AN ITERATIVE AUCTION MECHANISM FOR DATA TRADING

Xuanyu Cao†, Yan Chen†, and K. J. Ray Liu*  

†Dept. of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA  
*School of Electronic Engineering, University of Electronic Science and Technology of China, China

ABSTRACT

In the big data era, it is vital to allocate the vast amount of data to various users efficiently. However, the data agents (data owners, collectors and users) are self-interested and seek to maximize their own utilities instead of the overall system efficiency. In this paper, the data trading problem of a data market with multiple data owners, collectors and users is formulated and an iterative auction mechanism is proposed to coordinate the data trading. The proposed mechanism allows the selfish data agents to trade data efficiently and avoids direct access of the agents’ private information. We theoretically prove that the proposed mechanism can achieve the socially optimal allocation point. Moreover, we demonstrate that the mechanism satisfies appealing economic properties such as individual rationality and weakly balanced budget. Simulations as well as real data experiments validate the theoretical properties of the mechanism.

Index Terms— Iterative auction, data trading, optimization

I. INTRODUCTION

In the big data era, vast amount of data are generated and exploited by various agents. For example, numerous memes (such as Twitter hashtags) and advertisements are produced in online social networks. Many software/app developers may need certain online data (e.g., the click-through rate of some websites) to enhance the quality of their products. In such a circumstance, we face the problem of alloting data from the data owners (e.g., social networks/websites) to the data users (e.g., software developers). In fact, several data trading markets or companies have already emerged recently, such as the Data Marketplace and Big Data Exchange. However, these data markets are still at the incipient stage and lack appropriate regulations. Economically, the data agents are self-interested and seek to maximize their own utilities instead of the overall system efficiency. As such, a sophisticated mechanism is imperative to guide the agents to trade data efficiently.

The problem of coordinating data trading in a data market falls into the general topic of resource trading/allocation in networks, for which abundant works have been done in the past decades [1]–[9]. The most relevant resource allocation/trading problem to this paper is the privacy trading problem [10]–[14], in which a single data collector collects private data from multiple data owners. Various auction and contract theoretic mechanisms are proposed to maximize the utility of the single data collector. However, there are two limitations of existing models of data trading [12]–[14]. First, there is only one single data collector. This is not the case in most real-world data market, where multiple data collectors (such as many companies or groups like Big Data Exchange) often coexist and compete with each other. Second, in most data markets, the data collectors usually do not exploit the data by themselves. Instead, they often sell the data to data users, who are not capable of collecting and storing massive datasets but need data to develop projects or conduct research. In this paper, we resolve the these two limitations of existing works and investigate the data trading problem in a market with multiple data owners, collectors and users†.

In this work, we consider from a system designer’s perspective and are aimed at maximizing the overall social welfare. In practice, the data agents are usually self-interested and seek to maximize their own utilities instead of the overall system performance. In order to coordinate the data trading among multiple self-interested agents, we resort to the iterative auction mechanism [15], which has already been successfully applied to resource allocation in communication networks [16]–[19]. We propose an iterative auction mechanism for data trading and prove that it converges to the socially optimal operation point and satisfies economic properties including individual rationality and weakly balanced budget. Our theoretical results are validated through simulations and real data experiments.

II. MODEL

In this section, we describe the model of a data market with multiple data owners, collectors and users in detail. We consider a data market with $M$ data owners, $N$ data collectors and $L$ data users.

Suppose owner $m$ entitles collector $n$ to collect $x_{mn}$ amount of data, which is the maximum amount of data that collector $n$ can get from owner $m$. Due to the exposure of its data, the owner $m$ suffers a loss of $U_m(x_{mn})$, where $x_{mn} = [x_{m1}, ..., x_{mN}]$. This loss may stem from compromise of privacy or leakage of lucrative information/technologies. We assume that the data here are exclusive, i.e., the same data can only be assigned to one collector and one user. For example, software companies (data users) may need tailored data (e.g., click-through rate of specific web pages to monitor users’ feedback) to develop their own softwares. These data are useful only to this user and are useless for others, i.e., these data are exclusive. An extension to non-exclusive data trading (i.e., the same data can be used by multiple users) is presented in [20] and omitted here due to space limitation. We assume that owner $m$ has $C_m$ amount of data in total and the loss function $U_m$ is a convex function. Suppose collector $n$ collects $y_{mn}$ data from owner $m$. Clearly, $y_{mn}$ is no larger than $x_{mn}$. We assume that the collecting procedure incurs a loss of $V_n(y_{mn})$ for collector $n$, where $y_{mn} = [y_{mn1}, ..., y_{mnn}]^T$ since it costs efforts to collect data. We assume that $V_n$ is a convex function. Lastly, data user $l$ buys $z_{nl}$ amount of data from collector $n$. The gain of user $l$ is $W_l(z_{nl})$.

†In the following, we use the term data agents to refer to data owners, collectors and users.
where \( z_l = [z_{1l}, ..., z_{Nl}]^T \). The gain function \( W_l \) is assumed to be a concave function.

As the interests of the data agents conflict with each other, a system designer is needed to coordinate the agents’ behaviors to maximize overall system efficiency or social welfare. The corresponding social welfare maximization problem SWM can be formulated as follows.

\[
\text{Maximize } X, Y, Z = \sum_{m=1}^{M} U_m(x_m) - \sum_{n=1}^{N} V_n(y_n) + \sum_{l=1}^{L} W_l(z_l) \\
\text{s.t. } \sum_{n=1}^{N} x_{mn} \leq C_m, \quad \forall m, \\
\sum_{l=1}^{L} z_{nl} \leq \sum_{m=1}^{M} y_{mn}, \quad \forall n, \\
y_{mn} \leq x_{mn}, \quad \forall m, n.
\]

The first constraint is the total data constraint at each data owner. The second constraint is the data constraint at each collector where the total amount of sold data is no larger than the amount of total collected data. The third constraint means that the data collected by a collector from an owner is no bigger than the data that owner entitles collector to collect.

SWM is a convex optimization problem and can be solved in a centralized manner by using state-of-the-art optimization toolbox such as CVX [21]. However, in real-world applications, we cannot directly solve the SWM to coordinate the data trading due to the following two reasons. First, data agents (data owners, collectors and users) are selfish and seek to maximize their own utilities instead of the social welfare. As a result, even if the system designer computes the socially optimal point by solving SWM, the optimal solution cannot be enforced given the selfishness of the data agents. Second, the utility functions \( U, V, W \) are private information of the agents which is unknown to the system designer. Thereby, SWM cannot be solved at the system designer’s side in a centralized fashion. In order to elicit the private information of the agents and guild the selfish agents to cooperate to achieve social optimum, we resort to iterative auction mechanism [15].

### III. MECHANISM DESIGN

In this section, we design an iterative auction mechanism for the data trading problem formulated in Section II. Our design goal is to guild the selfish agents to trade data at a socially optimal point while avoiding direct inquiry of the agents’ utility functions. The proposed iterative auction mechanism is illustrated in Fig. 1. The system designer serves as the auctioneer and the data agents are the bidders. Analogous to many auction mechanisms in the literature [22], the agents submit bids to signal their valuations of the resources, or data in this context. The first step of the mechanism is that the system designer announces the data allocation and pricing/reimbursement rules to the agents. In the second step, based on these rules, each agent calculates and submits an appropriate bid in order to maximize her own utility in accordance with her selfishness. In the third step, the system designer computes the data allocation result according to the submitted bids and the data allocation rule. The aforementioned three steps are common in auction theory. The unique feature of iterative auction lies in the fourth step, in which the system designer adjusts the data allocation and pricing/reimbursement rules based on the data allocation results. Then, the system designer announces these new rules and another auction begins. This iterative process continues until the system designer observes convergence. In the following subsections, we describe each step of the mechanism in more detail.

As explained in Section II, a difficulty for the system designer to solve the SWM is that she is unaware of the loss and gain functions \( U, V, W \), which are private information of the agents. Thus, the system designer has to replace these unknown functions with some known functions. In addition, denote \( s_m = [s_{m1}, ..., s_{mN}] \geq 0 \) the bid that owner \( m \) submits to the system designer, where \( \geq \) denotes componentwise inequality. Similarly, denote \( t_n = [t_{n1}, ..., t_{MN}]^T \geq 0 \) the bid of collector \( n \) and \( r_l = [r_{1l}, ..., r_{NL}]^T \geq 0 \) the bid of user \( l \). The bids signal the agents’ valuations of the data and should be incorporated into the loss and gain functions in the system designer’s perspective. In the iterative auction mechanism, by replacing the unknown utility functions with some known functions, the system designer transforms SWM into the following designer’s allocation problem DAP.

\[
\text{Maximize } X, Y, Z = \sum_{n=1}^{N} \sum_{l=1}^{L} r_{nl} \log z_{nl} - \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{s_{mn}^2}{2} x_{mn}^2 \\
\text{s.t. } \text{the constraints (2), (3) and (4)}
\]

Denote \( \lambda \in \mathbb{R}^M, \mu \in \mathbb{R}^N, \eta \in \mathbb{R}^{M \times N} \) the dual variables associated with constraints (2), (3) and (4), respectively. The Karush-Kuhn-Tucker (KKT) conditions of DAP include the following stationarity conditions.

\[
s_{mn} x_{mn} + \lambda_m - \eta_{mn} = 0, \forall m, n, \\
t_{mn} y_{mn} - \mu_n + \tau_{mn} = 0, \forall m, n, \\
\frac{r_{nl}}{z_{nl}} + \mu_n = 0, \forall n, l.
\]
Therefore, we set the reimbursement rule of owner

The data allocation rule prescribes how the data are allocated given the submitted bids $S = [s_{mn}]_{m \times N}$, $T = [t_{mn}]_{m \times N}$, $R = [r_{nl}]_{N \times l}$. The allocation rule is parameterized by the Lagrangian multipliers $\lambda$, $\mu$, $\eta$. Given a set of $(\lambda, \mu, \eta)$, an allocation rule is defined according to Eq. (8), i.e., a relationship between the data allocation and the bids is specified. As stated in the first step of the mechanism in Fig. 1, besides data allocation rule, the system designer also needs to specify the data pricing/reimbursement rule, i.e., the price and reimbursement of data as functions of the bids of the agents. In other words, for owner $m$, given its bid $s_m$, the system designer needs to reimburse $f_m(s_m)$ amount of money to compensate her loss due to privacy compromise. Similarly, the system designer will reimburse $g_n(t_n)$ amount of money to collector $n$ given her bid $t_n$. Furthermore, the system designer will charge user $l$ $h_l(r_l)$ amount of money given her bid $r_l$. As a mechanism designer, we need to appropriately design the pricing/reimbursement functions $f_m, g_n, h_l$ so that the data allocation will gradually converge to the socially optimal point, i.e., the optimal point of SWM. In the following subsections, we specify how to design these pricing/reimbursement functions in detail.

For owner $m$, if she bids $s_m$, she will get an reimbursement of $f_m(s_m)$ as well as a loss of $U_m\left(\frac{\eta_{mn} - \lambda_m}{s_{mn}}, ..., \frac{-\eta_{mN} - \lambda_m}{s_{mN}}\right)$, according to the data allocation rule in Eq. (8). Hence, the utility maximization problem of owner $m$ can be written as:

$$\text{Maximize}_{s_m \geq 0} f_m(s_m) - U_m\left(\frac{\eta_{mn} - \lambda_m}{s_{m1}}, ..., \frac{-\eta_{mN} - \lambda_m}{s_{mN}}\right).$$

The first order optimality condition of owner $m$’s problem is:

$$\frac{\partial f_m(s_m)}{\partial s_{mn}} + \frac{\partial U_m}{\partial x_{mn}} \frac{\eta_{mn} - \lambda_m}{s_{mn}^2} = 0, \forall n.$$ (10)

In order to design a suitable $f_m$ such that the data allocation will converge to the socially optimal point, we need to compare Eq. (10) with the optimality condition of SWM. The stationarity conditions in the KKT conditions of SWM are:

$$\frac{\partial U_m(x_m)}{\partial x_{mn}} + \lambda_m - \eta_{mn} = 0, \forall m, n,$$ (11)

$$\frac{\partial U_m(x_m)}{\partial y_{mn}} - \mu_n + \eta_{mn} = 0, \forall m, n,$$ (12)

$$-\frac{W_l(z_l)}{z_{nl}} + \mu_n = 0, \forall n, l.$$ (13)

Combining equations (10) and (11), we derive:

$$\frac{\partial f_m(s_m)}{\partial s_{mn}} = \frac{\lambda_m - \eta_{mn}}{s_{mn}^2} \frac{\partial U_m}{\partial x_{mn}} = -\frac{(\lambda_m - \eta_{mn})^2}{s_{mn}^2}.$$ (14)

Therefore, we set the reimbursement rule of owner $m$ to be $f_m(s_m) = \sum_{n=1}^{N} -\frac{(\lambda_m - \eta_{mn})^2}{s_{mn}}$. Through analogous derivations, we can formulate the utility maximization problems of the collectors and users. We can similarly design the reimbursement rule of collector $n$ to be $g_n(t_n) = \sum_{m=1}^{M} \frac{\mu_n - \eta_{mn}}{t_{mn}}$ and the pricing rule of user $l$ to be $h_l(r_l) = \sum_{n=1}^{N} r_{nl}$.

The owners’ problem (9), the collectors’ problem and the users’ problem (which can be similarly derived) together specify how the bids are chosen in the second stage of the mechanism in Fig. 1.

Then, in the third stage, the system designer computes the new data allocation result based on these submitted bids and the data allocation rule in Eq. (8). In the fourth stage, we update the dual variables $\lambda, \mu, \eta$ (or equivalently, update the data allocation rule and data pricing/reimbursement rule) by invoking the subgradient method. The proposed iterative auction mechanism is summarized in Algorithm 1. We remark that Algorithm 1 is a distributed algorithm: each data agent solves its own utility maximization problem in a parallel manner and the interactions between the agents. Algorithm 1 clearly resolves the two difficulties for directly solving SWM in Subsection II: (i) each agent maximizes her own utility in accordance with her selfishness; (ii) the system designer does not directly access the private information of the agents, i.e., the loss/gain functions $U, V, W$. Instead the system designer gradually and implicitly elicit this information through iterative auctions.

Algorithm 1 The proposed iterative auction mechanism

1: Initialize $X^{(0)}, Y^{(0)}, Z^{(0)}, \lambda^{(0)}, \mu^{(0)}, \eta^{(0)}$ to be non-negative.
2: Set the time index $\tau$ to be 0.
3: Repeat the following until convergence:
4: The system designer announces $\lambda^{(\tau)}, \mu^{(\tau)}, \eta^{(\tau)}$.
5: $\tau \leftarrow \tau + 1$.
6: Each owner $m$ solves its problem (9) to get $s_m^{(\tau)}$.
7: Each collector $n$ solves its utility maximization problem to get $t_n^{(\tau)}$.
8: The system designer computes the new $X^{(\tau)}, Y^{(\tau)}, Z^{(\tau)}$ according to the current allocation rule (8) and the submitted bids $S^{(\tau)}, T^{(\tau)}$ and $R^{(\tau)}$.
9: The system designer updates the dual variables:

$$\lambda_m^{(\tau)} = \left(\lambda_m^{(\tau-1)} + \alpha \left(\sum_{n=1}^{N} x_{mn}^{(\tau)} - c_m\right)\right)^+, (15)$$

$$\mu_n^{(\tau)} = \left(\mu_n^{(\tau-1)} + \alpha \left(\sum_{l=1}^{L} z_{nl}^{(\tau)} - \sum_{m=1}^{M} y_{mn}^{(\tau)}\right)\right)^+, (16)$$

$$\eta_m^{(\tau)} = \left(\eta_m^{(\tau-1)} + \alpha \left(y_{mn}^{(\tau)} - z_{mn}^{(\tau)}\right)\right)^+. (17)$$

We theoretically show that the proposed iterative auction mechanism converges to the socially optimal point, which is formally stated in the following.

**Theorem 1.** Suppose the step size $\alpha$ in Algorithm 1 is small enough. Then, the data allocation $(X, Y, Z)$ of Algorithm 1 converges to the optimal point of SWM. Moreover, the dual variables $(\lambda,\mu,\eta)$ of Algorithm 1 converge to the dual optimal point of SWM.

Implementation of the proposed mechanism in real-world data trading market necessitates brilliant economic properties of the mechanism. The proposed mechanism is clearly **efficient** since it converges to the socially optimal point. To ensure that each agent complies to the mechanism voluntarily, the mechanism needs to guarantee that every agent has non-negative utility, i.e., the mechanism should be **individually rational**.

**Proposition 1.** Assume that $U_m(0) = 0, V_n(0) = 0, W_l(0) = 0, \forall m, n, l$. Then, when Algorithm 1 converges, every data agent
has non-negative utility, i.e., the proposed mechanism is individually rational.

We can further show that the system designer has weakly balanced budget, i.e., the income (through the data reimbursement/pricing) of the system designer in the mechanism is non-negative when Algorithm 1 converges. In other words, the system designer does not need to inject any money into the data market in order to implement the mechanism.

**Proposition 2.** When Algorithm 1 converges, the income of the system designer through data reimbursement/pricing in the mechanism is non-negative. In other words, the mechanism has weakly balanced budget.

**IV. SIMULATIONS AND REAL DATA EXPERIMENTS**

In this section, we present simulations as well as real data experiments to validate the theoretical results for the proposed iterative auction mechanism.

Consider a data market with $M = 2$ data owners, $N = 2$ data collectors and $L = 4$ data users. The total data amount of owners 1 and 2 are set to be 2 and 4, respectively. The owners’ convex loss functions are defined as follows:

$$U_m(x_m) = a_m \left( \sum_{n=1}^{2} e^{x_{mn}} - 2 \right), \quad m = 1, 2,$$

where $a_1 = 0.1, a_2 = 0.3$. The collectors’ convex loss functions are defined as:

$$V_n(y_n) = b_n \sum_{m=1}^{2} y_{mn}^2, \quad n = 1, 2,$$

where $b_1 = 0.5, b_2 = 1$. The users’ concave gain functions are:

$$W_l(z_l) = c_l \sum_{n=1}^{2} \log(1 + z_{nl}), \quad l = 1, 2, 3, 4,$$

where $c_1 = \frac{3}{2}, c_2 = \frac{2}{3}, c_3 = \frac{2}{4}, c_4 = \frac{1}{2}.$

We simulate the proposed iterative auction mechanism in Algorithm 1. In Fig. 2-(a), we validate the convergence behavior of the mechanism. The relative error used in Fig. 2-(a) is

$$\text{max} \left( \frac{||X - X^*||}{||X^*||}, \frac{||Y - Y^*||}{||Y^*||}, \frac{||Z - Z^*||}{||Z^*||} \right),$$

where $|| \cdot ||$ denotes the Frobenius norm. As guaranteed by Theorem 1, the mechanism converges to the socially optimal point, i.e., the mechanism is efficient. We further investigate the economic properties of the mechanism through simulations in Fig. 2-(b). We report the utilities of the owner 1, collector 1 and user 1 as the algorithm gradually converges. As asserted in Proposition 1, the mechanism is individually rational: the three data agents in Fig. 2-(b) have non-negative utilities when the algorithm converges. Furthermore, we show the budget balance (income) of the system designer and find that as assured by Proposition 2, the budget balance is non-negative when the algorithm converges.

Next, we use real data to get the loss/gain functions of the data agents and investigate the performance of the proposed mechanism on them. We use real data prices of Microsoft Azure Marketplace [23] to estimate the users’ gain functions to be:

$$W_l(z_l) = c_l \sum_{n=1}^{2} x_{mn}^{\beta_n}, \quad l = 1, 2, 3, 4$$

**Fig. 2:** Simulation results

**Fig. 3:** Real data experiments

where $\alpha_1 = 0.821, \alpha_2 = 1.267, \beta_1 = 0.9131, \beta_2 = 0.5329, c_1 = 1/2, c_2 = 5/6, c_3 = 7/6, c_4 = 3/2.$ We further estimate the loss functions of the owners from the relation between privacy protection level and information loss in [14], which itself is obtained by real data [24]. The loss of functions of the owners are set to be:

$$U_m(x_m) = a_m \sum_{n=1}^{2} (\theta_n x_{mn})^{0.5855}, \quad m = 1, 2,$$

where $\theta_1 = 1.5816, \theta_2 = 2.5816, a_1 = 5, a_2 = 15.$ Details of obtaining the estimated gain functions of users and loss functions of owners are omitted due to space limitation.

With the loss/gain functions estimated from real data, we test the performance of the proposed iterative auction mechanism. We first consider the exclusive data trading. The total data amounts of owner 1 and owner 2 are 0.25 and 0.5, respectively. As shown in Fig. 3-(a), the mechanism still converges to the socially optimal point. In Fig. 3-(b), we further observe that the individual rationality and weakly balanced budget still hold as the utilities of owner 1, collector 1 and user 1 as well as the budget balance of the system designer are all non-negative.

**V. CONCLUSION**

In this paper, we study the data trading problem with multiple data owners, collectors and users. The corresponding social welfare maximization problem is formulated. We present an iterative auction mechanism to guild the selfish agents to behave in a socially optimal way without direct access of their private information. We theoretically prove the convergence as well as economic properties of the mechanism, which is corroborated by simulations and real data experiments.
VI. REFERENCES


