Competitive Data Trading Model with Privacy Valuation for Multiple Stakeholders in IoT Data Markets

Hyeontaek Oh, *Student Member, IEEE,* Sangdon Park, *Member, IEEE,* Gyu Myoung Lee, *Senior Member, IEEE,* Jun Kyun Choi, *Senior Member, IEEE* and Sungkee Noh

***Abstract*—With the widespread of Internet of Things (IoT) environment, a big data concept has emerged to handle a large number of data generated by IoT devices. Moreover, since data-driven approaches now become important for business, IoT data markets have emerged, and IoT big data are exploited by major stakeholders such as data brokers and data service providers. Since many services and applications utilize data analytic methods with collected data from IoT devices, the conflict issues between privacy and data exploitation are raised, and the markets are mainly categorized as privacy protection markets and privacy valuation markets, respectively. Since these kinds of data value chains (which are mainly considered by business stakeholders) are revealed, data providers are interested in proper incentives in exchange for their privacy (i.e., privacy valuation) under their agreement. Therefore, this paper proposes a competitive data trading model that consists of data providers who weigh the value between privacy protection and valuation as well as other business stakeholders. Each data broker considers the willingness-to-sell of data providers, and a single data service provider considers the willingness-to-pay of service consumers. At the same time, multiple data brokers compete to sell their dataset to the data service provider as a non-cooperative game model. Based on the Nash Equilibrium analysis (NE) of the game, the feasibility is shown that the proposed model has the unique NE that maximizes the profits of business stakeholders while satisfying all market participants.**

***Index Terms*—Internet of Things, Data market, Profit maximization, Non-cooperative game, Privacy valuation**

I. Introduction

With the development of Internet of Things (IoT), various data sources (e.g., not only the massive number of connected devices but also numerous services/applications) generate a huge amount of data. According to the reports from Cisco, the total number of connected devices becomes 28.5 billion by 2021 [1], and the total amount of data created by these devices will reach 847 ZB per year by 2021 [2]. Now, it becomes hard to search, discover, process, and analyze the proper data from the whole. As a result, big data technology has emerged to extract the fine value of the data.

With the emerging big data concept, data-driven approaches now become essential for numerous IoT based services and applications with the support of various cloud computing technologies [3]–[5], and data becomes a new valuable asset for the fourth industrial revolution. Typically, a big data market consists of three major players: i) data broker (or data vendor); ii) data service provider (or data consumer); and iii) service consumer [6]. Specifically, a data broker collects raw data from various data sources (e.g., publicly available data, nonpublic data obtained through private contracts, online tracking data, etc.) and sells big data to other third party data service providers [7], [8]. Moreover, the data service providers utilize big data from the data brokers to raise revenue for improving the quality of their services to satisfy their customers. There are complex value chains for various ecosystems exploiting big data. According to the report from International Data Corporation [9], the worldwide revenue of big data analytics markets will grow up to $260 billion in the year 2022.

Many IoT services and applications require a detailed analysis from collected data through IoT devices. According to the study related to the relationship between privacy concerns and data innovation through new services and applications in IoT environments [10], it should be considered not only technical aspects but also regulatory and economics aspects. These kinds of heterogeneous characteristics increase the complexity of assessing the impacts on privacy; therefore, the potential (intended or unintended) privacy issues may be incurred.

In keeping with this trend, today, IoT data markets can be categorized into two major approaches regarding privacy [11]: i) privacy protection market and ii) privacy valuation market. In the privacy protection market, privacy-enhancing technologies are provided to data providers, which minimize potential privacy infringement risks and protect possible privacy violations caused by IoT services/applications [12], [13]. On the other hand, in the privacy valuation market, data brokers offer proper benefits or incentives to data providers (who consider the value between privacy protection and valuation) for collecting IoT datasets under agreement or consent. Data brokers share revenue from the data service providers as data consumers to data providers, which makes data providers motivated to participate in IoT data markets more.

H. Oh and J. K. Choi are with School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea (hyeontaek@kaist.ac.kr and jkchoi59@kaist.edu).

S. Park (corresponding author) is with Information and Electronics Research Institute, KAIST, Daejeon, Korea (sangdon.park@kaist.ac.kr)

G. M. Lee is with Department of Computer Science, Liverpool John Moores University, Liverpool, United Kingdom (g.m.lee@ljmu.ac.uk).

S. Noh is with Blockchain Technology Research Center, Electronics and Telecommunications Research Institute, Korea (sknoh@etri.re.kr).

This work was supported by an Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government [19ZH1200, Core Technology Research on Trust Data Connectome].

In IoT data markets, the needs for the privacy valuation market are increased because many data providers think their data as a financial asset [14]. Therefore, issues for privacy valuation have focused on providing proper economic benefits to data providers, and one of the important concepts for this market is a willingness-to-sell (WTS) data with offered prices from data brokers (simply, the WTS of data providers). According to the surveys [15], [16], many data providers have their WTS data with the proper benefits or incentives. In addition, studies in [17], [18] investigated that data providers have different WTS depending on the privacy sensitivity of data types by showing cumulative distribution of the portion of data providers who wanted to sell IoT data with various offered prices.

From the above backgrounds, the data provider should also be considered as a major stakeholder in IoT data markets. However, conventional studies for IoT data markets have mainly targeted interactions between business stakeholders (e.g., data broker and data service provider). Many studies only consider a willingness-to-pay (WTP) of consumers for provided services, which are the results of big data exploitation, because IoT data markets are mainly controlled by business stakeholders. Only a few studies focused on the behavior of data providers as a WTS data concept for privacy valuation (i.e., willingness-to-accept with the offered price for selling data to a data broker).

The previous work of the authors tackled a data trading model that jointly considered both WTP of data consumers and WTS of data providers in a data brokers' perspective [19]. It showed that the data trading model was feasible even if the data broker spent costs for buying data from data providers. However, it only showed a single data broker model without considering the behavior of a third party data service provider that is actually exploiting big data from the data brokers. Therefore, as an extension of the authors' previous work, this paper proposes an extended data trading model considering a competition among multiple data brokers and a data service provider as well as the behavior of data providers (considering the privacy valuation) and service consumers (considering the quality of a provided service) in IoT data markets. The contributions of this paper are summarized as follows:

- This paper designs a competitive data trading model with four major players: data providers, multiple data brokers, a single data service provider, and service consumers to cover various data value chains in IoT environments (i.e., data production, data exploitation, and data consumption). In the proposed model, they are organically formulated in four hierarchical levels with competitiveness.
- The proposed trading model among four stakeholders is analyzed by describing their behavior to maximize their own benefits. Each data broker competes to sell the dataset, and the data service provider decides the optimal budget allocation within the limited budget by considering a unit price of a dataset offered by each data broker. This paper proposes a unified method to decide the unit price of dataset for each data broker, which makes it possible to compare the competitiveness of each data broker even if it handles different data types. With the unified measure, it is formulated as a non-cooperative game between the data service provider and the data brokers. The existence and the uniqueness of the Nash Equilibrium (NE) of the proposed model are shown by utilizing similar analysis results from the previous work [20].
- The data service provider decides the optimal budget to maximize its profit with the consideration for both revenues from the service consumers (which is decided by the service quality obtained by exploiting the dataset and their WTP for the service) and costs for buying dataset from the data brokers. On the other hand, with the payment from the data service provider, each data broker minimizes costs for obtaining dataset to achieve the required dataset quality (measured by the correlation between each data type and the amount of collected data) from the data providers by considering their privacy sensitivity and WTS. The proposed WTP and WTS are also designed based on literature (which are [21] and [18], respectively) to reflect real-world behavior.
- Based on the theoretical and experimental analysis, this paper shows the impacts of important parameters of the proposed data trading model as well as the behavior of the data brokers (which aware data providers with the needs for their privacy valuation) and the data service providers (which take their service consumers' satisfaction into account). In addition, with real-world datasets, it shows that the results with data brokers have different competitiveness in the market.

With the best of our knowledge, this paper is the first paper that jointly considers not only a competitive data trading model as a game-theoretic approach between data brokers and a data service provider but also profit maximization problems of data brokers and the data service provider by considering the behavior of data providers that take the value between privacy protection and valuation into account (i.e., WTS of data providers with privacy sensitivity of data types) and service consumers (i.e., WTP and service quality) with characteristics of data types and dataset quality.

The rest of this paper is organized as follows. Section II introduces literature regarding IoT data markets, data trading models, and privacy valuation schemes. Section III presents an overview of the proposed IoT data trading model that consists of four major players. Section IV formulates a data trading model between the data brokers and the data service provider as a non-cooperative competition game, and it shows the existence of the unique NE. Section V and Section VI formulate profit maximization problems of the data service provider and the data brokers by considering the WTP of service consumers and the WTS of data providers, respectively. Section VII shows some numerical and experimental results, including the analysis based on a real-world dataset. Finally, this paper is concluded in Section VIII.

II. Related Work

The reports [7], [8] identified various big data ecosystems driven by data brokers. Moreover, *Cavanillas et al.* introduced

an overall big data ecosystem in Europe, including big data value chains, various real-world services/applications, and a future roadmap for data-driven economy [6]. Various data market structures and data trading models are also identified in [22].

Data markets and data trading issues have recently motivated, which mainly considered data brokers and data service providers. *Niyato et al.* [21] proposed a simple IoT data market model that considered WTP of service consumers depending on the service quality of a data service provider with IoT data quantity. *Ren et al.* [23] proposed a data purchasing and data placement model for a cloud-based data market. Meanwhile, competitive data trading models also have been studied. *Jiao et al.* [24] proposed an auction-based big data trading model that service consumers bid a service fee to a data service provider. In this paper, the data service provider utilized the Bayesian-optimal mechanism to maximize profits. *Jang et al.* [20] proposed a data trading model with a single data service provider and multiple data sources in an IoT data market. This paper modeled the interaction between the data service provider and the data sources as a non-cooperative game and showed the existence and the uniqueness of the NE point. Moreover, *Shen et al.* [25] proposed a profit optimization model using a Stackelberg game approach with the relationships among data sources, service providers, and service users. Since these studies mainly targeted IoT environments, they did not consider the characteristics of data as well as the behavior of data providers that are key factors for data trading markets.

On the other hand, many studies have focused on privacy valuation schemes as well as WTS of data providers with proper incentives for data value chains (e.g., literature in behavior economics). *Elvy* [11] introduced various privacy pricing models for data economy. *Malgieri and Custers* [26] investigated that the monetary value of data can be quantified with various data pricing factors by considering both characteristics of data quality and data themselves. People have different privacy concerns depending on data types, so privacy sensitivity of data (i.e., privacy attitude of the data provider) should be considered for modeling WTS [17], [18]. The concept of WTS (willingness-to-sell or -share) with privacy valuation considering incentives or rewards of data providers also have been studied. *Jai and King* [27] investigated a trend of WTS data to the data brokers in online services, and it showed that WTS significantly increased with proper rewards. In addition, *Kim et al.* [28] examined factors affecting WTS based on the privacy calculus theory for various IoT services, and it showed that perceived benefits had a positive effect on WTS data.

In the field of engineering, there are few studies considered the behavior and characteristics of data providers regarding privacy. *Parra-Arnau* [29] investigated the trade-off between privacy and money of data providers and proposed an optimization model for profile-disclosure risks and economic rewards. *Su et al.* [30] proposed an incentive-based crowdsourcing scheme for collecting various data in cyber-physical-social systems with an auction-based price bidding scheme for data providers. *Tian et al.* [31] proposed a contract-based mechanism for data trading. In this model, the data seller considered a utility via balancing the trade-off between data trading benefits and data privacy costs. *Ghosh and Roth* [32] proposed an auction-based privacy trading model using differential privacy techniques. It designed an auction model between data providers (who considered the chance to reveal their privacy) and data buyers (who considered their costs and the accuracy of data analytics using datasets). *Oh et al.* [19] (N.B., the authors' previous work) proposed a data trading model that jointly considered WTP of data consumers and WTS of data providers for maximizing profits of the data broker.

This paper proposes a data trading model with various stakeholders (i.e., data providers, data brokers, a data service provider, and service consumers to cover the entire data value chain) that behave to maximize their own benefits in a data value chain by jointly considering WTS of the data providers with their privacy considerations and WTP of the service consumers with the required quality of service as well as a competition between the data brokers and the data service provider.

III. System Model

A data trading model considered in this section consists of four groups that behave for their own benefits with K types of data: R data brokers $(d_i)_{i=1}^{R}$ participating in the market with $(N_i)_{i=1}^{R}$ potential data provider groups, who are interested in selling their data with privacy considerations under agreement, and a single data service provider that provides a service to M potential service consumer groups. Under four major players (or stakeholders), an overview of the proposed data trading model is described in Figure 1, and the entire flow is illustrated in Figure 2. In addition, the major symbols used in this paper are listed in Table I. In this paper, the proposed data trading model considers that each data broker deals with similar data

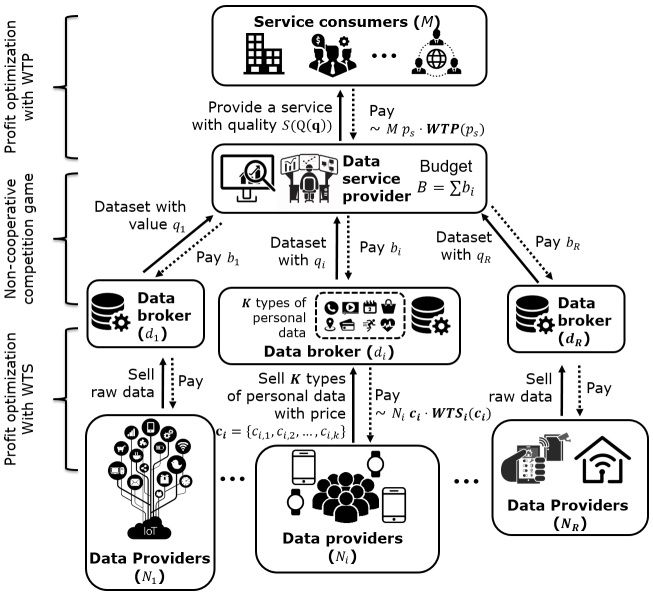


Fig. 1. The proposed competitive data trading model with multiple stakeholders in an IoT data market

TABLE I
MAJOR SYMBOLS

Symbols	Definition
K	Number of data types in the market
R	Number of data brokers in the market
N_i	Number of data providers in the ith data broker ($i \in R$)
d_i	The ith data broker ($i \in R$)
$c_{i,k}$	Cost for each data type $k \in K$ in d_i
$\mathbf{c}_i$	Cost vector for all data types in d_i $(c_{i,k})_{k=1}^{K}$
$\Phi_{i,k}$	The provider's willingness-to-sell function for the kth data type in d_i
$\vec{\Phi}_i$	The provider's willingness-to-sell function in the ith data broker
U_i	Profit function of the d_i
q_i	Quality of dataset provided by the d_i
M	Number of data consumers
p_s	Subscription fee of each data consumer
S	Service quality of the data service provider
Q	Expected dataset quality of the data service provider from $(q_i)_{i=1}^{R}$
Ψ	The consumer's willingness-to-pay function
B	Total amount of budget for the data service provider
b_i	Allocated budget to each data broker d_i

types; that is, each data broker competes to sell their dataset to the data service provider.

In this proposed data trading model, there are two optimizations (i.e., one between the data service provider and the service consumers and the other between the data brokers and the data providers) and two competitions (i.e., one among the data brokers and the other between the data service provider and the data brokers) among players.

First, the data service provider decides a budget B to buy dataset with quality $\mathbf{q} = (q_i)_{i=1}^{R}$, which are measured by considering both the amount of collected data and the correlation of them (in equation (18) of Section VI), from all data brokers. It anticipates the expected revenue by considering the number of the paid service consumers that can be decided by the WTP of the service consumers with the offered price of the service p_s and the expected service quality S (which can be obtained by the expected dataset quality $Q(\mathbf{q})$). Since the budget B is consumed as the cost, the data service provider finds the optimal required budget B to maximize its own profit.

After deciding the budget B, the data service provider requests a bid to gather dataset from data brokers. The interaction between these two groups can be explained as a game model. The first competition is within the group of data brokers. Each data broker buys data (e.g., location, service usage log, etc.) from data providers and sells collected dataset to the data service provider. Simultaneously, data brokers compete by bidding for their dataset. This can be described as a normal form game. In other words, each data broker chooses a bid without knowledge of others' bids.

The second competition happens between the data brokers and the data service provider. The data service provider spends the limited budget B for obtaining datasets with the quality $\mathbf{q}$ to maximize its quality of service S with the expected dataset quality $Q(\mathbf{q})$. On the other hand, each data broker receives a revenue b_i from the data service provider by selling a dataset with quality q_i. As a result, the revenue of each data broker is determined.

Note that this competitive trading model consists of multiple sellers (i.e., data brokers) and a single buyer (i.e., the data service provider). The reason why authors adopt this

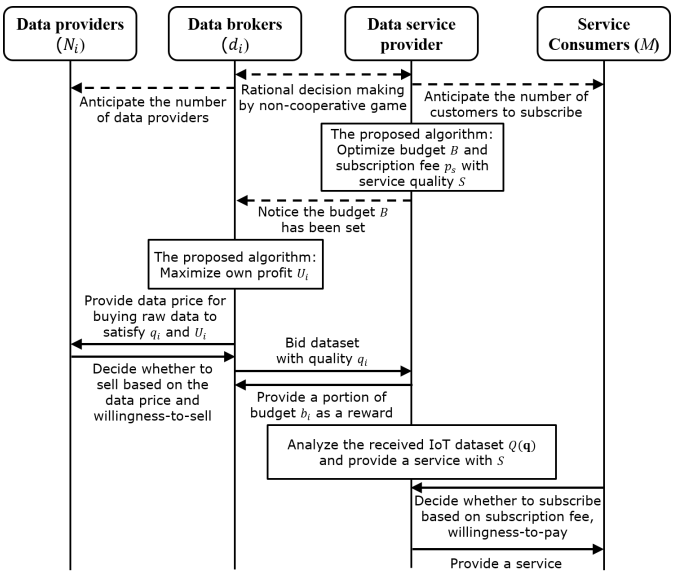


Fig. 2. Operational flows of the proposed data trading model

model is related to the characteristics of data and budget. The budget is a tangible and limited resource that cannot be multiplied by spending the budget; however, data have different characteristics. Data are intangible goods that are easily duplicated and copied without losing those inherent values; that is, the competition model between data service providers is negligible because the required goods (i.e., IoT datasets) are not limited in general, especially when multiple sellers are handling similar data types. Therefore, from the data brokers' perspective, they need to sell their collected dataset as much as possible; meanwhile, the data service provider needs to find the most efficient way to spend its limited budget.

During the competition, each data broker d_i also optimizes its profit for selling dataset with quality q_i by considering both the revenue b_i from the data service provider and the costs for buying data from their data provider group N_i. Since the revenue is decided by the non-cooperative game, the data broker focuses on minimizing the entire cost $\mathbf{c}_i = (c_{i,k})_{k=1}^{K}$ to obtain required dataset with quality q_i by taking the WTS of the data providers for each data type k into account.

This paper, firstly, describes the proposed non-cooperative game model between the data service provider and the data brokers in Section IV, and then the profit optimization problems for the data service provider and the data brokers are proposed in Section V and Section VI, respectively.

IV. NON-COOPERATIVE COMPETITION GAME

This section focuses on an analysis of a non-cooperative game among the data brokers in the proposed model to prove the existence and the uniqueness of Nash Equilibrium (NE). Since each data broker handles similar data types for selling datasets in this paper, the data service provider allocates the budget to the data brokers proportionally as in [20] and [33]. In other words, the revenue b_i of the data broker (d_i) is defined as

$$b_i = \frac{q_i}{\sum_{j=1}^{R} q_j} B. \tag{1}$$

Note that the sum of all b_i cannot exceed the budget B.

This resource allocation model is one of the possible models for a bidding-based competition among multiple players (i.e., data brokers) for the single resource (i.e., the budget B). In this model, each player is able to obtain the resource with the fixed unit price decided by the demand of the entire players (i.e., the entire resources can be distributed to all players in the market). Moreover, even if a player with an extreme demand participates in the market, it is possible to make a balance between the entire supply of the resource and the entire demand of players because the player makes the unit price of the resource higher. Therefore, this budget allocation model is reasonable and feasible for the competition proposed in this paper. To this end, an expected profit function U_i is defined for the data broker d_i, and a non-cooperative game among the data brokers with a NE.

Definition 1 (Expected Profit Function)**.** The expected profit function U_i of the data broker d_i is

$$U_i(q_i, \mathbf{q}_{-i},) = \frac{q_i}{\sum_{j=1}^{R} q_j} B - \delta_i q_i, \tag{2}$$

where δ_i is the unit price for dataset with quality $q_i \in [0, \infty)$.

As shown in Fig. 1, the data brokers with the expected profit function compete with each other to get the bidding budget of the data service provider; thus, it can be considered as a non-cooperative game among them as defined in Definition 2 and 3. The proposed non-cooperative competition game among data brokers (CGDB) with the models in equations (1) and (2) is suitable that players dynamically participate in the market because it requires the minimum information about other players. That is, the players need the information about the budget B and the unit price δ, which means each player does not need to care about detailed strategies of other players, unlike leader-follower game models (proposed in the previous studies [20], [33]). Since data trading is performed in a real-time manner (e.g., real-time bidding in the online advertisement market [34], [35]), the proposed CGDB is feasible for real-world applications.

Definition 2 (Competition Game among Data Brokers)**.** A competition game among data brokers with the data service provider is formulated as a non-cooperative strategic form game $G = (\mathbf{Q}, U_k)_{k \in \{1, \cdots, R\}}$ and is denoted by the Competition Game among Data Brokers (CGDB). Here, $\mathbf{Q} = \prod_{i=1}^{R} [0, \infty)$ is the domain for all data brokers, and U_k is the expected profit function given by the Definition 1.

Definition 3 (NE of CGDB)**.** A NE of the CGDB G is a profile of strategies $\mathbf{q}^*$ satisfying

$$U_i(q_i^*, \mathbf{q}_{-i}^*) \geqslant U_i(q_i, \mathbf{q}_{-i}^*), \qquad \forall q_i \in [0, \infty).$$

Now, the existence and the uniqueness of NE ($\mathbf{q}^*$) for maximizing $(U_i)_{i=1}^{R}$ are proved. There have been similar works to find the NE in various models. Especially, the proposed game model is similar to the models, studied by *Jang et al.* [20] and *Park et al.* [33], that are originated by *Hajek and Gopal* [36] and *Johari and Tsitsiklis* [37]; therefore, this paper adopts existing analysis models and presents the main theorems for the game model. First, Lemma 1 shows the feasibility of the proposed game model.

Lemma 1. The vector $\mathbf{q}$ is a NE (i.e., $\mathbf{q}^*$ of the game to maximize $(U_i)_{i=1}^{R}$ if and only if at least two components of $\mathbf{q}$ are positive, and the vector $\mathbf{q}$ satisfies a following conditions:

$$\begin{cases} \frac{1}{\delta_i}(1 - \frac{q_i}{\sum_{j=1}^{R} q_j}) = \frac{\sum_{j=1}^{R} q_j}{B} & \text{if } q_i > 0; \\ \frac{1}{\delta_i} \leqslant \frac{\sum_{j=1}^{R} q_j}{B} & \text{if } q_i = 0. \end{cases} \tag{3}$$

Proof. It can be proved by an argument similar to the proof of Lemma [33] and [20] by showing that the necessity and the sufficient conditions hold. Note that the conditions can be derived by taking partial derivative for q_i (i.e., $\frac{\partial}{\partial q_i} U_i(\mathbf{q})$). For the necessity condition, it can be proved by contradiction with the cases of zero and only one participant. The sufficient condition can be shown by considering $\mathbf{q}$ with at least two points (i.e., at least two participants for the market). □

Since the original CGDB is hard to find the NE solution, similar to the previous studies [20], [33], this paper transforms the original game to the modified game model with the modified utility function $\hat{U}_i$ as shown in Lemma 2.

Lemma 2. Consider an optimization problem given by,

$$\max_{\mathbf{b}} \sum_{i=1}^{R} \hat{U}_i(b_i), \tag{4}$$

$$\text{s.t.} \sum_{i=1}^{R} b_i \leqslant B; \tag{5}$$

$$b_i \geqslant 0, \text{ for all } i, \tag{6}$$

where

$$\hat{U}_i(b_i) = \frac{1}{\delta_i}(b_i - \frac{b_i^2}{2B}).$$

Then this problem has the unique solution given by

$$b_i^* = \begin{cases} B(1 - \upsilon\delta_i) \ , & \text{if } \upsilon < \frac{1}{\delta_i} \\ 0 \ , & \text{if } \upsilon \geqslant \frac{1}{\delta_i} \end{cases} \tag{7}$$

where υ is a real value satisfying $\sum_{i \in \{1, \cdots, R\}} b_i^* = B$.

Proof. It can be proved by an argument similar to the proof of Theorem 1 and 3 of [20] and Lemma 1 of [33] by showing the first order derivative of the modified utility function and then applying the Karush-Kuhn-Tucker(KKT) conditions [38] to find the exact b_i^*. Then, the equation (7) can be obtained. During the applying KKT conditions, the condition $\sum_{i=1}^{R} b_i^* = B$ also can be obtained. □

Note that the bidding of the data brokers is non-negative by Definition 1. A data broker with zero bidding (i.e., $q_i = 0$) means that the data broker does not participate in the market. Therefore, it can be assumed that the whole participants (i.e., data brokers) actually participate in a bidding with $b_i > 0$ and $q_i > 0$ for all $i \in R$. Then, the result of Lemma 2 can be simplified as follows:

$$b_i^* = B(1 - \upsilon\delta_i), \text{ for all } i \in R. \tag{8}$$

Moreover, with the condition (8) and the condition $\sum_{i=1}^{R} b_i^* = B$, υ can be obtained by evaluating $\sum_{i=1}^{R} B(1-\upsilon\delta_i) = B$; that is, $\upsilon = (R-1)/(\sum_{i=1}^{R}\delta_i)$.

Lemma 3. The CGDB G has the unique NE $\mathbf{q}^*$ with the corresponding optimal budget allocation $\mathbf{b}^* = \{b_i^*\}_{i\in\{1,\cdots,R\}}$ satisfying $q_i^* = \upsilon\delta_i b_i^*$, given by the solution from Lemma 2.

Proof. It can be proved by an argument similar to the proof of Theorem 2 of [20] and Lemma 2 of [33]. It can be shown as the modified game model in the equation (4) has the unique NE point with the conditions from its first order derivative, and then it can be obtained that the modified game model is actually the same as Lemma 1 of the CGDB.

Here is the sketch of the proof. First, it is shown that there exists the unique $\mathbf{b}^*$ and scalar υ such that $\hat{U}_i(b_i)' = \upsilon$ if $b_i^* > 0$; $\hat{U}_i(b_i)' = 1/\delta_i \leqslant \upsilon$ if $b_i^* = 0$; $\sum_{i\in I} b_i^* = B$ by showing that the problem in Lemma 2 has the same unique solution. Then, it can be shown that the vector $\mathbf{q}^* = (\upsilon\delta_i b_i^*)_{i\in I}$ is a NE, and that it is the unique solution using Lemma 1. Finally, it can be concluded that there exists the unique NE $\mathbf{q}^*$ with the corresponding optimal budget allocation $\mathbf{b}^* = \{b_i^*\}_{i\in\{1,\cdots,R\}}$ satisfying $q_i^* = \upsilon\delta_i b_i^*$. □

Finally, the unique NE point $\mathbf{q}^*$ of the original game in Definition 2 can be obtained by the following theorem.

Theorem 1. The CGDB has the unique NE $\mathbf{q}^*$, given by

$$q_i^* = \frac{B(R-1)(\sum_{j=1}^{R}\delta_j - \delta_i R + \delta_i)}{(\sum_{i=j}^{R}\delta_j)^2}. \tag{9}$$

Proof. From Lemma 2 and Lemma 3, the NE solution q_i^* is given by $q_i^* = \upsilon\delta_i b_i^*$ where $\upsilon = (R-1)/(\sum_{j=1}^{R}\delta_j)$. Then, substituting it in the equation (8), the equation (9) can be derived. □

With the closed form of the unique NE (equation (9)), all players can easily anticipate others' strategies in the market. From the data service provider's point of view, it is possible to decide the amount of dataset with its own budget and sellers' unit price for the dataset. On the other hand, from the data brokers' point of view, it is possible to estimate their competitiveness in the market, which is directly related to their profit.

Since this paper assumes the positive bidding of the data brokers, the NE $\mathbf{q}^*$ in equation (9) must be positive. Thus, the proposed data market model must satisfy the following two inequalities,

$$R-1>0 \quad \text{and} \quad \sum_{j=1}^{R}\delta_j - \delta_i R + \delta_i > 0. \tag{10}$$

The first condition is related to the feasibility of the CGDB game; that is, the CGDB game needs at least two data brokers (i.e., $R \geqslant 2$), which also is identified in Lemma 1. The second condition is related to the behavior of each data broker d_i, and it can be interpreted as follows.

$$\sum_{j=1}^{R}\delta_j - \delta_i R + \delta_i \Rightarrow \begin{cases} \delta_i > 0, & R = 2, \\ \delta_i < \dfrac{\sum_{j\neq i}^{R}\delta_j}{R-2}, & R > 2. \end{cases} \tag{11}$$

The first case ($\delta_i > 0$) shows the basic condition for a unit price of a dataset that is provided by each data broker. The unit price for each data broker should be larger than zero. The second case ($\delta_i < \frac{\sum_{j\neq i}^{R}\delta_j}{R-2}$) indicates that the unit price of each data broker should be competitive enough to participate in the CGDB game; that is, if one data broker has an extremely higher unit price value than other participants, the data broker has no chance to be bid from the data service provider. The detailed methods to decide the unit price δ_i for the dataset with quality q_i for each data broker will be discussed in Section VI.

V. Profit Optimization for the Data Service Provider

This section introduces an optimization problem for the data service provider by using the result of NE analysis. It defines a profit function P of the data service provider by considering the expected service quality S obtained by a gathered dataset with the expected quality Q, which is the function of the gathered dataset from each data broker $\mathbf{q}$ with the budget B. With the expected service quality, at the same time, the data service provider anticipates the expected revenue considering WTP (Ψ) of the service consumers (M) with the service quality S. On the other hand, the budget B is considered as the expected cost for achieving the service quality S.

Note that the proposed models about the expected dataset quality Q and the expected service quality S (based on real-world observations) are quantification methods for the data service provider to anticipate the amount of dataset to buy for maximizing its revenue, and these can be used as some of the criteria for deciding the market participation.

A. Service quality of the data service provider

Before defining the expected service quality, first, this paper newly defines an expected dataset quality Q of the data service provider that is obtained by the entire dataset $\mathbf{q} = (q_i)_{i=1}^{R}$ from all data brokers.

Three principles (obtained by various existing studies [26], [39]–[41] regarding dataset quality models) are applied to define the expected dataset quality with all gathered datasets.

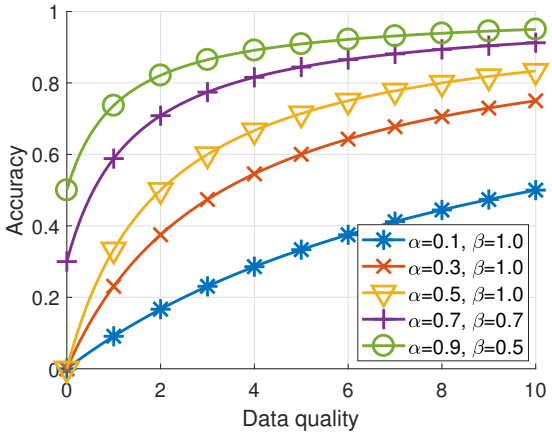


Fig. 3. The trend of expected service quality function S w.r.t. parameters

The first is *data quantity* (i.e., size or volume) related to inherent data quality [39] such as completeness, accuracy. The second is *characteristics of data* [26]; that is, the *combination of dataset* increases the chance for obtaining identifiable information. For example, in an online environment, the cookie syncing technique is widely used to identify and track users' behavior for the better quality of services and customization. One study showed that there is a 99.5% chance that a user will be tracked by all top 10 trackers within 30 clicks on search results [40]. The last is *the law of diminishing marginal utility* that means the marginal utility decreases as the supply increases and the decreasing rate is inversely proportional to the amount of dataset [41]. Then, the expected dataset quality Q is defined as follows.

Definition 4 (Expected dataset quality of the data service provider)**.**

$$Q(\mathbf{q}) = \log(1 + \sum_{i \in R} \sum_{j \in R, j \neq i} \eta_{ij} \sqrt{q_i q_j}), \tag{12}$$

where η_{ij} is the correlation between the dataset q_i and q_j.

To define the dataset quality function, the geometric mean (root terms) is the most popular and manageable function satisfying the first and second principles. That is, the dataset quality (which the data service provider has) increases when i) datasets are tightly correlated, ii) the amount of dataset is large, and iii) the amount of dataset from each data broker is balanced. Note that when $\eta_{ij} = 1$ $(i = j)$ and $\eta_{ij} = 0$ $(i \neq j)$, it forms $\sum_{i \in R} q_i$ which is the same as the linear sum of the dataset quality from each data broker. Moreover, the natural logarithm function is used to apply the third principle, which is usually proposed for modeling the law of diminishing marginal utility (the function decreases inversely proportional to the amount of dataset, namely, $du = x^{-1}dx$ [42]).

With the expected dataset quality Q, this paper proposes a service quality S for the data service provider as the accuracy of the prediction by exploiting the gathered dataset. This paper adopts some models from the field of machine learning that is now widely used for improving the quality of services and applications [43]. Many studies showed that the accuracy of the machine learning analysis increases when the number of training samples increases [21], [44], [45]. The study showed that an error rate curve of the machine learning technique with respect to training size follows power-law distribution (i.e., decreasing curve with long-tail) [45]. Based on the literature survey, this paper defines the service quality function S as follows.

Definition 5 (Expected service quality by dataset exploitation)**.**

$$S(Q(\mathbf{q})) = 1 - \frac{\beta}{1 + \alpha Q(\mathbf{q})}, \tag{13}$$

$\alpha \in (0, 1]$ and $\beta \in (0, 1]$ are parameters. Then, $S(Q) \in [0, 1)$.

Note that the adopted service quality model is the same as the one proposed in [21], which mathematically modeled several accuracy curves of the machine learning results by analyzing real-world dataset.

Figure 3 shows an illustration of the proposed service quality function S for variant parameters α and β. The α is used to set the initial slope of the function; that is, the higher α value makes the sharper S. On the other hand, the β means the achievable minimum accuracy rate without any data analysis when $Q(\mathbf{q}) = 0$. In this case, the lower β means the higher achievable minimum accuracy.

B. Willingness-to-pay of the service consumers

Based on the service quality model, the data service provider should anticipate the number of actual paid service consumers to maximize its revenue and to decide the amount of budget to buy a dataset from the market. Therefore, the trend of the service consumers' willingness-to-pay should be proposed.

Each service consumer has his/her own criteria to pay the offered price for the service (i.e., the service consumer decides whether to pay the offered price for the service or not). Therefore, it is hard to separately formulate each individual's behavior for the data service provider's perspective. In other words, the WTP should be modeled as a macro level by considering the cumulative distribution as the portion of the paid consumers from the entire service consumer group.

To formulate the WTP function, this paper considers the basic economic principles for demand: i) WTP decreases when the offered price increases, ii) the service consumers prefer to pay the service with higher quality, which is shown by many studies related to the relationship with the service quality (or user satisfaction) [46], [47]. With the analysis of the previous studies about WTP [48], [49] and the proposed principles, the WTP of the service consumers is defined as Definition 6.

Figure 4 shows the trends of the proposed WTP model with various service quality factors (S). Note that the curves of WTP from the real-world experiments [48], [49] and the proposed function are similar. Moreover, the proposed WTP function is well-fitted to reflect the actual results in the references. It basically decreases when the price of dataset increases, and it also rapidly decreases with lower quality and slowly decreases with higher quality, respectively.

Definition 6 (Willingness-to-pay of the service consumers)**.** The WTP function Ψ of the service consumers is defined by the cumulative distribution function (CDF) as the portion of the paid service consumers in the potential service consumer group that decides whether or not to pay based on the offered price p_s and the service quality S, and given by,

$$\Psi(p_s, S) = e^{-p_s(1-S)}, \tag{14}$$

where $p_s \geqslant 0$ and $0 \leqslant S < 1$.

Based on the proposed models for the service quality S and the WTP function Ψ, the profit function P of the data service provider can be modeled as follows:

$$\begin{aligned} P(p_s, B) &= p_s M \Psi(p_s, S(Q(\mathbf{q}(B)))) - B \\ &= p_s M e^{-p_s(1 - S(Q(\mathbf{q}(B))))} - B, \end{aligned}$$

where M is the number of service consumers, p_s is the offered price, and B is the budget to buy dataset $\mathbf{q}$ from all data brokers in the market.

Note that $M\Psi(p_s, S(Q(\mathbf{q}(B))))$ means the expected number of service consumers with the offered price p_s and the offered service quality Q. Therefore, $p_s M e^{-p_s(1-S(Q(\mathbf{q}(B))))}$ means the expected revenue from service consumers. On the other hand, the budget B is the expected cost for buying dataset. Then, the profit maximization problem can be formulated as follows:

Problem 1 (Profit maximization of the data service provider)**.**

$$\begin{aligned} \max_{p_s, B} \quad & P(p_s, B) \\ \text{s.t.} \quad & B > 0, \quad \mathbf{q}^*(B) > 0, \quad R > 0, \quad p_s > 0. \end{aligned}$$

where

$$q_i^*(B) = \frac{B(R-1)(\sum_{j=1}^{R} \delta_j - \delta_i R + \delta_i)}{(\sum_{j=1}^{R} \delta_j)^2}.$$

Note that $\mathbf{q}^*$ with the fixed budget B (equation (9)) that maximizes the service quality Q can be decided by the NE of the CGDB from Theorem 1.

C. Profit maximization for the data service provider

In this section, the optimal strategy of the data service provider that maximizes its profit $P(p_s, B)$ is solved (Problem 1). To this end, p_s^* is firstly obtained from the following theorem with the fixed budget B (Theorem 2), and the optimal budget B^* of the data service provider is finally obtained from the sequel theorem (Theorem 3).

Theorem 2. The optimal price is p_s^* that maximizes the profit function of the data service provider is given by,

$$p_s^* = \frac{1}{1 - S(B)}$$

where

$$S(B) = 1 - \frac{\beta}{1 + \alpha Q(\mathbf{q}^*(B))}, \text{ and}$$

$$q_i^*(B) = \frac{B(R-1)(\sum_{j=1}^{R} \delta_j - \delta_i R + \delta_i)}{(\sum_{j=1}^{R} \delta_j)^2}.$$

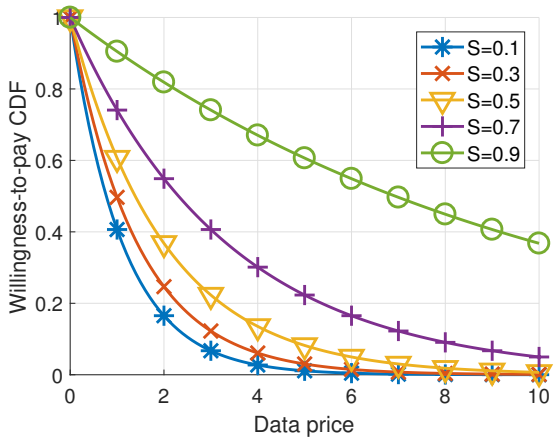


Fig. 4. The trend of willingness-to-pay (Ψ) w.r.t. the service quality

Proof. First, the concaveness of the P with respect to p_s is checked by considering whether the second order derivative of P is less than zero or not. Note that $S \in [0, 1)$.

$$\begin{aligned} \frac{\partial}{\partial p_s} P(p_s, B) &= M e^{p_s(S-1)} + M p_s e^{p_s(S-1)}(S-1) \\ \frac{\partial}{\partial^2 p_s} P(p_s, B) &= 2M e^{p_s(S-1)}(S-1) + M p_s e^{p_s(S-1)}(S-1)^2 \\ &\Rightarrow - M e^{p_s(S-1)}(1-S)(1+S) < 0. \end{aligned}$$

Since the profit function P is a concave function, the maximum point obtained by checking the first order derivative with p_s becomes zero.

$$\begin{aligned} \frac{\partial P(p_s, B)}{\partial p_s} &= M e^{p_s(S-1)}(1 + p_s(S-1)) = 0 \\ &\Rightarrow 1 - p_s(1-S) = 0 \\ &\therefore \; p_s^* = \frac{1}{1-S}. \end{aligned}$$

□

From this theorem, the optimal solution of Problem 1, which is the optimal strategy of the data service provider, is finally obtained as the following theorem.

Theorem 3. The optimal strategy (p_s^*, B^*) (for Problem 1) of the data service provider that maximizes the profit based on Theorem 2, is given by,

$$B^* = \frac{\alpha}{\beta} M e^{-1} - \frac{X^2}{Y(R-1)}$$

where

$$\begin{aligned} X &= \sum_{l=1}^{R} \delta_l, \\ Y &= \sum_{i \in R} \sum_{j \in R} \eta_{ij} \sqrt{(X - \delta_i R + \delta_i)(X - \delta_j R + \delta_j)}, \end{aligned}$$

and p_s^* is given by Theorem 2.

Proof. For the fixed budget B, the optimal price p_s^* of the data service provider is a function of B, obtained from Theorem 2. Now, denote the profit function $P(p_s, B)$ under the condition $p_s = p_s^*$ as $P_{s*}(B)$, that is, $P_{s*}(B) = P(p_s, B)|_{p_s = p_s^*}$.

$$\begin{aligned} & P_{s*}(B) \\ &= M e^{-1} \frac{(1 + \alpha Q(B))}{\beta} - B \\ &= \frac{M e^{-1}}{\beta} + \frac{\alpha}{\beta} M e^{-1} \log \left[1 + \sum_{i \in R} \sum_{j \in R} \eta_{ij} \sqrt{q_i q_j} \right] - B \\ &= - B + \frac{M e^{-1}}{\beta} + \frac{\alpha}{\beta} M e^{-1} \log \Bigg[1 + B(R-1) \\ &\quad \times \sum_{i \in R} \sum_{j \in R} \eta_{ij} \sqrt{\frac{(X - \delta_i R + \delta_i)}{X^2}} \sqrt{\frac{(X - \delta_j R + \delta_j)}{X^2}} \Bigg] \end{aligned}$$

$$\therefore P_{s*}(B) = \frac{M e^{-1}}{\beta} + \frac{\alpha}{\beta} M e^{-1} \log \left(1 + \frac{Y(R-1)}{X^2} B \right) - B,$$

Similarly, the concaveness of the modified profit function $P_{s*}(B)$ is checked by taking the second order derivative. If the function $P_{s*}(B)$ is concave, then it has the unique maximum point B^* [38].

$$\frac{\partial P_{s*}(B)}{\partial B} = \frac{\alpha}{\beta} M e^{-1} \frac{Y(R-1)}{X^2 + Y(R-1)B} - 1,$$
$$\frac{\partial^2 P_{s*}(B)}{\partial B^2} = \frac{\alpha}{\beta} M e^{-1} \frac{-(R-1)^2 Y^2}{(X^2 + Y(R-1)B)^2} < 0.$$

Since the modified profit function $P_{s*}(B)$ is concave, there exists the global maximum point B^* that maximizes the entire profit $P_{s*}(B^*)$ as follows.

$$\frac{\partial P_{s*}(B)}{\partial B} = \frac{\alpha}{\beta} M e^{-1} \frac{Y(R-1)}{X^2 + Y(R-1)B} - 1 = 0$$
$$\therefore B^* = \frac{\alpha}{\beta} M e^{-1} - \frac{X^2}{Y(R-1)}.$$

□

The required budget B^* to maximize the profit of the data service provider is obtained as a closed-form solution; therefore, it can be easily applicable in the dynamic market. Moreover, the main part of the result $\frac{\alpha}{\beta} M e^{-1}$ is only depended upon its own characteristics (i.e., service quality and the number of service consumers); that is, the data service providers can estimate their required budget without any information about other data brokers in the market.

Based on the analysis about the data service provider, an algorithm (Algorithm 1) is proposed for obtaining key results (i.e., the required dataset quality (q_i^*) and the allocated budget (b_i^*) for each data broker). The algorithm takes basic input parameters: the number of services consumers (M), the parameters for the service quality function (α, β, and η), the number data brokers, and the unit price of each data broker $(\delta)_{i=1}^R$. With the input parameters, the algorithm initializes common variables X and Y (Theorem 3), and then it calculates the total budget (B^*) that maximizes the profit of the data service provider. Based on the total budget (B^*) and the unit price of each data broker ($(\delta)_{i=1}^R$), it calculates the required dataset quality q_i^* (Theorem 1) and the allocated budget b_i^* (Equation (1)) for each data broker ($i \in R$). Finally, the data service provider can get the required dataset quality $(q_i^*)_{i=1}^R$ and the allocated budget $(b_i^*)_{i=1}^R$ for all data brokers.

Algorithm 1 Algorithm for Data Service Provider

Input:
M: the number of service consumers
α, β, η: the parameters for the service quality function
R: the number of data brokers
$(\delta_i)_{i=1}^R$: the unit price of each data broker
Initialization:
$X = \sum_i^R \delta_i$
$Y = \sum_{i \in R} \sum_{j \in R} \eta_{ij} \sqrt{(X - \delta_i R + \delta_i)(X - \delta_j R + \delta_j)}$
Start algorithm:
$B^* = \frac{\alpha}{\beta} M e^{-1} - \frac{X^2}{Y(R-1)}$ (Theorem 3)
Loop i to R:
$q_i^* = \frac{B^*(R-1)(X - \delta_i R + \delta_i)}{X^2}$ (Theorem 1)
$b_i^* = \frac{X - \delta_i R + \delta_i}{X} B^*$ (Equation (1))
Output:
$(q_i^*)_{i=1}^R$: the required dataset quality for each data broker
$(b_i^*)_{i=1}^R$: the allocated budget for each data broker

VI. Profit Optimization for the Data Brokers

This section describes an optimization problem for the data brokers by defining an actual profit function ($\bar{U}_i$) that considers a quality of dataset (q_i) and a WTS function of the data providers, who take their privacy into account as privacy sensitivity of data types, for each data broker (d_i). The income can be obtained by the price from the data service provider (i.e., b_i, the result of NE in Section IV), and the outcome can be obtained by costs ($\mathbf{c}_i$) for buying data from the data providers (N_i) with their WTS.

A. Willingness-to-sell data and the expected costs of the data broker

Data brokers should anticipate the number of data providers who actually participate in the market to minimize their cost. Therefore, the WTS of the data providers should be modeled.

First, this paper defines a WTS function Φ also similar to the authors' previous work in [19] based on the real-world experiment [18]. Since each data provider has their own WTS with different privacy concerns, it is hard to directly formulate the behavior of each data provider (i.e., with a certain offered price, the person will decide whether to sell his/her data or not). Therefore, similar to the WTP function, the WTS function should be defined as a macro level; that is, the cumulative distribution of WTS for the entire data provider group. To define the WTS function, two principles are applied. First, the more money is offered, the more people participate. The second is the privacy sensitivity of data. People hesitate to share or sell privacy sensitive data (e.g., many people think that information about credit card usage is more privacy sensitive than that of location). Moreover, people may have different WTS depending on the characteristics (e.g., popularity, reputation, etc.) of the data brokers even though they handle the same data type. The survey showed that the WTS to the trustworthy stakeholder is five times higher than that to the untrustworthy one [50]. By fitting an appropriate function with these principles and the real-world experiment in [18], the WTS function is defined as Definition 7.

Definition 7 (Willingness-to-sell data)**.** Willingness-to-sell of a certain type of data ($k \in K$) from data providers ($N_{i \in R}$) in a data broker ($d_{i \in R}$) is defined as follows:

$$\Phi_{i,k}(c_{i,k}) = 1 - e^{-\rho_{i,k} c_{i,k}} \tag{15}$$

where $\rho_{i,k}$ is the privacy sensitivity parameter and $c_{i,k}$ is the offered price for the kth type of data of ith data provider group, respectively.

Note that the WTS function is affected not only the characteristics of data (k) but also the characteristics of the data broker (d_i).

Figure 5 shows the WTS function for variant parameter ρ. Note that the data type with a smaller ρ is the more privacy

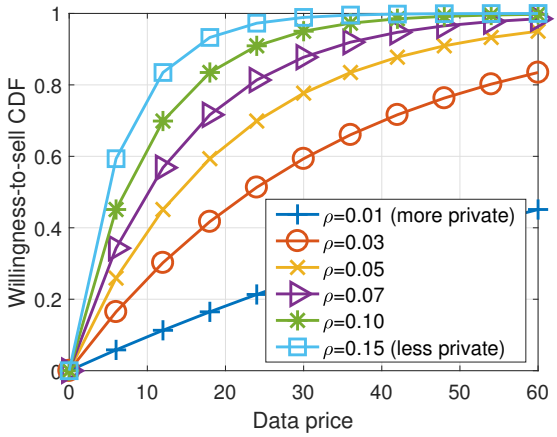


Fig. 5. Willingness-to-sell as a function of data price

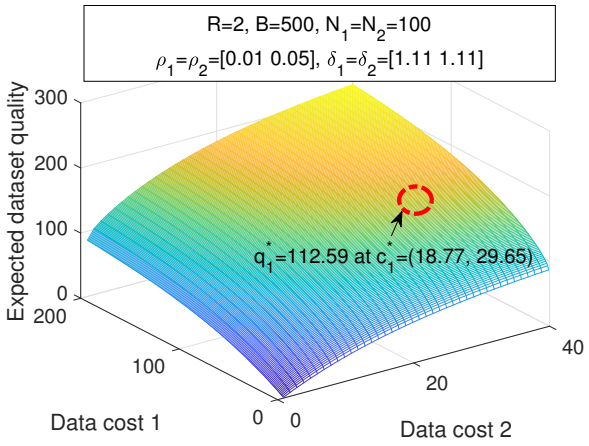


Fig. 6. An example of q_1 w.r.t $\forall \mathbf{c}_1$ with $R = 2$ and $K = 2$ for Problem 2.

sensitive than that with a larger one. The WTS of less sensitive data types increases more rapidly. This result is well-fitted to the real-world experiment performed in [18].

With the WTS function, $n_{i,k}$, the amount of the kth data type collected by the data broker d_i from ith data provider group (N_i) with a offered cost $c_{i,k}$ can be defined as follows:

$$n_{i,k} = N_i \Phi_{i,k}(c_{i,k}) \tag{16}$$

Note that the unit price for kth data type is $c_{i,k}/n_{i,k}$. Then, the total expected cost for buying all data in the ith data broker E_i follows:

$$E_i(\mathbf{c}_i) = \sum_{k \in K} c_{i,k}, \tag{17}$$

where $\mathbf{c}_i = (c_{i,k})_{k=1}^{K}$.

B. Dataset quality function of the data broker

From the data brokers' perspective, the method to measure and quantify the value of the gathered dataset is needed to decide whether they have good enough datasets for selling in the market. Therefore, the concept of quality measure for the dataset is from the authors' previous work [19] by considering both inherent data quality and the characteristics of data similar to Definition 4. Then, a quality of dataset provided by the data broker d_i is defined as follows.

Definition 8 (Quality of dataset provided by the data broker d_i)**.** The quality of dataset provided by the data broker d_i with N_i data provider can be collected by,

$$q_i = \sum_{x \in K} \sum_{y \in K, y \neq x} r_{xy} \sqrt{n_{i,x} n_{i,y}}, \tag{18}$$

where $n_{i,x}$ and $n_{i,y}$ are the expected amount of the xth and yth data type which can be collected by the data broker d_i, respectively. r_{xy} is the correlation between the data types.

Note that the geometric mean is used for quantifying quality of dataset to consider not only characteristics of big data (e.g., accuracy, completeness, etc.) but also characteristics of privacy data (e.g., identifiability, etc.), which is similar to Definition 4. As a result, the proposed model can measure not only the amount of data themselves but also their synergy effect with correlation. If $r_{xy} = 1$ $(x = y)$ and $r_{xy} = 0$ $(x \neq y)$, then it forms $\sum_{x \in K} 1 - e^{-c_x \rho_x}$ which is the same as data quality functions proposed in previous works [20], [21], [24].

C. Profit optimization for the data broker

Since the expected revenue is directly paid by the data service provider as b_i by equation (1) for the bidding dataset with quality q_i from the CGDB by equation (9) in Section IV, and the expected cost is E_i by equation (17), the actual profit function $\bar{U}_i$ of the data broker d_i can be represented as

$$\bar{U}_i(b_i, \mathbf{c}_i) = b_i - E_i(\mathbf{c}_i). \tag{19}$$

Since b_i^* (with q_i^*), the maximum revenue, is decided by the CGDB between the data broker d_i and the data service provider, the profit function of the data broker can be reduced as follows.

$$\bar{U}_i^*(\mathbf{c}_i) = b_i^* - \sum_{k=1}^{K} c_{i,k}.$$

Then, the profit maximization problem can be transformed to the cost minimization of the data broker d_i. Therefore, the cost minimization problem is solved. From the definition of the cost function E_i of the data broker d_i (equation (17)) the optimization problem can be defined as follows:

Problem 2 (Cost minimization of the data broker)**.**

$$\min_{\mathbf{c}_i^*} E_i(\mathbf{c}_i), \tag{20}$$

$$\text{s.t.} \quad 0 < c_{i,k} \leqslant \frac{2}{\rho_{i,k}} \quad \text{for all } k \in K, \tag{21}$$

$$q_i^* = N_i \sum_{x \in K} \sum_{y \in K} r_{xy} \sqrt{(1 - e^{-c_{i,x}\rho_{i,x}})(1 - e^{-c_{i,y}\rho_{i,y}})}, \tag{22}$$

$$N_i > 0, \quad \rho_{i,k} \in (0, 1), \quad \forall k \in K,$$
$$R \geqslant 2, \quad B > 0, \tag{23}$$

where $(c_{i,k}^*)_{k=1}^{K} \in \mathbf{c}_i^*$ is the data cost vector that minimizes the cost function $E_i(\mathbf{c}_i^*)$.

Note that q_i^* is obtained from equation (9) in Section IV. Since the optimization problem is bounded within the condition (21) (the detailed reason is described in Assumption 1), it is possible to apply any constrained nonlinear optimization algorithms (e.g., Sequential Quadratic Programming (SQP)) that satisfy the conditions (21)-(23) [51]. Therefore, this paper adopts the Sequential Least SQuares Programming (SLSQP) method, which is widely used in SciPy (a de facto standard for scientific Python) [52], originated from [53]. Figure 6 shows an illustration of a problem set for the cost minimization problem (Problem 2). With sample parameter settings listed in Figure 6, it is verified that the result of the SLSQP solver is the same as that of the exhaustive search[1].

The SLSQP solver is one of trust-region SQP methods [54], and it is hard to directly obtain the theoretical time complexity of the SLSQP because it depends on detailed implementations for finding internal parameters. Therefore, many benchmarking experiments for various optimization algorithms have been performed [55], [56]. Particularly, *Varelas and Dahito* showed performance benchmarking results of various multivariate solvers in SciPy, and it showed that the SLSQP solver has the best performance in terms of the average runtime and the statistical significance.

Since the WTS function is an asymptotic function, it is needed to define a tangible problem space for the proposed cost minimization problem. Based on the various surveys related to WTS (which are [14], [15], [27], [57], and [58]), the condition (21) is obtained by the following assumption.

Assumption 1 (The upper bound of the unit cost for the kth data type)**.** In order to consider a tangible problem space for the cost minimization, the upper bound of the cost for each data type is $2/\rho$ that makes the WTS value about 86% (i.e., $\Phi_i(2/\rho_{i,k}) \approx 0.86$ for each $i \in R$ and $k \in K$). In other words, about 14% of the data provider candidates refuses any privacy valuation methods.

The report [57] surveyed the relationship between individuals' preferences and their behavior for privacy, and it classified the individuals into three categories based on their privacy concerns, which are widely used in many areas related to privacy:

- Privacy Fundamentalists (PF): the group simply refuses any offer regarding privacy valuation (25% of the public);
- Privacy Pragmatists: the group weighs the value between privacy valuation and privacy protection (57% of the public);
- Privacy Unconcerned: the group less concerns about privacy violation or abuse (18% of the public).

Assumption 1 is obtained by surveys related to the portion of the PF groups in various areas. *Woodruff et al.* investigated WTS privacy of data providers using the categories with Google Consumer Surveys [58], and it showed that 35% of respondents are fit into the PF category. *Ponemon Institute* investigated a privacy profile similar to the categories above, and it showed that 26% of the respondents are categorized as the PF group [14]. Similarly, *Jai and King* conducted a survey for finding the relationship between the WTS data of individuals and the offered reward by a loyalty program in the data brokering market for online advertisement [27]. It showed that 15% of respondents are classified as the PF group. *Growth from Knowledge* also surveyed the WTS data of individuals in exchange for benefits or rewards, and it showed that 17% of the respondents strongly disagree (i.e., the PF group) [15].

Assumption 1 gives in some intuitions to find the expected unit price δ_i for obtaining dataset with quality q_i of the data broker d_i for the bidding market in Section IV. Since each data broker chooses a bid without knowledge of others' activities, each data broker needs to assume all possible scenarios (e.g., the data service provider buys the entire dataset that can be provided by the data broker); that is, it should consider the maximum capability of each data broker. Therefore, based on Assumption 1, the following is obtained.

Assumption 2. The unit price δ_i for dataset with quality q_i of the data broker d_i can be obtained by,

$$\delta_i = \frac{\max \sum_k^K c_{i,k}}{\max q_i} = \frac{\sum_k \frac{2}{\rho_{i,k}}}{N_i \sum_{x \in K} \sum_{y \in K} r_{xy}(1 - e^{-2})} \quad (24)$$

Assumption 2 gives the unified unit price for each data broker even if it has a different number of data providers with different data types. This factor can be used for measuring the relative competitiveness of each data broker in the market, and a data service provider can predict the capability of data brokers.

Note that, in this paper, the unit price for dataset δ is decided with Assumption 1. For practical applications, each data broker can decide the proper upper bound of the unit cost considering the portion of the PF group in its domain as already explained different statistics about PF groups above.

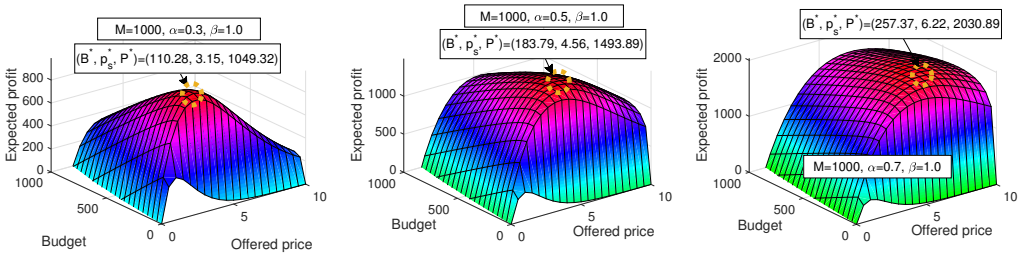
VII. Numerical Results

This section shows numerical results based on the analysis in the previous sections. First, this paper analyzes the optimal solutions for a simple case that both the number of the data brokers and the number of data types are two (i.e., $R = 2$ and $K = 2$) in Section VII-A. Based on the analytic results of the simple case, this paper shows the results for cases of multiple brokers with multiple data types (i.e., $R > 2$ and $K > 2$) with factors/parameters from the real-world dataset in Section VII-B.

In order to perform numerical analysis, this paper configures parameters as follows:

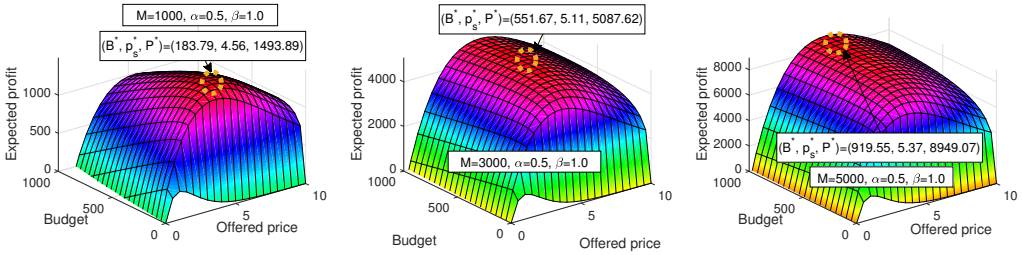
- Since the unit price δ_i is decided by the number of data providers (N_i) and the privacy sensitivity ρ_i for each data broker, similar to M, the reasonable values are chosen for N_i. Note that if the gaps between $(\delta_i)_{i=1}^R$ are too large, it is not able to satisfy the condition (10) for the NE point, that is, $\sum_{j=1}^R \delta_j - \delta_i R + \delta_i > 0$. In other words, it is not possible to put the brokers with a large difference of δ into the competition in the proposed CGDB model because the gap of their unit prices is too large.

[1]An example of SLSQP application is available in https://github.com/Hyeontaek-Oh/IEEE-IoTJ-2020/blob/master/slsqp-broker-opt-example.py



(a) $\alpha = 0.3$ (b) $\alpha = 0.5$ (c) $\alpha = 0.7$

Fig. 7. The expected profits of the data service provider w.r.t. various service quality parameter (α)



(a) $M = 1000$ (b) $M = 3000$ (c) $M = 5000$

Fig. 8. The expected profits of the data service provider w.r.t. the number of service consumers (M)

- For the correlation factors of data types r_{xy} for each data broker and the correlation factors of datasets η_{ij} are set as follows:

$$r_{xy} = \begin{cases} 1 & \text{if } x = y \\ 0.5 & \text{if } x \neq y \end{cases} \quad \text{and} \quad \eta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0.5 & \text{if } i \neq j. \end{cases} \tag{25}$$

- For parameter β for the service quality function (S), it is set as $\beta = 1$ that means the data service provider has no knowledge about their customers without dataset from the brokers.

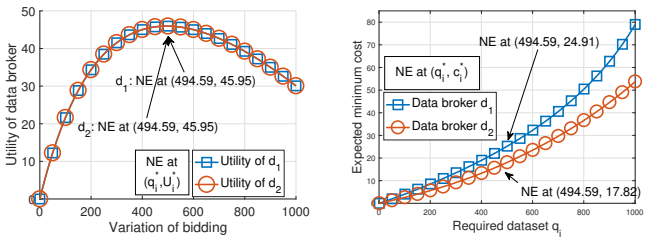
A. Theoretical experiment

This section checks various parameters for the proposed model with the simple case ($K = 2$ and $R = 2$). In this experiment, it is assumed that each data broker (d_1 and d_2) has the same unit price for obtaining dataset (i.e., $\delta_1 = \delta_2$) with different number of data providers and different data types (i.e., $N_1 \neq N_2$ and $\rho_1 \neq \rho_2$). One data broker (d_1) handles two data types with similar privacy sensitivity, and the other data broker (d_2) handles them with different privacy sensitivity. The detailed parameters are explained in Table II.

Before analyzing the behavior of the data brokers, this section verifies the data service provider's side related to WTP of the service consumers by checking various parameters like α for the service quality S and the number of service consumers M that are directly related to the optimal budget allocation B^* for the data service provider.

TABLE II
PARAMETERS FOR THEORETICAL EXPERIMENT

Parameters for service quality (S)	$\beta = 1.0$
# of data brokers	$R = 2$
# of data providers	$N_1 = 506$, $N_2 = 664$
# of data types	$K = 2$
Privacy sensitivity factor	$\rho_1 = (0.035, 0.045)$, $\rho_2 = (0.020, 0.060)$
Unit price for dataset	$\delta_1 = 0.0929$, $\delta_2 = 0.0929$



(a) The change of profits with variation of bidding (b) The change of costs with variation of bidding

Fig. 9. The results of theoretical analysis for the data brokers with parameters in Table II

First, the parameter α related to the service quality S of the data service provider is checked. Note that the higher α means that the service quality function S increases more faster; in other words, the data service provider is able to perform a much better service with the same amount of dataset. Figure 7 shows the expected profits of the data service provider P for the two variables, an offered service price p_s and budget for buying dataset B, with the same number of service consumers $M = 1000$ are follows:

$$\begin{cases} \alpha = 0.3 : P^* = 1049.32 \text{ at } (p_s^*, B^*) = (3.15, 110.28), \\ \alpha = 0.5 : P^* = 1493.89 \text{ at } (p_s^*, B^*) = (4.56, 183.79), \\ \alpha = 0.7 : P^* = 2030.89 \text{ at } (p_s^*, B^*) = (6.22, 257.37). \end{cases}$$

Since the parameter α directly affects the service quality

that is also directly related to WTP of the service consumers, the increment of α makes the offered price for the service p_s higher; that is, the data service provider earns higher profits, in other words, the number of paid service consumers increases (i.e., WTP of the service consumers increases).

Similarly, Figure 8 shows the trends of the data service provider's expected profit with respect to various number of service consumers M with the fixed parameter $\alpha = 0.5$ as follows:

$$\begin{cases} M = 1000 : P^* = 1493.89 \text{ at } (p_s^*, B^*) = (4.56, 183.79), \\ M = 3000 : P^* = 5087.62 \text{ at } (p_s^*, B^*) = (5.11, 551.67), \\ M = 5000 : P^* = 8949.07 \text{ at } (p_s^*, B^*) = (5.37, 919.55). \end{cases}$$

Similar to α, the higher M makes the higher expected profit P^*. However, the impact of M is different from that of α. The α affects the offered service price p_s. In contrast to α, the M affects the available budget B; in other words, the total number of paid service consumers increases.

The impacts of α and M have been analyzed from the data service provider's perspective so far. From now on, the detailed results for the data brokers are analyzed. Figure 9 shows the results regarding the data brokers with parameters in Table II.

With the budget B^* (from Figure 8(a)), the NE analysis of the data brokers is performed in Figure 9(a). Note that both data brokers have the same NE point (i.e., the dataset bidding $q^* = 494.59$ and the expected profit $U^* = 45.95$) that maximizes their profits because they have the same unit price δ. However, the actual profits of each data broker are different because they have different data providers and handle data types with different privacy sensitivity. Figure 9(b) shows the actual expected costs for buying data to satisfy the required dataset quality q_i. At the NE point, the data broker d_1 spends more costs than d_2; in other words, the actual profit of d_2 is larger than that of d_1. It shows that the actual profit of data brokers depends on the behavior of their data providers groups.

B. Experiments with the real-world dataset

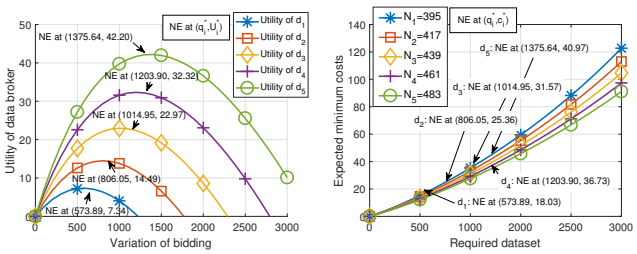
The previous section has mainly analyzed the impacts of parameters related to the data service provider. In this section, the detailed analysis of data brokers with real-world datasets is performed.

In order to perform the analysis, this paper configures parameters as follows (also summarized in Table IV):

- Since the number of service consumers M affects only the size of the budget B, there are relatively few restrictions on the parameter selection. Based on the detailed analysis of the impacts of α and M performed in the previous section (Section VII-A), this paper chooses the reasonable number as $M = 3000$ to valid the entire trading market. Similarly, parameters α and β for the service quality function (S) also affect the size of the budget B; therefore, they are set as $\alpha = 0.5$ and $\beta = 1.0$, which means the data service provider has no knowledge about their customers without dataset from the brokers.
- For the number of data brokers, this paper chooses $R = 3$ for verifying the proposed CGDB model based on the real-world surveys performed in [14], [59], [60]. These surveys contain various results from respondents that are related to the average price for each data type in the data providers' perspective. Based on the data in the surveys, this paper chooses the number of data providers of each data broker as $N_i = (282, 439, 1078)$ that are the number of respondents in the surveys (which are [59], [60], and [14], respectively).
- For privacy sensitivity factor ρ, this paper also adopts the surveys in the previous bullet item related to the average price for each data type in the data providers' perspective. It sets each ρ_k that makes the proposed WTS value 0.5 (50%) with the average price from the survey.
  This paper chooses six data types ($K = 6$) (i.e., payment details, purchase histories, hobbies & preferences, photos & videos, physical location: GPS, browsing histories) that are commonly available from smartphones. Note that each data type is available in the real-world dataset [61], [62]. The unit price δ_i is decided by the number of data providers (N_i) and the privacy sensitivity ρ_i for each data broker.
  For example, for Case 1, the average price value **c**=(payments details($20.8), purchase histories($20.7), preferences ($17.8), photos ($5.9), physical location: GPS ($5.9), browsing histories($5.1)) can be mapped into privacy sensitivity factor ρ=(0.0333, 0.0335, 0.0389, 0.1175, 0.1175, 0.1359), respectively.
- For the correlation factors of data types r_{xy} for each data broker and the correlation factors of datasets η_{ij} are set as equation (25).

Before directly checking the proposed CGDB with three different cases with real-world parameters, this section analyzes the impacts of the number of data providers when each data broker has the same privacy sensitivity factors. To analyze the trends and the behavior of data brokers, Case 2 parameters (in Table IV) are chosen for baseline. It is assumed that five data brokers (note that each data broker labeled as d_i) have different number of data providers (N_i) with the same privacy sensitivity factors (i.e., each data broker has $\rho = (0.0154, 0.0181, 0.0265, 0.0307, 0.0415, 0.0889)$. For setting the number of data providers (N), 439 (the number of respondents in Case 2) $\pm$ 5% and $\pm$ 10% values are taken, and the number of data providers N_i are set as $N_1 < N_2 < N_3 < N_4 < N_5$.

Table III shows the detailed analysis results. Since each data broker has a different number of data providers (N), available



(a) The change of utilities (b) The change of costs

Fig. 10. The result with a variation of bidding for each data broker (Case 2)

TABLE III
RESULTS OF THE CGDB WITH CASE 2 PARAMETERS IN TABLE IV

	N_i	δ_i	q_i^*	b_i^*	U_i^*(CGDB)	U_i^*(OPT)	$\mathbf{c}_i^*$
d_1	395	0.0981	573.89	63.66	7.34	45.63	(1.00, 1.24, 2.14, 2.65, 3.98, 7.01)
d_2	417	0.0929	806.05	89.41	14.49	64.04	(1.63, 2.02, 3.40, 4.11, 5.77, 8.43)
d_3	439	0.0883	1014.95	112.58	22.97	81.01	(2.25, 2.77, 4.53, 5.39, 7.19, 9.43)
d_4	461	0.0841	1203.90	133.54	32.33	96.81	(2.81, 3.45, 5.51, 6.46, 8.32, 10.17)
d_5	483	0.0802	1375.64	152.59	42.20	111.62	(3.30, 4.05, 6.34, 7.33, 9.20, 10.74)

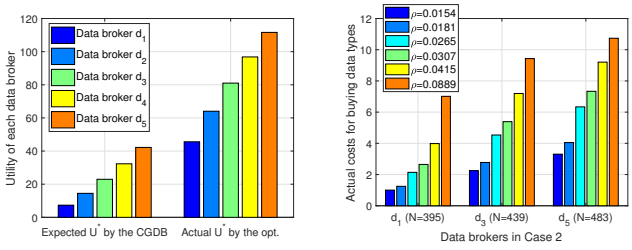
TABLE IV
PARAMETERS FOR EXPERIMENT WITH THE REAL-WORLD DATASET

M	3000		
S Param.	(α, β) = (0.5, 1.0)		
R	3		
	Broker Case 1	Broker Case 2	Broker Case 3
N	282	439	1078
K	6		
δ	0.066793	0.088289	0.023893
$\mathbf{c}$	(20.8, 20.7, 17.8, 5.9, 5.9, 5.1)	(45.1, 38.4, 26.2, 22.6, 16.7, 7.8)	(36.0, 20.6, 16.1, 12.2, 12.2, 7.1)
ρ	(0.0333, 0.0335, 0.0389, 0.1175, 0.1175, 0.1359)	(0.0154, 0.0181, 0.0265, 0.0307, 0.0415, 0.0889)	(0.0193, 0.0336, 0.0431, 0.0568, 0.0568, 0.0976)
δ	0.066793	0.088289	0.023893

unit prices δ_i, offered to the data service provider, are different. Note that a data broker with a lower δ is relatively more competitive in the market than that with a higher δ; that is, the data broker d_5 is more competitive than d_1 in this case. Based on the unit price δ, each data broker takes different required dataset quality q_i^* and budget b_i^*, which are directly related to the entire profit U_i^*. The more competitive data broker takes the more budget with a higher dataset quality requirement.

Moreover, similar to the previous section, the behavior of each data broker with variant bidding strategies, including the NE strategy, is analyzed in Figure 10(a). Each plot shows the profit trends of each data broker when all other data brokers already take bidding with NE strategies. It shows that the profit of each data broker is only maximized at the point of NE; in other words, other bidding strategies result in each data broker losing some profits. The more competitive data broker has a greater chance to get a higher profit with bidding because it has a greater chance to satisfy the required dataset quality at a lower cost. Figure 10(b) shows the expected minimum cost of each data broker to achieve the required dataset quality from the data service provider. It shows that the data broker with the more data providers spends fewer costs to the collected dataset to satisfy the data service provider because it has a higher chance of collecting more datasets with the same costs. It means that the data broker becomes more competitive in terms of dataset quality when the number of data providers increases.

Next, the difference between the expected profit by the CGDB and the actual profit by the proposed additional optimization is analyzed. Figure 11(a) shows the expected profit by the proposed CGDB (left) and the actual profit by the proposed cost minimization results (right) at the NE point for each data broker. The detailed results are marked as U^*(CGDB) and U^*(OPT) in Table III, respectively. It shows that the actual profit (U^*(OPT)) is higher than the expected profit (U^*(CGDB)) because each data broker participates in the market with the expected unit price δ defined in Assumption 2 to cover various cases in the market. However, when the bidding budget is set by the CGDB, each data broker can target its data providers to minimize costs for buying a dataset. Therefore, each data broker can achieve a higher profit than the expected profit in the CGDB.



(a) Expected and actual profits (b) Costs for collecting dataset

Fig. 11. The result at the NE point for each data broker (Case 2)

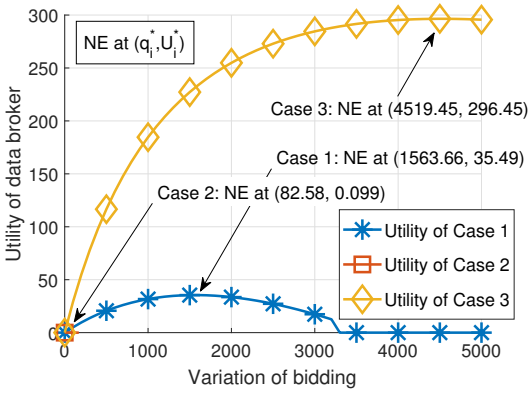


Fig. 12. The change of profits w.r.t. a variabtion of bidding for each data broker (Table IV)

Moreover, the costs $\mathbf{c}_i^*$ of each data broker are also analyzed in Figure 11(b). It shows that each data broker spends different costs even if it deals with the same data types with the same privacy sensitivity factor to maximize their utilities while satisfying the required dataset q_i^*. The data broker with a higher profit requires more dataset for the data service provider, i.e., dataset buyer, that means it spends more costs for collecting dataset.

Finally, with the configured parameters from the real-world dataset in Table IV, this paper analyses NE of each data broker with totally different characteristics (i.e., all three cases have different numbers of data providers and different values of privacy sensitivity factor). Figure 12 shows the detailed NE analysis for the proposed CGDB with three different data broker cases:

$$\begin{cases} \text{Broker Case 1} : U^* = 35.49 \text{ at } q_i^* = 1563.66, \\ \text{Broker Case 2} : U^* = 0.099 \text{ at } q_i^* = 82.58, \\ \text{Broker Case 3} : U^* = 296.45 \text{ at } q_i^* = 4519.45. \end{cases} \tag{26}$$

Since the difference of unit price values δ among three brokers are quite large (note that the δ values of data broker cases are 0.067, 0.088, 0.024, respectively), the behavior of data brokers is quite extreme. Particularly, the data broker of Case 3 is dominant in the market, but the data broker of Case 2 has no competitiveness in the market. Figure 12 shows the change of profits with a variation of bidding when other data brokers already take optimized NE strategies. The expected profit by the CGDB is maximized only at the point of NE for each data broker. Note that the profit of the data broker of Case 2 becomes zero after the bidding 3000 because it reaches the maximum achievable dataset quality; in other words, the Case 2 cannot bid more than its maximum achievable dataset quality. In addition, at the NE points, Figure 13 shows the expected profit of data brokers by the proposed CGDB model and the actual profit of them by the proposed cost minimization model. Similar to Figure 11(a), the actual profit by the cost minimization model is higher than the expected profit by the CGDB model.

In summary, the results show that data brokers should estimate their relative competitiveness in the data market before bidding datasets, and to increase their market competitiveness, they need to increase the number of data providers who actually sell their data to data brokers. Since many surveys identified the trust gap between data providers and other business stakeholders (e.g., data brokers and data service providers) [17], [50], [63], [64] (i.e., data providers hesitate to participate the data market due to lack of trust), it is needed to ensure trust of data providers by providing necessary actions (e.g., provide transparency of data flow, control of data usage, etc.) that ultimately increase their market participation.

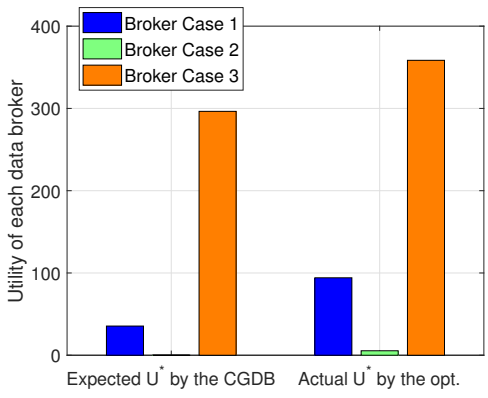


Fig. 13. The expected and actual profits of each data broker (Table IV)

VIII. Conclusion

With the widespread of IoT devices, data-driven services and applications become more popular. In keeping with these trends, much research has focused on IoT data markets and data trading issues from the data business stakeholders' perspective (i.e., data brokers, data consumers, etc.). Since the conflict issues are raised between privacy protection and innovation of IoT services/applications through data analytics, current IoT data markets are mainly categorized as privacy protection markets and privacy valuation markets, respectively. For the privacy valuation markets, data providers also have the need for proper benefits in exchange for their privacy; hence, they should also be considered as an important player in IoT data markets. Therefore, this paper has proposed a competitive data trading model with privacy valuation for multiple stakeholders in IoT data markets while considering not only the characteristics of data consumers but also those of data providers, who weigh the value between privacy protection and valuation and have the willingness to participate in the market with proper benefits. To model the market, this paper has considered four major stakeholders to cover various IoT data value chains (i.e., data providers (or data sources), data brokers, a data service provider (or data consumer), and service consumers). Particularly, this paper has proposed the CGDB model (a non-cooperative game model) between data brokers and a data service provider with the unified measure of the unit price of the dataset from data brokers for comparing the competitiveness of them with different data providers. This paper has also proposed the optimization models considering the relationship between a data broker and data providers (i.e., willingness-to-sell (WTS) data with privacy sensitivity factors) as well as the relationship between a data service provider and service consumers (i.e., willingness-to-pay (WTP) provided service with service quality). Based on the Nash Equilibrium and the optimization analysis of the proposed model, this paper has showed the feasibility of the proposed model with the parameters from real-world dataset while showing the existence of unique Nash Equilibrium point that maximizes benefits of business stakeholders while satisfying the requirements from all market participants (e.g., WTS, WTP, dataset quality, etc.). Since each proposed model (e.g., dataset quality, WTS, WTP, etc.) is one of the possible models to analyze the behavioral characteristics based on the observation of real-world experiments, as future work, various WTS and WTP mathematical models can be considered to design the behavior of data providers and services consumers more realistically, along with various cost models for data management (e.g., computing, storage, etc.).

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their insightful comments and suggestions.

REFERENCES

[1] Cisco, *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*, 2019.

[2] ——, *Cisco Global Cloud Index: Forecast and Methodology, 2016–2021 White Paper*, 2018.

[3] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, Feb 2018.

[4] C. Stergiou, K. E. Psannis, B. B. Gupta, and Y. Ishibashi, "Security, privacy & efficiency of sustainable cloud computing for big data & iot," *Sustainable Computing: Informatics and Systems*, vol. 19, pp. 174–184, 2018.

[5] S. P. Ahuja and N. Wheeler, "Architecture of Fog-Enabled and Cloud-Enhanced Internet of Things Applications," *International Journal of Cloud Applications and Computing*, vol. 10, no. 1, pp. 1–10, 2020.

[6] J. M. Cavanillas, E. Curry, and W. Wahlster, Eds., *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*. Springer International Publishing, 2016.

[7] A. Rieke, H. Yu, D. Robinson, and J. von Hoboken, *Data brokers in an open society*. Upturn and Open Society Foundations, 2016.

[8] E. Ramirez and et al., *Data brokers: A call for transparency and Accountability*. Federal Trade Commission, 2014.

[9] J. Goepfert and M. Shirer. (2018) Revenues for Big Data and Business Analytics Solutions Forecast to Reach $260 Billion in 2022, Led by the Banking and Manufacturing Industries, According to IDC. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=prUS44215218

[10] K. R. Sollins, "IoT Big Data Security and Privacy Versus Innovation," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1628–1635, April 2019.

[11] S.-A. Elvy, "Paying for privacy and the personal data economy," *Columbia Law Review*, vol. 117, no. 6, pp. 1369–1459, 2017.

[12] S. Cha, T. Hsu, Y. Xiang, and K. Yeh, "Privacy enhancing technologies in the internet of things: Perspectives and challenges," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2159–2187, April 2019.

[13] C. Li and B. Palanisamy, "Privacy in internet of things: From principles to technologies," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 488–505, Feb 2019.

[14] Ponemon Institute, *Privacy and Security in a Connected Life : A Study of US, European and Japanese Consumers*. Trend Micro and Ponemon Institute, 2015.

[15] Global GFK Survey. (2017) Willingness to share personal data in exchange for benefits or rewards. [Online]. Available: https://www.gfk.com/fileadmin/user_upload/country_one_pager/NL/images/Global-GfK_onderzoek_-_delen_van_persoonlijke_data.pdf

[16] EMC. (2014) EMC Privacy Index: Executive Summary. [Online]. Available: https://www.emc.com/collateral/brochure/privacy-index-executive-summary.pdf

[17] R. S. John Rose, Christine Barton and J. Platt, *The Trust Adventage: How to Win with Big Data*. Boston Consulting Group, 2013.

[18] V. Benndorf and H.-T. Normann, "The willingness to sell personal data," *The Scandinavian Journal of Economics*, vol. 120, no. 4, pp. 1260–1278, 2018.

[19] H. Oh, S. Park, G. M. Lee, H. Heo, and J. K. Choi, "Personal Data Trading Scheme for Data Brokers in IoT Data Marketplaces," *IEEE Access*, vol. 7, pp. 40 120–40 132, 2019.

[20] B. Jang, S. Park, J. Lee, and S. Hahn, "Three Hierarchical Levels of Big-Data Market Model Over Multiple Data Sources for Internet of Things," *IEEE Access*, vol. 6, pp. 31 269–31 280, 2018.

[21] D. Niyato, M. A. Alsheikh, P. Wang, D. I. Kim, and Z. Han, "Market model and optimal pricing scheme of big data and Internet of Things (IoT)," *2016 IEEE International Conference on Communications, ICC 2016*, pp. 1–6, 2016.

[22] F. Liang, W. Yu, D. An, Q. Yang, X. Fu, and W. Zhao, "A Survey on Big Data Market: Pricing, Trading and Protection," *IEEE Access*, vol. 6, pp. 15 132–15 154, 2018.

[23] X. Ren, P. London, J. Ziani, and A. Wierman, "Datum: Managing data purchasing and data placement in a geo-distributed data market," *IEEE/ACM Transactions on Networking*, vol. 26, no. 2, pp. 893–905, April 2018.

[24] Y. Jiao, P. Wang, S. Feng, and D. Niyato, "Profit maximization mechanism and data management for data analytics services," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 2001–2014, June 2018.

[25] B. Shen, Y. Shen, and W. Ji, "Profit optimization in service-oriented data market: A Stackelberg game approach," *Future Generation Computer Systems*, vol. 95, pp. 17–25, 2019.

[26] G. Malgieri and B. Custers, "Pricing privacy – the right to know the value of your personal data," *Computer Law & Security Review*, vol. 34, no. 2, pp. 289–303, 2018.

[27] T.-M. C. Jai and N. J. King, "Privacy versus reward: Do loyalty programs increase consumers' willingness to share personal information with third-party advertisers and data brokers?" *Journal of Retailing and Consumer Services*, vol. 28, pp. 296–303, 2016.

[28] D. Kim, K. Park, Y. Park, and J.-H. Ahn, "Willingness to provide personal information: Perspective of privacy calculus in iot services," *Computers in Human Behavior*, vol. 92, pp. 273 – 281, 2019.

[29] J. Parra-Arnau, "Optimized, direct sale of privacy in personal data marketplaces," *Information Sciences*, vol. 424, pp. 354 – 384, 2018.

[30] Z. Su, Q. Qi, Q. Xu, S. Guo, and X. Wang, "Incentive scheme for cyber physical social systems based on user behaviors," *IEEE Transactions on Emerging Topics in Computing*, pp. 1–11, 2017.

[31] L. Tian, J. Li, W. Li, B. Ramesh, and Z. Cai, "Optimal contract-based mechanisms for online data trading markets," *IEEE Internet of Things Journal*, vol. Early Access, pp. 1–2, March 2019.

[32] A. Ghosh and A. Roth, "Selling privacy at auction," *Games and Economic Behavior*, vol. 91, pp. 334–346, 2015.

[33] S. Park, J. Lee, G. Hwang, and J. K. Choi, "Event-driven energy trading system in microgrids: Aperiodic market model analysis with a game theoretic approach," *IEEE Access*, vol. 5, pp. 26 291–26 302, 2017.

[34] L. Olejnik, M. Tran, and C. Castelluccia, "Selling off user privacy at auction," in *21st Annual Network and Distributed System Security Symposium, NDSS 2014, San Diego, California, USA, February 23-26, 2014*, 2014.

[35] J. Wang, W. Zhang, and S. Yuan, "Display advertising with real-time bidding (rtb) and behavioural targeting," *Foundations and Trends® in Information Retrieval*, vol. 11, no. 4-5, pp. 297–435, 2017.

[36] B. E. Hajek and D. Gopal, "Do greedy autonomous systems make for a sensible internet," 2002.

[37] R. Johari and J. N. Tsitsiklis, "Efficiency loss in a network resource allocation game," *Mathematics of Operations Research*, vol. 29, no. 3, pp. 407–435, 2004.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[39] ISO 25012, "Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model," International Organization for Standardization, Geneva, CH, Standard, Dec 2008.

[40] R. Gomer, E. M. Rodrigues, N. Milic-Frayling, and M. C. Schraefel, "Network analysis of third party tracking: User exposure to tracking cookies through search," in *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 1, Nov 2013, pp. 549–556.

[41] O. Lange, "The Determinateness of the Utility Function," *Review of Economic Studies*, vol. 1, no. 3, pp. 218–225, 1934.

[42] Y. Lengwiler, *The Origins of Expected Utility Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 535–545.

[43] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 2923–2960, Fourthquarter 2018.

[44] P. Domingos, "A few useful things to know about machine learning," *Commun. ACM*, vol. 55, no. 10, pp. 78–87, Oct. 2012.

[45] M. Johnson, P. Anderson, M. Dras, and M. Steedman, "Predicting accuracy on large datasets from smaller pilot data," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 450–455.

[46] R. Casidy and W. Wymer, "A risk worth taking: Perceived risk as moderator of satisfaction, loyalty, and willingness-to-pay premium price," *Journal of Retailing and Consumer Services*, vol. 32, pp. 189–197, 2016.

[47] C.-H. Lien, Y. Cao, and X. Zhou, "Service quality, satisfaction, stickiness, and usage intentions: An exploratory evaluation in the context of WeChat services," *Computers in Human Behavior*, vol. 68, pp. 403–410, 2017.

[48] T. O. Kamoto and T. Hayashi, "Analysis of service provider's profit by modeling customer's willingness to pay for ip qos," in *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol. 2, Nov 2002, pp. 1549–1553.

[49] Y. Kim, S. Lim, and J. Choi, "Estimation of willingness to pay for smart home service by contingent valuation method," *Journal of the Korean Society for Quality Management*, vol. 44, pp. 833–843, Dec 2016.

[50] A. L. John Rose and E. Baltassis, *Bridging the Trust Gap in Personal Data*. Boston Consulting Group, 2018.

[51] L. D. BERKOVITZ, *Convexity and Optimization in Rn*. John Wiley & Sons, Inc., 2002.

[52] Scipy.org. Scipy reference guide – optimization and root finding (scipy.optimize.minimize). [Online]. Available: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html

[53] D. Kraft, "A software package for sequential quadratic programming," *Forschungsbericht. Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt, DFVLR*, vol. 88-28, 1988.

[54] J. Nocedal and S. J. Wright, *Numerical Optimization*, 1st ed. New York, NY, USA: Springer, 1999.

[55] R. E. Perez, P. W. Jansen, and J. R. R. A. Martins, "pyOpt: A Python-based object-oriented framework for nonlinear constrained optimization," *Structures and Multidisciplinary Optimization*, vol. 45, no. 1, pp. 101–118, 2012.

[56] K. Varelas and M.-A. Dahito, "Benchmarking Multivariate Solvers of SciPy on the Noiseless Testbed," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '19. New York, NY, USA: ACM, 2019, pp. 1946–1954.

[57] P. Kumaraguru and L. F. Cranor, "Privacy Indexes: A Survey of Westin's Studies," in *CMU-ISRI-5-138*. Institute for Software Research International, School of Computer Science, Carnegie Mellon University, Dec. 2005.

[58] A. Woodruff, V. Pihur, S. Consolvo, L. Brandimarte, and A. Acquisti, "Would a Privacy Fundamentalist Sell Their DNA for $1000...If Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences," in *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. Menlo Park, CA: USENIX Association, 2014, pp. 1–18.

[59] Ponemon Institute, *Privacy and Security in a Connected Life : A Study of European Consumers*. Trend Micro and Ponemon Institute, 2015.

[60] ——, *Privacy and Security in a Connected Life : A Study of US Consumers*. Trend Micro and Ponemon Institute, 2015.

[61] H. Sain, *data_sms*. Kaggle, 2017. [Online]. Available: https://www.kaggle.com/moose9200/data-sms

[62] N. Eagle and A. (Sandy) Pentland, "Reality mining: Sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, March 2006.

[63] W. E. Forum and A. Kearney, *Rethinking Personal Data: A New Lens for Strengthening Trust*. World Economic Forum, 2014. [Online]. Available: https://www.weforum.org/reports/rethinking-personal-data

[64] M. Little. (2014, January) Personal Data Futures: The Disrupted Ecosystem – How user-control of personal data could 'game-change' Big Data and reconfigure the Internet. [Online]. Available: http://www.crc.gov.mn/file/newfile/TMW21557%20M_Little_final.pdf

Hyeontaek Oh (S'14) is currently a PhD candidate in School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST). He received his BS degree in computer science and MS degree in electrical engineering from KAIST in 2012 and 2014, respectively. His research interests in trust in ICT environments, personal data ecosystem, Internet of Things (IoT), and Web technologies. He has actively participated in several nationally-funded research projects for ICT environment as a research assistant. He also has contributed the International Telecommunication Union Telecommunication Standardization Sector Study Group 13/20 as contributors and editors since 2015. He received the Outstanding Demo Award from IEEE 3rd Global Conference on Consumer Electronics in 2014 and Outstanding Paper Award from the 15th International Conference on Advanced Communication Technology in 2013.

Sangdon Park (S'16, M'17) is currently Brain Plus 21 Postdoctoral Researcher in the Information and Electronics Research Institute at Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea,. He received the B.S., M.S. and Ph.D. degrees in KAIST in 2011, 2013, and 2017, respectively. He has contributed several articles to the International Telecommunication Union Telecommunication (ITU-T). He received a Best Student Paper Award at the 11th International Conference on Queueing Theory and Network Applications in 2016. His current research interests lie in optimizing wireless networks or smart grids, which hold great potential for practical applications to industries and he has focused on processing energy big data via various machine-learning methodologies and optimizing network economics of the edge cloud computing.

Gyu Myoung Lee (S'02, M'07, SM'12) received his BS degree from Hong Ik University, Seoul, Korea, in 1999 and his MS and PhD degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2000 and 2007. He is with the Liverpool John Moores University (LJMU), UK, as Reader from 2014 and with KAIST Institute for IT convergence, Korea, as Adjunct Professor from 2012. Prior to joining the LJMU, he has worked with the Institut Mines-Telecom, Telecom SudParis, France, from 2008. Until 2012, he had been invited to work with the Electronics and Telecommunications Research Institute (ETRI), Korea. He also worked as a research professor in KAIST, Korea and as a guest researcher in National Institute of Standards and Technology (NIST), USA, in 2007. His research interests include Internet of things, future networks, multimedia services, and energy saving technologies including smart grids. He has been actively working for standardization in ITU-T, IETF and oneM2M, etc. In ITU-T, he served as the chair of FG-DPM and currently serves as a WP chair in SG13 as well as the Rapporteur of Q16/13 and Q4/20. He is a Senior Member of IEEE.

Jun Kyun Choi (M'88, SM'00) received the B.Sc. (Eng.) from Seoul National University in electronics engineering, Seoul, Korea in 1982, and M.Sc. (Eng.) and Ph.D. degree in 1985 and 1988, respectively, in electronics engineering from Korea Advanced Institute of Science and Technology (KAIST). From June 1986 until December 1997, he was with the Electronics and Telecommunication Research Institute (ETRI). In January 1998, he joined the Information and Communications University (ICU), Daejeon, Korea as Professor. At the year of 2009, he moved to Korea Advanced Institute of Science and Technology (KAIST) as Professor. He is a Senior Member of IEEE, the executive member of The Institute of Electronics Engineers of Korea (IEEK), Editor Board of Member of Korea Information Processing Society (KIPS), Life member of Korea Institute of Communication Science (KICS).

Sungkee Noh received the M.S. degree from Postech, Korea, in 1992, and the Ph.D. degree from Chungnam National University, Korea, in 2004. He is currently a Principal Researcher of Electronics and Telecommunications Research Institute, Korea. His current research interests include Internet of Things and Blockchain.