Optimal Mechanism Design for Fresh Data Acquisition

Meng Zhang†, Ahmed Arafa*, Ermin Wei†, and Randall A. Berry†
†Department of Electrical and Computer Engineering, Northwestern University
*Department of Electrical and Computer Engineering, University of North Carolina at Charlotte

***Abstract*—In this paper, we study a fresh data acquisition problem to acquire fresh data and optimize the age-related performance when strategic data sources have private market information. We consider an information update system in which a destination acquires, and pays for, fresh data updates from a source. The destination incurs an age-related cost, modeled as a general increasing function of the *age-of-information* (AoI). The source is strategic and incurs a sampling cost, which is its private information and may not be truthfully reported to the destination. To this end, we design an optimal (economic) mechanism for timely information acquisition by generalizing Myerson's seminal work. The goal is to minimize the sum of the destination's age-related cost and its payment to the source, while ensuring that the source truthfully reports its private information and will voluntarily participate in the mechanism. Our results show that, under some distributions of the source's cost, our proposed optimal mechanism can lead to an unbounded benefit, compared against a benchmark that naively trusts the source's report and thus incentivizes its maximal over-reporting.**

I. Introduction

The rapidly growing number of mobile devices and the dramatic increase in real-time applications have driven interest in fresh data as measured by the *age-of-information* (AoI) [1], [2]. Real-time applications in which fresh data is critical include real-time monitoring, data analytics, and vehicular networks. For example, real-time traffic information and the speed of vehicles is crucial in autonomous driving and unmanned aerial vehicles. Another example is real-time mobile crowd-sensing (or mobile crowd-learning [3]) applications, in which a platform is fueled by mobile users' participatory contribution of real-time data. This class of examples includes real-time traffic congestion and accident information on (e.g., Google Waze [4]) and real-time location information for scattered commodities and resources (e.g., GasBuddy [5]).

Keeping data fresh relies on frequent data generation, processing, and sampling, which can lead to significant (sampling) costs for the data source. In practice, data sources (i.e., fresh data contributors) are *self-interested* in the sense that they may have their own interests different from those of data destinations (i.e., fresh data requestors). Consequently, the participation of sources relies on proper incentives from the destination. The resulting economic interactions between sources and destinations constitute *fresh data markets*, which have been studied in [3], [6]–[8].

The existing studies on fresh data markets [3], [6]–[8] designed incentives assuming complete information. A crucial economic challenge not addressed in these works is dealing with *market information asymmetry*. Specifically, sources in practice may have private (market) information (e.g., sampling cost and data freshness) that is unknown by others. Therefore, they may manipulate the outcome of the system (e.g., their subsidies and the scheduling policies) by misreporting such private information to their own advantages. To the best of our knowledge, no existing work has addressed fresh data markets with such asymmetric information. Motivated by the above issue, this work aims to solve the following key questions:

Question 1. *How bad can self-interested reports (of sources' cost) be in a fresh data acquisition system?*

Question 2. *How should a destination acquire fresh data from self-interested sources with market information asymmetry?*

A. Challenges and Solution Approach

Existing related studies on information asymmetry in data markets (without considering data freshness) have identified two different levels of possible manipulation [9]–[14], depending on whether data is *verifiable*, i.e., whether the destination can verify the authenticity (or freshness) of data. These two levels of manipulation are:

1) *Market information misreporting.* For *verifiable data*, a source may benefit from misreporting its cost and quality information (as in, e.g., [9]–[13]).
2) *Data fraud.* For *unverifiable data*, a source may even fake the data itself, e.g., by sending dummy data to avoid incurring corresponding costs (as in, e.g., [14]).

As a first step towards tackling a fresh data market with asymmetric information, this work focuses on the first type of manipulation due to misreporting private cost information and assumes *verifiable fresh data.* Even this level of misreporting is challenging and may lead to an arbitrarily bad loss, as we will analytically show in Section III-C.

In the economics literature, a standard approach for designing markets with asymmetric information is via the *optimal mechanism design* approach of Myerson [32]. Many standard optimal mechanism design problems are linear and can be reduced to computing a "posted price" (e.g., [32]). Different from the standard setting, our fresh data market framework features a non-linear age-related cost. This nature of AoI requires a new design of optimal mechanisms and problem formulations.

We summarize our contributions as follows:

- *Fresh Data Market Modeling with Private Cost Information.* We develop a new analytical model for a fresh data

This work was supported in part by NSF grants CNS-1908807, AST-2037838, and ECCS-2030251.

market with private cost information and allow a source to strategically misreport this information. To the best of our knowledge, this is the first work in the AoI literature to address market information asymmetry.
- *Mechanism Design.* We first show that the existence of an optimal mechanism with special structures and then reformulate the mechanism design problem to find such an optimal solution. The infinite-dimensional nonlinear nature of the problem makes it different from the standard setting. We then solve the problem using tools from infinite dimension functional optimization and analytically derive the optimal solution.
- *Performance Comparison.* Our analytical and numerical results show that, when the sampling cost is exponentially distributed, the performance gains of our optimal mechanism can be unbounded compared against a benchmark that naively trusts whatever the source reports.

II. Related Work

Age-of-Information: The AoI metric has been introduced and analyzed in various contexts in the recent years, see, e.g., [1], [2], [21]–[31]. Of particular relevance to this work are those pertaining to the economics of fresh data and information [3], [6]–[8]. The most closely-related studies to ours are in [3], [7], which consider systems with destinations using dynamic pricing schemes to incentivize sensors to provide fresh updates. The sources in [3], [7] are *myopic* instead of *forward-looking*, i.e., in our case the source considers its longer term payoff. None of this prior work has considered the role of private market information as we do here.

Optimal Mechanism Design: There exists a rich economics literature on optimal mechanism design (e.g., [32] and surveys in [33], [34]). Our approach is based on Myerson's characterization of incentive compatibility and optimal mechanism [32]. However, existing mechanisms cannot be directly applied here due to differences in the problem setting, in particular, the infinite dimension nature in our setting.

Information Acquisition: There has been a recent line of work on viewing data as an economic good. A growing amount of attention has been placed on understanding the interactions between the strategic nature of data holders and the statistical inference and learning tasks that use data collected from these holders (e.g., [9]–[14]). In this line of research, a data collector designs mechanisms with payments to incentivize data holders to reveal data. Other related studies include strategic information transmission (e.g., [15]–[20]). However, none of the studies in this line considered data freshness.

III. System Model and Problem Formulation

A. System Model

We consider an information update system in which one data source (such as an Internet-of-Things device) generates data packets and sends them to one destination.

1) Data Updates and AoI: We consider a generate-at-will model (as in, e.g., [22], [26]), in which the source is able to generate and send a new update when requested by the destination. We assume instant update arrivals at the destination, with negligible transmission delay (as in, e.g. [26]). Let $\mathcal{X} \triangleq \{x_k\}_{k\in\mathbb{N}}$ be the update policy requested by the destination, where every $x_k \geq 0$ denotes the interarrival time between the $(k-1)$-th and k-th updates.

The AoI at time t is defined as [2]

$$\Delta_t(\mathcal{X}) = t - U_t, \tag{1}$$

where U_t is the time stamp of the most recently received update before time t, i.e., $U_t = \max_k[\sum_{j=1}^{k} x_j \leq t]$.

2) Destination's AoI Cost: The destination's AoI cost is given by $g(\Delta_t(\mathcal{X}))$, which is a general increasing function of $\Delta_t(\mathcal{X})$ and satisfies $g(0) = 0$.

3) Source's Sampling Cost and Private Information: We denote the source's unit *sampling cost* by c for each update, which is the source's *private information*. We consider a Bayesian setting in which the sampling cost is drawn from $\mathcal{C} = [\underline{c}, \bar{c}]$. Let $\Gamma(c)$ be the cumulative distribution function (CDF) and $\gamma(c)$ be the probability density function (PDF); we assume that this prior distribution is known by the destination.

B. Mechanism Design and Problem Formulation

1) Mechanism Overview: The destination designs an (economic) mechanism for acquiring the source's truthful report of its sampling cost and data updates. The source reports (potentially misreports) its sampling cost just once at the beginning. A mechanism takes the source's report of its sampling cost as the input of the update policy and the monetary reward to the source. Mathematically, a general mechanism $m = (\mathcal{P}, \mathcal{X})$ is a tuple of a payment rule $\mathcal{P}$ and the update policy $\mathcal{X}$. The prices (i.e., rewards) can be different across different updates. That is, $\mathcal{P} = \{p_k\}_{k\in\mathbb{N}}$, where $p_k : \mathcal{C} \to \mathbb{R}_+$. Functions $\mathcal{P}$ and $\mathcal{X}$ are functions of the source's reported cost $\tilde{c}$.

2) Source's Payoff: A strategic source aims to maximize its (long-term time average) payoff, defined as

$$P(\tilde{c}, m) = \lim_{K\to\infty} \frac{\sum_{k=1}^{K} p_k(\tilde{c}) - c \cdot K}{\sum_{k=1}^{K} x_k(\tilde{c})}. \tag{2}$$

Note that, this differs from [3], [7], in which the considered sources are not *strategic*. Instead, they are *price-taking*, i.e., not maximizing their respective long-term objectives.

3) Destination's Overall Cost: Since the destination only knows the statistical information of c, its objective is its expected (long-term time average) overall cost:

$$J(m) = \mathbb{E}_c\left[\lim_{K\to\infty} \frac{\sum_{k=1}^{K} G\left(x_k(\tilde{c}^*(m))\right) + p_k\left(\tilde{c}^*(m)\right)}{\sum_{k=1}^{K} x_k(\tilde{c}^*(m))}\right], \tag{3}$$

where $G(x)$ is the destination's cumulative AoI cost for an interarrival time x, defined as

$$G(x) \triangleq \int_0^x g(\Delta_t) d\Delta_t, \tag{4}$$

and $\tilde{c}^*(m)$ is the source's optimal reporting strategy in response to m, i.e., $\tilde{c}^*(m) \in \arg\max_{\tilde{c}\in\mathcal{C}} P(\tilde{c}, m)$.

The source may have an incentive to misreport its private information $\tilde{c}$. However, according to the *revelation principle* [32], for any mechanism m, there exists an incentive compatible (i.e., truthful) equivalence $\tilde{m}$, such that $J(m) = J(\tilde{m})$. This allows us to replace all $\tilde{c}^*(m)$ in (3) by c, restrict our attention to incentive compatible (IC) mechanisms, and impose the following constraint:

$$\text{IC}: \quad c \in \arg\max_{\tilde{c}\in\mathcal{C}} P(\tilde{c}, m), \ \forall c \in \mathcal{C}. \tag{5}$$

Further, a mechanism should satisfy an *(interim) individual rationality (IR)* constraint. An IC mechanism satisfies IR if:

$$\text{IR}: \quad P(c, m) \geq 0, \ \forall c \in \mathcal{C}. \tag{6}$$

4) Problem Formulation: The destination seeks to find a mechanism m to minimize its overall cost:

$$\min_{m} \quad J(m) \tag{7a}$$
$$\text{s.t.} \quad \text{IC in (5) and IR in (6).} \tag{7b}$$

This is a challenging optimization problem as the space of all mechanisms is infinite dimensional and further the constraints in (5) *and* (6) *are non-trivial.*

We will now show that a special, simplified, class of mechanisms, m satisfying (5) and (6) is optimal.

Definition 1 (Equal-Spacing and Flat-Rate Mechanism)**.** *A mechanism $m = (\mathcal{P}, \mathcal{X})$ is **equal-spacing** and **flat-rate** if*

$$p_k(\cdot) = p(\cdot) \text{ and } x_k(\cdot) = x(\cdot) \ ,\forall k \in \mathbb{N}, \tag{8}$$

for some functions $p : \mathcal{C} \to \mathbb{R}_+$ and $x : \mathcal{C} \to \mathbb{R}_+$.

Lemma 1. *There exists an optimal mechanism $m^* = (\mathcal{P}^*, \mathcal{X}^*)$ that is equal-spacing and flat-rate, satisfying* (8)*.*

Due to space limits, we present the detailed proofs of all lemmas and theorems in the extended version of this work [38]. The proof of Lemma 1 involves showing that, for any optimal mechanism m^*, we can always construct an equal-spacing and flat-rate mechanism that yields at most the same objective value. This is mainly done by leveraging the convexity of $G(\cdot)$. By Lemma 1, we can now drop the index k in p_k and x_k.

C. Naive Mechanism

In this subsection, we introduce a *naive mechanism* that satisfies Definition 1. We use this to show that such a mechanism can lead to an arbitrarily large cost for the destination when $g(x) = x^\alpha$, $\alpha > 0$.

Example 1 (Naive Mechanism)**.** *The destination subsidizes the source's reported cost; the update policy rule $x_N(\tilde{c})$ aims at minimizing its overall cost in* (3)*, naively assuming the source's report is truthful, i.e.,*

$$p_N(\tilde{c}) = \tilde{c}, \tag{9a}$$
$$x_N(\tilde{c}) = \arg\min_{x\geq 0} \frac{x^{\alpha+1}/(\alpha+1) + p^N(\tilde{c})}{x}. \tag{9b}$$

Solving (9b) further gives $x_N(\tilde{c}) = \left[\left(1 + \frac{1}{\alpha}\right) \cdot \tilde{c}\right]^{\frac{1}{1+\alpha}}$. Given this naive mechanism, the source solves the following reporting problem:

$$\tilde{c}^* = \arg\max_{\tilde{c}\in\mathcal{C}} \ \frac{\tilde{c} - c}{\left[\left(1 + \frac{1}{\alpha}\right) \cdot \tilde{c}\right]^{1/(1+\alpha)}}, \tag{10}$$

whose solution can be shown to be given by $\tilde{c}^* = \bar{c}$, i.e., the optimal reporting strategy is to report the maximal possible value. This makes the destination's overall cost be given by $\left[\bar{c}\left(1 + \frac{1}{\alpha}\right)\right]^{\frac{\alpha}{1+\alpha}}$. Note that the ratio of the destination's objectives under the source's optimal report and the true cost is $\left(\frac{\bar{c}}{c}\right)^{\frac{\alpha}{1+\alpha}}$, which can be arbitrarily large as $\bar{c}$ approaches infinity. Misreports leading to an arbitrarily large cost to the destination motivates the optimal mechanism design in the next section.

IV. Optimal Mechanism Design

In this section, we use the results of Lemma 1 to reformulate (7) and characterize the IC and the IR constraints in (5) and (6). The optimal mechanism design problem is then reduced to an infinite-dimensional optimization problem, which we analytically solve and derive useful insights.

A. Problem Reformulation

Lemma 1 allows us to focus on the equal-spacing and flat-rate mechanism (i.e., $m = (p, x)$). To further facilitate our analysis, we use $f(\tilde{c})$ to denote the *update rate rule* and $h(\tilde{c})$ to denote the *payment rate rule* such that

$$h(\tilde{c}) \triangleq \frac{p(\tilde{c})}{x(\tilde{c})} \text{ and } f(\tilde{c}) \triangleq \frac{1}{x(\tilde{c})}, \ \forall \tilde{c} \in \mathcal{C}. \tag{11}$$

Since (11) defines a one-to-one mapping between (p, x) and (f, h), we can focus on $m = (f, h)$ in the following and then derive the optimal (p^*, x^*) based on the optimal (f^*, h^*).

B. Characterization of IC and IR

1) Incentive Compatibility: We can characterize the IC constraint in (5) based on Myerson's work [32].

Theorem 1. *A mechanism $m = (f, h)$ is incentive compatible if and only if the following two conditions are satisfied:*

1) *$f(c)$ is non-increasing in $c \in \mathcal{C}$;*
2) *$h(c)$ has the following form:*

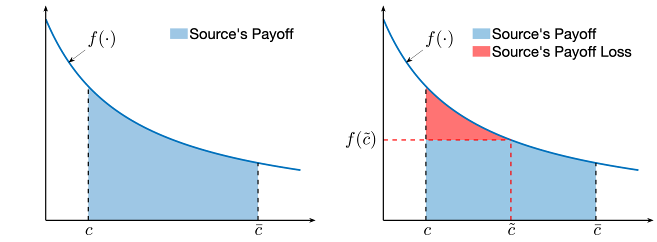
$$h(c) = c \cdot f(c) - \int_{\underline{c}}^{c} f(z)dz + A, \tag{12}$$

for some constant $A \in \mathbb{R}$ (here, A does not depend on c but may depend on $f(\cdot)$.)

2) Individual Rationality: Given an arbitrary incentive compatible mechanism satisfying (12), to further satisfy the IR constraint in (6), we have that the minimal A for the incentive compatible mechanism in Theorem in 1 is

$$A = \int_{\underline{c}}^{\bar{c}} f(z)dz. \tag{13}$$

We will assume that this choice of A is used in the following.



(a) (b)

Fig. 1: Illustration of IC and IR under a mechanism satisfying (12) and (13): The source's payoff comparison between (a) a truthful report ($\tilde{c} = c$) and (b) an over-report ($\tilde{c} > c$).

We present an example in Fig. 1 to illustrate (12) and (13). Under a non-increasing $f(\cdot)$ and $h(\cdot)$ satisfying (12) and (13), a truthfully reporting source receives a payoff of $\int_c^{\bar{c}} f(t)dt$, as shown in Fig. 1 (a); when the source reports $\tilde{c}$, its payoff is $(\tilde{c}-c)f(\tilde{c})+\int_{\tilde{c}}^{\bar{c}} f(t)dt$. As shown in Fig. 1 (b), such an over-report incurs a payoff loss. Similarly, an under-report would also incur a payoff loss. This demonstrate IC. In addition, the source's payoff is always non-negative for any c and becomes 0 only when c approaches $\bar{c}$. This demonstrates IR.

C. Mechanism Optimization Problem

Based on (12) and (13), we can focus on optimizing the update rate function $f(c)$ only in what follows. Since $\int_{\underline{c}}^{\bar{c}} |f(c)|^2 d\Gamma(c) < +\infty$, the update rate function $f(\cdot)$ lies in the Hilbert space $L^2(\Gamma)$ associated to the CDF $\Gamma(c)$.

By Theorem 1, we transform the destination's problem into

$$\min_{f(\cdot)} J(f) \triangleq \mathbb{E}_c\left[G\left(\frac{1}{f(c)}\right)f(c) + c\cdot f(c) + \int_c^{\bar{c}} f(z)dz\right] \tag{14a}$$

$$\text{s.t. } f(\cdot) \in \mathcal{F} \triangleq \{f(\cdot) : f(c) \geq 0, f'(c) \leq 0,\ \forall c \in \mathcal{C}\}. \tag{14b}$$

This is a functional optimization problem. To derive insightful results, we first relax the constraint in (14b) and then show when such a relaxation in fact leads to a feasible solution $f^*(\cdot)$ (i.e., when it automatically satisfies (14b)).

We introduce the definition of the source's *virtual cost* analog to the standard definition of virtual value in [32]:

$$\phi(c) \triangleq c + \frac{\Gamma(c)}{\gamma(c)}, \tag{15}$$

which allows us to transform the destination's problem as in the following lemma:

Lemma 2. *The objective in* (14a) *can be rewritten as*

$$J(f) = \mathbb{E}_c\left[G\left(\frac{1}{f(c)}\right)f(c) + f(c)\phi(c)\right]. \tag{16}$$

The proof of Lemma 2 simply involves changing the order of integration. If we relax the constraint $f'(c) \leq 0$, Lemma 2 makes the problem in (14) decomposable across every $c \in \mathcal{C}$. Each subproblem is given by

$$\min_{f(c)} G\left(\frac{1}{f(c)}\right)f(c) + f(c)\phi(c), \tag{17}$$

which can be solved separately. We are now ready to introduce the solution to problem (14):

Theorem 2. *If $\phi(c)$ is non-decreasing, the optimal mechanism $m^* = (f^*, h^*)$ satisfies* (12), (13)*, where $f^*(\cdot)$ satisfies*

$$\underbrace{g\left(\frac{1}{f^*(c)}\right)\frac{1}{f^*(c)} - G\left(\frac{1}{f^*(c)}\right)}_{\text{Marginal AoI Cost Reduction}} = \underbrace{\phi(c)}_{\text{Virtual Cost}},\ \forall c \in \mathcal{C}. \tag{18}$$

To comprehend the above results, the optimal $f^*(\cdot)$ equalizes the marginal AoI cost reduction and the virtual cost $\phi(c)$ for all $c \in \mathcal{C}$. To see when (18) yields a feasible solution satisfying (14b), note that there always exists a unique positive value of the optimal $f^*(c)$ for each c in (18), and so the optimal $f^*(c)$ for each c in (18) is well defined. In addition, if $\phi(c)$ is non-decreasing in c, $f^*(c)$ is non-increasing in c.[1] A non-decreasing virtual cost is in fact satisfied for a wide range of distributions of the source's sampling cost. We will focus on such specific distributions in Section V and generalize Theorem 2 to the more general (potentially not monotonic) virtual cost case in [38].

V. Performance Comparison

In this section, we present analytical and numerical studies to understand when the optimal mechanism in Theorem 2 is most beneficial, compared against the naive mechanism in (9).

We assume a power age cost function [22]: $g(x) = x^\alpha$ for some $\alpha \in (0,\infty)$. In the following, we consider both a uniform distribution and a truncated exponential distribution of the source's sampling cost.[2]

A. Uniform Distribution

We first compare the performance under a uniform distribution of the sampling cost on the interval $[\underline{c}, \bar{c}]$. The naive mechanism leads to an overall cost of the destination of

$$J_N = \left[\bar{c}\left(1+\frac{1}{\alpha}\right)\right]^{\frac{\alpha}{1+\alpha}}. \tag{19}$$

The complete information benchmark leads to an overall cost:

$$J_C = \frac{\bar{c}^{\frac{1+2\alpha}{1+\alpha}} - \underline{c}^{\frac{1+2\alpha}{1+\alpha}}}{\bar{c}-\underline{c}}\,\frac{1+\alpha}{1+2\alpha}\left(1+\frac{1}{\alpha}\right)^{\frac{\alpha}{1+\alpha}}. \tag{20}$$

Hence, we have

$$\frac{J_N}{J_C} = \left(1+\frac{\alpha}{1+\alpha}\right)\left[\frac{(\bar{c}-\underline{c})\bar{c}^{\frac{\alpha}{1+\alpha}}}{\bar{c}^{\frac{1+2\alpha}{1+\alpha}} - \underline{c}^{\frac{1+2\alpha}{1+\alpha}}}\right] \leq 1+\frac{\alpha}{1+\alpha}, \tag{21}$$

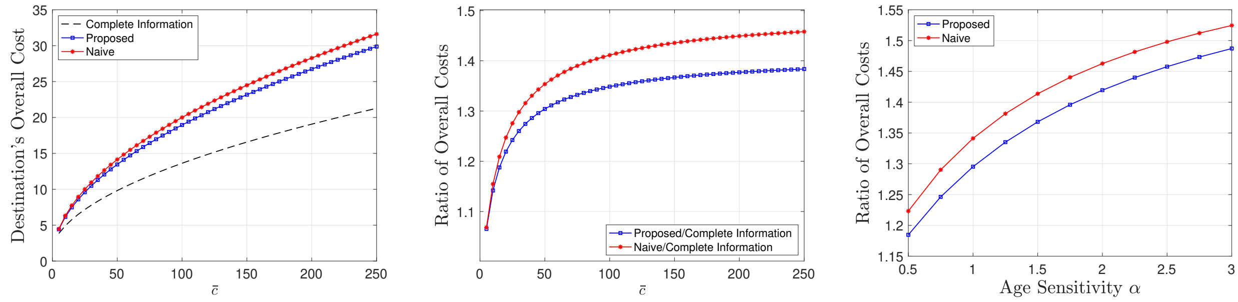
indicating that, under the uniform distribution, the naive mechanism incurs a bounded loss due to private information.

On the other hand, the optimal mechanism in Theorem 2 leads to an overall cost of

$$J^* = \left[2\left(1+\frac{1}{\alpha}\right)\right]^{\frac{\alpha}{1+\alpha}}\left(\frac{1+\alpha}{1+2\alpha}\right)\frac{\left\{(c-\frac{\underline{c}}{2})^{\frac{1+2\alpha}{1+\alpha}}\right\}\Big|_{\underline{c}}^{\bar{c}}}{\bar{c}-\underline{c}}. \tag{22}$$

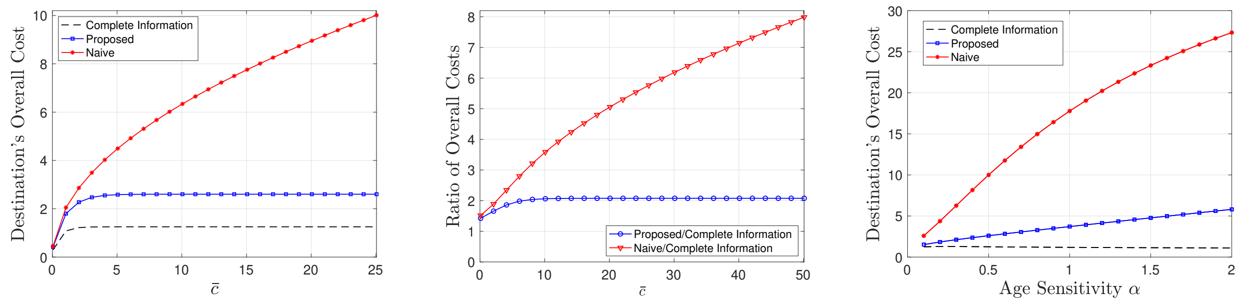
[1] The condition of the virtual cost $\phi(c)$ being non-decreasing is known as the regularity condition in [32].

[2] These two distributions of costs are also considered in [37].



(a) (b) (c)

Fig. 2: Performance comparison under the uniform distribution with different upper bounds of the sampling cost $\bar{c}$ in (a), (b) and with different age sensitivity coefficients α in (c). We set $\alpha = 1$ in (a) and (b) and $\bar{c} = 30$ in (c).



(a) (b) (c)

Fig. 3: Performance comparison under the truncated exponential distribution with different upper bounds of the sampling cost $\bar{c}$ in (a), (b) and with different age sensitivity coefficients α in (a), (c).

It then follows that

$$\frac{J^*}{J_C} = \frac{\left\{ \left(c - \frac{c}{2}\right)^{\frac{1+2\alpha}{1+\alpha}} \right\} \Big|_{\underline{c}}^{\bar{c}}}{\bar{c}^{\frac{1+2\alpha}{1+\alpha}} - \underline{c}^{\frac{1+2\alpha}{1+\alpha}}} 2^{\frac{\alpha}{1+\alpha}} \leq 2^{\frac{\alpha}{1+\alpha}}. \tag{23}$$

Equations (21) and (23) imply that, under the uniform distribution, the performance gain of the optimal mechanism compared to the naive mechanism is limited.

In Fig. 2, we numerically compare the performances of the proposed optimal mechanism, the naive mechanism, and the complete information benchmark. We observe a relatively small gap between the proposed optimal mechanism and the naive mechanism in Fig. 2(a). In Fig. 2(b), both the proposed optimal mechanism and the naive mechanism approach their upper bounds in (21) and (23).

B. Truncated Exponential Distribution

In this subsection, we consider an exponential distribution of the sampling cost truncated on the interval $[0, \bar{c}]$, i.e., assuming $\underline{c} = 0$. The corresponding PDF is

$$\gamma(c) = \frac{\frac{1}{\lambda}\exp(-c/\lambda)}{1 - \exp(-\bar{c}/\lambda)}, \tag{24}$$

and we fix $\lambda = 1$. Note that the performance of the naive mechanism only depends on $\bar{c}$ instead of the specific distribution of c. Hence, the overall cost is the same as in (19).

The lower bound of the overall cost under complete information is:

$$J_C = \frac{(1+\frac{1}{\alpha})^{\frac{\alpha}{1+\alpha}}}{1 - e^{-\bar{c}}} \left[\Gamma\left(\frac{2\alpha}{1+\alpha} + \frac{1}{\alpha}, 0\right) - \Gamma\left(\frac{2\alpha+1}{1+\alpha}, \bar{c}\right) \right], \tag{25}$$

where $\Gamma(s,x) = \int_x^{\infty} t^{s-1} \exp(-t)\, \mathrm{d}t$. Note that (25) converges to a finite value when $\bar{c} \to \infty$. Finally, the optimal mechanism leads to an overall cost of:

$$J^* = \frac{(1+\frac{1}{\alpha})^{\frac{\alpha}{1+\alpha}}}{1 - e^{-\bar{c}}} \int_0^{\bar{c}} (t - 1 + \exp(t))^{\frac{a}{1+a}} \exp(-t) dt. \tag{26}$$

Fig. 3(a) shows that overall costs of the destination under both the optimal mechanism and the complete information benchmark converge to some constants as $\bar{c}$ increases. The naive mechanism in this case leads to an unbounded overall cost as $\bar{c}$ increases. In Fig. 3(b), we observe a small gap between the proposed optimal mechanism and the complete information benchmark. In particular, we have $J^*/J_C \approx 2$ when $\bar{c} \geq 10$. Therefore, we have shown that under the truncated exponential distribution, the proposed optimal mechanism can lead to unbounded benefits, compared against the naive mechanism. Fig. 3(c) shows that the proposed optimal mechanism becomes more beneficial when the destination is more sensitive to AoI, compared with the naive mechanism.

VI. Conclusions

We have studied the fresh information acquisition problem in the presence of private information. We have designed an optimal mechanism to minimize the destination's AoI cost and its payment to the source, while satisfying the truthful reporting and individual rationality constraints. Our analysis revealed that the proposed optimal mechanism may lead to an unbounded benefit, compared against a naive benchmark, though this gain depends on the distribution of the sampling cost.

References

[1] R. D. Yates, Y. Sun, D. R. Brown III, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Area Comm.*, vol. 39, no. 5, pp. 1183-1210, May 2021.

[2] S. Kaul, R. D. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, 2012.

[3] B. Li, and J. Liu, "Achieving information freshness with selfish and rational users in mobile crowd-learning?" *IEEE J. Sel. Area Comm.*, vol. 39, no. 5, pp. 1266-1276, May 2021.

[4] Waze Mobile App. [Online]. Available: https://www.waze.com/

[5] GasBuddy Mobile App. [Online]. Available: https://www.gasbuddy.com/

[6] X. Wang and L. Duan, "Dynamic pricing for controlling age of information," in *Proc. IEEE ISIT,* 2019.

[7] S. Hao and L. Duan, "Regulating competition in age of information under network externalities," *IEEE J. Sel. Area Comm.*, vol. 38, no. 4, pp. 697-710, 2020.

[8] M. Zhang, A. Arafa, J. Huang and H. V. Poor, "Pricing Fresh Data," *IEEE J. Sel. Area Comm.*, vol. 39, no. 5, pp. 1211-1225, May 2021.

[9] Y. Chen, N. Immorlica, B. Lucier, V. Syrgkanis, and J. Ziani, "Optimal data acquisition for statistical estimation," in *Proc. ACM EC*, 2018.

[10] J. Abernethy, Y. Chen, C.-J. Ho, and B. Waggoner. "Low-cost learning via active data procurement," in *Proc. ACM EC*, 2015.

[11] A. Roth and G. Schoenebeck, "Conducting truthful surveys, cheaply," in *Proc. ACM EC*, 2012.

[12] Y. Cai, C. Daskalakis, and C. H. Papadimitriou. "Optimum statistical estimation with strategic data sources," in *Proc. COLT*, 2015.

[13] M. Babaioff, R. Kleinberg, and R. Paes Leme. "Optimal mechanisms for selling information." in *Proc. ACM EC*, 2012.

[14] B. Waggoner, and Y. Chen, "Output agreement mechanisms and common knowledge," in *AAAI HCOMP*, 2014.

[15] V. Crawford and J. Sobel, "Strategic information transmission," *Econometrica*, vol. 50, pp. 1431–1451, 1982.

[16] M. Gentzkow and E. Kamenica, "Bayesian persuasion," *Amer. Econ. Rev.*, vol. 101, no. 6, pp. 2590–2615, 2011.

[17] P. Milgrom, "What the seller won't tell you: Persuasion and disclosure in markets," *J. Econ. Perspectives,* vol. 22, no. 2, pp. 115–131, 2008.

[18] H. Yamamoto, "A rate-distortion problem for a communication system with a secondary decoder to be hindered," *IEEE Trans. Inf. Theory,* vol. 34, no. 4, pp. 835–842, 1988.

[19] F. Farokhi, A. M. H. Teixeira and C. Langbort, "Estimation With Strategic Sensors," *IEEE Trans. Automatic Control*, vol. 62, no. 2, pp. 724-739, Feb. 2017.

[20] E. Akyol, C. Langbort, and T. Başar, "Information-theoretic approach to strategic communication as a hierarchical game," *Proc. IEEE*, vol. 105, no. 2, pp. 205-218, 2016.

[21] Q. He, D. Yuan, and A. Ephremides, "Optimal link scheduling for age minimization in wireless systems". *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5381-5394, July 2018.

[22] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff, "Update or wait: How to keep your data fresh," *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7492-7508, Nov. 2017.

[23] R. Talak, S. Karaman, and E. Modiano. "Optimizing information freshness in wireless networks under general interference constraints," in *Proc. ACM Mobihoc*, 2018.

[24] I. Kadota, A. Sinha, and E. Modiano. "Optimizing age of information in wireless networks with throughput constraints," in *Proc. IEEE INFOCOM*, 2018.

[25] X. Wu, J. Yang, and J. Wu, "Optimal status update for age of information minimization with an energy harvesting source," *IEEE Trans. on Green Commun. Netw.*, vol. 2, no.1, pp. 193–204, March 2018.

[26] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies", *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 534-556, Jan. 2020.

[27] B. T. Bacinoglu, Y. Sun, E. Uysal-Biyikoglu, and V. Mutlu. "Achieving the age-energy tradeoff with a finite-battery energy harvesting source," in *Proc. IEEE ISIT*, 2018.

[28] B. Zhou and W. Saad, "Optimal sampling and updating for minimizing age of information in the Internet of things", in *Proc. IEEE GLOBECOM*, 2018.

[29] M. Costa, M. Codreanu and A. Ephremides, "On the age of information in status update systems with packet management," in *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1897-1910, April 2016.

[30] M. A. Abd-Elmagid, N. Pappas and H. S. Dhillon, "On the Role of Age of Information in the Internet of Things," *IEEE Commun. Magazine*, vol. 57, no. 12, pp. 72-77, Dec. 2019.

[31] G. D. Nguyen, S. Kompella, C. Kam, J. E. Wieselthier, and A. Ephremides, "Information freshness over an interference channel: A game theoretic view," in *Proc. IEEE INFOCOM*, 2018.

[32] R. B. Myerson, "Optimal auction design." *Mathematics of operations research*, vol. 6, no. 1, pp. 58-73, Feb. 1981.

[33] T. Roughgarden, "Approximately optimal mechanism design: motivation, examples, and lessons learned." *ACM SIGecom Exch.* 2015.

[34] J. D. Hartline, "Mechanism design and approximation," Book draft, 2013.

[35] D. Butnariu and A. N. Iusem, "Totally convex functions for fixed points computation and infinite dimensional optimization." Norwell, Kluwer Academic Publisher, 2000.

[36] D. P. Bertsekas, Nonlinear Programming, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[37] J. Wang, J. Tang, D. Yang, E. Wang and G. Xue, "Quality-aware and fine-grained incentive mechanisms for mobile crowdsensing," in *Proc. IEEE ICDCS*, 2016.

[38] M. Zhang, A. Arafa, E. Wei, R. A. Berry, "Optimal and Quantized Mechanism Design for Fresh Data Acquisition", *IEEE J. Sel. Area Comm.*, vol. 39, no. 5, pp. 1226-1239, May 2021 [Online]. Available: https://arxiv.org/abs/2006.15751