PRIVATA: Differentially Private Data Market Framework using Negotiation-based Pricing Mechanism

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

Junkyu Lee
Computer Science and Engineering
Sogang University
Seoul Republic of Korea
ljk7776@sogang.ac.kr

Kunyoung Park
Computer Science and Engineering
Sogang University
Seoul Republic of Korea
pky3145@gmail.com

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

ABSTRACT

As the value of digital data increases, the data market is in the spotlight as a means of obtaining a personal information. However, the collection of personal information makes a serious privacy violation and it is a serious problem in the use of personal 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 should be determined between the provider and consumer. However, this matter has not yet been studied. In this work, we introduce a Privata which is a differentially private data market framework to set the appropriate price and noise parameter in the data market environment. The Privata is based on negotiation technique using Rubinstein bargaining considering social welfare to prevent unfair transactions. We explain the Privata overview and negotiation technique in Privata, and show the Privata implementation.

CCS CONCEPTS

• Security and privacy → Security services → Pseudonymity, anonymity and untraceability

KEYWORDS

Privacy, Differential privacy, Negotiation, Data market

ACM Reference format:
Kangsoo Jung, Junkyu Lee, Kunyoung Park, & Seog Park. 2019. PRIVATA: Differentially Private Data Market Framework using Negotiation-based Pricing Mechanism. In *Proceedings of ACM CIKM conference (CIKM'19), November 3–7, 2019, Beijing, China.* ACM. NY, NY, USA 4 pages. https://doi.org/10.1145/3357384.3357855

1 INTRODUCTION

With the growth of digital data volume and the development of data analysis technology, digital data are increasing constantly in value, 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. Data brokers, such as Axim, who collect personal information and resell the information, have emerged; however, as data ownership and 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. Datacoup is a prototype service that operates data market services. Through the data market, the data provider can obtain additional revenue from selling data and data consumers can obtain personal data for analysis.

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 often 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. In recent years, considerable research has been conducted to apply differential privacy to various fields in the real world [2-3]. [4] conducted a survey of user attitudes on privacy and data trading and summarized the key principles for data market.

In this demonstration, we introduce a Privata which is a differentially private data market framework which proposes a pricing mechanism that considers fair data trading between the data provider and the data consumer in differentially private manner. 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 ε value can be determined to be fair to the data provider and the consumer.

2 PRIVATA FRAMEWORK

3.1 Overview

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.

CIKM '19, November 3–7, 2019, Beijing, China.

ACM ISBN 978-1-4503-6976-3/19/11...$15.00.
DOI: https://doi.org/10.1145/3357384.3357855

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.,

$$\Pr[K(D1) \in S] \leq \exp(\varepsilon) \times \Pr[K(D2) \in S]$$

The privacy protection level in differential privacy is determined by the parameter ε; thus, the data price is also affected by the ε. That is, if the value of ε is small, and the amount of noise insertion increases, the value of the data also decreases. If the value of ε increases, the value of the data increases.

In the proposed framework, the market manager mediates the negotiation of ε unit price and ε value for applying differential privacy. The term "ε unit price" refers to the price per ε value (e.g., 0.1$ per ε value 0.01) and the final reward for the data provider is the product of the ε value and the ε unit price determined by negotiation. As described above, the proposed data market consists of a data provider, data consumer, and market managers who operate data markets [Figure 1].

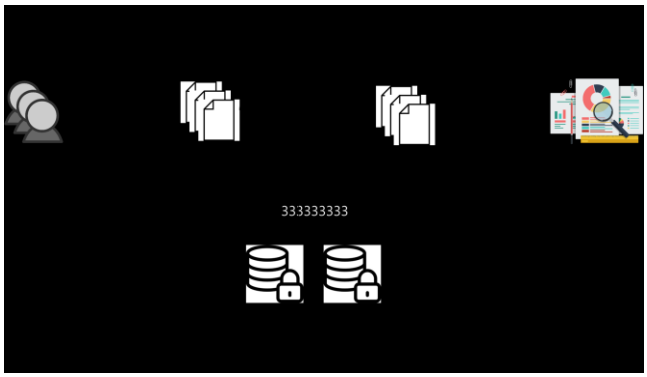

Figure 1: Privata framework overview

Data consumer: The data consumer registers the question they need, the required ε lower bound $\varepsilon_{min\_c,j}$ and unit price $\varepsilon_{price\_c,j}$, the desired number of data providers PN_j, and available budget $budget_j$ with the market manager. After question registration is completed, the unit price and ε value are determined through negotiations with the data provider.

Data provider: The data provider chooses a question from the data consumer that they can answer. They send their answer, the ε upper bound $\varepsilon_{max\_p,i}$, and the required unit price $\varepsilon_{price\_p,i}$ with the market manager. Then, once a negotiation is made with the data consumer, the data provider decides whether to provide their information at the negotiated price and ε value.

Market manager: The market manager is considered as a trusted-but-curious participant and aims to link a data provider with a consumer. The market manager only handles participant information which is complete information that enables Rubinstein bargaining for negotiation and management. In addition, the market manager encourages long-term trading participation by reflecting losses incurred in unfair trading. The data trading process is as follows.

Phase1 Registration

Step 1a: The data consumer registers the question they need, the lower bound of required ε $\varepsilon_{min\_c,j}$, the required ε unit price $\varepsilon_{price\_c,j}$, the desired number of data providers PN_j, and the available budget $budget_j$ with the market manager.

Step 1b: When the Market Manager notifies the consumer's question, the data provider chooses a question that they can answer. They send their ε upper bound $\varepsilon_{max\_p,i}$, and the required unit price $\varepsilon_{price\_p,i}$ with the market manager.

Phase2 Negotiation

Step 2: The matched data provider and consumer negotiate with the market manager to determine the ε unit price and the ε value using their $\varepsilon_{max\_p,i}$, $\varepsilon_{price\_p,i}$, $\varepsilon_{min\_c,j}$, $\varepsilon_{price\_c,j}$, and PN_j.

Step 3: After the ε value and unit price are determined using negotiation, the data provider and the consumer determine whether to accept the negotiated price and ε value.

Phase3 Data transaction

Step 4: If the negotiation result is mutually approved, the provider will permute their answer for applying differential privacy and encrypt it with the query-specific symmetric key. And then, the provider encrypts again the encrypted data and the symmetric key with the public key of the matched consumer. After encryption is completed, the provider sends the encrypted data to the market manager.

Step 5: when the matched consumer receives encrypted data from the market manager, the consumer decrypts the encrypted symmetric key with its own private key, and decrypts the encrypted data with a symmetric key. After decyption is completed, the consumer performs aggregation on query results using decrypted answer.

Step 6: The market manager delivers the consumer's payment to the provider for the data.

3.2 Pricing mechanism based on negotiation

After the data provider and consumer are matched, 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. The negotiation process is composed as follows.

First, the proposed Rubinstein-bargaining-based negotiation technique, which prevents participants from wasting unnecessary resources and time by repeating proposals is performed. If the unit price determined by the first 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 disadvantage, 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. Finally, our second negotiation considering social welfare is performed to solve the unfairness. The Algorithm for two-stage negotiation is as follows.

Algorithm 1 Negotiation

Input: Matching pair p_i,c_j

Output: $price_{i,j}$, $social\_price_{i,j}$, $additional\_\varepsilon_{i,j}$

//calculate p_i's privacy sensitivity θ_i

1: $\theta_i = \frac{rank_{last} - rank(\varepsilon_{price\_p,i}) + 1}{rank_{last} - rank_{first}}$

//calculate c_j's data necessity γ_i

2: $\gamma_j = \frac{rank_{last} - rank(\varepsilon_{price\_c,j}) + 1}{rank_{last} - rank_{first}}$

//calculate p_i and c_j's discount factor $\delta_{p,i}$ and $\delta_{c,j}$ *(r is constant)*

3: $\delta_{p,i} = \frac{\theta_i}{r} + \frac{credit_{p,i}}{\max(credit_p)}$

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

//calculate weight ω

5: $\omega = \frac{1-\delta_{c,j}}{1-\delta_{p,i}\delta_{c,j}}$

6: $1-\omega = \frac{\delta_{c,j}(1-\delta_{p,i})}{1-\delta_{p,i}\delta_{c,j}}$

7: $price_{i,j} = \omega \times \varepsilon_{price\_p,i} + (1-\omega) \times \varepsilon_{price\_c,j}$

//find ***OPT*** $social\_profit_c$,

8: $social\_price_{i,j}$, $additional\_\varepsilon_{i,j}$ =**Genetic_algorithm(**p_i,c_j**)**

9: **return** $price_{i,j}$, $social\_price_{i,j}$, $additional\_\varepsilon_{i,j}$

we weigh the requirements of both sides to determine the final unit price as follows:

$$price_{i,j} = \omega \times \varepsilon_{price\_p,i} + (1-\omega) \times \varepsilon_{price\_c,j}$$

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

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

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, the provider's privacy sensitivity θ_i is related to the discount factor $\delta_{p,i}$. 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 line 1 and line 2. In the equation, $rank_{first}$ is the highest ranking and value is 1, and $rank_{last}$ is the lowest ranking value, which is the number of all data providers and consumers. Through the discount factor $\delta_{p,i}$ and $\delta_{c,j}$, we can calculate the unit price $price_{i,j}$ and the ε value is $\varepsilon_{min\_c,j}$ as line 3-7.

Finally, we calculate the $social\_price_{i,j,}$ and $additional\_\varepsilon_{i,j}$. Although we consider both the provider and consumer's requirements to determine the price and ε, the profit gap increases as the trade progresses, and many providers make a loss because the number of providers is much larger than the consumer. Thus, the consumer can choose a more favorable provider to proceed with the trade. We propose a technique to decide the additory ε value above the minimum ε_{min_c} and determine the unit price $social\_price_{i,j}$ for additory ε value considering social welfare to solve the profit imbalance. By this two-phase negotiation, the consumer can obtain the minimum level of ε value within the budget through the first stage and then obtain the additional ε value considering social welfare for the provider in the second stage. 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:

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

$price(price_{c,j})$-$social\_price_{i,j}$ 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. 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 follows.

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

The consumer 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 line 8. 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.

4 DEMONSTRATION

We implemented the Privata framework as a web application. In this section, we describe the registration and the response of questions, negotiation and the transaction in the Privata framework with the screenshot.

The Privata consumer and provider register their email and password first. The consumer generates a question that he/she wants to collect the data with the minimum ε value, the desired price, the required number of provider, and the available budget for the negotiation. There are two types of data types that can be selected: categorical and numeric.

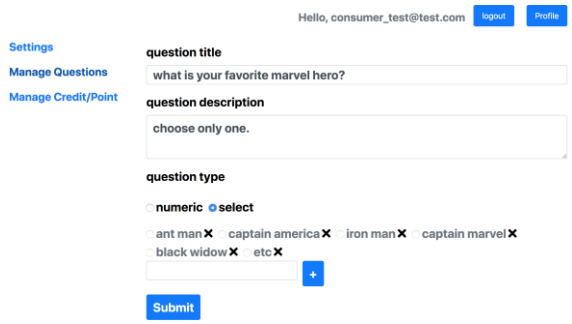

Figure 2: Consumer's question generation

The provider can check the entire registered questions on the main page, and select the question that they want to answer. Provider input the answers to the selected question, the maximum available ε value and the desired price for negotiation.

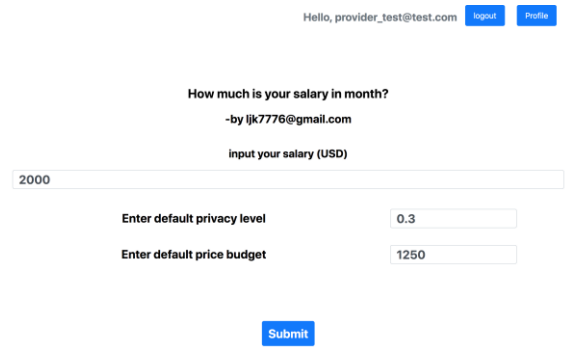

Figure 3: Provider's question response

Once the question response is completed, the market manager calculates the negotiation price and ε value through a negotiation-based pricing mechanism and presents it to the provider and the consumer. The provider and the consumer can confirm the negotiation price and ε to determine the approval/deny on their query management page. If either participant decides to deny the transaction, the transaction is deleted. When a participant decide to approve, the user permutes his/her data and provides it to the market manager with encryption. For permutation, we use the LDP algorithm proposed in [5] for numeric type, and the algorithm given in [6] for categorical type. The permutation and encryption algorithms are implemented in JavaScript and executed in the provider's browser.

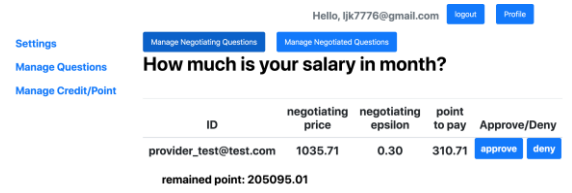

Figure 4: Provider's negotiated question management

The market manager keeps the encrypted data, and sends the encrypted data to the consumer when the user requests the question response. The consumer decrypts the encrypted data, and aggregates it to obtain the desired result.

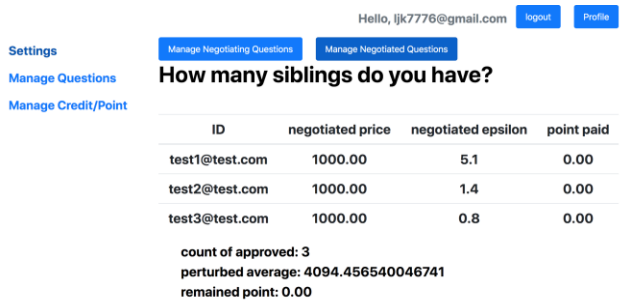

Figure 5: Consumer's question result aggregation

Since the aggregation algorithm is performed on the consumer side, the market manager does not have any information related to the question's response other than the information for the negotiation in the whole process. The market manager manages the entire provider and consumer information and can manage the parameters related to social welfare at an appropriate level.

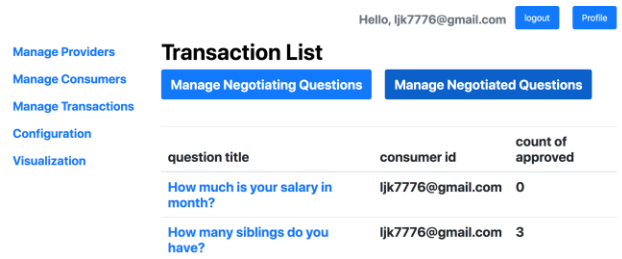

Figure 6: Market manager's management

ACKNOWLEDGMENTS

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

[3] 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.

[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] T. T. Nguyên, et al. "Collecting and analyzing data from smart device users with local differential privacy." arXiv preprint arXiv:1606.05053, pp. 1-11, 2016.

[6] R. Bassily, A. Smith, "Local, private, efficient protocols for succinct histograms", Proceedings of the 47th annual ACM symposium on Theory of computing, pp. 127-135, 2015.