2019 IEEE International Conference on Big Data (Big Data)

Privacy Bargaining with Fairness: Privacy-Price Negotiation System for Applying Differential Privacy in Data Market Environments

Kangsoo Jung
Dept. Computer Science and Engineering
Sogang University
Seoul, Republic of Korea
azure84@naver.com

Seog Park
Dept. Computer Science and Engineering
Sogang University
Seoul, Republic of Korea
spark@sogang.ac.kr

***Abstract*—Digital data are an essential resource for intelligent decision making. As the value of digital data increases, digital markets, where data owner and consumers can deal with data, have also been attracting attention as a means to obtain data. However, the collection of digital data can lead to privacy breaches, which are a substantial impediment that hinders an individual's willingness to provide data. Differential privacy, which is a de facto standard for privacy protection in statistical databases, can be applied to solve the privacy violation problem. To apply differential privacy to the data market, the amount of noise and corresponding data price must be determined; however, this matter has not yet been studied. In this work, we propose a fair negotiation method that can set the appropriate price and noise parameter in the differentially private data market environment. We suggest a data market framework with a market manager that acts as a broker between the data provider and consumer. We also propose a negotiation technique to determine the data price and noise parameter ε using Rubinstein bargaining considering social welfare to prevent unfair transactions. We validate that the proposed negotiation technique can determine an appropriate level of ε and unit price without unfair trade to either the data provider and the consumer.**

Keywords—privacy, differential privacy, negotiation, data market

I. Introduction

With the growth of digital data volume and the development of data analysis technology, demand for digital data will increase because they are an indispensable resource for product or service improvement. The data market concept that can sell or purchase digital data is designed to meet these requirements. In a data market, a data owner makes a profit by selling data, and a data consumer pays to obtain personal data. In the past, companies called data brokers such as Axim who collect personal information and resell the information were responsible for data distribution. However, as data ownership and right to control becomes important with privacy issues, the data market in which data owners sell their data directly has attracted attention as a channel for personal data acquisition.

However, as demonstrated by the cases of AOL or Netflix, personal data collection and analysis can lead to unintended disclosure of personal information. Particularly in the data market environment, the data providers are individuals and the consumers are corporations or government organizations. Hence, personal data can be abused easily because of power imbalance. This factor hinders an individual's voluntary participation in data trading. Therefore, implementing appropriate privacy protection techniques is an essential requirement for the data market environment.

Differential privacy, which is the existing de facto standard for privacy protection, is a mathematical model that can address the privacy violation problem in statistical databases. Considerable researches has been conducted to apply differential privacy to various fields in the real world.

In this study, we propose a pricing mechanism that considers fair data trading between the data provider and the data consumer in differentially private data. The proposed pricing mechanism is performed by a market manager that mediates between the data provider and the consumer. Through negotiation, the ε unit price and value of ε can be determined to be fair to the data provider and the consumer. The term "ε unit price" refers to the price per value of ε (e.g., 0.1$ per ε value 0.01). The final reward for the data provider is the product of the value of ε and the ε unit price determined by negotiation. The contributions of this study are as follows.

Market-manager-based data market framework: We propose a data market framework that can determine data prices and privacy protection levels fairly through negotiations by market managers.

Privacy–price negotiation based on the Rubinstein bargaining model: We propose a negotiation technique based on Rubinstein bargaining, in which the provider and the consumer who have different privacy and price requirements, can determine the appropriate noise parameter ε's unit price and value of ε. The proposed negotiation algorithm allows the data provider to reflect the risk of a privacy violation to the price and the data consumer can determine the appropriate price for the data, considering the required level of data accuracy and available budget.

Social welfare: The proposed price model determines the final noise parameter ε unit price and the value of ε, considering the social welfare function. Despite applying the proposed negotiation method, an individual data provider concludes a disadvantageous trade, because data consumers are generally corporations or governments. This property is a long-term disincentive for individuals to provide their data. To prevent this disincentive, the final data price and value of ε are determined by considering the social welfare function within the data consumer's budget. The social welfare function is a concept used to assess how desirable behavior in terms of social welfare. In the data market environment, we regard the fairness of the entire participant's benefit from a transaction as social welfare. Pricing based on the social welfare function

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE
978-1-7281-0858-2/19/$31.00 ©2019 IEEE

promotes an individual's participation in data trading by compensating for the unfairness of data trading.

II. RELATED WORKS

Differential privacy, which is the existing de facto standard for privacy protection, can satisfy the requirement by which users are restricted from obtaining additional information from the database. Differential privacy suggests a mathematical model that prevents information exposure and ensures privacy protection at a specified level ε, which is customized by the data owner. Given two neighboring databases D1 and D2, which differ by only one record, the definition and property of differential privacy are as follows:

Definition 1. Differential privacy [1]

A randomized function *K* provides ε-differential privacy if all datasets with D1 and D2 differing by one element only and all S ⊆ Range(K), i.e.,

$$\Pr[K(D1) \in S] \le \exp(\varepsilon) \times \Pr[K(D2) \in S] \quad \dots (1)$$

Differential privacy inserts random noise into the real output before returning the results to the user. According to the definition, the ε value affects the amount of added noise. Privacy protection is enhanced as ε decreases. Conversely, the degree of privacy protection decreases as ε increases.

Arguments on the proper value of the noise parameter ε have been raised since the introduction of the concept of differential privacy. These arguments continue because no criteria to determine the value of ε exist. Accordingly, this issue has undermined the claim that personal information can be protected by differential privacy. Disputes have surrounded the setting of the proper value of noise parameter ε since the presentation of differential privacy [15-18]. According to the definition of differential privacy, the amount of noise is determined by the sensitivity and the noise parameter ε. The higher the sensitivity and the smaller the ε, the more noise is inserted. A value of ε between 0.01 and 2 has been determined arbitrarily by domain experts considering the lack of a criterion for determining an appropriate value of ε [1].

To solve this problem, many studies have been conducted to set an appropriate level of ε [3-9]. The work of [9] proposes a data market mechanism that analyst pay individuals for the use of their data, show this mechanism would benefit both analysts and individuals. In the previous researches, a data pricing mechanism set the price according to the predefined query type or proceed auction. However, these methods have limitations in determining price only from the data consumer's perspective. [4], which suggested a mechanism to adjust the balance between privacy and cost in data market environment, is the most similar study to the proposed study. However, the proposed pricing mechanism is still at an early stage and needs further research.

As shown in previous studies, no gold standard has been established for determining the value of ε and the price. If the value of ε and the price are determined on the basis of the simple supply-demand or only considering the consumer's position, the data provider who have relatively less information than data consumers has a disadvantageous deal. This unfair pricing mechanism leads to poor participation of data providers in the long run. Therefore, a privacy–price negotiation mechanism that determines a fair ε value and price is needed to activate the data market.

In this study, we attempt to satisfy this requirement by proposing a pricing mechanism considering social welfare functions. This study is the first to consider the requirements of the data provider and the consumer through negotiation in the pricing mechanism with differential privacy.

III. DATA MARKET FRAMEWORK

A. Overview

As described above, the data market consists of a data provider, data consumer, and market managers who operate data markets [Fig. 1].

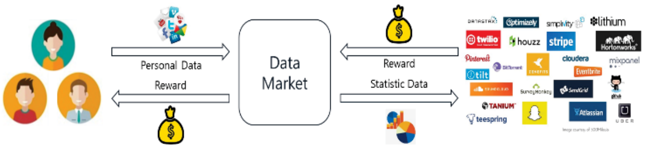


Fig. 1. Data market overview

In this study, we propose a data market framework with differential privacy. In the proposed framework, the market manager matches the data provider and the consumer and mediates the negotiation of ε unit price and value of ε for applying differential privacy.

B. Proposed data market framework

The privacy protection level in differential privacy is determined by the parameter ε; thus, the data price is also affected by the ε. In the proposed differentially private data market framework, the market manager mediates the negotiation between the provider and the consumer to determine the value of ε and the unit price. The proposed data market framework is shown in Fig. 2. The components of this framework are the data provider, data consumer, and market managers.

Data provider: The data provider registers the data type, the upper bound of ε, and the required unit price of the ε with the market manager.

Data consumer: The data consumer registers the data type they need, the lower bound of required ε and ε unit price, the desired number of data providers, and available budget with the market manager. After matching is completed with the data provider, the unit price and ε value are determined through negotiations with the data provider.

Market manager: The market manager is considered a trusted-but-curious participant and aims to link a data provider with a consumer by performing a match using the registered information. The market manager only handles participant information for matching and negotiation.

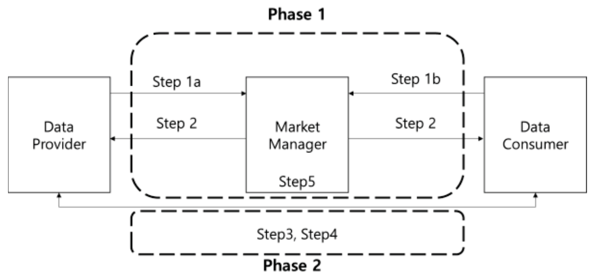


Fig. 2. Proposed data market framework

In the proposed data market framework, data trading is conducted through two phases. In Phase 1, a provider and a consumer are matched using their requirements. In Phase 2, the matched data provider and the consumer negotiate to determine the final ε value and unit price. The reason the trading process is performed in two phases is that all the data providers and consumers spend too much time when negotiation is performed on a one-to-one basis. The data trading process is as follows.

Registration

Step 1a: The data provider registers the data type, the upper bound of ε, and the required ε unit price with the market manager.

Step 1b: The data consumer registers the data type they need, the lower bound of required ε, the required ε unit price, the desired number of data providers, and the available budget with the market manager.

Matching Phase

Step 2: The market manager performs the matching based on the ε unit price reported by the provider and consumer and then notifies the provider and the consumer of the matching result.

Negotiation Phase

Step 3: The matched data provider and consumer negotiate with the market manager to determine the ε unit price and the ε value.

Step 4: After the final ε value and unit price are determined, the data provider adds the noise corresponding to the determined ε and sends it to the consumer, and the consumer pays the corresponding price to the provider.

Step 5: The market manager records the loss and profit amount of the provider and the consumer during negotiation to reflect future trading.

To maintain the continuity of data trading, the market manager forces the trade to be performed at the consumer's minimum requirement of ε and ε unit price determined by the negotiation. Afterward, the consumer can obtain the additional ε value above the minimum requirement considering the social welfare function.

IV. Bargaining-based Two-phase Negotiation

On the data market, data providers and consumers have different requirements. For example, a provider with a high privacy sensitivity has a low ε upper bound and a high ε unit pricee. Data consumers also have different ε values and unit prices as needed. Therefore, for data trading, an agreement must be made on noise insertion level and the unit price.

In this study, we propose a negotiation technique between the provider and the consumer to overcome the above shortcomings and to establish an agreement among the participants. The proposed negotiation technique is a trust-but-curious market manager that performs mediation and management roles for matching and negotiation to prevent additional privacy violations. The market manager adjusts the disadvantages that may arise in the negotiation process in consideration of the social welfare function.

A. Phase1: Matching

As mentioned above, matching is performed prior to the negotiation, considering the requirements of the provider and consumer. Each participant creates a set of candidates to be negotiated. This matching can solve the time delay problem due to one-to-one negotiation.

1) Stable matchings: In the data market environment, the number of data providers tend to be overwhelmingly larger than the number of data consumers. A data consumer must be matched to the desired number of data providers to obtain the required quantity of data. Thus, the matching (Phase 1) is in the form of a many-to-one matching, in which multiple providers are matched with one data consumer. The market manager converts the many-to-one matching form into a one-to-one matching, as follows, for convenience.

When a data consumer informs the market manager of the desired number of data providers at initial registration, the market manager creates virtual data consumers that are as numerous as the consumer's desired number of data providers. Then, one-to-one matching is performed between the virtual data consumers and the data providers. That is, the data consumer is $C_i=\{C_1,...,C_n\}$, and each data consumer C_i has the desired number of data providers PN_i. The market manager creates a virtual consumer $VC_i=\{VC_1,\cdots,VC_m\}(m=\sum_{i=1}^{n}C_i\times PN_i)$ by copying the data consumer C_i for one-to-one matching.

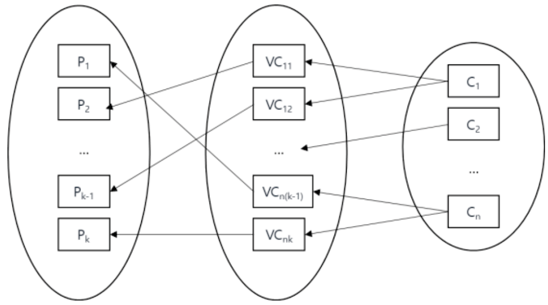


Fig. 3. Data provider–consumer one-to-one matching

We use the stable matching scheme proposed in [13] for matching, but other matching techniques can be used. For the matching technique proposed in [13], the market manager calculates each participant's preference using the ε unit price registered by the provider and the consumer for matching.

Definition 2. Matching preference

We have the data provider P_i ($P_i\in\{P_1,\cdots,P_n\}$, n=number of provider) and the consumer VC_j ($VC_j\in\{VC_1,\cdots,VC_m\},m=\sum_{i=1}^{n}C_i\times PN_i$) . The proposed ε unit prices of P_i and VC_j are defined as $price_{p,i}$ and $price_{c,j}$ ($1\le i\le n$, $1\le j\le m$). The preference operator $x<_i y$ means that participants i prefers participants y to x. At this time, the preference of P_i is calculated as follows: If $price_{c,j}>price_{c,k}$, the preference of P_i for VC_j , VC_k is $VC_j>_{P_i}VC_k$. In other words, P_i prefers VC_j to VC_k , which offers a more favorable price compared with VC_k . Meanwhile, if $price_{p,k}>price_{p,j}$, the preference of VC_j for P_i , P_k is $P_k<_{VC_j}P_i$ since the consumer prefers a provider who offer a lower prices.

B. Phase2: Negotiation

After the data provider and consumer are matched through first-phase matching, the final ε value and unit price should be determined by negotiation. Negotiation is divided into two stages. First, Rubinstein-bargaining-based negotiations are performed to determine the unit price for the minimum level of ε required by the consumer. Then, we determine the additional ε value above the minimum required ε and unit price for additional ε value considering social welfare. The reason for dividing the negotiation into two stages is to allow a consumer to obtain the minimum ε value through the first negotiation, while the second negotiation is intended to alleviate the unfairness that the trading is overly profitable for the consumer considering social welfare.

We introduce the Rubinstein-bargaining-based negotiation technique, which prevents participants from wasting unnecessary resources and time by repeating proposals. If the unit price determined by the negotiation is lower/higher than the required price, participants make a loss even though they report prices honestly. This loss eliminates the motivation to report their privacy price honestly. To prevent this occurrence, the market manager records the loss and profit of the participant as a credit and then compensates the loss by reflecting them in the negotiation process. We also describe how the market manager uses credits to offset participants' loss in the long-term trade. Finally, we shows that unfairness occurs when the proposed negotiation technique is applied to the trading, and we propose the second negotiation technique considering social welfare to solve this unfairness.

1) Negotiation for ε unit price: Data consumer C_j negotiates with data providers that belong to the matched set. Negotiation aims at determining the ε unit price. In this study, we weigh the requirements of both sides to determine the unit price as follows

$$price_{i,j} = \omega \times \varepsilon_{price\ p,i} + (1-\omega) \times \varepsilon_{price\ c,j}\ . \quad \ldots(2)$$

We apply Rubinstein bargaining as follows to determine reasonable weights while considering the requirements of both sides:

$$\omega = \frac{1-\delta_{c,j}}{1-\delta_{p,i}\delta_{c,j}} \quad \ldots(3)$$

$$1-\omega = \frac{\delta_{c,j}(1-\delta_{p,i})}{1-\delta_{p,i}\delta_{c,j}} \quad \ldots(4)$$

To obtain the weight value ω from these equations, the discount factor of the data provider $\delta_{p,i}$ and consumer $\delta_{c,j}$ must be determined. In our negotiation problem, the data provider and the consumer have a different requirement, which affects the discount factor. First, in terms of the data provider, a provider's privacy sensitivity is related to the discount factor. A provider who has high privacy sensitivity do not want to provide their data at low prices. Meanwhile, the consumer sets a higher unit price as the data necessity increases. Therefore, we can calculate privacy sensitivity and data necessity through the price suggested by the participant as follows:

In the case of the provider's privacy sensitivity, a high unit price means high sensitivity. Thus, the provider's privacy sensitivity θ can be calculated as follows:

$$\theta_i = \frac{rank_{last} - rank(\varepsilon_{price\ p,i}) + 1}{rank_{last} - rank_{first}} \quad \ldots(5)$$

In the equation, $rank_{first}$ is the highest ranking of all data providers, value is 1, and $rank_{last}$ is the lowest ranking value, which is the number of all data providers. $rank(\varepsilon_{price\ p,i})$ is the ranking value of the required price of P_i among all data providers.

The higher the unit price of ε of P_i, the higher the privacy sensitivity. The higher the sensitivity of privacy, the less motivated the data provider is to discount the price; thus, the discount factor $\delta_{p,i}$ is determined as θ_i/r in proportion to privacy sensitivity, where r is the normalization parameter for adjusting the discount factor difference among the providers.

In the case of consumer's data necessity sensitivity, a high unit price means high data necessity. Thus, the consumer's data necessity γ can be calculated as follows. In the equation, $rank(\varepsilon_{price\ c,j})$ means the ranking value of the required price of C_j among all data consumers.

$$\gamma_j = \frac{rank_{last} - rank(\varepsilon_{price\ c,j}) + 1}{rank_{last} - rank_{first}} \quad \ldots(6)$$

The higher the data necessity, the more motivation the data consumer has to discount the price; thus, the discount factor $\delta_{c,j}$ is determined by $(1-\gamma_j)/r$.

2) Compensation for losses due to negotiations: Depending on the results of the negotiations described in the previous section, participants make profits or losses. If the consumer is matched with a provider who proposes a greater unit price, then the consumer makes a loss; otherwise, the consumer makes profit. This loss or profit would undermine the proposed incentive mechanism for the truthful report of the participant's required unit price. To solve this problem, the market manager records this loss or profit as credit and reflects the accumulated credit for future negotiations to offset the profit and loss in the long term.

The provider makes a loss when his/her required unit price is greater than the negotiation price but makes profit when the required unit price is less than the negotiation price. At this time, the P_i's profit $Profit_{p,i}$ is calculated as follows:

$$Profit_{p,i} = \left(price_{i,j} - price(\varepsilon_{price\ p,i})\right) \times \varepsilon_{\min\ c,j} + \left(social\ price_{i,j} - price(\varepsilon_{price\ p,i})\right) \times additional\ \varepsilon_{i,j} \quad \ldots(7)$$

In Equation (12), $\left(price_{i,j} - price(\varepsilon_{price\ p,i})\right)$ is the profit in the first round of negotiations with Rubinstein bargaining, and $\left(social\ price_{i,j} - price(\varepsilon_{price\ p,i})\right) \times additional\ \varepsilon_{i,j}$ is the profit in pricing considering social welfare.

Moreover, the C_j's profit $Profit_{c,j}$ is calculated as follows:

$$Profit_{c,j} = \left(price(\varepsilon_{price\ c,j}) - price_{i,j}\right) \times \varepsilon_{\min\ c,j} + \left(price(\varepsilon_{price\ c,j}) - social\ price_{i,j}\right) \times additional\ \varepsilon_{i,j} \quad \ldots(8)$$

The credit recorded by the market manager is the cumulative value of the loss and profit obtained from the trade performed by the participant. P_i and C_j's credit are calculated as follows; t represents the number of trades the participant has participated in.

$$credit_{p,i} = -\sum_{k=1}^{t} Profit_{p,i} \quad \ldots(9)$$

$$credit_{c,j} = -\sum_{k=1}^{t} Profit_{c,j} \quad \ldots(10)$$

The market manager performs the negotiation by reflecting the credit of each participant to the discount factor of each participant. The discount factor considering the credit is calculated as follows:

$$\delta_{p,i} = \frac{\theta_i}{r} + \frac{credit_{p,i}}{\max(credit_p)} \quad \ldots(11)$$

$$\delta_{c,j} = \frac{(1-\gamma_j)}{r} + \frac{credit_{c,j}}{\max(credit_c)} \quad \ldots(12)$$

$\max(credit_p)$ is the maximum value of all the provider's credit, and $\max(credit_c)$ is the maximum value of all the consumer's credits. That is, a participant having accumulated credit greater than 0 may reduce the loss by proceeding to the next negotiation at a higher price than his/her original negotiation price. If the credit is less than 0, the trading proceeds at an unfavorable position in the next negotiation. Thus, as trades are repeated, the profits and losses from the negotiations are offset, and the participants remain motivated to report their required unit prices honestly.

3) Determinig unir price and ε value considering social welfare: The consumer can choose a more favorable provider to proceed with the trade because the number of providers is much larger than the consumer. Such a profit imbalance has a negative impact on the long-term maintenance of the data market, and it acts as a motivation for the provider to stop participating in the data trade. We propose a technique to increase the additory ε value above the minimum $\varepsilon_{\min c}$ and determine the unit price social $price_{i,j}$ for additory ε value considering social welfare to solve the profit imbalance.

First, P_i will agree to provide the additional ε values only if *social price$_{i,j}$* is greater than *price(price$_{p,i}$)*. In this case, the profit of C_j considering social welfare is as follows:

$$social\ profit_{c,j} = \sum_{i=1}^{PN_j} (1-\beta) \times \left(\frac{price(price_{c,j}) - social\ price_{i,j}}{\max(price_{c,j} - price_{i,j})} \right) \times \varepsilon_{i,j} + \prod_{i=1}^{PN_j} \beta \times \left(\frac{credit_{p,i} - (social\ price_{i,j} - price(price_{p,i})) \times additional\ \varepsilon_{i,j} + |\min(credit_p)| + 1}{\max(credit_{p,i} - (social\ price_{i,j} - price(price_{p,i}))\times additional\ \varepsilon_{i,j} + |\min(credit_p)| + 1)} \right)$$
…(13)

price(price$_{c,j}$)-social price$_{i,j}$ means the C_j's profit by the data trade and *credit$_{p,i}$-(social price$_{i,j}$-price(price$_{p,i}$))* × *additional $\varepsilon_{i,j}$+min(credit$_i$)+1* is the social welfare of the provider who deals with C_j. *min(credit$_p$)* is the minimum value of the provider's credit. The reason for adding one is to make sure that the value is not zero or negative when multiplying each credit. As the credit of the provider who trades with C_j becomes more equal, the value of social welfare increases. The parameter β is a weight that determines the degree of reflection on the profits obtained through the data trade and social welfare function. The larger the β, the greater the ratio of social welfare to the *social profit$_{c,j}$* of C_j. The parameter β is determined according to the credit value C_j, as shown in Equation (14). *min(credit$_c$)* denotes the minimum value among the credits of the consumer.

$$\beta = \frac{credit_{c,j} + \min(credit_c) + 1}{\max(credit_c) + \min(credit_c) + 1} \quad \ldots(14)$$

The buyer has to determine the optimal *social price$_{i,j}$* and *additional* $\varepsilon_{i,j}$ to maximize *social profit$_{c,j}$* within the budget given to him/her, It is expressed as follows:

OPT *social profit$_{c,j}$=Max(social profit$_{c,j}$)*
s.t *social price$_{i,j}$-price(price$_{p,i}$)>0,*
$\varepsilon_{\min c,j}$+*additional* $\varepsilon_{i,j} < \varepsilon_{\max p,i}$,
$$\sum_{i=1}^{PN_j} \left((price_{i,j} \times \varepsilon_{\min\ c,j}) + (social\ price_{i,j} \times additional\ \varepsilon_{i,j}) \right) < budget_j$$

Finding the combination of *social price$_{i,j}$* and *additional* $\varepsilon_{i,j}$ to obtain ***OPT*** *social profit$_{c,j}$* takes considerable computation time; hence, we calculate this using the genetic algorithm [14]. experimental results

V. Expereimental results

A. Experimental environment

To verify the proposed method, we perform the following experiments:

(1) Fairness of pricing

(2) Number of providers who succeeded in the transaction

The experimental variables are set as follows: the $\varepsilon_{max\,p,i}$ is selected randomly in the range of 1–2 in 0.1 unit, and the $\varepsilon_{min\ c,j}$ is randomly selected in the range of 0.1–1 in 0.1 unit. The $\varepsilon_{price\ p,i}$ and $\varepsilon_{price\ c,j}$ is set to 0.1 unit in the range of 1–10 in proportion to the $\varepsilon_{min\ c,j}$ in the case of the consumer, and in inverse proportion to $\varepsilon_{max\,p,i}$. The PN_j value of data consumer j is randomly selected from 10 to 20. The *budget$_j$* is assigned to the value of the highest ε value with the highest unit price for the PN_j. Finally, the number of consumers is set at 10, and the number of providers is set at 400. We use the balanced pricing mechanism proposed in [4] for comparison with the proposed scheme.

B. Fairness of pricing with proposed technique

The proposed method considers the negotiation based on Rubinstein bargaining and social welfare to determine the ε unit price and the value of ε. We designed an experiment to show the proposed technique adjusts the profit of both sides equally. In addition to the control group [4], we also perform experiments on Rubinstein-bargaining-based negotiation, credit application, and social welfare to compare the results. The parameter value used in the experiments are maintained with the default settings. The experimental result is the cumulative profit of the provider and consumer at the time of 100 transactions.

Table I shows the maximum/minimum/median values of provider and consumer profit for each technique. As shown in Table I, the difference in profits between the provider and the consumer is considerably large when only negotiation is applied. When credit is applied, the difference is reduced compared with the case in which only negotiation is applied. However, the provider still makes a loss in the trade compared with the provider, because as mentioned above, there is a larger number of providers than consumers, and the consumers can proceed in a more advantageous position by dealing with multiple providers. When social welfare is applied, the consumer still makes a favorable transaction compared with the provider, however, the difference between the provider and the consumer is considerably reduced.

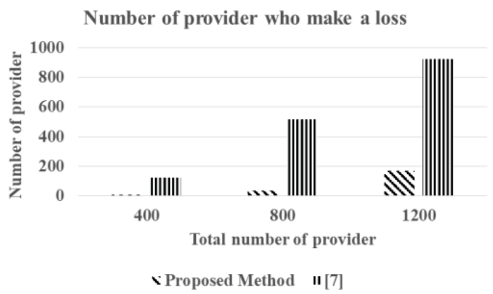
TABLE I Max/Min/Median of provider and consumer credit

		Max	Min	Median
Negotiation	provider	2033	−6530.2	160.5
	consumer	116,245.5	46,040.5	76070.8
Credit	provider	2420.2	-6457.2	222.11
	consumer	112,376.2	53,120.8	82,066.5
Social Welfare	provider	3976.8	-809	2028.3
	consumer	50,064.9	6281.8	26451.4
Balanced Price	provider	4548.8	-6206.4	2378.2
	consumer	117,012	56,269	84,468.8

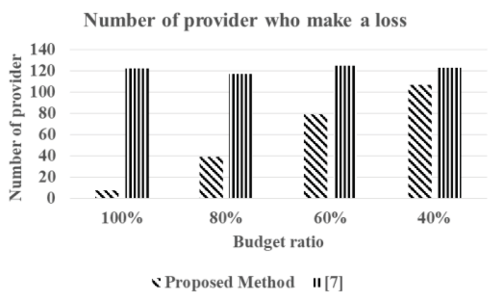
minimize the number of providers leaving the transaction in consideration of social welfare if sufficient budget is allowed.

C. *Number of providers who makes a loss with proposed technique*

The existing pricing technique determines the price considering only the consumer's position. Thus, the existing technique does not consider the case in which the consumer cannot obtain the necessarsy number of providers as the provider refuse to provide their data when they make a loss in transaction. We demonstrate through the experiment that the proposed technique can prevent an unfair transaction, thereby ensuring that the consumer can obtain the necessary number of providers. We assume a provider who makes a loss in transaction will not be participating in future data. We perform the transaction 100 times and compare the number of provider who makes a loss in transactions with the proposed technique and balanced price technique [4].



(a)



(b)

Fig. 4. Number of providers that make a loss after transaction 100 times

Fig. 4 (a) shows the results of experiments conducted by changing the total number of provider to 400, 800, and 1200 when the number of consumer is 20. The y-axis means that the number of provider who make a loss after transactions 100 times. As shown in the experiments, as the number of provider increases, the number of providers who make a loss increases. However, in the proposed technique considering social welfare, only 170 out of 1200 provider make a loss, whereas 923 provider make a loss in [4]. Fig. 4 (b) shows that when the total number of providers is fixed at 400, the consumer's available budget changes to 100%, 80%, 60%, and 40% from maximum budget*(the number of sellers required by each data buyer × the highest unit price × the largest ε value)*. The y-axis means that the number of provider who make a loss after transactions 100 times. As shown in the results of experiment (b), the lesser the consumer's budget is, the lesser capability to share the consumer's profit considering social welfare. In other words, experiments confirm the proposed technique can

VI. CONCLUSION

In this study, we propose a data market framework with a market manager to apply differential privacy in a data market environment and a Rubinstein-bargaining-based negotiation technique to determine the appropriate ε unit price and ε value. The most important contribution of this study is the proposed pricing model that reflects the position of provider, which is different from the existing method in which price is determined considering only the usefulness and cost of the consumer. Our future work aims to conduct negotiations without disclosing the participant's required ε unit price and ε value to the market manager for enhanced privacy protection. Query sensitivity according to the query types not considered in the proposed method will be considered in the negotiation.

ACKNOWLEDGMENT

This work was supported by Institute for Information& Communications Technology Promotion (IITP) grant funded by the Korea Government (MSIT) (No. 2017-0-00498, A Development of De-Identification Technique Based on Differential privacy).

REFERENCES

[1] C. Dwork, Cynthia, A. Roth. "The algorithmic foundations of differential privacy" Foundations and Trends® in Theoretical Computer Science Vol. 9. No.3–4 ,pp. 211-407, 2014

[2] J. Tang, A, et al, "Privacy loss in Apple's implementation of differential privacy on macOS 10.12", arXiv preprint arXiv:1709.02753, pp.1-12, 2017.

[3] J. Lee, C. Clifton, "How much is enough? Choosing Epsilon for Differential Privacy", Proceedings of the International Conference on Information Security, pp.325–340, 2011.

[4] R. Nget, Y. Cao,M. Yoshikawa, "How to balance privacy and money through pricing mechanism in personal data market", arXiv preprint arXiv:1705.02982, pp.1-10, 2018.

[5] J. Hsu, et al, "Differential privacy: An economic method for choosing epsilon", Proceedings of the 27th IEEE Computer Security Foundations Symposium, pp.1–29, 2014.

[6] A. Ghosh, A. Roth, "Selling privacy at auction", Games and Economic Behavior, Vol. 91, No.1, pp.334-346, 2015.

[7] A. Roth, "Buying private data at auction: the sensitive surveyor's problem", ACM SIGecom Exchanges, Vol.11, No.1, pp.1-8, 2012.

[8] L. K. Fleischer, Y. H. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data", Proceedings of the 13th ACM Conference on Electronic Commerce, pp.568-585, 2012.

[9] C. Aperjis, B. A. Huberman, "A market for unbiased private data: Paying individuals according to their privacy attitudes", Available at SSRN: https://ssrn.com/abstract=2046861, pp.1-17, 2012.

[10] H. Oh et al., "Personal data trading scheme for data brokers in IoT data marketplaces", IEEE ACCESS, Vol.7, 2019, pp.40120-40132, 2019.

[11] C. Li, D. Li, Y. D, G. Miklau, D. Suciu, "A theory of pricing private data", ACM Transactions on Database Systems, Vol.39, No.4, pp.34-60, 2013

[12] W. Li, C. Zhang, Z. Liu, Y. Tanaka, "Incentive mechanism design for crowdsourcing-based indoor localization", IEEE Access, Vol. 6, pp.54042-54051, 2018

[13] I. Giannakopoulos,et al, "An equitable solution to the stable marriage problem", Proceedings of the 27th International Conference on Tools with Artificial Intelligence, pp.989-996, 2015.

[14] J. McCall, J "Genetic algorithms for modelling and optimization", "Journal of Computational and Applied Mathematics", Vol. 184, No. 1, pp.205-222, 2015.

[15] Z. Jorgensen, T. Yu, G. Cormode, G, "Conservative or liberal? personalized differential privacy", Proceedings of the International Confefence on Data Engineering, pp.1023-1034, 2015