An Online Truthful Auction for IoT Data Trading with Dynamic Data Owners

Zhenni Feng[1(✉)], Junchang Chen[1], and Tong Liu[2]

[1] Donghua University, Shanghai, China
fzn@dhu.edu.cn, 2202500@mail.dhu.edu.cn
[2] Shanghai University, Shanghai, China
tong_liu@shu.edu.cn

Abstract. Data is an extremely import asset in modern scientific and commercial society. The life force behind powerful AI or ML algorithms is data, especially lots of data, which makes *data trading* significantly essential to unlocking the power of AI or ML. Data owners who offer personal data and data consumers who request data blocks negotiate with each other to make an agreement on trading prices via a big data trading platform; consequently both sides gain profit from data transactions. A great many existing studies have investigated to trade various kinds of data as well as to protect data privacy, or to construct a decentralized data trading platform due to untrustworthy participants. However, existing studies neglect an important characteristic, *i.e.*, dynamics of both data owners and data requests in IoT data trading. To this end, we first construct an auction-based model to formulate the data trading process and then propose an truthful online data trading algorithm which not only resolves the problem of matching dynamic data owners and randomly generated data requests, but also determines the data trading price of each data block. The proposed algorithm achieves several good properties, such as constant competitive ratio for near-optimal social efficiency, incentive-compatibility, individual rationality of participants, via rigorous theoretical analysis and extensive simulations.

Keywords: Data trading · Dynamic data owners · Near-optimal online auction

1 Introduction

Data is an extremely import asset in modern scientific and commercial society. Predicted by IDC, there will be 55.7 billion connected devices worldwide by 2025, 75% of which will be connected to an IoT platform [10]. Furthermore, data generated by these IoT devices is estimated to be 73.1 ZB then. Most of these data arise from security and video surveillance; industrial IoT applications may also

H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 406, pp. 554–571, 2021.
https://doi.org/10.1007/978-3-030-92635-9_32

take a significant portion of these data. Almost all companies are aggressively turning to artificial intelligence (AI) or machine learning (ML) technology to gain competitive advantages. Nevertheless, the life force behind these AI or ML algorithms is data, especially a vast amount of data. Consequently, *data trading* is significantly essential to unlocking the power of AI or ML.

Data trading is different from the general concept of data sharing. In mobile crowdsensing applications, participants or workers usually share their sensing data and get reward in return [13,14,19]. Application users generally share their data, *e.g.*, web browsing, online shopping orders, to application service providers in order to get accessible to their services. To get things moving, participants, which are generally application users, application service providers and third-party data consumers, go from the general concept of data sharing with others to the specifics of exactly what data they want and what they are willing to give of value in exchange.

Big data has fueled the emergency of data trading platforms which serve as bridges between sellers and buyers. Examples, such as Terbine [2] and GXS TradeWeb [1], have been designed to provide big data trading services. Data owners who offer personal data and data consumers who request data blocks negotiate with each other to make an agreement on trading prices via a big data trading platform; consequently both sides gain profit from data transactions.

Many recent studies [3,6,12,17,22] have paid attention to propose data trading algorithms or to design data trading platforms. Various kinds of datasets, such as raw data samples, range counts, aggregate statistic results, are traded between data owners and data consumers. To negotiate the data trading process between them, data brokers are usually introduced to assist transmission of data trading messages or traded data. To protect data privacy, some recent studies employ encryption algorithms and then to disclose encrypted data to data consumers; other studies introduce privacy-preserving schemes such as differential privacy or its variations to take control of the level of disclosed data privacy. Some other existing studies [4,5,8,9,18,21] consider participants or the data platform are untrustworthy; they propose decentralized data trading platforms based on blockchain technology. Almost all existing studies, however, ignore an important characteristic,*i.e.*, dynamics of both data owners and data requests in IoT data trading, where data owners are end devices and data consumers are application service providers.

In the paper, we model the dynamics of data owners and data requests in the scenario of IoT data trading. For example, when a data request of querying real-time noise level locating at a given block is submitted, a smartphone can serve as a data owner only if the owner of the smartphone passes by the block. Consequently, we assume that data owners are intermittent and only available to trade their data during a specific time period, which is called *active time*, because of limited resources or mobility. Furthermore, data requests are randomly generated by data consumers and then submitted to the data trading platform running on a edge server or a cloud server.

There are several main technical challenges to solve the IoT data trading problem. *First of all*, data owners are allowed to dynamically join in the data trading process and data requests are randomly generated according to the demand of applications. Such uncertain and unpredictable data requests make the data trading process relatively complicated. *Then*, it is very difficult for the data trading platform to make an efficient matching between data owners and data requests, because both the real value of data and active time are private information of data owners. *Finally*, rational and strategic data owners are not willing to offer their valuable data or reveal their private information truthfully except with enough compensation.

To this end, we propose a truthful online auction-based data trading algorithm containing two key components, which resolves two subproblems of how to match data owners with data requests and how to determine trading price of each data block. On the one hand, an auction model is constructed to formulate data trading process and an efficient online matching algorithm based on greedy scheme is further proposed to achieve the near-optimal system efficiency with a constant competitive ration of 1/2. On the other hand, the trading price of each data block is computed according to a *critical value*, which is the highest bidding price that a data owner would win a bid. Both rigorous theoretical analysis and extensive simulations demonstrate good properties of our proposed online data trading algorithm, *e.g.*, *individual rationality*, *incentive-compatibility*, *near-optimality* on system efficiency. Major technical contributions in this paper are summarized as follows.

- It is the first work, to the best of our knowledge, which takes account of both dynamic behaviors of data owners and randomly generated data requests in the problem formulation of the data trading process.
- We propose a truthful online auction-based data trading algorithm which not only determines the matching rule with incomplete information but also computes proper data trading prices between data owners and data consumers.
- We have demonstrated that the proposed algorithm achieves several good properties via rigorous theoretical analysis and extensive simulations.

The rest of the paper is organized as follows. The system model and problem formulation of data trading is presented in Sect. 2. Then, the proposed online data trading algorithm is discussed in Sect. 3 along with rigorous theoretical analysis. Section 4 provides extensive simulations and numerical results to demonstrate desirable properties of the proposed algorithm. We review relate work in Sect. 5 and finally conclude the paper in Sect. 6.

2 System Model and Problem Formulation

We first introduce participants in the data trading process and describe the data trading model between data owners and data consumers; the mathematical formulation of data trading is then provided.

2.1 System Model

In a data trading market for sharing IoT data, there are mainly two kinds of participants, *end devices* and *edge servers or cloud servers*. End devices who collect data are *data owners*; edge servers or cloud servers who buy data from data owners are *data consumers*.

We divide time into time slots of equal size. Auctions between data owners and data consumers are executed round by round. Data owners are short-sighted so that they expect to get as much profit as possible in the current round. Without loss of generality, we only consider the auction process in a single round.

Data requests are submitted randomly and dynamically. We assume that a practical data requirement can be decomposed into several smaller data requests, each of which can be satisfied by a single data block generated by a single data owner. Let τ_j denote the number of data requests arriving at time slot j. A data request submitted at the time slot j is denoted by $r_{j,k}, k \leq \tau_j$. The set of all data requests is denoted by $R = \{r_{j,k} | j = 1, 2, \cdots, T, k \leq \tau_j\}$, where T is the total number of time slots in each round. For simplicity, we assume that there are a sufficient number of data owners that every data request is matched to a data block at its arriving time slot.

Data owners are dynamic because their data are only accessible during their *active time*. The active time of a data owner i is a time period described by $[s_i, e_i)$, where s_i and e_i are the start time slot and end time slot (not included) of her active time. Each data owner can sell at most one data block or sell her data block at most once during her active time because of limited resources. A data owner i has a valuation v_i for her data block which indicates that she would not trade her data block with a trading price lower than v_i. Similarly, the reported active time is possibly different from her real active time.

2.2 Data Trading Model Based on an Auction Mechanism

The interaction between data owners and data consumers are modeled by an auction mechanism, as shown in Fig. 1. In a data trading market, data owners share their data blocks with others and then they are compensated according to trading prices; data consumers get data blocks and pay to data owners. There exists a third-party trading platform to manage the auction process. The auction process is described as follows:

1. The platform sends data requests to data owners.
2. Each data owner generates a bid B_i which reports the active time and valuation for her data block, and then sends her bid to the platform.
3. The platform matches bids to data requests and determines the trading time slot and trading price (payment) for each selected bid. Then platform return matching results to data owners and data consumers.
4. Each data owner whose bid is selected uploads her data block at a specific time slot.
5. The data consumer pays for the data owner according to the negotiated trading price.

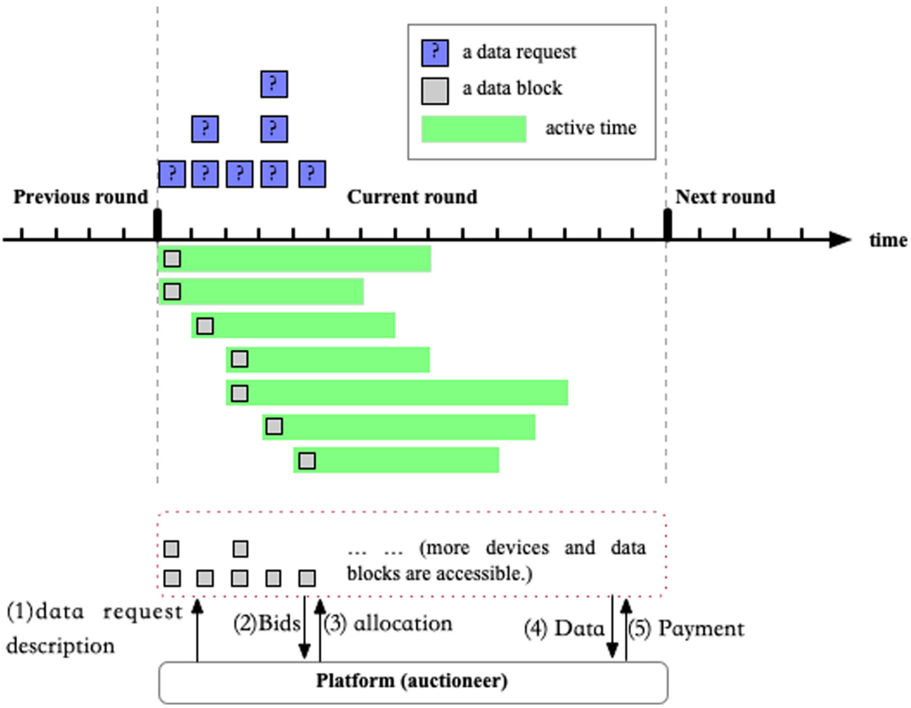


Fig. 1. The auction process repeated round by round. Data blocks offered by dynamic end devices and dynamic data requests arrives at the first five time slots are depicted in the picture.

The platform must determine a *matching rule* and a *trading price rule*. We use $B = \{B_i | i = 1, 2, \cdots, N\}$ to denote the set of all bids submitted by all data owners, where N is the total number of data owners in a round. The matching rule $\mathbf{X}$ is a matrix of indicator variables; each element, $x_{i,j} \in \{1, 0\}, 1 \leq i \leq N, 1 \leq j \leq T$, represents whether bid B_i is selected at time slot j or not. Actually, we should denote the matching rule by $x_{i,j}(B)$ since each $x_{i,j}$ is determined by B. For simplicity, we use $x_{i,j}$ instead of $x_{i,j}(B)$ in the rest of the paper. According to the trading price rule, payment to each bid B_i is denoted by $p_i(B) \in \mathbf{R}$.

In each round, a data owner i submits at most one bid, $B_i = (\tilde{s_i}, \tilde{e_i}, b_i), 1 \leq \tilde{s_i} < \tilde{e_i} \leq T+1, b_i \geq 0$, where $[\tilde{s_i}, \tilde{e_i})$ is the reported active time and b_i is the bid price. The bid price may be different from real valuation v_i, because data owners are usually *selfish*. Similarly, the reported active time period may be different from her real active time.

Next, we discuss utilities of data owners and explain why data owners are selfish.

Definition 1 ***(Utility of a data owner).*** *The utility of a data owner is the difference between the trading price and her valuation if the bid of the data owner is selected; Otherwise her utility is zero. The utility is computed as follows.*

$$u_i^o(B) = \sum_{j=s_i}^{e_i-1} x_{i,j}(p_i(B) - v_i), \tag{1}$$

where $\sum_{j=s_i}^{e_i-1} x_{i,j} \leq 1$ *holds because a data owner only trade her data block at most once during her active time.*

A data owner is selfish so that she probably selects strategies solely to maximize her utility. The data owner probably misreports start time or end time of her active time as well as to charge a higher price than her valuation. Data owners can not report earlier arrivals or delayed departure, because it is easy to detect their absence and thus they would be punished. Therefore, there are three of strategic behaviors for selfish data owners, *i.e.*, delayed arrival, earlier departure, and misreporting valuation.

Definition 2 ***(Utility generated by a data request).*** *The utility of a data requester who publishes a data request is the difference between the amount of profit that the data requester get from data and the trading price. If the k-th data request* $r_{j,k}$ *at time slot* j *is matched to the bid* B_i *of a data owner* i*, then the utility generated by the data request* $r_{j,k}$ *is computed as follows:*

$$u_{j,k}^r(B) = x_{i,j}(\varphi_j - p_i(B)), \tag{2}$$

where φ_j *is the amount of profit that the data requester get from the traded data block.*

Consequently, utilities of all data requests at time slot j is

$$\sum_{k=1}^{\tau_j} u_{j,k}^r(B) = \sum_{i=1}^{N} x_{i,j}(\varphi_j - p_i(B)), \tag{3}$$

where $\sum_{i=1}^{N} x_{i,j} = \tau_j$ holds because a data request can be satisfied by a single data block and there are τ_j data owners selected at time slot j.

Definition 3 ***(Social efficiency).*** *The social efficiency is defined as the sum of utilities of all participants. It is computed as follows:*

$$\begin{aligned}
\Delta &= \sum_{i=1}^{N} u_i^o(B) + \sum_{j=1}^{T}\sum_{k=1}^{\tau_j} u_{j,k}^r(B) \\
&= \sum_{i=1}^{N}\sum_{j=s_i}^{e_i-1} x_{i,j}(p_i(B) - v_i) + \sum_{j=1}^{T}\sum_{i=1}^{N} x_{i,j}(\varphi_j - p_i(B)) \\
&\overset{(a)}{=} -\sum_{i=1}^{N}\sum_{j=s_i}^{e_i-1} x_{i,j}v_i + \sum_{j=1}^{T}\tau_j\varphi_j + (\sum_{i=1}^{N}\sum_{j=s_i}^{e_i-1} x_{i,j}p_i(B) - \sum_{j=1}^{T}\sum_{i=1}^{N} x_{i,j}p_i(B)) \\
&\overset{(b)}{=} C - \sum_{i=1}^{N}\sum_{j=s_i}^{e_i-1} x_{i,j}v_i,
\end{aligned} \tag{4}$$

where (a) uses $\sum_{i=1}^{N} x_{i,j} = \tau_j$ *and (b) follows from exchanging two summations and replacing* $\sum_{j=1}^{T} \tau_j \varphi_j$ *with a constant parameter* C *whose value is not related to* $x_{i,j}$*.*

2.3 Problem Formulation

In the paper, we aim to design an auction mechanism for the data trading market so that the maximum social efficiency is achieved as well as following properties, *i.e.*, *individual rationality*, *incentive-compatibility*, *computation efficiency.*

Definition 4 ***(Individual rationality).*** *An auction mechanism satisfies the property of individual rationality if and only if every data owner has a non-negative utility,* i.e., $u_i^o \geq 0, i = 1, 2 \cdots N$.

Definition 5 ***(Incentive-compatibility).*** *An auction mechanism is incentive-compatible if and only if, for each data owner* i*, she cannot increase her utility by misreporting her private information,* i.e.,

$$u_i^o(B_i \cup B_{-i}) \geq u_i^o(\tilde{B}_i \cup B_{-i}), \tag{5}$$

where $B_i = (s_i, e_i, v_i)$ *and* $\tilde{B}_i = (\tilde{s}_i, \tilde{e}_i, b_i)$ *are not the same, which means any of* $s_i \leq \tilde{s}_i$*,* $e_i \geq \tilde{e}_i$*, and* $v_i \neq b_i$ *holds;* B_{-i} *denotes the set of all bids except* B_i*.*

In the following, we offer the mathematical formulation of our problem. We need to determine the matching rule of $\mathbf{X} = \{x_{i,j} | i = 1, 2, \cdots, N, j = 1, 2, \cdots, T\}$ by solving the optimization problem defined in (6) as well as the trading price rule $\{p_i | i = 1, 2, \cdots, N\}$ satisfying definitions of (4) and (5).

$$\begin{aligned}
\arg\max_{x_{i,j}} \Delta = \arg\min_{x_{i,j}} & \sum_{i=1}^{N} \sum_{j=s_i}^{e_i - 1} x_{i,j} v_i, \\
s.t. & \sum_{j=s_i}^{e_i-1} x_{i,j} \leq 1, \quad \sum_{j \in \{t|t<s_i\} \cup \{t|t \geq e_i\}} x_{i,j} = 0, \forall i, \\
& \sum_{i=1}^{N} x_{i,j} = \tau_j, \forall j, \\
& x_{i,j} \in \{0, 1\}, \forall i, \forall j.
\end{aligned} \tag{6}$$

From the objective of the problem, we can see that maximizing social efficiency is equivalent to minimizing sum of valuation from selected data owners.

Our problem cannot be solved by classic optimization algorithms solving linear programming for several reasons. *Firstly*, valuation v_i for each data owner is not accessible. *Secondly*, both data owners (with their data blocks) and data requests joins the data trading market dynamically, which means $x_{i,j}$ should be determined online without future information. *Thirdly*, the solution of Eq. (6) fails to provide any information about trading prices.

Algorithm 1: Online Matching Algorithm

Input : The set of bids B.
Output: The matching rule $\mathbf{X} = \{x_{i,j} | i = 1, 2, \cdots, N, j = 1, 2, \cdots, T\}$.

$A \leftarrow \emptyset$, $j \leftarrow 1$, $\mathbf{X} \leftarrow \mathbf{0}$; `//` A `is the set of all active bids which can provide accessible data block at current time slot.`
while $j \leq T$ **do**
  Remove expired bids (bids with $e_i = j$) from A;
  Add newly active bids (bids with $s_i = j$) to A;
  `/* Greedy select the first` τ_j `bids with lowest bid price at each time slot` j`. */`
  for $k \leftarrow 0$ **to** τ_j **do**
    `// Loop` τ_j `times.`
    Choose a bid $B_{(i')}$ with the lowest bid price and match it to the k-th data request at current time slot, *i.e.*, $x_{(i'),j} \leftarrow 1$;
    $A \leftarrow A - B_{(i')}$; `// Remove` $B_{(i')}$ `from` A`.`
  end for
  $j \leftarrow j + 1$;
end while
return X

3 Online Data Trading Algorithm

In the section, we propose an online data trading algorithm to determine both the matching rule and the trading price rule. Our online data trading algorithm contains two components, which solves the matching subproblem in Sect. 3.1 and trading price subproblem in Sect. 3.2. For simplicity, we assume that every data owner submits a bid which is exactly the same as her private information and then we prove that data owners would honestly report her private information under the given trading price rule in Sect. 3.3.

3.1 Online Matching Algorithm Based on a Greedy Strategy

We propose an online matching algorithm to match data blocks of data owners to data requests using a greedy strategy. The basic idea of this algorithm is to greedily select one with the lowest bid price from current available data blocks to satisfy the newly submitted data request. The selection is executed at the beginning of every time slot. As shown in Algorithm 1, the algorithm maintains a set of active bids which have not been matched to any data request; the set is updated at the beginning of each time slot, *i.e.*, appending or removing bids into the active set according to active time of bids. At each time slot j, the first τ_j bids with lowest bid prices are selected.

3.2 Computing Trading Prices Based on Critical Data Owners

Unfortunately, a VCG-based payment scheme [16] is inapplicable to the online auction mechanism, because the online matching algorithm is not optimal.

Algorithm 2: Trading Price Determination Algorithm (for a single selected bid)

```
Input  : A selected bid B_i = (s_i, e_i, b_i), time slot j̄ that B_i is selected
         (i.e., x_{i,j̄} = 1), the set of all bids B.
Output: The trading price p_i of the data block that is associated to B_i.
A ← ∅, j ← 1, p_i ← b_i;
B ← B − B_i; // Remove B_i from the set of all bids.
while j < e_i do
    Remove expired bids (bids with e_i = j) from A;
    Add newly active bids (bids with s_i = j) to A;
    for k ← 0 to τ_j do
        Choose a bid B_i' = (s_i', e_i', b_i') from A with the lowest bid price;
        A ← A − B_i';
        /* Find the highest bid price from all bids that are selected
           during [j̄, e_i − 1]                                          */
        if j ≥ j̄ and b_i' > p_i then
            p_i ← b_i';
        end if
    end for
    j ← j + 1;
end while
return p_i;
```

In the paper, we propose a trading price determination scheme based on *critical bid* which guarantees that each data owner reports private information truthfully. The basic idea is to set the trading price of a selected bid B_i as the bid price of the first bid that makes B_i fails. The first bid that makes B_i fails is the critical bid of B_i. Actually, if $B_i = (s_i, e_i, b_i)$ is selected at time slot $\bar{j}$ according to the matching rule, the critical bid $c(B_i)$ of B_i is the bid with the highest bid price which are selected during the time period of $[\bar{j}, e_i - 1]$ other than B_i.

Main steps of computing trading price for a selected bid are shown in Algorithm 2. Firstly, remove $B_i = (s_i, e_i, b_i)$ from the set of all bids B. Secondly, employ the matching rule proposed in Algorithm 1 to find all bids that are selected earlier than $\bar{j}$ and remove all of them from the active set of bids. Finally, find the bid (*i.e.*, the critical bid) with the highest bid price from all bids that are selected during the time period $[\bar{j}, e_i - 1]$ and return the bid price of the critical bid as the trading price. Similarly, we can repeat the procedure in Algorithm 2 for each selected bid. Besides, if a bid is not selected, then the trading price of the data block that is associated to the bid is zero.

3.3 Theoretical Analysis

In the subsection, we prove that the proposed online auction mechanism which contains two components of an online matching algorithm (Algorithm 1) and a trading price determination algorithm (Algorithm 2) satisfies several good properties aforementioned.

To prove the auction mechanism is incentive-compatible, it is equivalent to prove that it satisfies following two conditions: (i) the matching rule in Algorithm 1 is *monotonic*, and (ii) the trading price of the data block associated to each bid is equal to the *critical value*.

Definition 6 ***(Monotonicity).*** *The matching rule is monotonic if and only a data owner whose bid $B_i = (s_i, e_i, b_i)$ is selected would also win if she reports a more attractive bid $\tilde{B_i} = (\tilde{s_i}, \tilde{e_i}, \tilde{b_i})$ with a lower bid price or a longer active time period,* i.e.*, $\tilde{s_i} \leq s_i$, $\tilde{e_i} \geq e_i$, $\tilde{b_i} \leq b_i$.*

Definition 7 ***(Critical value).*** *For a data owner whose bid $B_i = (s_i, e_i, b_i)$ is selected, the critical value of the bid is the highest bid price b_i' that the data owner submits a bid $B_i' = (s_i, e_i, b_i' \geq b_i)$ and the new bid B_i' is still selected.*

Theorem 1 ***(Incentive-compatibility).*** *The proposed auction mechanism is incentive-compatible, because the matching rule is monotonic and the trading price is set as the critical value.*

Proof. First of all, we show that the matching rule is monotonic. Suppose a bid $B_i = (s_i, e_i, b_i)$ is selected at time slot j according to the matching rule. We replace the bid B_i with another bid $\tilde{B_i} = (\tilde{s_i}, \tilde{e_i}, \tilde{b_i})$, where $\tilde{s_i} \leq s_i$, $\tilde{e_i} \geq e_i$, $\tilde{b_i} \leq b_i$. Obviously, $\tilde{B_i}$ would be selected at time slot j or earlier. Therefore, the matching rule is proved to be monotonic.

Then we check whether the trading price computed by Algorithm 2 is exactly the critical value. Suppose a data owner whose bid $B_i = (s_i, e_i, b_i)$ is selected at time slot j and the trading price of this bid computed by Algorithm 2 is $\dot{p_i}$. Therefore, there must be another bid B^0 whose bid price is $\dot{p_i}$ and is selected during the time period of $[j, e_i - 1]$. If the data owner submits another bid $\hat{B_i} = (s_i, e_i, \dot{p_i} - \xi)$ instead of B_i, where $\xi > 0$, then $\hat{B_i}$ would be selected during its active time and makes B^0 fails. On the contrary, if the data owner submits another bid $\bar{B_i} = (s_i, e_i, \dot{p_i} + \zeta)$ instead of B_i, where $\zeta > 0$, then $\bar{B_i}$ would not be selected since its bid price is higher than all selected bids during its active time. So we have verified that the trading price computed by Algorithm 2 is the critical value. We therefore conclude that the proposed auction mechanism is incentive-compatible.

Theorem 2 ***(Individual rationality).*** *The proposed online auction mechanism is individually rational.*

Proof. For a data owner whose bid fails, her utility is zero. For a data owner whose bid is selected at any time slot, we can compute the trading price of her data block according to Algorithm 2. Suppose a bid $B_i = (s_i, e_i, b_i)$ is selected at time slot j, there must be another bid $B_{i'} = (s_{i'}, d_{i'}, b_{i'})$ chosen at time slot j and p_i is updated to $b_{i'}$ (Line 10 in Algorithm 2). According to the update rule of trading prices, the final trading price would be $p_i \geq b_{i'}$. We can see that $b_{i'} \geq b_i$; otherwise, $B_{i'}$ would be selected at time slot j instead of B_i according to the matching rule in Algorithm 1. Since we have demonstrated that the auction mechanism is incentive-compatible, *i.e.*, every data owner would report their valuation truthfully, we can get that $p_i \geq b_{i'} \geq b_i = v_i$. Therefore, the utility of data owners are always nonnegative.

Theorem 3 ***(Competitive ratio).*** *The online matching algorithm achieves a competitive ratio of $\frac{1}{2}$,* i.e., *$\Delta_{online}/\Delta^* \geq \frac{1}{2}$, where Δ_{online} and Δ^* denote the resulting social efficiency of the online matching algorithm and the optimal solution of Eq. (6), respectively.*

Proof. The competitive ratio is computed by introducing a parameter a whose value is 0 initially. For a bid $B_i = (s_i, e_i, v_i)$ that is selected both in the online matching rule and the optimal solution, increase a by $\varphi_j - v_i$ (suppose B_i is selected at time slot j). For a bid B_i that is selected in the optimal solution at time slot j to a data request but not in the online matching rule, suppose this data request is matched to another bid $B_{i'}$ with a bid price of $v_{i'}$ in the online case, i.e., $v_{i'} \leq v_i$, increase a by $\varphi_j - v_{i'}$. Then, we can get $a \geq \Delta^*$.

For each bid that is matched by the online matching rule, its matching value $\varphi_j - v_i$ is added to a at most twice, *i.e.*, $a \leq 2\Delta_{online}$. Therefore, we have $2\Delta_{online} \geq a \geq \Delta^*$.

4 Numerical Illustration

In the section, we perform extensive simulation and report simulation results to show performance of the proposed online auction mechanism for data trading with dynamic data owners.

4.1 Methodology and Simulation Settings

We compare our proposed online auction mechanism with an optimal data trading algorithm with an optimal matching rule based on complete information and a incentive-compatible trading price rule. Suppose that active time of data owners are known and submitted in advance, we employ the Hungarian algorithm to find the optimal matching and then introduce the VCG payment scheme to compute trading prices. That is to say, since we have complete information about all bids and data requests in the auction and make decisions off-line, we can solve the optimization problem in Eq. (6) using the Hungarian algorithm. The VCG scheme can be utilized to determine trading prices, stimulating data owners to report their private information truthfully.

In simulation experiments, we suppose that both data requests and data owners are generated randomly. Specifically, both arrivals of data owners and data requests are generated from Poisson distributions, *i.e.*, the number of data owners newly joining the data trading market at each time slot follows a Poisson distribution with a parameter λ_o; the number of data requests submitted at each time slot follows a Poisson distribution with a parameter λ_r. The active time length, *i.e.*, the number of slots of active time, follows a uniform distribution.

The valuation every data owner or data requester is randomly chosen from a uniform distribution. Settings of all parameters used in simulation experiments are listed in Table 1.

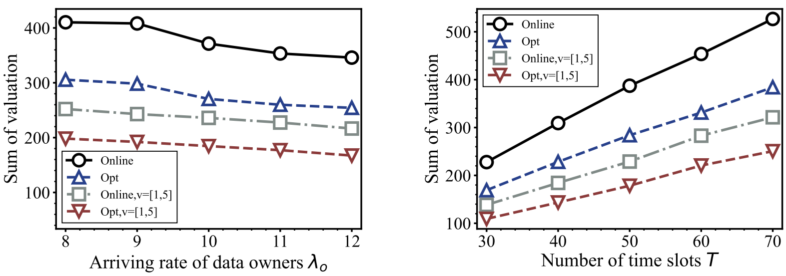
Table 1. Summary of default settings.

Parameter	Default value
Rate of data owners, λ_o	10
Rate of data requests, λ_t	3
Range of active time length	[5, 20]
Range of data owners' valuation	[1, 10]
Number of time slots in a round, T	50

In following pictures, the online auction mechanism and the optimal data trading algorithm are denoted by "Online" and "Opt" in the legends, respectively. Furthermore, these approaches under different valuations of data owners' are evaluated, denoted by "Online, v = [1,5]" and "Opt, v = [1,5]" for example.

These two approaches are evaluated with extensive simulations based on three metrics of *social efficiency*, *competitive ratio*, and *running time*. We conduct several groups of experiments and report comparison results of these two approaches. Each point in these figures is average result over 100 runs.

4.2 Numerical Results

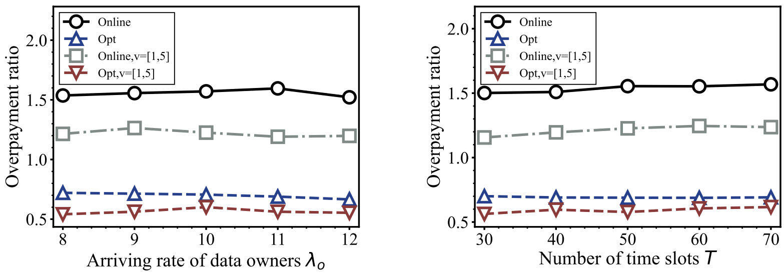


(a) Comparison on sum of valuation with different rates of data owners, λ_o.

(b) Comparison on sum of valuation with different numbers of time slots in a round, T.

Fig. 2. Performance of sum of valuation.

According to Eq. (4), C is a constant which is irrelevant with matching and trading prices of bids, we instead evaluate performance of sum of valuation of all data owners whose have been matching, *i.e.*, $\sum_{i=1}^{N}\sum_{j=s_i}^{e_i-1} x_{i,j}v_i$, in Fig. 2. As shown in Fig. 2a, there is a decrease in sum of valuation for all matched data owners when there are more available data owners with a larger arriving rate λ_o. It is obvious that sum of valuation is on the increase along with the number of time slots in a round, since there are more data requests are satisfied and more data owners are chosen. The performance of Opt outperforms Online in terms of sum of valuation.



(a) Comparison on overpayment ratio with different rates of data owners, λ_o.

(b) Comparison on overpayment ratio with different numbers of time slots in a round, T.

Fig. 3. Performance of overpayment ratio.

To stimulate data owners to honestly report their valuation, the trading price of a data block is usually no lower than the valuation that data owner claims. We further introduce a metric of *overpayment ratio* to show that data requests should pay extra money to ensure social efficiency. The overpayment ratio is the amount of extra expenditure (*i.e.*, the difference between the trading price and the valuation of a data block) to the valuation. The performance of overpayment ratio is shown in Fig. 3. Compared to Opt, the proposed approach of Online must pay higher prices to encourage data owners' cooperation since less information is known in the setting of Online. We can also get that overpayment ratio keeps stable with different arriving rates of data owners or different numbers of time slots in a round.

Evaluation on Competitive Ratio. We plot empirical CDF of competitive ratio of the proposed online auction with different parameters in Fig. 4. In each parameter setting, simulations are repeated 1,000 times. The result on competitive ratio of each run is regarded as a sample; all these samples are utilized to derive the empirical CDF. When the valuation range of data owners varies, we can see that the competitive ratio is always above the bound of 0.5. To simplify the simulation, we set the valuation of each data request to be the upper bound of the valuation of data owners and then compute social efficiency of Online.

To compare the proposed online auction mechanism with optimal algorithm in detail, we further compare ratio of sum of valuation between different methods

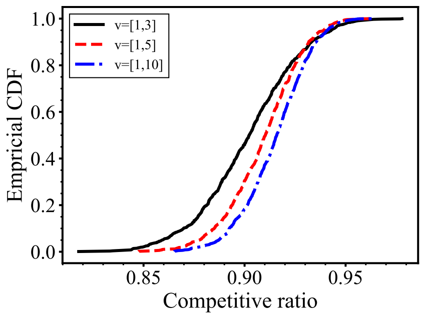


Fig. 4. Empirical CDF of competitive ratio derived from 1,000 simulations.

with different parameter settings; results are shown in Table 2 and Table 3. We can see that the sum of valuation of Online is a little higher than that of Opt. Results of ratio of sum of valuation range from 1.2 to 1.4. Ratio of sum of valuation remains stable, even if number of time slots increases or arriving rate of data owners increases. With smaller range of valuations, *i.e.*, v = [1, 5], the ratio is smaller since data owners participate in a more competitive auction and it is easier to induce their truthfulness.

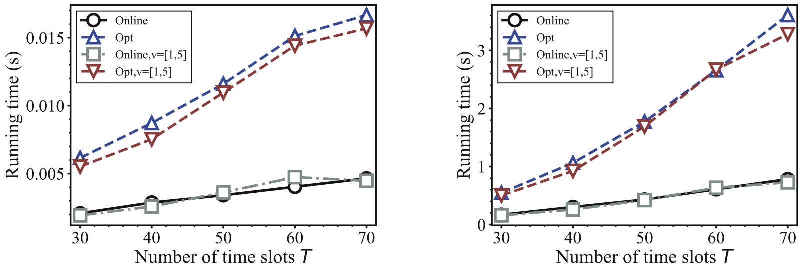
Table 2. Ratio of sum of valuation when the number of time slots in a round changes.

T	30	40	50	60	70
Ratio (Online)	1.35501	1.36143	1.36706	1.37421	1.37441
Ratio (Online, v = [1, 5])	1.27243	1.29097	1.28802	1.28590	1.28542

Table 3. Ratio of sum of valuation when arriving rate of data owners changes.

λ_o	8	9	10	11	12
Ratio (Online)	1.34804	1.37381	1.37681	1.36705	1.36356
Ratio (Online, v = [1, 5])	1.27757	1.26861	1.28496	1.28957	1.29698

Evaluation on Running Time. To show computation efficiency of proposed online auction mechanism, running time of both matching (Algorithm 1) and trading price determination (Algorithm 2) are recorded and plotted in Fig. 5. Plotted lines of Online shows modest growth in running time for both Algorithm 1 and Algorithm 2 as shown Fig. 5; lines of Opt increase dramatically with the number of time slots increases. Consequently, Online is much more computationally efficient than Opt.



(a) Running time of Algorithm 1. (b) Running time of Algorithm 2.

Fig. 5. Performance of running time.

5 Related Work

We review related work from the following two aspects and point out that existing solutions cannot be applied to solve our problem.

5.1 Decentralized Data Trading Based on the Blockchain Technology

A great many research papers have paid attention to design a data trading market or platform based on blockchain technology [4,8,9,18,21] because of the absence of a trustworthy and centralized data trading platform, single point of failure and DDoS. Dai et al. claim that both data brokers and buyers are dishonest and none of them is accessible for raw data; data processing and analysis algorithms encoded in smart contracts are deployed in a secure data trading platform supported by the hardware of Intel's Software Guard Extensions (SGX) based secure execution environment [4]. Another trusted data trading platform employing both the blockchain technology and trusted execution environment (TEE) is implemented by Su et al. where the trusted trading platform contains a special kind of nodes each of which supports TEE and serves as a trust exchange for exchanging data or payment between data sellers and data buyers [18]. Ha et al. introduce a decentralized private data trading marketplace called "Digital Me" based on the blockchain technology [8], where data sellers and data buyers trade personal data directly without trustworthy servers. The AI agent included in "Digital Me" serves as a trading assistant to recommend trading prices based on a user's personal data and data transaction history data. He et al. propose a distributed and trusted data trading platform based on blockchain technology to detect misbehavior of participants and a dataset similarity comparison scheme based on MinHash for detecting illegal resale efficiently is then employed [9]. Zheng et al. deploy smart contracts to solve the problem of data matching and reward distribution in a distributed data trading platform and then introduce proxy re-encryption to guarantee secure data transmission, where trading data are encrypted and only valid data requesters are allowed to decrypt trading data

[21]. Nguyen et al. design a distributed ledger based IoT data trading system along with three typical data trading protocols for city-level environmental monitoring using NB-IoT connections and further analyze the cost of data trading in terms of end-to-end transmission latency and energy consumption [15].

5.2 Trading Data with Different Levels of Privacy

Existing studies [3,12,17,22] investigate how to trade private data where the authors propose pricing functions for personal data or other private data to compensate data owners' different levels of privacy loss. Private data with different privacy loss is generally traded at different prices. Furthermore, data in different formats *e.g.*, data samples and range counts, are returned to data consumers. Higher prices should be paid to a higher level of privacy loss caused by traded data undoubtedly. A few desirable properties, *e.g.*, arbitrage-freeness, budget feasibility, performance accuracy based on traded data, are considered when designing pricing functions.

Gao et al. propose a pricing rule based on an auction-based model where both task description and bid prices in bids are possible to disclose sensitive information of data owners; they employ differential privacy schemes at both stages of data collection and trading price determination [7]. Another research paper [11] using the auction-based model introduces geo-indistinguishability to quantify privacy loss of geographical locations and then pays for their sensing cost as well as privacy breach. Zhang et al. point out that disclosure of raw social media data of users probably cause privacy leakage, because anonymous user IDs can be linked to real users and they propose a novel mechanism based on a notion of ϵ-text indistinguishability to guarantee different user privacy as well as to achieve high data utility [20].

All existing studies, however, neglect the dynamic behavior of data owners as well as randomly generated data requests of data consumers in IoT data trading. It is reasonable that end devices only trade their data intermittently due to limited resources or mobility.

6 Conclusion

In the paper we have investigated the data trading problem with dynamic data owners, aiming to share IoT data and to take full advantages of big data. Existing studies have neglected an important observation that a data owner is not always available to trade her data blocks except in her active time period. To this end, we have proposed a truthful and efficient online data trading algorithm which not only resolves the problem of matching dynamic data owners and randomly generated data requests with near-optimal social efficiency, but also determines the trading price of each data block which ensures the incentive-compatibility and individual rationality of participants.

Acknowledgment. This research is partially supported by Shanghai Sailing Program (Grant No. 19YF1402200) and the Fundamental Research Funds for the Central Universities (Grant No. 2232021D-23).

References

1. GXS TradeWeb - A service of GXS. https://gxstradeweb.gxsolc.com/pub-html/EdiServiceInfoFrameset.html. Accessed 11 Apr 2021
2. Terbine: The data exchange for advanced mobility and infrastructure. https://terbine.com/. Accessed 11 Apr 2021
3. Cai, Z., He, Z.: Trading private range counting over big IoT data. In: Proceedings IEEE International Conference Distributed Computing System (ICDCS), pp. 144–153 (2019)
4. Dai, W., Dai, C., Choo, K.K.R., Cui, C., Zou, D., Jin, H.: SDTE: a secure blockchain-based data trading ecosystem. IEEE Trans. Inf. Forensics Secur. **15**, 725–737 (2019)
5. Feng, Z., Chen, J.: Blockchain based mobile crowd sensing for reliable data sharing in IoT systems. In: Proceedings IFIP Networking, pp. 1–3 (2021)
6. Feng, Z., Chen, J., Zhu, Y.: Uncovering value of correlated data: trading data based on iterative combinatorial auction. In: Proceedings IEEE International Conference Mobile Ad-Hoc and Smart System (MASS) (2021, to appear)
7. Gao, G., Xiao, M., Wu, J., Zhang, S., Huang, L., Xiao, G.: DPDT: a differentially private crowd-sensed data trading mechanism. IEEE Internet Things J. **7**(1), 751–762 (2020)
8. Ha, M., Kwon, S., Lee, Y.J., Shim, Y., Kim, J.: Where WTS meets WTB: a blockchain-based marketplace for digital me to trade users' private data. Pervasive Mob. Comput. **59**, 101078 (2019)
9. He, Y., Zhu, H., Wang, C., Xiao, K., Zhou, Y., Xin, Y.: An accountable data trading platform based on blockchain. In: Proceedings IEEE International Conference Computing Communication Workshops (INFOCOM WKSHPS), pp. 1–6 (2019)
10. idc.com: IoT growth demands rethink of long-term storage strategies, says IDC, https://www.idc.com/getdoc.jsp?containerId=prAP46737220. Accessed 11 Apr 2021
11. Jin, W., Xiao, M., Li, M., Guo, L.: If you do not care about it, sell it: trading location privacy in mobile crowd sensing. In: Proceedings IEEE International Conference Computing Communication (INFOCOM), pp. 1045–1053 (2019)
12. Li, C., Li, D.Y., Miklau, G., Suciu, D.: A theory of pricing private data. Commun. ACM **60**(12), 79–86 (2017)
13. Liu, T., Li, D., Cao, C., Gao, H., Li, C., Feng, Z.: Joint location-value privacy protection for spatiotemporal data collection via mobile crowdsensing. In: Proceedings International Conference Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom) (2021, to appear)
14. Liu, T., Wu, W., Zhu, Y., Tong, W.: Accuracy-Guaranteed event detection via collaborative mobile crowdsensing with unreliable users. In: Wang, X., Gao, H., Iqbal, M., Min, G. (eds.) CollaborateCom 2019. LNICST, vol. 292, pp. 729–744. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30146-0_49
15. Nguyen, L.D., Leyva-Mayorga, I., Lewis, A.N., Popovski, P.: Modeling and analysis of data trading on blockchain-based market in IoT networks. IEEE Internet Things J. **8**(8), 6487–6497 (2021)

16. Nisan, N., Roughgarden, T., Tardos, E.V., Vazirani, V. (eds.): Algorithmic Game Theory. Cambridge University Press, Cambridge (2007)
17. Niu, C., Zheng, Z., Wu, F., Tang, S., Gao, X., Chen, G.: Unlocking the value of privacy: Trading aggregate statistics over private correlated data. In: Proceedings ACM International Conference on Knowledge Discovery and Data Mining (KDD), pp. 2031–2040 (2018)
18. Su, G., Yang, W., Luo, Z., Zhang, Y., Bai, Z., Zhu, Y.: BDTF: a blockchain-based data trading framework with trusted execution environment. CoRR abs/2007.06813 (2020)
19. Yang, C., et al.: Mobile data sharing with multiple user collaboration in mobile crowdsensing (short paper). In: Gao, H., Wang, X., Yin, Y., Iqbal, M. (eds.) CollaborateCom 2018. LNICST, vol. 268, pp. 356–365. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-12981-1_25
20. Zhang, J., Sun, J., Zhang, R., Zhang, Y., Hu, X.: Privacy-preserving social media data outsourcing. In: Proceedings IEEE International Conference Computing Communication (INFOCOM), pp. 1106–1114 (2018)
21. Zheng, S., Pan, L., Hu, D., Li, M., Fan, Y.: A blockchain-based trading platform for big data. In: Proceedings IEEE International Conference Computing Communication Workshops (INFOCOM WKSHPS), pp. 991–996 (2020)
22. Zheng, S., Cao, Y., Yoshikawa, M.: Money cannot buy everything: trading mobile data with controllable privacy loss. In: Proceedings IEEE International Conference Mobile Data Management (MDM), pp. 29–38 (2020)