IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 25, 2021, accepted August 10, 2021, date of publication August 16, 2021, date of current version August 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3104870

Deposit Decision Model for Data Brokers in Distributed Personal Data Markets Using Blockchain

HYEONTAEK OH[1], (Member, IEEE), SANGDON PARK[2], (Member, IEEE), JUN KYUN CHOI[2], (Senior Member, IEEE), AND SUNGKEE NOH[3]

[1]Institute for Information Technology Convergence, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
[2]School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
[3]Blockchain Research Section, Electronics and Telecommunications Research Institute, Daejeon 34129, Republic of Korea

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

This work was supported in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant through Korean Government [Ministry of Science and ICT (MSIT)] (Development of 5G-IoT Trustworthy AI-Data Commons Framework, 50%) under Grant 2020-0-00048, and in part by the Electronics and Telecommunications Research Institute (ETRI) Grant through Korean Government
(Core Technology Research on Trust Data Connectome, 50%) under Grant 21ZR1330.

ABSTRACT Since personal data becomes a valuable asset for IT industries, the necessity of data ecosystems with personal data trading methodologies is continuously increased, and the concept of data brokers is now widely used in the market. Moreover, utilizing the advantages of blockchain, distributed personal data markets have recently been considered because all participants behave selfishly for their profits in the markets. Since the participants may behave maliciously, it is necessary to prevent such behavior, and one of the considered approaches is putting additional deposits as a penalty into blockchain smart contracts. Since data brokers give various advantages to maintain personal data trading markets, many studies have considered both data brokers and blockchain at the same time. However, those studies have mainly focused on putting deposits into data buyers/sellers but not data brokers. However, since data brokers are also major players in the market, it is necessary to consider deposit decision models for data brokers. Hence, this paper proposes a deposit decision model for data brokers in distributed personal data markets using blockchain. Particularly, this paper proposes a profit model with deposits depending on their behavior for handling contracts and a credit level model that puts fewer deposits for a data broker with a higher credit level to motivate the data brokers' truthful behavior. With the analysis of the proposed models, this paper shows that the models are feasible to motivate data brokers' truthful behavior by allocating deposits for not only a large enough penalty but also a fair enough incentive.

INDEX TERMS Personal data market, data broker, blockchain, deposit.

I. INTRODUCTION

With emerging the importance of data, data-driven approaches are widely applied to both online and offline businesses by utilizing cutting edge technologies (including big data analysis, machine learning, artificial intelligence, etc.) [1]. As data-driven services and applications take the lead of both online and offline businesses, big data and the business analytic market size is projected to reach 512 Billion USD by 2026 with about 14% compound annual growth rate according to *Valuates Reports* [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Yilun Shang.

Since data has become valuable assets for IT industries, the necessity of data ecosystems with data trading methodologies is continuously increased. With these backgrounds, many reports investigated various big data ecosystems driven by data brokers [3]–[6]. Particularly, personal data draw huge attention from the data brokers because personal data contains highly valuable information from the business perspective. According to the reports, a personal data market typically consists of three major actors: i) data brokers, ii) data providers, and iii) data consumers. Data brokers collect (or buy) personal data from data providers, manage/analyze collected data if necessary, and sell personal data to data buyers.

Data brokers earn revenues from brokerage fees and/or any value-added services regarding data analysis and data management, and these revenues motivate data brokers to maintain personal data trading markets.

To motivate data brokers for performing truthful actions (e.g., fulfilling personal data trading contracts), two approaches have mainly been considered: i) providing additional incentive as compensation for truthful actions and ii) charging additional deposit as a penalty for malicious actions. Since the former approach requires trustworthy data brokers with various approaches measuring trust [7], [8] (i.e., the data brokers can receive additional incentives only if they behave truthfully), it has been considered in the centralized personal data brokering markets [9]–[12]. The latter approach considers a malicious data broker that may act the malicious behavior (from the other participants' point of view) depending on its profit in the distributed personal data brokering markets. It means that some systematic methods are necessary to prevent the malicious behavior of the data broker in the market. Particularly, it is mainly considered in distributed personal data markets with blockchain technology because all participants have chances to selfishly behave in the markets [13]–[21]. Since markets with incentive-motivated data brokers are widely applied in the real world, many studies about personal data trading markets consider data brokers as intermediaries [9]–[12]. Data brokers intermediate data trading transactions with the collected information about data buyers and data sellers; therefore, the market participants can utilize the advantages of data brokers (such as data discovery, data pricing, data quality managing, etc.). However, it requires (centralized) trustworthy data brokers that do not perform any malicious actions in the market.

For overcoming the issues about the trustworthiness of (centralized) data brokers, the other approach assumes that data brokers can selfishly behave to maximize their profit (i.e., they do not always perform truthful actions). Therefore, many studies have recently considered distributed data trading market with the emergence of blockchain technologies [13]–[21]. These works proposed distributed personal data trading markets by utilizing the advantages of the blockchain (i.e., transparency, immutability, etc.) [22]. Particularly, in these models, all contracts are written as a smart contract in the blockchain, so all participants (i.e., data buyers, data brokers, data sellers) check the validity of transactions at any time. Moreover, the development of deposit-based smart contracts makes it possible to enforce penalties for malicious actions of the participants by putting a certain amount of deposits in the contracts [23]–[25], which can form incentive-compatible personal data trading models. In other words, the participants put some amounts of deposit into the blockchain before making transactions to compensate for the results of any malicious actions.

For data trading markets with blockchain, not only purely peer-to-peer based models [13]–[15] (i.e., models without any intermediaries or agents) but also broker-based models [16]–[21] has been proposed. Particularly, for personal data trading markets, since one data buyer wants to buy personal data from multiple data sellers [10], [12]. In addition, many management issues should be considered due to the characteristics of personal data (such as consent management [26], privacy and security compliance [27], data price decision [28], etc.).

Therefore, data broker (i.e., agent or intermediary) based models are more suitable to form personal data trading markets, and many studies have considered distributed personal data trading market with data brokers while utilizing blockchain and its smart contracts. de la Vega *et al.* [16] proposed a peer-to-peer distributed data trading model with an IoT context broker to operate blockchain as ledger management and intermediates the participants. Schlarb *et al.* [17] proposed a distributed data market model with a market platform (i.e., data broker) to manage the market participants and various metadata to operate data trading transactions. Chen *et al.* [19] proposed a blockchain-based data trading approach for the Internet of Vehicles. The authors proposed a data broker as a manager for pricing and contracting data trading transactions. Liu *et al.* [20] proposed a blockchain-enhanced IoT data market with the assist of cloud computing to maintain blockchain networks. The authors considered a blockchain network as a market operator (market-agency or data broker) to form a data trading market and proposed smart contracts and pricing mechanisms for a data seller and a data buyer.

Although the studies mentioned above considered data brokers in distributed personal data trading market, the studies concentrated on blockchain based architectures without any considerations of deposit mechanisms for the market participants. Some studies considered distributed personal data trading market with participants' deposit model for ensuring honest behavior. Xiong and Xiong [18] proposed a smart contract-based data trading model that utilizes data hosting/aggregation centers (i.e., data brokering agents) to meditate transactions between data owners and data purchasers. This paper considered deposits models for the data owners and purchasers to enforce honest behavior. An *et al.* [21] proposed a reverse auction-based crowdsensed data trading model supported by blockchain. The study also proposed an agent to manage the proposed reverse auction model between data sellers and data buyers with deposits to compensate for any malicious actions during the transaction.

Even if the studies have utilized (deposit-based) smart contracts to establish distributed personal data markets, the studies have little considerations about putting deposits of data brokers (i.e., agents or intermediaries) and the decision mechanisms for the deposits during the market operation. Since a data broker is also one of the participants in the distributed personal data trading market, it is necessary to consider a deposit model for data brokers in the market. Hence, this paper proposes a deposit decision model with data brokers for blockchain-based distributed personal data markets.

The detailed contributions of this paper are summarized as follows.

- This paper considers a distributed personal data market model with four major players: i) a data buyer, ii) a data broker, iii) data sellers, and iv) a blockchain. The data buyers and data sellers register their information to the data broker, and the activities are recorded in the blockchain. With the considered market model, this paper also proposes transaction flows among players in three major phases: i) data curation phase, ii) data brokering phase, iii) data trading phase. Particularly, this paper mainly focuses on the data trading phase that makes personal data trading transaction contracts for the data buyer and data sellers through the data broker with proper data brokering deposits.
- By proposing a data broker's profit model, this paper describes the possible actions of the data broker with a deposit. For motivating data brokers' truthful behavior, this paper proposes a credit rating model that is tightly related to the amount of deposit for data brokering. Specifically, this paper proposes a credit rating model that contains a finite order in positive and negative directions of credits. Data brokers' credit level either increases or decreases one step according to their behavior. Moreover, to cover the proposed step-wise credit rating model, this paper also proposes a deposit factor model that decides the required amount of deposit for data brokers with different credit levels.
- With the proposed models, this paper shows that the proposed deposit and credit level model can motivate data brokers to behave truthfully, and it proposes methods to decide a proper amount of deposits for data brokers in the personal data market. This paper also verifies the proposed model through numerical analysis by showing the trends of credit levels and the average profits of data brokers with different characteristics. It is shown that the proposed deposit decision model is feasible to form an incentive-compatible distributed personal data trading model with data brokers by putting enough penalties on data brokers with malicious actions.

The rest of this paper is organized as follows. In Section II, this paper introduces an overall structure of a distributed personal data trading market and identifies major participants and their characteristics. In addition, it presents a general personal data trading procedure among market participants. Under the considered market model, this paper proposes both a deposit decision model and a credit rating model for data brokers. In Section III, it is analyzed that the conditions for validating the proposed deposit decision model, and it is shown that the proposed model is suitable to form incentive-compatible personal data trading models. Based on the analysis, in Section IV, various results are presented to check the validity and feasibility of the proposed model. Particularly, it shows that a data broker that always performs truthful actions can earn more profit than any other data broker that sometimes performs malicious actions. Finally, this paper is concluded in Section VI.

II. SYSTEM MODEL

This section introduces a considered data trading market model. Then, it also describes a major information flow to explain interactions among market participants. With the considered system model, this section also proposes analytical models for data brokers, including a profit function and a credit rating model, to decide a proper deposit decision model.

A. DATA TRADING MARKET MODEL

The considered market consists of four major players as shown in Figure 1: i) a data buyer, ii) a data broker, iii) data sellers, and iv) a blockchain. A data buyer wants to buy personal datasets from the data sellers. The buyer requests relevant information about potential data sellers with its dataset requirements. Data sellers want to sell their personal data (e.g., from a single person to data warehouse). They register information about their personal datasets (including data types, data volumes, or data prices, etc.) and sell their datasets to the buyer if conditions from the buyer satisfy their price requirements.

A data broker acts as an intermediary for personal data trading transactions. It collects and manages information about potential data buyers and sellers for curating each personal data trading request. In addition, the data broker collects and manages information about the market and disseminate the information to the market participants (i.e., data buyers and data sellers) for helping their internal decisions about personal data trading transactions. For example, a data broker manages statistics of a certain data type and provides a market price to the market participants as a guideline. With the statistics, the market participants can decide whether they trade personal datasets or not.

A blockchain system manages, verifies, and executes personal data trading contracts among data buyers, brokers, and sellers. During a personal data trading transaction, each market participant also puts deposit into the blockchain for

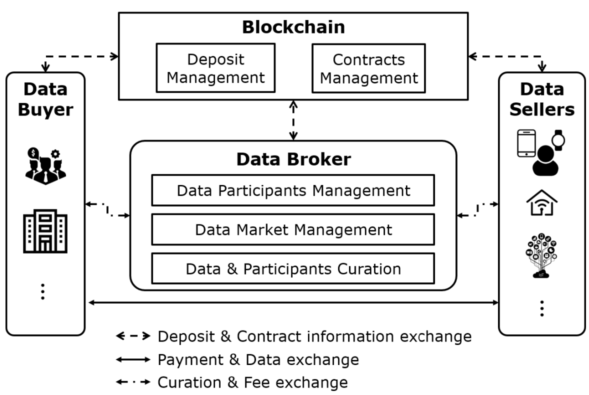


FIGURE 1. A considered distributed data trading model.

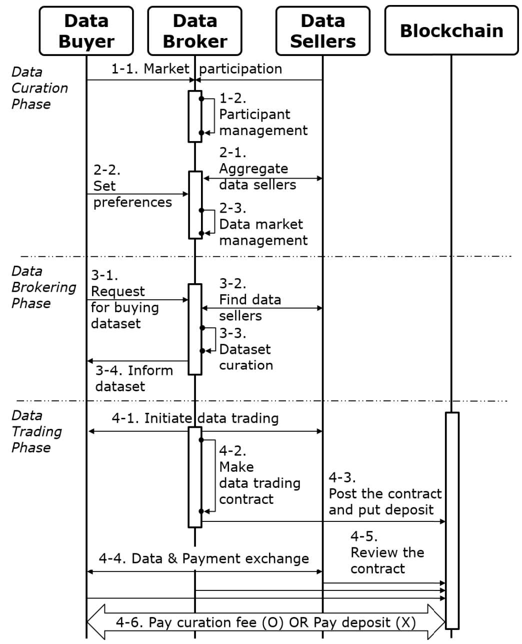


FIGURE 2. The proposed transaction flow.

TABLE 1. Major symbols.

Symbol	Definition
P	Payment from a data buyer
U_R	Profit of a data broker
D_R	Total deposit of a data broker
D_{Base}	Baseline of deposit
V	Private valuation of a data broker
γ	Interest rates
Σ_O, Σ_X	Set of truthful and malicious actions
σ_O	Any truthful actions (i.e., $\sigma_O \in \Sigma_O$)
σ_X	Any malicious actions (i.e., $\sigma_X \in \Sigma_O$)
σ_R	An action of the data broker R
t	Time indices (one transaction at one time)

ensuring its trust to others. If one participant violates the contract, the blockchain enforces a deposit distribution to the other participants. If the transaction is successfully ended, all deposits are returned to each participant.

As shown in Figure 1, data buyers and data sellers directly trade both personal datasets and payments with the intermediation of the data broker (i.e., payment and data exchange flow). The data broker provides relevant information for personal data trading to the buyers and sellers, which can help their decision and take curation fees from both participants (i.e., curation and fee exchange flow). When data buyers and data sellers make a personal data trading transaction with a data broker, their contracts should be managed in secure and trust manner. Therefore, the blockchain is utilized to ensure the transaction among data buyers, brokers, and sellers. During the transaction, each participant puts deposit into the blockchain system to prevent any malicious activities (i.e., deposit and contract information exchange flow).

A detailed flow is shown in Figure 2. The flow consists of three major phases: i) data curation phase, ii) data brokering phase, and iii) data trading phase. In the data curation phase, data buyers and data sellers participate in personal data trading market by registering their information to the data broker. Then, the data broker manages each participant for further data trading transactions (e.g., categorizing participants as data buyers and data sellers, etc.). Particularly, the data broker aggregates information about the data buyers and data sellers for establishing a personal data market. For example, the data sellers notify the amount and the type of datasets, and the data buyers set their dataset preferences.

In the data brokering phase, a data buyer initiates the request for buying a certain personal dataset to the data broker. Then, according to the request, the data broker finds the potential data sellers from the market, and it curates a personal dataset for the data buyer (by checking consent, privacy compliance, data price, etc.). After the data curation, the data broker informs the curation result to the data buyer. With the curation result, the data buyer decides whether it buys dataset or not.

When the data buyer decides to buy dataset, the data broker initiates the data trading phase for realizing a personal data trading transaction between the data buyer and the data sellers. With the agreed information, the data broker creates a new transaction contract and posts the contract into the blockchain. While posting the contract, the data broker also put deposit for assuring the results of the transaction brokering. According to the contract in the blockchain, the data buyer and the data sellers perform the personal data trading transaction. Specifically, the data sellers provide their dataset to the data buyer, and the data buyer makes a payment to each data seller. After the data and payment exchange, they (including the data broker) review the contract for validating the transaction. At this time, third party entities, which are not stakeholder of the contract, can also participate to review the contract. Based on the review, the data buyer pays data curation fee to the data broker if the transaction is successfully done. If not, the participants receive the deposit as a compensation of unsuccessful transaction.

Based on the considered market model, this paper mainly considers the data trading phase that each market participant makes personal data trading contracts with deposit. Particularly, this paper focuses on the behavior of the data broker that plays a key role to maintain distributed personal data trading market more scalable and feasible. Note that the behavior of the data buyers and sellers is out of scope of this paper for the simplicity of the analysis.

B. DATA BROKER MODEL

In a distributed personal data trading market, a data broker may (intentionally or unintentionally) perform malicious actions (from the data buyers and the data sellers points of view) when it intermediates personal data trading transactions between the buyers and sellers. Therefore, the data broker has two action choices.

- A data broker R performs a truthful action (i.e., fulfilling contracts) $\sigma_R \in \Sigma_O$, where Σ_O is defined as a set of all actions that evaluate to true from all participants in a transaction;
- A data broker R performs a malicious action (i.e., violating contracts) $\sigma_R \in \Sigma_X$, where Σ_X is defined as a set of all actions that evaluate to false from any participants in a transaction.

Each of the two action choices results in profits for the data broker expressed by U_R (for $\sigma_R \in \Sigma_O$) and $\overline{U}_R$ (for $\sigma_R \in \Sigma_X$), respectively. The profit of an action depends on various parameters as follows (the mathematical symbols used in this paper are listed in Table 1).

- Payment P determines how much a data buyer B has to pay for a personal data trading transaction t to buy datasets from data sellers (i.e., the size of the personal data trading transaction). Note that methods for pricing dataset is not a scope of this paper. Many studies proposed dataset pricing models (e.g., [6], [10], [28], [29]);
- Curation incentive α determines how much the data broker can get rewards for its curation activity for the data buyers;
- Deposit D means the deposits by the participants for the personal data trading transaction. In this paper, D_R indicates the deposit from the data broker R for brokering the transaction;
- Expected future return γD describes the opportunity costs for locking deposit D within the personal data trading transaction that could be used in other purposes to earn interest;
- Valuation V denotes the private preference of the data broker for outcome depending on the personal data trading transaction, which implies private values when the data broker takes a malicious action.

With the parameters, the following represents the profits of a data broker R brokering a personal data trading transaction between data buyers and data sellers.

$$\begin{cases} U_R = \alpha P - \gamma D_R & \text{if } \sigma_R \in \Sigma_O \\ \overline{U}_R = V_R - D_R - \gamma D_R & \text{if } \sigma_R \in \Sigma_X \end{cases} \tag{1}$$

For each data trading transaction, the data broker tries to maximize its profit by choosing an appropriate decision. A profit of the data broker with fulfilling the contract (U_R) is a margin from a payment from a data buyer (P) with a curation incentive (α) for brokering the data trading transaction. Similarly, a profit of the data broker with violating the contract ($\overline{U}_R$) is considered with a privately achievable value V_R as revenue and a deposit D_R as cost (or loss) for brokering the data trading transaction. With the contract violation, the data buyers and data sellers receive compensations based on the deposit from the data broker as $D_{R\to B}$ and $D_{R\to S}$, respectively. Note that the expected loss of interest during the transaction is adopted as γD_R for both cases.

In this paper, by considering equation (1) from the data broker's perspective, violating a contract becomes a rational choice if $\overline{U}_R > U_R$. A data broker chooses to fulfill a contract if the profit of the truthful action is higher than that of the malicious action, i.e., $U_R > \overline{U}_R$. Then, the data broker can be categorized into three types depending on the achievable profits as follows (inspired by [30]).

- Type R_O: the data broker will always perform a truthful action as its profit resulting from a transaction is larger than the valuation of a malicious action, i.e., $V_R < \alpha P + D_R$;
- Type R_X: the data broker will always perform a malicious action as $V_R > \alpha P + D_R$. The deposit D_R is never large enough to prevent a malicious action;
- Type R_Δ: the data broker is undecided in a transaction which decision to take as $V_R \approx \alpha P + D_R$. It performs either a truthful action or a malicious action by considering the profits.

Data brokers categorized as R_X cannot be economically motivated to perform any truthful actions. These brokers will always perform the malicious actions as their private profit from the outcome of the malicious actions (from the other participants' perspective) is much larger than the economic damage due to the loss of their deposit.

Therefore, the proposed model can ensure that economically rational data brokers (type R_Δ) and honest data brokers (type R_O) behave in truthful actions. Then, a personal data trading model is an incentive-compatible if an economically rational data brokers would not choose a malicious action because the profit of a truthful action is always greater than that of a malicious one, i.e., $U_R > \overline{U}_R$.

C. DATA BROKER CREDIT AND DEPOSIT MODELS

For motivating data brokers to behave truthfully, this paper proposes a credit rating model. Based on a credit level of a data broker, the amount of deposit for brokering a personal data trading transaction is decided. A data broker with a higher credit level requires a smaller amount of deposit. A credit is a set of data brokers that provide the same relative deposits for a personal data trading transaction.

Since a credit model should be able to put the same level of penalty for any credit level, this paper proposes a symmetric credit rating model for data brokers. In other words, this paper defines a finite order in positive and negative direction (i.e., the maximum value M for the highest credit level and the minimum $-M$ for the lowest credit level) of credits $\{C_{-M} \prec c \cdots \prec C_{-1} \prec C_0 \prec C_1 \prec \cdots \prec C_M\} \in \mathcal{C}$, where each credit $C_m \in \mathcal{C}$ has a deposit factor $\lambda_m \in \mathbb{R}$. Each credit maps to a deposit level in order, $\mathcal{C} \to \mathcal{D}$; for example, C_3 requires deposit D_3 (in general, C_m requires a deposit D_m where m is a credit level ($-M \leq m \leq M$)). Note that one of the possible ways to satisfy the above condition is mapping a credit level C_m to a real number m. Therefore, this paper adopts this model for achieving credit levels of data brokers.

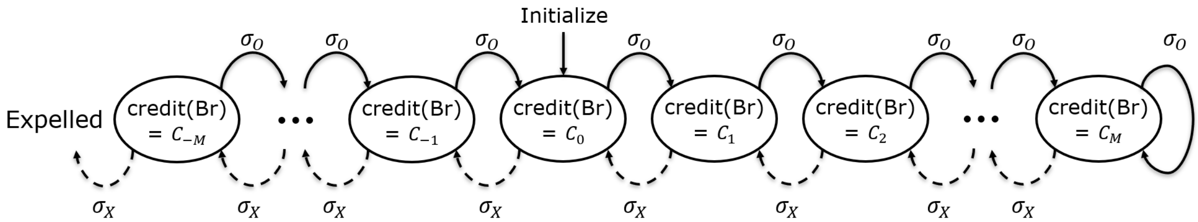


FIGURE 3. The proposed credit rating model.

With the proposed credit rating model, each data broker is only mapped into a single credit level at any point in time. The *credit* function returns the credit level of a data broker that is currently assigned. This paper defines the credit(R) as the current credit of a data broker R. Then, credit(R)-1 is a lower credit with a higher deposit factor, and credit(R)$+1$ is a higher credit with a lower deposit factor. In other words, a data broker with credit(R)-1 needs a larger amount of deposit than that with credit(R), and a data broker with credit(R)-1 needs a smaller amount of deposit than that with credit(R).

To provide the same amount of advantages and penalties, this paper proposes that a credit level of a data broker has increased one step whenever it fulfills a contract and has decreased one step whenever it violates contracts as shown in Figure 3. In addition, a data broker with no record (i.e., a new data broker that participates in the market) is assigned to the initial credit level C_0. If a data broker with the lowest credit level violates a contract again, then the data broker is expelled from the market.

The credit levels are used to calculate the deposit for a data broker that needs to provide to mediate a transaction. This paper defines a baseline deposit D_{base} for a personal data trading transaction and a deposit factor λ_m that are used to calculate D_m, where $C_m \in \mathcal{C}$ (i.e., the required deposit with a credit level m) as shown in the following equations:

$$D_{\text{base}} = \alpha P, \tag{2}$$
$$D_m = D_{\text{base}}\lambda_m. \tag{3}$$

The baseline deposit D_{base} is the original required deposit for brokering personal data trading transaction with the payment αP. Since, at each time, the data broker intermediates different contracts with different amounts of payment, the baseline deposit should also be relatively decided by the amount of payment. Therefore, the major objective of this deposit model is to put incentives to higher credit levels. In other words, based on the credit level, a data broker requires a different amount of deposit even if it handles the same contract.

Deposit factors λ are ordered similarly to the credits and an finite order $\{\lambda_M < \cdots < \lambda_0 < \cdots < \lambda_{-M}\}$ where the factor corresponding to the lowest credit level λ_{-M} is the largest. By reducing the required amount of deposit, data brokers are motivated to truthful behavior because their opportunity costs are kept decreased. With this principle, any deposit factor decision model f with a credit level m can be utilized if and only if it satisfies the following condition.

$$\lambda_m = f(m), \tag{4}$$
$$s.t. f(M) = \lambda_M \text{ and } f(m) < f(n), \quad \forall m > n,$$
$$\times\, where - M \leq m \leq M. \tag{5}$$

Condition (5) indicates the lower bound of the deposit factor (λ_M) and condition for monotonically decrements. In the next section, this paper analyzes how the proposed credit rating model and a corresponding deposit decision function f make an incentive-compatible personal data trading mechanism.

III. NUMERICAL ANALYSIS

The purpose of a deposit is to enforce a penalty when a data broker violates its personal data brokering contract from the perspective of data buyers and data sellers. A penalty for a data broker includes both the immediate deposit loss and the loss of the future expected profit caused by its credit level decrement. This section describes the validity of the proposed credit rating model and the deposit model by showing relevant lemmas and theorems, and it also proposes a method to set deposit factors for data brokers for motivating truthful actions.

A. THEORETICAL ANALYSIS

Firstly, this section shows that the proposed credit rating model works well as designed; in other words, a data broker with a higher credit level has more benefits than that with a lower credit level by the following lemma.

Lemma 1: If a data broker's credit level is the highest, the penalty for violating a contract is the least.

Proof: Consider two data brokers with two different credit levels at time t_0. For example, a data broker R_1 with the highest credit level C_M and a data broker R_2 with any lower credit level C_m ($m < M$). To this end, the loss of future expected profit is checked by a penalty from the contract violation at t_0.

By violating the contract, both data brokers' credit levels become one step lower. In other words, the credit level of R_1 becomes C_{M-1} and that of R_2 becomes C_{m-1}. However, if they fulfill the contract, then the credit level of R_1 becomes C_M and that of R_2 becomes C_{m+1}. That is, R_1 only loses penalty of one credit level (from C_M to C_{M-1}), but R_2 loses penalty of two credit levels (from C_{m+1} to C_{m-1}). It means a required amount of deposit for brokering the future personal data trading transaction becomes different.

In other words, R_2 requires more deposits (from D_{m+1} to D_{m-1}) than R_1 (from D_M to D_{M-1}) for brokering the same personal data trading transaction. It means the expected loss of R_2 becomes a larger than that of R_1. Therefore, a data broker with the highest credit level gets the least penalty in future. □

It is shown for the highest credit level only. Based on this analysis, it is possible to generalize Lemma 1 to data brokers with any other credit level in the market as follows.

Theorem 1: If a data broker's credit level is higher, then a penalty for contract violation is lower.

Proof: Without loss of generality, consider two data brokers R_1 with a credit level C_n $(n < M)$ and R_2 with a credit level C_m $(m < n)$. Similar to Lemma 1, it is checked that the loss of future expected profit caused by the penalty from the contract violation at t_0. If $n = M$ for R_1, it is proved by Lemma 1. Therefore, it only considers the case with $n < M$.

In this case, the difference of the credit level can be compared. In other words, for the contract breach case, the credit level of R_1 becomes C_{n-1} and that of R_2 becomes C_{m-1}. Similarly, for the contract fulfillment case, the credit level of R_1 becomes C_{n+1} and that of R_2 becomes C_{m+1}.

From now on, to analyze the future expected profits of two brokers, it is assumed that the probability for contract fulfillment and violation is the same for both brokers R_1 and R_2. Since the credit level of R_1 is higher than that of R_2 (i.e., $C_n > C_m$), the probability of R_1 for reaching the highest credit level C_M is higher than that of R_2. Then, by Lemma 1, the loss of future expected profit of R_1 caused by the penalty from the contract breach is lower than that of R_2. In other words, R_1 earns more profits than R_2 in future. □

To generalize the idea, it is considered that one data broker with the credit level C_n $(n < M)$ has a chance to intermediate a new personal data trading transaction at time t_0. The data broker chooses either contract fulfillment or violation depending on its profit function. Now, consider two random variables $X_{\sigma_{t_0}\in\Sigma_X}(t)$ and $X_{\sigma_{t_0}\in\Sigma_O}(t)$ to express the profit of the data broker, where σ_{t_0} means an action performed at time t_0. One random variable $X_{\sigma_{t_0}\in\Sigma_X}(t)$ means the profit at time t when the data broker violates the current contract at time t_0. Similarly, the other random variable $X_{\sigma_{t_0}\in\Sigma_O}(t)$ means the profit at time t when the data broker fulfills the current contract at time t_0. Then, it is assumed that the data broker fulfills all contracts from $t_0 + 1$. In this case, it is obvious that

$$\sum_{t=t_0+1}^{\infty} E[X_{\sigma_{t_0}\in\Sigma_X}(t)] < \sum_{t=t_0+1}^{\infty} E[X_{\sigma_{t_0}\in\Sigma_O}(t)].$$

In addition, by Theorem 1, the following can be obtained for setting a deposit for data brokers in the market.

Theorem 2: If a deposit is set to large enough for putting a penalty to a data broker with the highest credit level (i.e., the data broker's future expected profit is less than zero when it violates a single contract), then the data broker cannot violate contracts for earning more profit with any lower credit level.

Proof: By Lemma 1, a data broker with the highest credit level has the least loss of future expected profit. Therefore, if the data broker loses a larger amount of future expected profit (as a penalty), then the data broker will get more penalty when it has a lower credit level. Therefore, if a penalty is large enough to punish the data broker with the highest credit level, then the penalty can enforce that the data broker with any lower credit levels behaves truthfully.

Now, to find the condition where a data broker with the highest credit level violates a contract, it is assumed that the data broker violates the contract at time t_0. The data broker takes such a decision because it can earn a higher profit (i.e., $\overline{U} > U$).

In this case, compared to the case with fulfilling the contract, the data broker loses some future expected profit caused by the decrease of the credit level. If the data broker cannot increase its credit level, it cumulatively loses some profit caused by extra deposits (compared to the case with the highest credit level). That is, if the data broker increases its credit level back to the highest by fulfilling a contract at time $t_0 + 1$, the future expected profit from $t_0 + 1$ will be the same as the contract fulfilling case.

Therefore, if the data broker violates a contract once, the best strategy for the next contract is fulfilling the contract, which means the amount of loss of future expected profit caused by the decrease of the credit level be minimized. Finally, it can be concluded that if a deposit is set to large enough for putting a penalty to a data broker with the highest credit level, then the data broker with any lower credit level cannot violate the contract for earning more profit. □

According to Theorem 2, it can achieve the expected profit of a data broker with the highest credit level C_M for both contract violation $(X_{R,1})$ and fulfillment$(X_{R,2})$ cases as follows by using equation (1),

$$\begin{aligned}
&E[X_{R,1}(t)|\sigma_{R,t_0} \in \Sigma_X, \sigma_{R,t} \in \Sigma_O] \\
&\quad = V_{R,t_0} - D_{R,M,t_0} - \gamma D_{R,M,t_0} \\
&\qquad + \alpha P_{R,t_0+1} - \gamma D_{R,M-1,t_0+1} \\
&\qquad + \sum_{t=t_0+2}^{\infty} (\alpha P_{R,t} - \gamma D_{R,M,t}), \\
&E[X_{R,2}(t)|\sigma_{R,t_0} \in \Sigma_O, \sigma_{R,t} \in \Sigma_O] \\
&\quad = \sum_{t=t_0}^{\infty} (\alpha P_{R,t} - \gamma D_{R,M,t}).
\end{aligned}$$

Since the expected revenue (αP) depends on the size of the contract, at this time, it is assumed that $\alpha = 0$ to get baseline condition. Then, the difference of the expected profit of the data broker is follows.

$$\begin{aligned}
&E[X_{R,2}(t)|\sigma_{R,t_0} \in \Sigma_O, \sigma_{R,t} \in \Sigma_O] \\
&\quad - E[X_{R,1}(t)|\sigma_{R,t_0} \in \Sigma_X, \sigma_{R,t} \in \Sigma_O] \\
&\quad = -V_{R,t_0} + D_{R,M,t_0} - \gamma D_{R,M,t_0+1} + \gamma D_{R,M-1,t_0+1} \geq 0
\end{aligned} \tag{6}$$

Note that if the profits of a data broker R earned by fulfilling and violating a contract are equal (i.e., $E[X_{R,2}(t)|\ \sigma_{R,t_0} \in \Sigma_O, \sigma_{R,t} \in \Sigma_O] - E[X_{R,1}(t)|\sigma_{R,t_0} \in \Sigma_X, \sigma_{R,t} \in \Sigma_O] = 0$), then the data broker chooses the contract fulfillment action because the future expected profit earned by fulfilling the contract is larger (i.e., it can avoid the decrease of the credit level).

Since the expected profits are asymptotically analyzed, condition (6) can be represented as follows,

$$V_R \leq D_{R,M} - \gamma D_{R,M} + \gamma D_{R,M-1}, \tag{7}$$

where V_R is a private valuation obtained when the data broker R violates the contract, $D_{R,M}$ is the required amount of deposit for the data broker R with the credit level M, and γ is an interest rate of the deposit in the system.

B. NUMERICAL APPLICATION

As shown in Theorem 2, any deposit factor function that satisfies the condition (5) can be used to form an incentive-compatible data trading model. To analyze effects of deposit factor, this paper selects a linear-based deposit factor that satisfies condition (5) as follows:

$$f(m) = \lambda_M + \eta \frac{(M-m)}{M}, \tag{8}$$

where η is an increasing rate of the deposit factor, M is the highest credit level in the market, and m is the current credit level of a data broker. Note that λ_M is the lowest deposit factor. The deposit factor λ_m is tightly related to the credit rating model proposed in this paper. Since this paper proposes the step-wise credit rating model (i.e., a credit level is changed one step up or down), a potential penalty for violating a contract should be equal at any credit level; therefore, this paper chooses a linear-based deposit factor model. Note that depending on a credit rating model, a deposit factor may have a different form. For example, *Harz et al. [30]* proposed an exponential-based deposit factor model based on its credit rating model.

With the deposit factor function, by applying equations (3) and (8) to equation (7), the following can be achieved

$$V_R \leq D_{R,\text{base}}(\lambda_M + \frac{\gamma}{M}\eta) \Rightarrow \frac{M(\frac{V_R}{D_{R,\text{base}}} - \lambda_M)}{\gamma} \leq \eta, \tag{9}$$

where $D_{R,\text{base}}$ is the baseline deposit for the data broker R, λ_M is a deposit factor for data brokers with the highest credit level M, and η is a deposit increasing step. Note that η should be larger than zero; therefore, one condition is obtained by the equation as follows,

$$\lambda_M < \frac{V_{avg}}{D_{\text{base},avg}} < 1, \tag{10}$$

where V_{avg} is the average size of payment that all brokers earn in the market ($V_{avg} = \sum_R V_R/R$, and $D_{\text{base},avg} = \sum_R D_{R,\text{base}}/R$) is the average size of deposits that all brokers put while brokering contracts in the market.

Note that it is not possible to assign different deposit factors for every data broker in the market. Therefore, it should be considered the macroscopic information to manage the market (i.e., the average value such as V_{avg} and $D_{\text{base},avg}$) and to provide a common guideline for setting proper deposits. In addition, $\frac{V_{avg}}{D_{\text{base},avg}}$ indicates the amount of deposits in the market is enough to control data brokers for enforcing their truthful behavior by establishing the incentive-compatible data trading model. If $\frac{V_{avg}}{D_{\text{base},avg}} > 1$, it means that, in average, data brokers' profit earned by violating contracts are larger than that earned by fulfilling contracts. Therefore, to stabilize the market, a market operator should set a larger enough amount of deposit to maintain $\frac{V_{avg}}{D_{\text{base},avg}} < 1$.

Equation (9) gives various information to manage the entire personal data trading market from the perspective of the blockchain. Since all transactions are recorded and published in the blockchain, for a distributed personal data trading market, all participants in the market can obtain V_{avg}, $D_{\text{base},avg}$, and γ through analyzing records in the blockchain.

By using equation (9) and (10), the following Algorithm 1 can be obtained. The main purpose of Algorithm 1 dynamically decides the required deposit factors and penalty rates for data brokers with various credit levels. First, the algorithm takes various input parameters regarding market information (V_{avg}, $D_{\text{base},avg}, \gamma, M$). Then, it decides the deposit factor with the highest credit level M by considering the condition (10). Since all relevant information are recorded in the blockchain, all participants can know the baseline deposit factor for participating in the market. After deciding the deposit factor with the highest credit level λ_M, the algorithm decides the increasing factor of the deposit according to the credit level. Then, the algorithm provides two outputs (λ_M and η) for setting deposit guidelines in the market. One of the issues is deciding the proper amount of λ_M and η; therefore, the detailed effect of the λ_M and η is analyzed in the next section.

Then, with Algorithm 2, the data broker with the credit level m can immediately know the amount of deposit for a contract at time t. Algorithm 2 takes various parameters: M is the maximum credit level of the market, and m is the current

Algorithm 1 Deposit Factor and Penalty Decision

Input:
V_{avg}: the average size of the market
$D_{\text{base},avg}$: the average size of the deposit
γ: the interest rate of deposit in the market
M: the highest credit level of the market
Start algorithm:
(1) Decide λ_M where $0 < \lambda_M < \frac{V_{avg}}{D_{\text{base},avg}}$.
(2) Decide η where $\frac{M(\frac{V_{avg}}{D_{\text{base},avg}} - \lambda_M)}{\gamma} \leq \eta$.
Output:
λ_M: the deposit factor for the height credit level M
η: the penalty factor of the deposit

TABLE 2. Simulation parameters.

Symbol	Value
αP	100
γ	0.01 / 365 / 24 / 60 * 10
$V_{avg}/D_{\text{base},avg}$	0.9
M	100
λ_M	$0.9 * V_{avg}/D_{\text{base},avg}$
V	$\alpha P + \{\sim N(\alpha P, 1), \sim N(\alpha P, 3), \sim N(\alpha P, 5)\}$

credit level of the data broker. η is the penalty factor of the deposit decided by Algorithm 1. P is the payment of a data buyer at time t, and α is the rate of curation incentive. Then, using equation (3) and (8), the required deposit for brokering data contract at time t is decided. Note that a baseline deposit $D_{\text{base},t}$ is equal to the amount of curation fee αP which is a 100% of its potential income through contract fulfillment (by equation (2)).

IV. NUMERICAL RESULTS

Based on the analysis for deciding deposit policies to manage incentive-compatible data trading models with data brokers in distributed personal data markets, this section shows various numerical results to check the feasibility of the proposed models. The detailed parameters for the numerical analysis are described in Table 2. The amount of payment only affects the size of the profit, so the total amount of curation incentive is set to 100 for this simulation. γ means an interest rate of the deposit, which is set to 10 minutes (block time for Bitcoin blockchain) with 1% of annual interest rate $(0.01/365/24/60 * 10)$. For the market statistics, $V_{avg}/D_{\text{base},avg}$ is set to 0.9, which is close enough to 1 according to the condition (10).

First, this section checks a basic trend of the proposed models. Particularly, the decision variables from Algorithm 1 (i.e., λ_M and η) are checked. According to the algorithm, the market participants can choose any λ_M that satisfies the condition (10); therefore, the market participants must check the feasible range of λ_M to set the proper amount of deposits for data brokers in the market. Figure 4 shows trends of

Algorithm 2 Deposit Decision for a Contract

Input:
M: the maximum credit level of the market
m: the current credit level of the data broker
α: the curation incentive of the data broker
η: the deposit penalty factor
P_t: the payment of a data buyer at time t
Start algorithm:
(1) $D_{\text{base},t} = \alpha P_t$
(2) $\lambda_m = \lambda_M + \eta \frac{(M-m)}{M}$
(3) $D_{m,t} = D_{\text{base},t} \lambda_m$
Output:
$D_{m,t}$: the required deposit for brokering the contract at time t

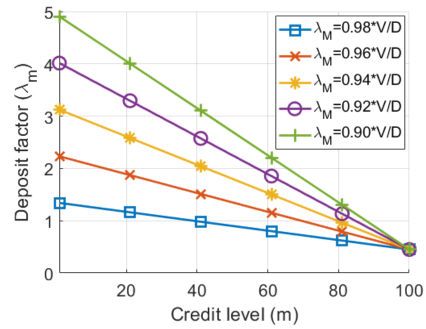


FIGURE 4. The trends of deposit factor λ_m.

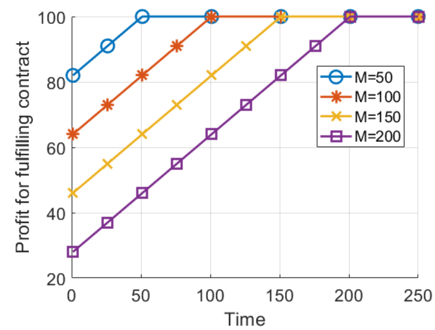
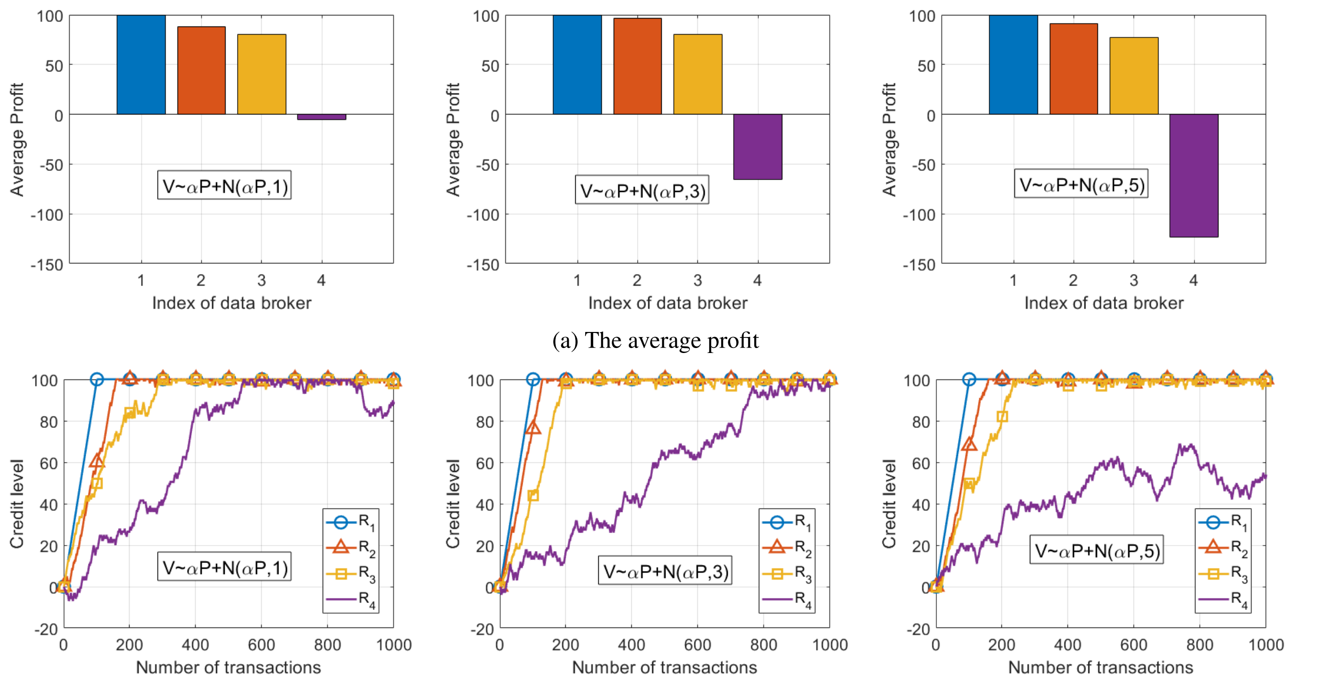


FIGURE 5. The trends of profit of a data broker type R_O.

deposit factor at each credit level λ_m with respect to the deposit factor at the highest credit level λ_M from Algorithm 1. To see the trends for the relationship between the deposit factor (λ_m) and the highest credit level (λ_M), this experiment chooses five different λ_M (i.e., $\{0.90, 0.92, 0.94, 0.96, 0.98\} * V/D$), and the remaining parameters such as γ, M are set as mentioned in Table 2. It shows that if the difference between $V_{avg}/D_{\text{base},avg}$ and λ_M becomes larger, then the deposit penalty factor η also becomes larger according to condition (9). It means that if the deposit factor λ_M is set to lower then the market circumstance ($V_{avg}/D_{\text{base},avg}$), it puts more penalty to data brokers with lower credit levels. Note that to motivate the honest behavior of data brokers in the market, the deposit factor for above certain credit levels should be less than 1.0. For example, a data broker with the highest credit level needs a deposit less than the baseline (i.e., $D_M < D_{base}$). This phenomenon was also represented in [30].

Next, it is checked that the profit of a data broker type R_O (i.e., the data broker always fulfills its contract at any time) concerning various maximum credit levels as shown in Figure 5. This experiment chooses four different maximum credit level (M) values (i.e., $\{50, 100, 150, 200\} \in M$) with the same factors for the deposit (λ_M and η), and the remaining parameters are set as mentioned in Table 2. The trends of profit for the data broker are observed by calculating U_R in equation (1). It is shown that the profit of the data broker reaches the maximum after it reaches the maximum credit level because the required amount of deposit becomes the



(a) The average profit

(b) The trends of credit level

FIGURE 6. The behavior of various data brokers.

minimum. Moreover, for each case, profits of the data broker with lower credit level are all different because factors of deposit λ_m and η are linearly applied to calculate the required amount of deposit D (equation (3) and (8)).

From now on, this section analyzes the behavior of data brokers concerning the deposit factors decided in the market. To check the results, it is assumed that there exist four brokers (from R_1 to R_4) in the market with different privacy valuation values V. This simulation considers three different cases, which the privacy valuation values are set as basic revenue (αP) plus a normal distribution with the average value αP and three different variances 1, 3, 5, respectively (i.e., $V = \alpha P + \{\sim N(\alpha P, 1), \sim N(\alpha P, 3), \sim N(\alpha P, 5)\}$). Note that the proposed model cannot cover the extreme case such as a data broker with the extreme amount of private valuation for violating contracts (as already mentioned in Section II regarding three types of data brokers (i.e., type R_O, R_X, and R_Δ)); therefore, to check the behavior of the proposed model with a proper range, this paper uses a normal distribution with relatively high variance to cover the cases.

First, this section checks the behavior of data brokers when the data brokers violate contracts in the market. To check that the proposed model can put penalties to data brokers with contract violations, the probability of contract violations is set as $(0\%, 15\%, 30\%, 45\%)$ to each data broker (R_1, R_2, R_3, R_4), respectively, regardless of the result of profit decision (U and $\overline{U}$) (i.e., data brokers R_1, R_2, R_3 are categorized as type R_Δ that violate a contract with the given probability). Then, the average profit and the trends of credit levels for each data broker are checked as shown in Figure 6(a) and (b), respectively.

It is checked that a data broker with a higher number of contract fulfillment cases can take more profits in the market with the proposed model. The result of the data broker R_1 is a baseline profit because it always fulfills contracts during the simulation. Depending on the number of contract violations, the profits of data brokers are different. Particularly, for the case with the data broker R_4, it keeps lower credit levels and pays many deposits as penalties for violating contracts. Therefore, it is shown that the proposed model can put enough penalties on data brokers with type R_Δ. Consequently, the proposed model is safe enough (from the perspective of data buyers and sellers) for handling data brokers with an incentive-compatible data trading model in distributed data trading markets.

In addition, it is checked that the proposed deposit decision model is working well for data brokers by forming an incentive-compatible data trading model. By setting $M = 100$, $V_{avg}/D_{\text{base},avg} = 0.9$, and $\lambda_M = 0.9 * V_{avg}/D_{\text{base},avg}$ (obtained by the result from Figure 4), Figure 7 shows the trends of credit level of all four data brokers in the market with the proposed model at time from 0 to $1,000$. At the beginning (at time 0), their credit levels are set to 0 as defined in Figure 3. During the simulation, at each time, each data broker checks profits for both fulfilling contracts (U) and violating contracts (V) and decides its action. Since this simulation already sets proper deposit factors (λ_M and η) based on the proposed model, all data brokers always behave

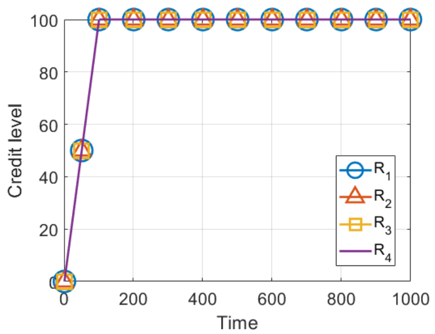


FIGURE 7. The trends of credit level of data brokers type R_O with the proposed deposit decision model.

truthfully by fulfilling contracts. Therefore, at time 100, all data brokers reach the highest credit level and keep the credit level to time 1, 000. It is shown that the proposed deposit decision model can motivate data brokers' truthful behavior.

V. DISCUSSION

This paper has proposed a deposit decision model for data brokers in distributed personal data trading market. In this section, a couple of points are discussed.

- Even though this paper has proposed a credit rating model and a deposit decision model for data brokers, the proposed models can be applied to other participants such as data buyers and data sellers in the personal data trading market. In other words, the proposed models can be used in markets not only with data brokers but also without data brokers (i.e., pure peer-to-peer model). However, as mentioned in the introduction section, the existence of data brokers in distributed personal data trading market is inevitable, and this paper has tackled that the behavior of data brokers should be also systematically managed.
- Since this paper has only focused on the theoretical analysis of the proposed credit rating and deposit decision models for data brokers in distributed personal data markets with blockchain, the detailed off-chain-based data and payment mechanisms have been considered as out of scope. Since off-chain-based (personal) data trading frameworks and platforms proposed similar procedures utilizing smart contracts by putting payment and off-chain data storage addresses in a contract, this model also can adopt such mechanisms for personal data trading. In addition, this paper has a lack of consideration about benchmark analysis that shows the actual performance in the real world. Such implementation and validation of the proposed model will be considered as future work.

VI. CONCLUSION

Distributed personal data markets have now actively studied with the widespread of blockchain technologies. Since data brokers have various advantages to form more scalable and feasible personal data markets, many studies have considered distributed personal data markets that consist of both data buyers/sellers and data brokers. However, the studies have mainly focused on deposit models for data buyers and sellers. Since the data broker is also a participant in the market, this paper has proposed a deposit decision model for data brokers in distributed personal data markets. Data brokers put a certain amount of deposit for brokering personal data trading transactions between data buyers and sellers. With the analysis of the proposed model, this paper has shown the proposed model setting the proper deposit level based on the credit level of each data broker to motivate data brokers' truthful action. For future work, joint analysis with the behavior of data brokers, data buyers, and data sellers should be considered.

ACKNOWLEDGMENT

The authors would like to thank the editors and anonymous reviewers who kindly reviewed this article and provided valuable suggestions and comments.

REFERENCES

[1] UNCTAD. (2019). *Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries*. [Online]. Available: https://unctad.org/webflyer/digital-economy-report-2019
[2] Valuates Reports. (2020). *Big Data and Business Analytics Market Size is Projected to Reach USD 512.04 Billion by 2026*. [Online]. Available: https://www.prnewswire.com/news-releases/big-data-and-business-analytics-market-size-is-projected-to-reach-usd-512-04-billion-by-2026–valuates-reports-301002664.html
[3] E. Ramirez, *Data Brokers: A Call for Transparency and Accountability*. Washington, DC, USA: Federal Trade Commission, 2014.
[4] A. Rieke, H. Yu, D. Robinson, and J. von Hoboken, *Data Brokers in an Open Society*. London, U.K.: Open Society Foundation, 2016.
[5] 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*. Cham, Switzerland: Springer, 2016.
[6] 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.
[7] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Syst.*, vol. 43, no. 2, pp. 618–644, 2007.
[8] J.-H. Cho, K. Chan, and S. Adali, "A survey on trust modeling," *ACM Comput. Surv.*, vol. 48, no. 2, pp. 1–40, Nov. 2015.
[9] 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.
[10] 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. 40120–40132, 2019.
[11] E. Curry, W. Derguech, S. Hasan, C. Kouroupetroglou, and U. U. Hassan, "A real-time linked dataspace for the Internet of Things: Enabling 'pay-as-you-go' data management in smart environments," *Future Gener. Comput. Syst.*, vol. 90, pp. 405–422, Jan. 2019.
[12] H. Oh, S. Park, G. M. Lee, J. K. Choi, and S. Noh, "Competitive data trading model with privacy valuation for multiple stakeholders in IoT data markets," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3623–3639, Apr. 2020.
[13] 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.
[14] S. Zheng, L. Pan, D. Hu, M. Li, and Y. Fan, "A blockchain-based trading platform for big data," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 991–996.
[15] W. Dai, C. Dai, K.-K.-R. Choo, C. Cui, D. Zou, and H. Jin, "SDTE: A secure blockchain-based data trading ecosystem," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 725–737, 2020.

[16] F. de la Vega, J. Soriano, M. Jimenez, and D. Lizcano, "A peer-to-peer architecture for distributed data monetization in fog computing scenarios," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–15, Sep. 2018.

[17] S. Schlarb, R. Karl, R. King, T. J. Lampoltshammer, L. Thurnay, B.-P. Ivanschitz, and V. Mireles, "Using blockchain technology to manage membership and legal contracts in a distributed data market," in *Proc. 6th Int. Conf. Softw. Defined Syst. (SDS)*, Jun. 2019, pp. 272–277.

[18] W. Xiong and L. Xiong, "Smart contract based data trading mode using blockchain and machine learning," *IEEE Access*, vol. 7, pp. 102331–102344, 2019.

[19] C. Chen, J. Wu, H. Lin, W. Chen, and Z. Zheng, "A secure and efficient blockchain-based data trading approach for internet of vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 9110–9121, Sep. 2019.

[20] K. Liu, X. Qiu, W. Chen, X. Chen, and Z. Zheng, "Optimal pricing mechanism for data market in blockchain-enhanced Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9748–9761, Dec. 2019.

[21] B. An, M. Xiao, A. Liu, G. Gao, and H. Zhao, "Truthful crowdsensed data trading based on reverse auction and blockchain," in *Database Systems for Advanced Applications*, G. Li, J. Yang, J. Gama, J. Natwichai, and Y. Tong, Eds. Cham, Switzerland: Springer, 2019, pp. 292–309.

[22] J. Xie, H. Tang, T. Huang, F. R. Yu, R. Xie, J. Liu, and Y. Liu, "A survey of blockchain technology applied to smart cities: Research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2794–2830, 3rd Quart., 2019.

[23] P. Das, L. Eckey, T. Frassetto, D. Gens, K. Hostáková, P. Jauernig, S. Faust, and A.-R. Sadeghi, "FastKitten: Practical smart contracts on bitcoin," in *Proc. 28th USENIX Secur. Symp. (USENIX Security)*, Santa Clara, CA, USA, Aug. 2019, pp. 801–818.

[24] BigchainDB Ocean Protocol Foundation. (2020). *Ocean Protocol: Tools for the Web3 Data Economy*. [Online]. Available: https://oceanprotocol.com/tech-whitepaper.pdf

[25] Y. Li, L. Li, Y. Zhao, N. Guizani, Y. Yu, and X. Du, "Toward decentralized fair data trading based on blockchain," *IEEE Netw.*, vol. 35, no. 1, pp. 304–310, Jan. 2021.

[26] J. Oh, J. Hong, C. Lee, J. J. Lee, S. S. Woo, and K. Lee, "Will EU's GDPR act as an effective enforcer to gain consent?" *IEEE Access*, vol. 9, pp. 79477–79490, 2021.

[27] K. R. Sollins, "IoT big data security and privacy versus innovation," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1628–1635, Apr. 2019.

[28] 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, Apr. 2018.

[29] X.-B. Li and S. Raghunathan, "Pricing and disseminating customer data with privacy awareness," *Decis. Support Syst.*, vol. 59, pp. 63–73, Mar. 2014.

[30] D. Harz, L. Gudgeon, A. Gervais, and W. J. Knottenbelt, "Balance: Dynamic adjustment of cryptocurrency deposits," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*. New York, NY, USA: Association for Computing Machinery, 2019, pp. 1485–1502.

HYEONTAEK OH (Member, IEEE) received the B.S. degree *(summa cum laude)* in computer science from Korea Advanced Institute of Science and Technology (KAIST), in 2012, and the M.S. and Ph.D. degrees in electrical engineering from KAIST, in 2014, and 2020, respectively. He is currently a Team Leader of the Institute for IT Convergence, KAIST. He has actively participated in nationally-funded research projects for ICT environment. He has been contributing the International Telecommunication Union Telecommunication Standardization Sector Study Group 13/20, as a Contributor and an Editor, since 2015. His research interests include trust in ICT environments, personal data ecosystems, the Internet of Things (IoT), and web technologies.

SANGDON PARK (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 2011, 2013, and 2017, respectively. He is currently a Brain Plus 21 Postdoctoral Researcher with the School of Electrical Engineering, KAIST. He has contributed several articles to the International Telecommunication Union Telecommunication (ITU-T). His current research interests include optimizing wireless networks or smart grids, which hold great potential for practical applications to industries. He has focused on processing energy big data via various machine-learning methodologies and optimizing network economics of the edge cloud computing. He received the Best Student Paper Award at the 11th International Conference on Queueing Theory and Network Applications, in 2016.

JUN KYUN CHOI (Senior Member, IEEE) 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 Korea Advanced Institute of Science and Technology (KAIST), in 1985 and 1988, respectively. From June 1986 to December 1997, he was with the Electronics and Telecommunication Research Institute (ETRI). In January 1998, he joined the Information and Communications University (ICU), Daejeon, South Korea, as a Professor. From the year of 2009, he has moved to the School of Electrical Engineering, KAIST, as a Professor. He is also currently an Executive Member of The Institute of Electronics Engineers of Korea (IEEK), an Editorial Board Member of Korea Information Processing Society (KIPS), and a Life Member of Korea Institute of Communication Science (KICS).

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