IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 23, 2019, accepted March 7, 2019, date of publication March 11, 2019, date of current version April 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2904248

Personal Data Trading Scheme for Data Brokers in IoT Data Marketplaces

HYEONTAEK OH[1], (Student Member, IEEE), SANGDON PARK[2], (Member, IEEE), GYU MYOUNG LEE[3], (Senior Member, IEEE), HWANJO HEO[4], AND JUN KYUN CHOI[1], (Senior Member, IEEE)

[1]School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea
[2]Information and Electronics Research Institute, Korea Advanced Institute of Science and Technology, Daejeon, South Korea
[3]Department of Computer Science, Liverpool John Moores University, Liverpool, U.K.
[4]Graduate School of Information Security, Electronics and Telecommunications Research Institute, Daejeon, South Korea

Corresponding author: Sangdon Park (sangdon.park@kaist.ac.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].

ABSTRACT With the widespread use of the Internet of Things, data-driven services take the lead of both online and off-line businesses. Especially, personal data draw heavy attention of service providers because of the usefulness in value-added services. With the emerging big-data technology, a data broker appears, which exploits and sells personal data about individuals to other third parties. Due to little transparency between providers and brokers/consumers, people think that the current ecosystem is not trustworthy, and new regulations with strengthening the rights of individuals were introduced. Therefore, people have an interest in their privacy valuation. In this sense, the willingness-to-sell (WTS) of providers becomes one of the important aspects for data brokers; however, conventional studies have mainly focused on the willingness-to-buy (WTB) of consumers. Therefore, this paper proposes an optimized trading model for data brokers who buy personal data with proper incentives based on the WTS, and they sell valuable information from the refined dataset by considering the WTB and the dataset quality. This paper shows that the proposed model has a global optimal point by the convex optimization technique and proposes a gradient ascent-based algorithm. Consequently, it shows that the proposed model is feasible even if the data brokers spend costs to gather personal data.

INDEX TERMS Data brokers, profit maximization, willingness-to-buy, willingness-to-sell.

I. INTRODUCTION

As data-driven services and applications take the lead of both online and offline businesses, data have become ubiquitous and are now considered as new oil for the emerging fourth industrial revolution, as a new valuable asset and an indispensable driving force. Accordingly, data are wildly overwhelming in not only its volume but also its diversity due to a number of Internet of Things (IoT) devices have rapidly increased [1]. It becomes hard to search, discover, process, and analyze the proper data from the whole. As a result, the big-data technology has emerged to find the true value of the data.

The associate editor coordinating the review of this manuscript and approving it for publication was Md Fazlul Kader.

With the emerging big-data technology, a data broker (also called an information broker or an information reseller) appears, which collects, analyzes, and sells data about individuals (i.e., data providers), to other third parties (i.e., data consumers) [2], [3]. Especially, personal data (any information relating to an identified or identifiable individual) are the main target for data brokers because it can be used to value-added services for customers (e.g., customization, recommendation, etc.) [4], [5]. Data brokers gather personal data from various IoT devices (e.g., sensors, smartphones, or wearable devices) and service providers (e.g., social network services, calendars, or mails, etc.) and perform big-data analytics to extract new information and knowledge. And then, they sell reproduced information to many other third party services to improve service quality

(e.g., potential customer detection, identity verification, fraud detection, etc. [2]).

Because more personal data result in more revenue, data brokers try to collect personal data from IoT environment as much as possible. Even though data brokers provide terms and conditions for privacy as opt-in based agreement (or consent), data providers (i.e., data subjects which provide their personal data) have no chance to know how their personal data are processed, delivered, and used. According to the survey [6], 67% of the respondents said that organizations, companies and agencies ask for too much personal information online. Moreover, less than 40% of the respondents trust online business actors (e.g., search engine companies, social network service providers, online marketers and advertisers). Many people think the current personal data ecosystem has various risks and is not trustworthy because the current personal data market has little transparency between data providers and data brokers/data consumers [6], [7]. These kinds of little transparent environments discourage data providers to share or sell any personal data to the market.

On the other hand, many studies (e.g., literature in behavior economics [8]–[11]) show that people also have an interest in valuation of their privacy with proper incentives or benefits [12]–[14]. This behavior is also considered as the concept of willingness-to-sell (WTS) personal data (or willingness-to-accept offered price from data brokers). According to the survey [10], individuals are willing to share personal data for deals and better customer services. In addition, Grossklags and Acquisit [12] investigated that the average number of people with WTS their personal data is dramatically higher than the average of people with willingness-to-protect them. Moreover, Benndorf and Normann [13] studied a incentive based WTS personal data based on an assessment of individuals. This study showed cumulative distribution of the WTS personal data of individuals in a social network service.

However, conventional studies have mainly focused on the relationship between data brokers and data consumers. Many studies only consider WTB of data consumers for provided services (i.e., willingness-to-pay for services of data brokers), and data providers are not well-considered because the current personal data market environment is mainly controlled by data brokers and third party consumers [15], [16]. However, according to reports [8]–[11], a data provider's perspective is also important and should be well-considered for future personal data market development. Consequentially, the WTS of data providers is one of the important aspects, but, only few studies have considered the WTS of data providers for the data market in the field of engineering.

Therefore, in this paper, we propose a novel personal data trading model in which a data broker buys multiple types of personal data from data providers by providing proper incentives based on privacy awareness and WTS of each personal data type through IoT interfaces; and the data broker sells valuable information from refined personal data as a service to data consumers by considering WTB and quality of gathered personal dataset. We consider that data providers have WTS of each personal data type, and they sell their personal data if and only if their WTS are satisfied. Similarly, data consumers also decide to buy personal data based on proposed prices and their WTB. The contributions of this paper are summarized as follows:

- This paper designs a personal data trading model for data brokers in IoT data marketplace with multiple types of personal data (e.g., physical location, brand preferences, purchase histories, etc.) from IoT environment by considering economic benefits of personal data providers as well as satisfaction of personal data consumers. Specifically, in order to satisfy requirements of each market participant, this paper considers not only WTB of the consumers but also WTS of the providers.
- The proposed WTS is designed based on a real-world experience performed in [13] with different privacy awareness factors because people feel different privacy violations depending on personal data types. The proposed WTB is also designed based on literature [17] to reflect real-world behaviors.
- This paper also proposes a personal data quality model for the collection of heterogeneous types of personal data by considering correlations of multiple personal data types as well as quantity of personal dataset based on literature [18] related to personal data pricing factors. The correlation of multiple personal data types is important to measure quality of personal dataset because when personal data is combined with other personal data, personal data providers can potentially be identified through additional processing. It means personal data consumers are able to create more value with personal dataset by targeting proper stakeholders for their business.
- This paper models a profit maximization problem with expected revenue and cost by considering WTB and WTS, respectively. It also shows that the objective function of the proposed profit maximization problem is concave. Based on the concaveness, the multivariate gradient ascent (MGA) algorithm is proposed to find the global maximum point. With some numerical and experimental analysis, this paper also shows the proposed personal data trading model is feasible to real-world applications.

With the best of our knowledge, this paper is the first paper which mathematically solves the profit maximization problem of a data broker by jointly considering both WTS of personal data providers and WTB of personal data consumers in IoT data marketplaces. This paper shows that the proposed personal data trading model is feasible even if a data broker spends costs to gather personal data from providers, which implies that personal data providers can actively participate in the IoT data market for their own benefits.

The rest of this paper is organized as follows. Section II introduces previous works regarding the data market and pricing schemes for privacy. Section III presents an overview of

the proposed personal data trading model with the proposed WTS, WTB, and personal data quality models. Section IV formulates a profit function of a data broker satisfying both WTS and WTB and solves the profit maximization problem. Section V shows some numerical and experimental results including analysis based on a real-world dataset, and Section VI discusses the feasibility of the proposed model in detail. Finally, this paper is concluded in Section VII.

II. RELATED WORK

Data market and data trading issues have recently motivated studies to maximize revenues and profits of data brokers (or data service providers) while satisfying data consumers (i.e., data buyers) [19]–[26]. Specifically, various data market structures for data trading like monopoly market, oligopoly market, and strong competition market were introduced in [19]. Niyato *et al.* [20] proposed a simple data market model for IoT environments. This study considered WTB of data consumers depending on data quality. Zhao *et al.* [21] proposed a machine learning based privacy-preserved data trading market with blockchain for preventing single-point-failure. Yu *et al.* [22] proposed a mobile data trading model based on the prospect theory in behavior economics to trade mobile data as quantity between mobile users by considering data demands and demand uncertainty. Al-Fagih *et al.* [23] proposed a data pricing scheme for public sensing framework considering delay, quality of services, and trust factors. Competitive data markets also have been studied. Jang *et al.* [24] modeled a data market with multiple independent data sources in IoT environments. In this model, a data service provider (i.e., a data broker) has limited budget to buy data from all data sources with the non-cooperative data trading model. This paper showed the existence and the uniqueness of the Nash equilibrium for the proposed trading model. Jiao *et al.* [25] proposed an auction based data trading model between a data service provider and data consumers. Hui *et al.* [26] proposed a sensing service system considering utilities of data providers and data service providers with a data pricing scheme in vehicle sensor networks. However, these studies did not consider behaviors of data providers and/or characteristics of personal data which are important factors in personal data brokering/analytic services.

In a personal data ecosystem, people have different privacy concerns regarding personal data types (i.e., less private or more private), so privacy awareness should be considered for the personal data market [27], [28]. Personal data markets and valuations of privacy (i.e., personal data pricing or privacy pricing models) have also been studied [18], [28]–[33]. Malgieri and Custers [18] investigated that the monetary value of personal data can be quantified with various personal data pricing factors. Gkatzelis *et al.* [29] proposed a bundle based personal data trading scheme with the linear pricing model. Shen *et al.* [30] also proposed a linear function based the personal data pricing scheme by considering information entropy, credit of data providers, and data reference index. Xu *et al.* [31] proposed a dynamic personal data pricing scheme using the multi-armed bandit approach under privacy-preserved personal data environments. This paper modeled a cumulative distribution of willingness-to-accept proposed price to buy personal data (i.e., WTS) of data providers as simple counting (i.e., in viewpoint of a data broker, the number of successful data gathering divided by total number of requests.). Wang *et al.* [32] proposed a query based data pricing scheme. In this paper, the data pricing scheme modeled for approximate aggregate queries from data consumers by considering pricing points from data sellers and the accuracy of query results. In addition, there are few studies tackled the personal data market with data providers' (or data owners') incentives. Li and Raghunathan [28] proposed a incentive-compatible mechanism for data owners to price and disseminate their private data based on the privacy sensitivity level. Parra-Arnau [33] investigated the trade-off between privacy and money of data providers and proposed an optimization model for profile-disclosure risk and economic reward. Su *et al.* [34] proposed a incentive based crowd sourcing scheme for collecting various data in cyber-physical-social systems. It also proposed an auction based price bidding scheme for data providers. Li and Raghunathan [28], Parra-Arnau [33], and Su *et al.* [34] only considered the relationship between data providers and data brokers.

In this paper, to maximize profit of a data broker, we jointly consider WTS of personal data providers as well as WTB of personal data consumers with different privacy awareness factors and personal data quality, respectively.

III. SYSTEM MODEL

This section describes a personal data trading model. We consider a IoT data market consisting of three groups that behave for their own benefits:

- $G_P = \{\omega_1, \cdots, \omega_N\}$, a group of candidates who may provide their own personal data;
- d_B, a single data broker who processes personal data and provides it as a service;
- $G_S = \{\tau_1, \cdots, \tau_M\}$ a group of candidates who may subscribe a service from the data broker.

This market handles K personal data types (e.g., gender, location, e-mail, purchase history, etc.). The overview of the proposed personal data trading model is described in Figure 1.

Normally, the data broker obtains opt-in based agreement for personal data collection from each personal data provider, and then the data broker continuously gathers personal data based on the agreement. This paper models that the data broker buys personal data from the providers under the subscription-based model with unit prices $\mathbf{c} = (c_1, c_2, \ldots, c_k)$ (i.e., each personal data type k has the unit price c_k.). The providers sell personal data to the data broker based on their own WTS, and each personal data type has different WTS. On the other hand, personal data consumers buy personal data from the data broker based on their own WTB. Then, the expected profit U is decided by expected revenue and cost depending on WTB and WTS, respectively.

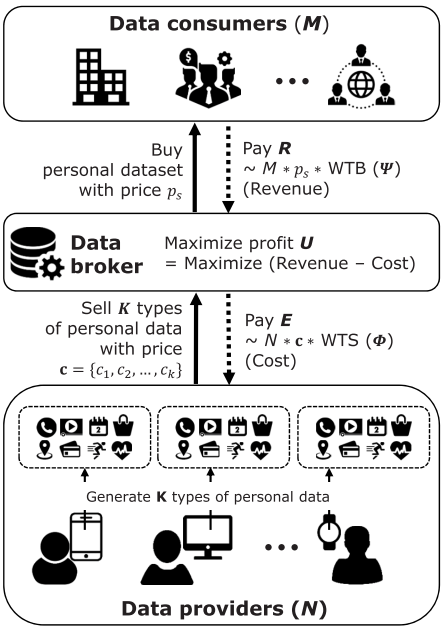


FIGURE 1. The proposed personal data trading model with multiple types in the IoT data market.

TABLE 1. Major symbols.

Symbols	Definition
K	Number of personal data types in the market
c_k	Cost of each personal data type $k \in K$
$\mathbf{c}$	Cost vector for all personal data types $\{c_1, c_2, ..., c_k\}$
ρ_k	Privacy awareness factor of each personal data type $k \in K$
N	Number of personal data providers
Φ	The provider's willingness-to-sell function
M	Number of personal data consumers
p_s	Subscription fee of each personal data consumer
Ψ	The consumer's willingness-to-buy function
U	Profit function of the data broker
Q	Personal data quality function
r	Correlations between different personal data types

Table 1 lists major symbols in this paper for the convenience of readers.

A. WILLINGNESS-TO-SELL OF DATA PROVIDER

First, the following real-valued random variable of kth personal data type should be defined in order to provide a well-defined WTS function of the data provider.

Definition 1 (Limit Selling Price of Data Provider): Let ω be an arbitrary candidate in G_P, who may provide their own personal data ($\omega \in G_P$). A random variable X_k, called a limit selling price of kth personal data type, is defined by the limit price at which ω decides to sell the kth personal data type. In other words, ω will not sell his/her own kth personal data type with the price lower than X_k, but sell it otherwise.

Since each individual has their own personal opinions regarding the value of personal data, a WTS function of the kth personal data type is defined by the cumulative distribution function (CDF) of X_k ($\Phi_k(c_k) = \mathbb{P}\{X_k \leq c_k\}$). To define the WTS function, following two principles are applied. The first is the more money offered, the more people accept to sell their personal data. The second is privacy awareness of each personal data type. Privacy awareness (denoted as ρ) means that people consider private of personal data differently, so the higher price is needed to buy the more private personal data. Table 2 shows some parts of statistics from the survey [27], which are related to privacy awareness of different personal data types categorized by various age groups. It shows that people have different privacy concerns depending on types of personal data. For example, people think data of credit card usage is more private than that of brand preferences. Finally, the definition of the WTS function for the kth personal data type is provided as following.

TABLE 2. Privacy awareness from [27].

Personal data types	Age group 18 to 24	Age group 35 to 48	Age group 68 above
Credit card data	89	88	89
Health/genetic information	70	74	71
Exact location	71	65	59
Purchase history	50	53	43
Media usage/preference	44	43	37
Brand preference	24	21	17

[1] Survey Question: "How private do you consider the following types of personal?"
[2] Result: Percentage(%) of respondents who consider the data type moderately or extremely private

Definition 2 (Willingness-To-Sell (WTS) Function): The WTS function Φ_k of the kth personal data type is defined by the CDF of the limit selling price of the kth personal data type, that is, $\Phi_k(c_k) = \mathbb{P}\{X_k \leq c_k\}$, and given by,

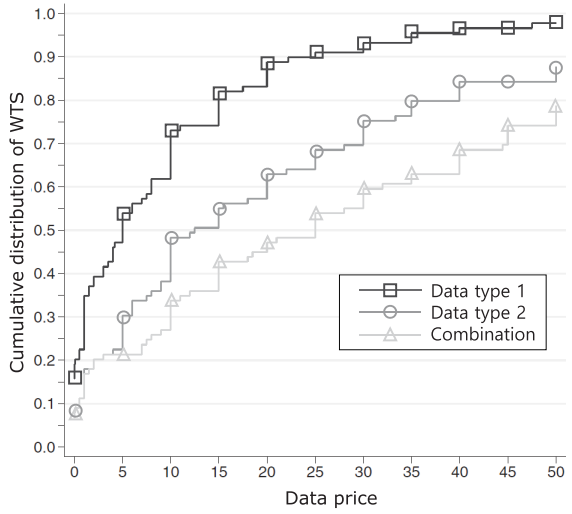
$$\Phi_k(c_k) = 1 - e^{-c_k \rho_k}. \tag{1}$$

Figure 2 shows an empirical WTS as cumulative distribution from [13] (Figure 2(a)) and the proposed WTS personal data function curves with various privacy-awareness factors (ρ) (Figure 2(b)). Note that the curves of WTS from the reference and the proposed function are similar. WTS value reaches to 1.0 rapidly in less private case (with larger ρ) and reaches slowly in more private case (with smaller ρ). The details to decide proper ρ for each personal data type are discussed in numerical analysis part (Section V).

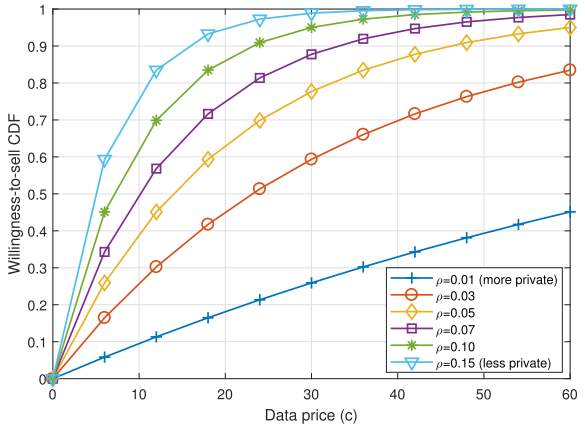
B. PERSONAL DATA QUALITY MODEL

Big-data analytics can be performed to extract new valuable information from raw data using various techniques including data mining, machine learning, etc. However, this paper does not consider data analytic methods; this paper is interested in quality of data analytic results. Thus, we define the quality function Q to estimate improved personal data quality.

Quantifying the value and quality of personal data is an important factor to decide price of personal data. We choose several metrics to define the personal data quality function Q based on [18]. By analyzing various privacy regulations and researches, this work investigated privacy pricing factors as



(a)



(b)

FIGURE 2. **Willingness-to-sell as a function of data price. (a) Empirical WTS [13]. (b) Proposed WTS.**

follows: *size* of dataset, *completeness* of dataset, *a number of data types* and their *combinations*, the level of *identifiability* of personal data.

Based on previous works, we choose two principles for personal data quality. The first is data quantity related to *size* and *completeness* of dataset, which are also related to general data quality measurements [35], [36]. More data equates the high chance to estimate each individual correctly. Data quantity is also related to *accuracy* for data processing. Several studies showed that when the number of data increases, the accuracy of machine learning analysis increases [20], [24], [25].

The second is the correlation of personal data types (i.e., *a number of personal data types* and their *combination*, which are characteristics of personal data). Personal data types with higher correlations have more valuable information to data consumers. When personal data is combined with other personal or identifiable information which is linked or linkable to a specific individual, the data providers can potentially be identified through additional processing of other attributes [37] (the level of *identifiability*). Note that privacy awareness of each personal data type is already considered in the WTS function. Based on these principles, the definition of personal data quality can be provided by the following definition.

Definition 3 (Personal Data Quality Function): The personal data quality function Q is defined by the quality of service that can be provided by the personal dataset $\mathbf{c} = (c_1, \cdots, c_K)$ and given by,

$$Q(\mathbf{c}) = N\Big[w_1\Big\{\sum_{i\in K}\Phi_i(c_i)\Big\} + w_2\Big\{\sum_{i\in K}\sum_{\substack{j\in K\\ i\neq j}} r_{ij}\sqrt{\Phi_i(c_i)\Phi_j(c_j)}\Big\} + w_3\Big\{\sum_{i\in K}\sum_{\substack{j\in K\\ i\neq j}}\sum_{\substack{k\in K\\ i\neq j\neq k}} r_{ijk}\sqrt[3]{\Phi_i(c_i)\Phi_j(c_j)\Phi_k(c_k)}\Big\} + \cdots\Big],$$

where w is the weight for each order term ($\sum w_i = 1$) and r is the correlation among personal data types ($\forall r \in [0, 1]$).

To define the personal data quality function, the geometric mean (root terms) is used to consider personal data relativity. The personal data quality, which the data broker has, increases when i) personal data types are tightly correlated, ii) the amount of personal data is large, and iii) the amount of data for each personal data type is balanced. Note that if $w_1 = 1$ and $w_2 = w_3 = \cdots = 0$, then it forms $\sum_{k\in K} 1 - e^{-c_k\rho_k}$ which is the same as data quality functions proposed in previous works [20], [24], [25]. For simplicity, we consider the personal data quality (Q) up to second order terms as follows:

$$Q = N\sum_{i\in K}\sum_{j\in K} r_{ij}\sqrt{\Phi_i(c_i)\Phi_j(c_j)}. \tag{2}$$

C. WILLINGNESS-TO-BUY OF DATA CONSUMERS

Similar to the WTS function, a random variable about candidates who may buy personal dataset should be provided first in order to define the WTB function of personal data consumers.

Definition 4 (Limit Buying Price of Data Consumer): Let τ be an arbitrary candidate in G_S who may buy personal dataset form the data broker ($\tau \in G_S$). A random variable Y, called a limit buying price, is defined by the limit price at which τ decides to buy the personal dataset from the data broker d_B. In other words, τ will not buy the service larger than the price Y, but buy it otherwise.

Then, the WTB function Ψ of the data consumers is defined by the CDF of the random variable Y defined as above, that is, $\Psi(p_s) = \mathbb{P}\{Y \geq p_s\}$.

It remains to model the WTB function rationally. We consider basic economic principles for demand as follows: i) WTB decreases when data price increases, ii) data consumers prefer to buy personal data with higher quality rather than that with lower quality [17]. There are many ways to define the demand curve depending on the price elasticity of demand (e.g., linear, polynomial, exponential, etc.). From the assumptions, the WTB function of personal data consumers

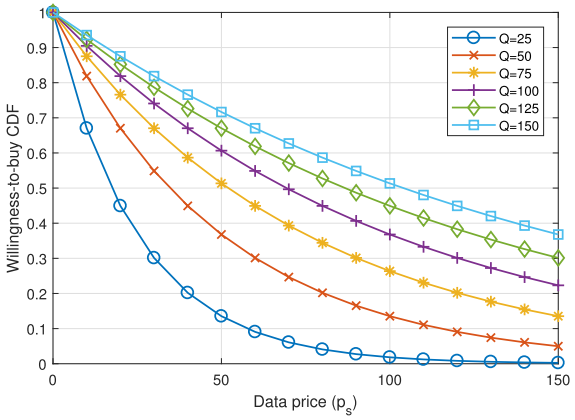


FIGURE 3. Willingness-to-buy function with personal data quality (Q).

is defined as an exponential based function, given by the following definition.

Definition 5 (Willingness-To-Buy (WTB) Function): The WTB function Ψ of the data consumers is defined by the CDF of the limit buying price of the kth personal data type, that is, $\Psi_k(p_s) = \mathbb{P}\{Y \leq p_s\}$, and given by,

$$\Psi(p_s) = e^{-p_s/Q}. \tag{3}$$

Figure 3 shows the curves of the WTB function with respect to various personal data quality Q. It basically decreases when price of personal dataset increases, and it also rapidly decreases with lower quality and slowly decreases with higher quality, respectively.

IV. OPTIMAL PERSONAL DATA PRICING SCHEME

In this section, the data broker's profit maximization problem is addressed from models for expected cost and revenue of the data broker based on the proposed WTS and WTB functions, respectively.

A. EXPECTED COST AND REVENUE

To define the profit function U, we consider the cost and revenue of the data broker. As explained in the previous section, personal data providers determine whether they sell personal data or not based on their own WTS requirements. The cost for buying one personal data type is $c_k\Phi_k(c_k)$, and the cost for buying all personal data types is $\sum_{k\in K} c_k\Phi_k(c_k)$. If the number of the providers is N, then the total cost (E) for buying personal data follows:

$$E(\mathbf{c}) = N\sum_{k\in K} c_k\Phi_k(c_k). \tag{4}$$

Similarly, personal data consumers buy personal data from the data broker based on their WTB requirements. The revenue for selling personal dataset to one personal data consumer is $p_s\Psi(p_s)$. If the number of the consumers is M, then the total revenue (R) for selling personal dataset follows:

$$R(p_s) = Mp_s\Psi(p_s). \tag{5}$$

B. DATA BROKER PROFIT FUNCTION

Based on expected cost and revenue functions from equations (4) and (5), the profit function U of the data broker is modeled as follows:

$$\begin{aligned} U(p_s, \mathbf{c}) &= R(p_s) - E(\mathbf{c}) \\ &= Mp_s\Psi(p_s) - N\sum_{c_k\in\mathbf{c}} c_k\Phi_k(c_k) \end{aligned}$$

The profit of the data broker is the total revenue by selling personal dataset minus the total cost by buying individuals' personal data. In this paper, we assume that additional costs for data processing, storage, and management are negligible. Then, the profit maximization problem can be formulated as follows:

$$\begin{aligned} \mathbf{P1}: \ & \max_{p_s,\mathbf{c}} U(p_s, \mathbf{c}) \\ & \text{such that } U > 0, \quad N > 0, \ M > 0, && (6) \\ & p_s > 0, \quad \forall c_k \in \mathbf{c} > 0, && (7) \\ & r \in [0, 1], \quad \rho \in [0, 1]. && (8) \end{aligned}$$

The condition (6) means a profit of the data broker should be larger than zero to validate this problem, and the number of data providers and consumers should be larger than zero. Similarly, the cost for buying personal data $\mathbf{c}$ and the price for selling personal data p_s should be larger than zero (condition (7)). The conditions for privacy awareness factors (ρ) and personal data correlations (r) are explained in Sections III-A and III-B, respectively (condition (8)).

C. OPTIMIZATION

In this section, the profit optimization problem is solved. We check that the existence of p_s^* which maximizes the revenue of the data broker when all $c_k \in \mathbf{c}$ are determined.

Theorem 1: The optimal price that maximizes the profit function of the data broker is exactly the same as the personal data quality, that is, $p_S^* = Q$.

Proof: It is checked that the revenue function is concave, thus the maximum point can be obtained by checking the first order derivative of the revenue function R becomes 0.

$$\begin{aligned} \frac{d}{dp_s}R &= M[e^{-p_s/Q} + p_s\{e^{-p_s/Q} - \frac{1}{Q}\}] = 0 \\ &\Rightarrow M[e^{-p_s/Q}(1 - \frac{p_s}{Q})] = 0 \\ &\Rightarrow 1 - \frac{p_s^*}{Q} = 0 \\ &\Rightarrow p_s^* = Q \end{aligned} \tag{9}$$

□

Since p_s^* is also the function of $\mathbf{c}$ based on the equations (2) and (9), the profit function $U(p_s^*, \mathbf{c})$ can be represented as $U^*(\mathbf{c})$:

$$U^*(\mathbf{c}) = NMe^{-1}\sum_{i\in K}\sum_{j\in K} r_{ij}\sqrt{\Phi_i(c_i)\Phi_j(c_j)} - N\sum_{c_k\in\mathbf{c}} c_k\Phi_k(c_k).$$

Then, the optimization problem for the profit function (**P1**) can be reduced as follows:

$$\mathbf{P1}^*: \max_{p_s,\mathbf{c}} U^*(\mathbf{c})$$
$$\text{such that } U^* > 0, \quad N > 0,\ M > 0, \tag{10}$$
$$p_s > 0, \quad \forall c_k \in \mathbf{c} > 0, \tag{11}$$
$$r \in [0,1], \quad \rho \in [0,1]. \tag{12}$$

We try to prove that $U^*(\mathbf{c})$ is a concave function for all $c_k \in \mathbf{c}$ with the analytical method by showing the second order derivative of the function $U^*(\mathbf{c})$ is negative. If the function $U^*(\mathbf{c})$ is concave, the profit function U^* has the unique global maximum point, which is the largest value of the function. Checking convexity of multivariable functions can be done by checking convexity of functions of one variable [38]. If we take the second derivative of $U^*(\mathbf{c})$ with respect to c_k, then we have

$$\frac{\partial^2}{\partial c_k^2}U^*(\mathbf{c}) = Nc_k\rho_k^2 e^{-c_k\rho_k} - 2N\rho_k e^{-c_k\rho_k},$$
$$- NMe^{-1}\rho^2 e^{-c_k\rho_k}(1 + \frac{A}{\sigma} + \frac{Ae^{-c_k\rho_k}}{2\sigma^{3/2}})$$

where $A = \sum_{i\in K, i\neq k}\sqrt{1-e^{-c_i\rho_i}}$ and $\sigma = \sqrt{1-e^{-c_k\rho_k}}$.

Unfortunately, it is hard to confirm the concaveness of the function U^* based on the result of second derivative, so we empirically check that the function U^* is concave with all c_k that make the U^* value positive with conditions from (10) to (12). Detailed graphs are described in Section V.

D. MGA ALGORITHM

The gradient ascent method is a well-known method to find the maximum point of a concave function, which iteratively takes steps proportional to the positive of the gradient. Since the concave function U^* has a k number of variables, we can apply the multivariate gradient ascent (MGA) algorithm as shown in Algorithm 1.

For the MGA algorithm to find the maximum point of the profit function U^*, first, this algorithm gets basic input parameters: the number of data providers (N), data consumers (M), personal data types (K), and privacy awareness factors (ρ). And then, it initializes several variables. The variable i, the number of iteration, is set to zero. The ϵ, threshold for the ending iteration, is set to 0.00001. If ϵ is small enough, the algorithm has a higher chance to find the global maximum point, however, if the ϵ is too small, the algorithm needs too many iterations to reach the maximum point. h is the constant to calculate the first order derivative based on the definition. h needs to be small enough to find the proper gradient at the given point $\mathbf{c}_i$. Then, it initializes the starting point $\mathbf{c}_0$ as $\{h, h, \ldots, h\}$ and calculates the expected profit U^* at $\mathbf{c}_0$. We set $\mathbf{c}_0$ with the value of h (i.e., very small value) because $\mathbf{c}$ should have non-zero values from the condition (11).

Next, it decides the step size of each iteration. The step size δ is the most important factor of this MGA algorithm because it can only reach the proper optimal point with the proper step size. Any positive value can be selected for the step size; however, it is very difficult to choose an arbitrary value for the fixed step size because the $U^*(\mathbf{c})$ is convex if and only if it satisfies conditions from (10) to (12). If this algorithm chooses a wrong step size value, there is a chance to violate the conditions during iterations. Therefore, to choose δ, we check the first order derivative of the profit function (U^*) as follows:

$$\frac{\partial U^*(\mathbf{c})}{\partial c_k} = N(e^{-c_k\rho} - 1) - Nc_k\rho e^{-c_k\rho}$$
$$+ NMe^{-1}\rho e^{-c_k\rho}(1 + \frac{A}{\sigma_k})$$

where $A = \sum_{i\in K, i\neq k}\sqrt{1-e^{-c_i\rho_i}}$ and $\sigma_k = \sqrt{1-e^{-c_k\rho_k}}$. It shows that both N and M are related to $\frac{\partial}{\partial c_k}U^*(\mathbf{c})$. Based on the experiment, this algorithm chooses δ as $\frac{1}{min(N,M)}$. The detailed analysis about the step size δ including time complexity (i.e., the number of iterations for various cases) is shown in Table 4.

After the initialization, the algorithm starts to find the maximum point. First, it increases the number of iteration (i) (step 1), and then finds the first order derivative of the profit function (U^*) for $\forall c_k \in \mathbf{c}$ (step 2). To check the first order derivative, this algorithm uses $\nabla U^*(\mathbf{c}_i) = \frac{U^*(\mathbf{c}_i+h)-U^*(\mathbf{c}_i-h)}{2h}$. Then, the algorithm updates the new cost point ($\mathbf{c}_i$) by finding the slope of the function ($\delta\nabla U^*(\mathbf{c}_i)$) at the previous cost point ($\mathbf{c}_{i-1}$) (step 3). Finally, it calculates the new profit value (U_i^*) in the newly updated cost point ($\mathbf{c}_i$) (step 4). If the difference

Algorithm 1 Multivariate Gradient Ascent (MGA)

Input:
N: the number of data providers
M: the number of data consumers
K: the number of data types
ρ: the privacy-awareness factor
Initialization:
(a) The number of iteration $i = 0$
(b) The threshold for the ending iteration $\epsilon = 0.00001$
(c) The constant for derivative $h = 0.000001$
(d-1) Allocate cost $\mathbf{c}_0 = \{c_k |\ c_k = h,\ \forall k \in K\}$
(d-2) Calculate the expected profit U_0^* at $\mathbf{c}_0$
(e) The step size $\delta = 1/min(N, M)$
Start algorithm:
do
(1) $i = i + 1$
(2) Find derivative of U^*: $\nabla U^*(\mathbf{c}_i)$
which means
$\frac{\partial}{\partial c_k}U^*(\mathbf{c}_i) = \frac{U^*(\mathbf{c}_i+h)-U^*(\mathbf{c}_i-h)}{2h}, \forall c_k, k \in K$
(3) $\mathbf{c}_i = \mathbf{c}_{i-1} + \delta\nabla U^*(\mathbf{c}_i)$
(4) $U_i^* = U^*(\mathbf{c}_i)$
while $\{|U_i^* - U_{i-1}^*| < \epsilon\}$
Output:
$\mathbf{c}$: a set of allocated cost to buy each data types
U^*: the maximum profit value

between the new profit value (U_i^*) and the previous profit value (U_{i-1}^*) is less then ϵ, then the algorithm stops and returns the allocated cost ($\mathbf{c}$) and the maximum profit (U^*), otherwise it goes to step 1.

V. NUMERICAL RESULTS WITH KAGGLE SMS DATASET

This section analyzes the proposed personal data trading model using both theoretical and empirical approaches. Moreover, by using 3-dimensional graphs, we will verify visually whether the optimal point of the profit maximization problem with multiple cost variables is correct.

In order to analyze the proposed personal data trading model, we configure parameters N, M, ρ, and r as follows:

- Since the N affects only the size of the resulting values, there are relatively few restrictions on the parameter selection. Thus, we just set the number of data providers $N = 1000$. For the number of data consumer M, we choose a reasonable number as $M = 200$ to valid the profit maximization problem.
- For privacy awareness factors in the proposed WTS function, we choose six ($K = 6$) personal data types (i.e., health condition ($35.0), payment details: credit cards ($20.8), purchase histories ($17.8), hobbies, tastes & preferences ($9.1), photos & videos ($5.9), and physical location: GPS ($5.1)) based on the real-world survey in [10] which introduces the average prices for each personal data type. We find each ρ that makes WTS value 0.5 with the proposed price for each personal data type from the survey. Then, each ρ_k is decided as $\rho = \{0.0198, 0.0333, 0.0389, 0.0762, 0.1175, 0.1359\}$, respectively. Note that the smaller ρ means more private and the larger ρ means less private.
- For the personal data quality function, correlations between two different personal data types r is randomly chosen in the range of [0, 1] (i.e., $r \in [0, 1]$) with uniform distribution unless conditions are separately mentioned.
- For the cost value $\mathbf{c}$, it represents the vector for cost of each personal data type as $\mathbf{c} = (c_1, c_2, \ldots, c_k)$.

First, we check the optimal values with a single personal data type ($K = 1$) and with two different personal data types ($K = 2$), and then we check the optimal values with various numbers of personal data types ($K \geq 3$). The results of the proposed algorithm (MGA) and the global optimum are also discussed. Moreover, we also show that how the average profit of the data broker changes with respect to the number of personal data providers and consumers.

A. THEORETICAL EXPERIMENT

This section checks the optimal values with a single personal data type ($K = 1$) and with two different personal data types ($K = 2$). Figure 4 shows the data broker's profit function U^* with the single data type ($K = 1$) with various privacy awareness factors (ρ) with the same personal data correlation factor ($r_{ii} = 1, \forall i \in K$). We can see the $U^*(\mathbf{c})$ is a concave function for all variables (N, M, ρ, r), and has the globally

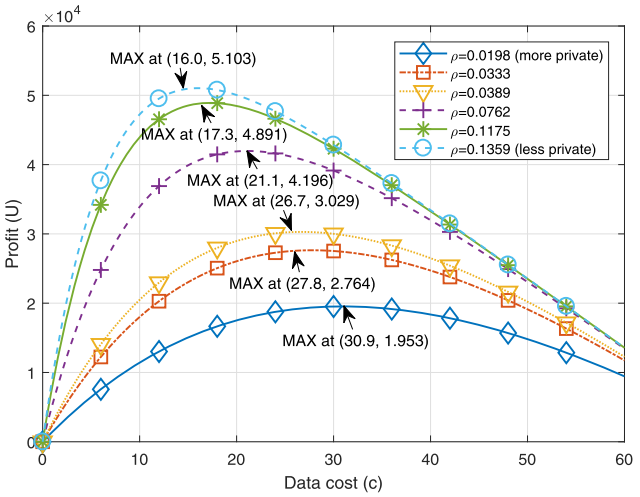


FIGURE 4. Profit values of the data broker with single personal data type w.r.t. privacy awareness factors (ρ).

unique maximum value for each ρ. The maximum values are follows:

$$\begin{cases} \rho = \{0.0198\} : 1.953 \times 10^4 & \text{at } \mathbf{c} = (30.9), \\ \rho = \{0.0762\} : 4.196 \times 10^4 & \text{at } \mathbf{c} = (21.1), \\ \rho = \{0.1359\} : 5.103 \times 10^4 & \text{at } \mathbf{c} = (16.0). \end{cases}$$

The profit value of larger ρ (i.e., less private personal data) is higher than that of smaller ρ (i.e., more private personal data) because the number of participants for selling data is different. A unit price for more private personal data is larger than less private personal data, however, it is hard to collect more private personal data because WTS of that is lower. It means the data broker spends much money to buy enough amount of more private ones. As a result, the amount of personal data collected is different; that is, the personal data quality is different, and it is directly applied to the profit value.

Next, we verify the profit values when the data broker handles two different personal data types ($K = 2$). Figure 5 shows the data broker's profit function U^* with the combination of two different privacy awareness factors (ρ) and personal data correlation constants (r) as follows:

$$r_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0.5 & \text{if } i \neq j. \end{cases}$$

It shows various results with different combinations of personal data types. The maximum profits are follows:

$$\begin{cases} \rho = \{0.0333, 0.0198\} : 6.772 \times 10^4 & \text{at } \mathbf{c} = (32.2, 37.8), \\ \rho = \{0.0762, 0.0198\} : 8.496 \times 10^4 & \text{at } \mathbf{c} = (23.4, 38.6), \\ \rho = \{0.1359, 0.0198\} : 9.518 \times 10^4 & \text{at } \mathbf{c} = (17.3, 38.9). \end{cases}$$

The profit increases when the data broker sells the combination of less private data and more private data. When the profit increases, the cost for buying the same data also increases (see c_2 with $\rho = 0.0198$ in $\mathbf{c} = (c_1, c_2)$).

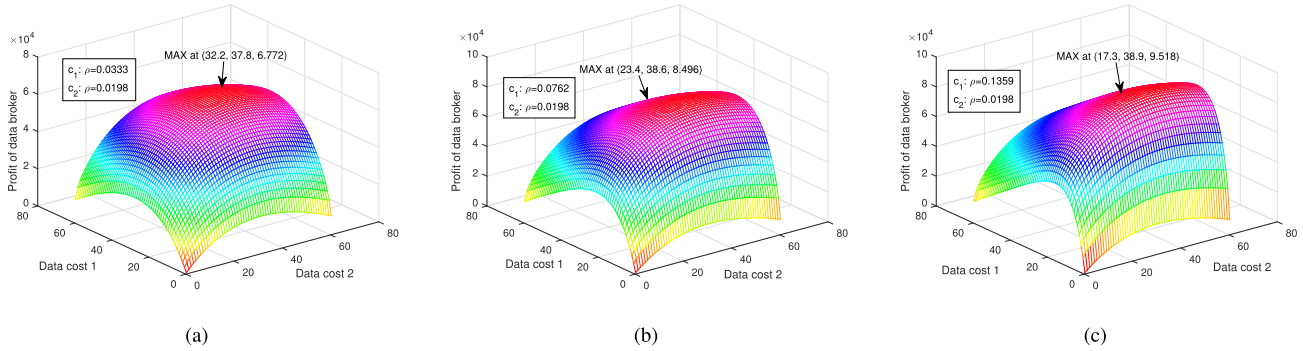


FIGURE 5. Profit values of the data broker with two different personal data types w.r.t. various privacy awareness factors (ρ). (a) $\rho_1 = 0.0333$ and $\rho_2 = 0.0198$. (b) $\rho_1 = 0.0762$ and $\rho_2 = 0.0198$. (c) $\rho_1 = 0.1359$ and $\rho_2 = 0.0198$.

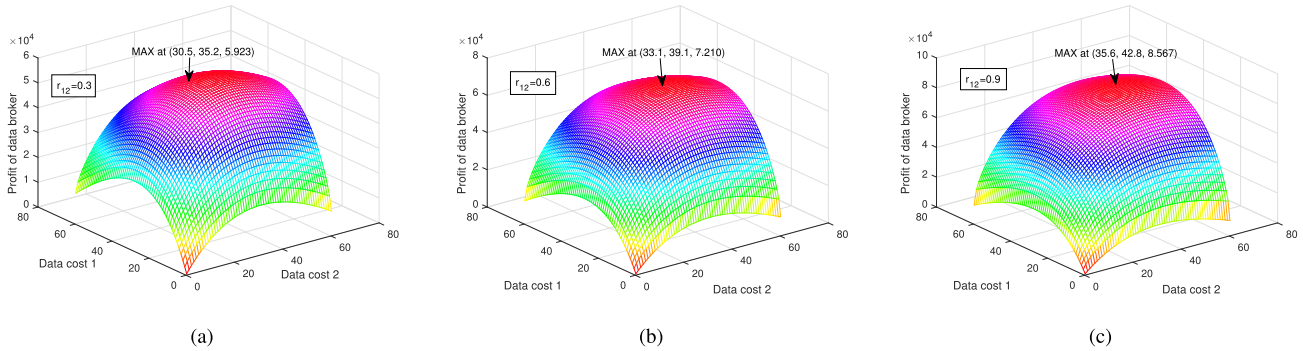


FIGURE 6. Profit values of the data broker with two different personal data types w.r.t. various personal data correlations (r) $\rho = \{0.0198, 0.0333\}$. (a) $r_{12} = 0.3$. (b) $r_{12} = 0.6$. (c) $r_{12} = 0.9$.

Note that the profit with two different personal data types is higher than that with a single personal data type (comparing Figure 4 and 5(a)) as follows:

$$\begin{cases} \rho = \{0.0198\} & : 1.953 \times 10^4 \quad \text{at } \mathbf{c} = (30.9), \\ \rho = \{0.1359\} & : 5.103 \times 10^4 \quad \text{at } \mathbf{c} = (16.0). \\ \rho = \{0.1359, 0.0198\} & : 9.518 \times 10^4 \quad \text{at } \mathbf{c} = (17.3, 38.9). \end{cases}$$

If the data broker handles each personal data type separately, the profit is 7.056×10^4 at $\mathbf{c} = (16.0, 30.9)$; however, if the data broker handles two data types together, the profit is 9.518×10^4 at $\mathbf{c} = (17.3, 38.9)$ because the personal data quality of combined dataset is more higher.

We also verify the profit value with different personal data correlations r. Figure 6 shows the data broker's profit value (U^*) for the combination of two personal data types ($\rho = 0.0198, 0.0333$) with different personal data correlation values ($r_{ij} = 1$ if $i = j$ and $r_{ij} = \{0.3, 0.6, 0.9\}$ if $i \neq j$). The maximum profits are follows:

$$\begin{cases} r_{12} = 0.3 : 5.923 \times 10^4 & \text{at } \mathbf{c} = (30.5, 35.2), \\ r_{12} = 0.6 : 7.210 \times 10^4 & \text{at } \mathbf{c} = (33.1, 39.1), \\ r_{12} = 0.9 : 8.567 \times 10^4 & \text{at } \mathbf{c} = (35.6, 42.8). \end{cases}$$

It shows that higher personal data correlation makes higher profit because the personal data quality increases when the personal data correlation increases. Similar to Figure 5, personal data costs (c_1 and c_2) also increase when the profit increases.

B. EXPERIMENT: KAGGLE SMS DATASET

In the previous section, we have checked that the proposed personal data trading model is well defined, and the optimization problem is truly and globally optimized for the profit maximization problem. From now on, we check the validity of the proposed MGA algorithm (Algorithm 1 in Section IV-D) with various numbers of personal data types ($K \geq 3$) using a real-world dataset from Kaggle [39], which is the public data platform for data-science, to implement the proposed MGA algorithm.

The data are short message service (SMS) texts with various personal data from mobile phones, which consist of 42,000 messages from year 2010 to 2017. These data are categorized into 12 different types including payment, reservation, medical appointment, delivery, etc. From these data, we choose 6 different personal data types used in the previous section (i.e., $\rho = \{0.0198, 0.0333, 0.0389, 0.0762, 0.1175, 0.1359\}$) and select N number of SMS text data for each personal data type. We assume that this dataset collected from N number of data providers and M number of the data consumers try to buy personal dataset

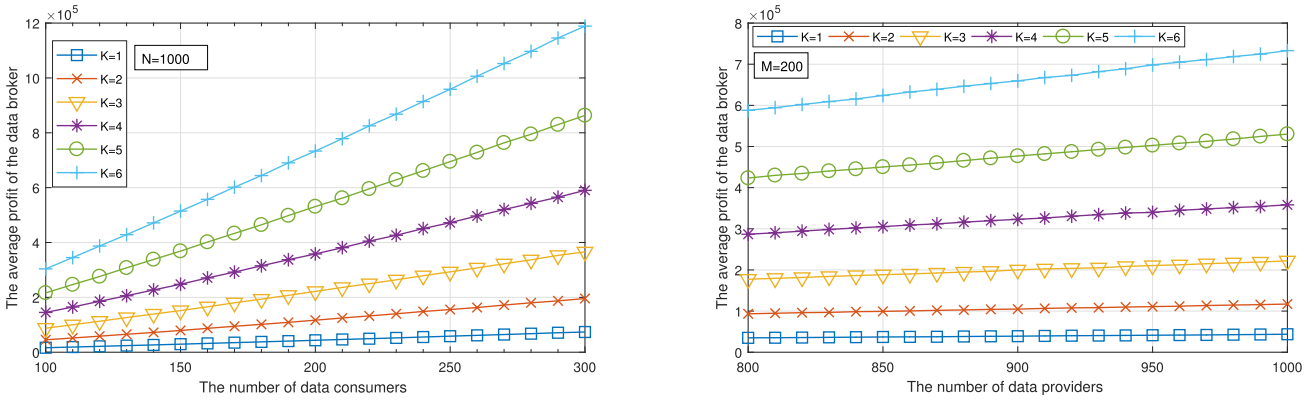


FIGURE 7. The average profit of the data broker with respect to the number of market participants. (a) Various numbers of personal data consumers (*M*). (b) Various numbers of personal data providers (*N*).

TABLE 3. The comparison of the proposed MGA and the global optimum.

K	Privacy awareness (ρ)	U_{MGA}	$\mathbf{c}_{MGA}$	$U_{Optimal}$	$\mathbf{c}_{Optimal}$
1	{0.0198}	19530.07	(30.92)	19530.07	(30.92)
2	{0.0198, 0.0333}	67723.11	(37.83, 32.23)	67723.11	(37.83, 32.23)
3	{0.0198, 0.0333, 0.0389}	146307.14	(44.11, 36.81, 34.64)	146307.14	(44.11, 36.81, 34.64)
4	{0.0198, 0.0333, 0.0389, 0.0762}	272946.37	(50.10, 41.13, 38.50, 27.75)	272946.37	(50.10, 41.13, 38.50, 27.75)
5	{0.0198, 0.0333, 0.0389, 0.0762, 0.1175}	444853.69	(55.53, 44.98, 41.94, 29.70, 22.97)	-	-
6	{0.0198, 0.0333, 0.0389, 0.0762, 0.1175, 0.1359}	656452.48	(60.45, 48.42, 44.99, 31.41, 24.11, 21.94)	-	-

because this paper assumes the subscription based personal data trading model as explained in Section III.

Table 3 shows that the comparison of the proposed MGA and the global optimum for profit values of the data broker using the dataset. Note that cases of $K = 5$ and $K = 6$ for the global optimum are not obtained because it takes too much time for finding optimal points with large K. From $K = 1$ to 4, the results of the proposed MGA and that of the global optimum are same. It shows that the proposed MGA algorithm is properly designed.

Next, using the proposed MGA algorithm, we calculate the average profit values of the data broker with respect to the number of data consumers and data providers as shown in Figure 7. To obtain average values, we simulate 10,000 times for each case with uniformly distributed random values $r \in [0, 1]$ and $\rho \in [0.01, 0.2]$ (i.e., the average data price c_k that makes WTS value 0.5 within $c_k \in [3, 70]$) for each personal data type in the dataset.

First, we check the profit of the data broker with different number of data consumers (M) as shown in Figure 7(a). It shows that as the number of data consumer increases, the average profit also increases. Note that the average profit increases rapidly when the number of personal data types increases because the personal data quality becomes higher which is directly related to revenue. Similarly, we also check the profit with different number of data providers (N) as shown in Figure 7(b). Since the number of data providers is related with both expected revenue and cost as described in Section IV-A, the profit slowly increases.

At last, we analyze the performance of the proposed MGA algorithm with respect to various parameters including the step size δ and input parameters N, M, and K as shown in Table 4. The left side of the table shows the results of the MGA algorithm with respect to various step sizes. If the step size is large enough, it can reach the optimal point very fast (i.e., less than 40 iterations with the step size $1/100$); however, it may not be proper to find the optimal point if the step size is too large. As shown in the bottom of the left side of the table, the fixed step size with an arbitrary value may not be applicable to some cases. Consequently, we need to choose either $1/N$ or $1/M$ as the step size of the algorithm for robustness and stability, so the proposed MGA algorithm chooses the step size as $\delta = 1/min(N, M)$. The right side of the table shows the results of the proposed algorithm with the step size $1/N$ or $1/M$ with different cases. Both $1/N$ and $1/M$ can find the same optimal point (U_{MGA}), but the required number of iterations is different.

VI. PRACTICAL INSIGHTS

In this section, we discusses the detailed feasibility of the proposed personal data trading model for real-world applications.

TABLE 4. The step size and time complexity analysis for the MGA algorithm.

N	M	K	U_{MGA}	step size δ	# of iter.	N	M	K	U_{MGA}	step size δ	# of iter.
1000	200	6	-	$< 1/50$	Not applicable	1000	200	3	146307.14	(1/N, 1/M)	(331, 67)
1000	200	6	656452.48	1/100	37	1000	200	4	272946.37	(1/N, 1/M)	(338, 69)
1000	200	6	656452.48	1/M	71	1000	200	5	444853.69	(1/N, 1/M)	(346, 70)
1000	200	6	656452.48	1/N	353	1000	400	6	1546256.61	(1/N, 1/M)	(415, 170)
1000	200	6	656452.48	1/10000	3172	2000	400	6	3092513.22	(1/N, 1/M)	(429, 84)
4000	400	6	-	1/100	Not applicable	4000	400	6	6185026.45	(1/N, 1/M)	(440, 57)
4000	800	6	-	1/100	Not applicable	4000	800	6	13717067.41	(1/N, 1/M)	(489, 108)
4000	8000	6	-	1/100	Not applicable	4000	8000	6	155439207.62	(1/N, 1/M)	(739, 983)

As introduced in the beginning of this paper, the personal data brokering market has been continuously growing, and data brokers make a huge amount of money with the personal data analytic; but many personal data providers think that the current ecosystem is not trustworthy due to little transparency and individual's empowerment for personal data usage. Many literature surveys have identified the trust gap between personal data brokers/consumers and providers [6], [7], [40], and one of the major items to improve trust of the personal data ecosystem is the empowerment of personal data providers.

Since new regulations and legislation with strengthening the rights of personal data providers for the personal data ecosystem were introduced and already applied to the real-world [40]–[42], it is necessary for data brokers to accept a new paradigm about empowering providers. However, these new regulations become obstacles for conventional data brokers' business. A typical example is the European General Data Protection Regulation (GDPR) which affects personal data exploitation and hugely impacts on personal data brokering businesses [40]–[46]. One of the major articles related to data brokers in GDPR is *conditions for consent* (Article 7). Before GDPR enforcement, data brokers were able to collect and process personal data with opt-in based agreement with long illegible terms and conditions. However, data brokers now can collect personal data only after acquiring consent from individuals by specifying the purpose of personal data processing. For example, if the personal data provider refuses to use his/her personal data for other purposes rather than the original purpose, the personal data cannot be used even though it is collected and used for the original purpose. Now, data brokers must consider how to acquire the consent from personal data providers for their businesses.

There are many possible ways to increase the possibility for market participation of personal data providers; and in this paper, we have focused on a possible approach with personal data providers' incentives. Since personal data now become a new financial asset for individuals [10], this approach also has investigated in previous works [18], [28], [33]. However, it was still ambiguous about the feasibility of the IoT data market with personal data providers' incentives; in other words, how is possible to maintain the current IoT data market structure by satisfying stakeholders (i.e., providers, consumers, and especially brokers) because the previous studies did not jointly consider major stakeholders in the market. In this paper, with some numerical and experimental analysis in Section V, we have shown that the proposed personal data trading model is feasible for brokers as well as providers/consumers while satisfying their requirements (i.e., profit maximization, WTS, and WTB, respectively).

The proposed trading model will have positive effects for stakeholders by invigorating the personal data ecosystem. Personal data providers will be more inclined to provide their personal data. So far, IoT data markets have not been well-formed due to the lack of transparency. However, if the proposed trading model is applied, each individual may feel comfortable to provide their personal data at a fair value. Personal data consumers do not need to spend budget on unnecessary personal dataset to meet the desired purpose and quality requirements.

VII. CONCLUSION

In this paper, we have proposed a novel personal data trading model that consists of personal data providers, a data broker, and personal data consumers while considering heterogeneous personal data types. We have designed the realistic WTS and WTB personal data functions based on the literature, so the expected number of providers and consumers are obtained practically. We also have designed the personal data quality model taking into consideration multiple types of personal data. The analytic results have demonstrated that the proposed approach is guaranteed to find a global maximum point for the profit function of the data broker, and it is feasible to the current IoT data marketplaces. As a future work, multiple personal data stores and multiple data brokers can be considered to extend the proposed trading model. In addition, various cost models for data computing and storage can be considered [28]. Moreover, an auditable ledger technique, which is able to record personal data transactions among the market stakeholders [47], can also be applied to design a personal data trading model more realistically.

ACKNOWLEDGEMENT

We would like to thank the anonymous reviewers for their insightful suggestions.

REFERENCES

[1] C. Perera, C. H. Liu, and S. Jayawardena, "The emerging Internet of Things marketplace from an industrial perspective: A survey," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 4, pp. 585–598, Dec. 2015.

[2] Federal Trade Commission, *Data Brokers: A Call for Transparency and Accountability*. 2014.

[3] A. Rieke, H. Yu, D. Robinson, and J. von Hoboken. (2016). Data Brokers in an Open Society. Open Society Foundations. [Online]. Available: https://www.opensocietyfoundations.org/reports/data-brokers-open-society

[4] A. Roosendaal, M. van Lieshout, and A. F. van Veenstra, "Personal data markets," Earth, Life Social Sci., Delft, The Netherlands, Tech. Rep. TNO 2014 R11390, 2014. [Online]. Available: https://publications.tno.nl/publication/34612412/riZsP9/TNO-2014-R11390.pdf

[5] S. Spiekermann, R. Böhme, A. Acquisti, and K. L. Hui, "Personal data markets," *Electron. Markets*, vol. 25, no. 2, pp. 91–93, 2015.

[6] W. E. Forum and A. Kearney, "Rethinking personal data: A new lens for strengthening trust," World Econ. Forum, Cologny, Switzerland, Tech. Rep., 2014. [Online]. Available: https://www.weforum.org/reports/rethinking-personal-data

[7] J. Rose, A. Lawrence, and E. Baltassis, "Bridging the trust gap in personal data," Boston Consulting Group, Boston, MA, USA, Tech. Rep., 2018.

[8] S. Schudy and V. Utikal, "'You must not know about me'—On the willingness to share personal data," *J. Econ. Behav. Org.*, vol. 141, pp. 1–13, Sep. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167268117301580

[9] W. Bizon and A. Poszewiecki, "The willingness to trade privacy in the context of WTA and WTP," *Int. J. Trade, Econ. Finance*, vol. 7, no. 4, pp. 121–124, 2016.

[10] *Privacy Security a Connected Life: A Study US*, Trend Micro Ponemon Inst., 2015.

[11] G. G. Survey, "Willingness to share personal data in exchange for benefits or rewards," Gesellschaft Konsumforschung, Tech. Rep., 2017. [Online]. Available: https://www.gfk.com/fileadmin/user_upload/country_one_pager/NL/images/Global-GfK_onderzoek_-_delen_van_persoonlijke_data.pdf

[12] J. Grossklags and A. Acquisti, "When 25 Cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information," in *Proc. 6th Workshop Econ. Inf. Secur. (WEIS)*, 2007, pp. 1–22.

[13] V. Benndorf and H.-T. Normann, "The willingness to sell personal data," *Scandin. J. Econ.*, vol. 120, no. 4, pp. 1260–1278, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/sjoe.12247

[14] S.-A. Elvy, "Paying for privacy and the personal data economy," *Columbia Law Rev.*, vol. 117, no. 6, pp. 1369–1459, 2017. [Online]. Available: http://www.jstor.org/stable/44392955

[15] D. Kifer, "Privacy and the price of data," in *Proc. 30th Annu. ACM/IEEE Symp. Logic Comput. Sci.*, Jul. 2015, p. 16.

[16] S. Spiekermann and J. Korunovska, "Towards a value theory for personal data," *J. Inf. Technol.*, vol. 32, no. 1, pp. 62–84, Mar. 2017. doi: 10.1057/jit.2016.4.

[17] L. Rittenberg and T. Tregarthen, *Principles of Economics*. FlatWorld, 2014.

[18] G. Malgieri and B. Custers, "Pricing privacy—The right to know the value of your personal data," *Comput. Law Secur. Rev.*, vol. 34, no. 2, pp. 289–303, 2018.

[19] 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. 15132–15154, 2018.

[20] 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)," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[21] Y. Zhao, Y. Yu, Y. Li, G. Han, and X. Du, "Machine learning based privacy-preserving fair data trading in big data market," *Inf. Sci.*, vol. 478, pp. 449–460, Apr. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025518309174

[22] J. Yu, M. H. Cheung, J. Huang, and H. V. Poor, "Mobile data trading: Behavioral economics analysis and algorithm design," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 4, pp. 994–1005, Apr. 2017.

[23] A. E. Al-Fagih, F. M. Al-Turjman, W. M. Alsalih, and H. S. Hassanein, "A priced public sensing framework for heterogeneous IoT architectures," *IEEE Trans. Emerg. Topics Comput.*, vol. 1, no. 1, pp. 133–147, Jun. 2013.

[24] B. Jang, S. Park, J. Lee, and S. G. Hahn, "Three hierarchical levels of big-data market model over multiple data sources for Internet of Things," *IEEE Access*, vol. 6, pp. 31269–31280, 2018.

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

[26] Y. Hui, Z. Su, and S. Guo, "Utility based data computing scheme to provide sensing service in Internet of Things," *IEEE Trans. Emerg. Topics Comput.*, to be published.

[27] J. Rose, C. Barton, R. Souza, and J. Platt, "The trust advantage: How to win with big data," Boston Consulting Group, Boston, MA, USA, Tech. Rep., 2013.

[28] X.-B. Li and S. Raghunathan, "Pricing and disseminating customer data with privacy awareness," *Decis. Support Syst.*, vol. 59, pp. 63–73, Mar. 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167923613002534

[29] V. Gkatzelis, C. Aperjis, and B. A. Huberman, "Pricing private data," *Electron. Markets*, vol. 25, no. 2, pp. 109–123, Jun. 2015.

[30] Y. Shen, B. Guo, Y. Shen, X. Duan, X. Dong, and H. Zhang, "A pricing model for big personal data," *Tsinghua Sci. Technol.*, vol. 21, no. 5, pp. 482–490, 2016.

[31] L. Xu, C. Jiang, Y. Qian, Y. Zhao, J. Li, and Y. Ren, "Dynamic privacy pricing: A multi-armed bandit approach with time-variant rewards," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 2, pp. 271–285, Feb. 2017.

[32] X. Wang, X. Wei, Y. Liu, and S. Gao, "On pricing approximate queries," *Inf. Sci.*, vol. 453, pp. 198–215, Jul. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025518302901

[33] J. Parra-Arnau, "Optimized, direct sale of privacy in personal data marketplaces," *Inf. Sci.*, vol. 424, pp. 354–384, Jan. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025517310022

[34] Z. Su, Q. Qi, Q. Xu, S. Guo, and X. Wang, "Incentive scheme for cyber physical social systems based on user behaviors," *IEEE Trans. Emerg. Topics Comput.*, to be published.

[35] N. Askham *et al.*, "The six primary dimensions for data quality assessment," DAMA UK Working Group, Tech. Rep., 2013.

[36] C. Batini, M. Palmonari, and G. Viscusi, *Opening the Closed World: A Survey of Information Quality Research in the Wild*. Cham, Switzerland: Springer, 2014. doi: 10.1007/978-3-319-07121-3_4.

[37] *Unlocking the Value of Personal Data: From Collection to Usage*, World Econ. Forum, Cologny, Switzerland, 2013.

[38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[39] H. Sain. (2017). *Kaggle*. [Online]. Available: https://www.kaggle.com/moose9200/data-sms

[40] P. Hacker and B. Petkova, "Reining in the big promise of big data: Transparency, inequality, and new regulatory frontiers," *Northwestern J. Technol. Intellectual Property*, vol. 15, no. 1, pp. 1–47, 2017.

[41] C.-L. Yeh, "Pursuing consumer empowerment in the age of big data: A comprehensive regulatory framework for data brokers," *Telecommun. Policy*, vol. 42, no. 4, pp. 282–292, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0308596117304743

[42] M. Oostveen and K. Irion, "The golden age of personal data: How to regulate an enabling fundamental right?" in *Personal Data in Competition, Consumer Protection and Intellectual Property Law* (MPI Studies on Intellectual Property and Competition Law), vol. 28. 2018, pp. 7–26.

[43] V. Janeček, "Ownership of personal data in the Internet of Things," *Comput. Law Secur. Rev.*, vol. 34, no. 5, pp. 1039–1052, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0267364918300487

[44] N. Duch-Brown, B. Martens, and F. Mueller-Langer, "The economics of ownership, access and trade in digital data," JRC, Digit. Economy, Tokyo, Japan, Working Paper 2017-01, 2017. [Online]. Available: https://ssrn.com/abstract=2914144

[45] H. J. Pandit, P. Petkov, D. O'Sullivan, and D. Lewis, "Investigating conditional data value under GDPR," in *Proc. 14th Int. Conf. Semantic Syst. (SEMANTiCS)*, 2018, pp. 1–5.

[46] A. Katwala. (2018). *Forget Facebook, Mysterious Data Brokers are Facing GDPR Trouble*. [Online]. Available: https://www.wired.co.uk/article/gdpr-acxiom-experian-privacy-international-data-brokers

[47] E. Kokoris-Kogias *et al.*, "Calypso: Auditable sharing of private data over blockchains," IACR Cryptol. ePrint Arch., Tech. Rep. 2018/209, 2018. [Online]. Available: https://eprint.iacr.org/2018/209

HYEONTAEK OH (S'14) received the B.S. degree in computer science and the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), in 2012 and 2014, respectively, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering. His research interests include trust in ICT environments, personal data ecosystems, the Internet of Things (IoT), and web technologies. He has actively participated in several national funded research projects for ICT environment as a Research Assistant. He has also been contributing to the International Telecommunication Union Telecommunication Standardization Sector Study Group 13/20 as a Contributor and an Editor, since 2015.

SANGDON PARK (S'16–M'17) received the B.S., M.S., and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2011, 2013, and 2017, respectively, where he is currently the Brain Plus 21 Postdoctoral Researcher with the Information and Electronics Research Institute. His current research interest includes 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. He has contributed several articles to the International Telecommunication Union Telecommunication (ITU-T). He received the Best Student Paper Award at the 11th International Conference on Queueing Theory and Network Applications, in 2016.

GYU MYOUNG LEE (S'02–M'07–SM'12) received the B.S. degree from Hongik University, Seoul, South Korea, in 1999, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2000 and 2007, respectively. He was a Research Professor with KAIST and a Guest Researcher with the National Institute of Standards and Technology (NIST), USA, in 2007. He has been an Adjunct Professor with the KAIST Institute for IT Convergence, South Korea, since 2012. Until 2012, he had been invited to work with the Electronics and Telecommunications Research Institute (ETRI), South Korea. He has been a Reader with Liverpool John Moores University (LJMU), U.K., since 2014. Prior to joining LJMU, he was with the Institut Mines-Télécom, Télécom SudParis, France. His research interests include the 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, oneM2M, and so on. In ITU-T, he currently serves as a WP chair in SG13, the Rapporteur of Q16/13 and Q4/20, as well as the Chair of FG-DPM.

HWANJO HEO received the B.S. degree in electrical engineering from Korea University, Seoul, South Korea, in 2004, and the M.S. degree in computer science from Purdue University, West Lafayette, IN, USA, in 2009. He is currently pursuing the Ph.D. degree with the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea. He is currently a Senior Researcher with ETRI. His research interests include network security, network measurement, and distributed systems.

JUN KYUN CHOI (M'88–SM'00) received the B.Sc. (Eng.) degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1982, and the M.Sc. (Eng.) and Ph.D. degrees in electronics engineering from the Korea Advanced Institute of Science and Technology (KAIST), in 1985 and 1988, respectively. From 1986 to 1997, he was with the Electronics and Telecommunication Research Institute (ETRI). In 1998, he joined the Information and Communications University (ICU), Daejeon, South Korea, as a Professor. In 2009, he was moved to KAIST as Professor. He is currently an Executive Member of The Institute of Electronics Engineers of Korea (IEEK), the Editor Board of Member of the Korea Information Processing Society (KIPS), and a Life Member of the Korea Institute of Communication Science (KICS).