 2017.

[4] H. Oh, S. Park, G. M. Lee, J. K. Choi, and S. Noh, "Competitive data trading model with privacy valuation for multiple stakeholders in iot data markets," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3623–3639, 2020.

[5] W. Mao, Z. Zheng, and F. Wu, "Pricing for revenue maximization in iot data markets: An information design perspective," in *IEEE INFOCOM*. IEEE, 2019, pp. 1837–1845.

[6] Y. Sun, I. Kadota, R. Talak, and E. Modiano, "Age of information: A new metric for information freshness," *Synthesis Lectures on Communication Networks*, vol. 12, no. 2, pp. 1–224, 2019.

[7] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1183–1210, 2021.

[8] A. Kosta, N. Pappas, and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, pp. 162–259, 2017.

[9] S. K. Kaul, R. D. Yates, and M. Gruteser, "Status updates through queues," in *Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2012, pp. 1–6.

[10] Y. Sun, E. Uysal-Biyikoglu, R. Yates, C. Koksal, and N. Shroff, "Update or wait: How to keep your data fresh," *IEEE INFOCOM*, pp. 1–9, 2016.

[11] A. Bedewy, Y. Sun, S. Kompella, and N. Shroff, "Age-optimal sampling and transmission scheduling in multi-source systems," *ACM Mobihoc*, pp. 121–130, 2019.

[12] Y.-P. Hsu, E. Modiano, and L. Duan, "Scheduling algorithms for minimizing age of information in wireless broadcast networks with random arrivals," *IEEE Transactions on Mobile Computing*, vol. 19, no. 12, pp. 2903–2915, 2019.

[13] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2637–2650, 2018.

[14] N. Lu, B. Ji, and B. Li, "Age-based scheduling: Improving data freshness for wireless real-time traffic," in *ACM Mobihoc*, 2018, pp. 191–200.

[15] M. Bastopcu and S. Ulukus, "Information freshness in cache updating systems," *IEEE Transactions on Wireless Communications*, 2020.

[16] P. R. Jhunjhunwala and S. Moharir, "Age-of-information aware scheduling," in *International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2018, pp. 222–226.

[17] C. Joo and A. Eryilmaz, "Wireless scheduling for information freshness and synchrony: Drift-based design and heavy-traffic analysis," *IEEE/ACM transactions on networking*, vol. 26, no. 6, pp. 2556–2568, 2018.

[18] B. T. Bacinoglu, Y. Sun, E. Uysal, and V. Mutlu, "Optimal status updating with a finite-battery energy harvesting source," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 280–294, 2019.

[19] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," *IEEE International Symposium on Information Theory (ISIT)*, pp. 1122–1126, 2017.

[20] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 534–556, 2019.

[21] B. Li and J. Liu, "Achieving information freshness with selfish and rational users in mobile crowd-learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1266–1276, 2021.

[22] X. Wang and L. Duan, "Dynamic pricing for controlling age of information," in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2019, pp. 962–966.

[23] M. Zhang, A. Arafa, J. Huang, and H. V. Poor, "Pricing fresh data," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1211–1225, 2021.

[24] M. Zhang, A. M. Arafa, E. Wei, and R. Berry, "Optimal and quantized mechanism design for fresh data acquisition," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 5, pp. 1226–1239, 2021.

[25] Z. Wang, L. Gao, and J. Huang, "Taming time-varying information asymmetry in fresh status acquisition," *IEEE INFOCOM*, pp. 1–9, 2021.

[26] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 204–219, 2019.

[27] W. Zhao and Y.-S. Zheng, "Optimal dynamic pricing for perishable assets with nonhomogeneous demand," *Management science*, vol. 46, no. 3, pp. 375–388, 2000.

[28] J. He, Q. Ma, M. Zhang, and J. Huang, "Optimal fresh data sampling and trading: Online appendix," 2021. [Online]. Available: https://www.dropbox.com/s/c4z9bflpdp5g777/appendix.pdf?dl=0