Hindawi
Journal of Electrical and Computer Engineering
Volume 2024, Article ID 7486890, 6 pages
https://doi.org/10.1155/2024/7486890

Research Article

The Formal Analysis on Negative Information Selections for Privacy Protection in Data Publishing

Ping Chen,[1] **Jingjing Hu**,[1] **Zhitao Wu**,[1] **Ruoting Xiong**,[2] **and Wei Ren**[3,4,5]

[1]*China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou, China*
[2]*School of Computing Sciences, University of East Anglia, Norfolk, Norwich NR47TJ, UK*
[3]*State Key Laboratory of Geo-Information Engineering and Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology of MNR, CASM, Beijing, China*
[4]*Key Laboratory of Data Protection and Intelligent Management (Sichuan University), Ministry of Education, Beijing, Sichuan, China*
[5]*School of Computer Science, China University of Geosciences, Wuhan 430078, Hubei, China*

Correspondence should be addressed to Wei Ren; weirencs@cug.edu.cn

Received 11 August 2023; Revised 22 January 2024; Accepted 8 February 2024; Published 29 February 2024

Academic Editor: Susana Ortega-Cisneros

Negative information selection is an approach to protect the privacy by using negative information to replace original information. In this paper, we prove some bounds for negative information selection. Those bounds reveal the privacy protection strength of quantitative probability analysis. We also analyzed the reconstruction probability of original information from available negative information. The formal analysis can specify the bound on the strength of security and utility for negative information selection. Besides, we simulate brute force attacks under different data leakage ratios. Specifically, we calculate the attacker's guess times before and after the data leakage. Experimental results indicate that the data leakage of over 30% can put the original information in a dangerous situation. Furthermore, we found that the leakage possibility has little relevance to the number of elements in the full set, but it is influenced by the ratio of the leaked information.

1. Introduction

Negative information selection is a general method for privacy protection, as negative information is selected to be out instead of original information so as to protect the privacy of the original information [1]. The negative information selection can be widely employed in various scenarios such as data publication with data privacy protection, privacy-aware data mining or machine learning, and federated learning [2, 3].

Supposing the full set is S, which could be a set of words, sentences, or objects. The user information is collected in set A, but the information is stored in the negative format A', which is the subset of $\widetilde{A}$ ($\widetilde{A}$ is the complementary set of A). The relationship of A, A', and S is shown in Figure 1. As shown, users store A locally and only expose A' to the database. Since many systems suffer from the central database attack which causes the data leakage frequently [4], attackers can steal the A' information from database due to the data leakage, and they try to guess A from the A'. Users only pick some elements from the set A, and attackers will launch brute force attacks to guess the exact combination of elements chosen by users (a brute force attack is a hacking method that uses trial and error to crack passwords and encryption keys). There is no doubt that the A' leakage will reduce the guess times of attackers and increase the successful rate of cracking. The major concern is, thus, to what extent the negative information selection can protect the privacy of original information and/or to what extent the negative information can recover the original information [5].

Negative surveys (NSs) were first proposed in 2006 [6], which is a privacy-preserving method for cryptography, anonymity, and in legal guarantees. Take a direct working hours survey:

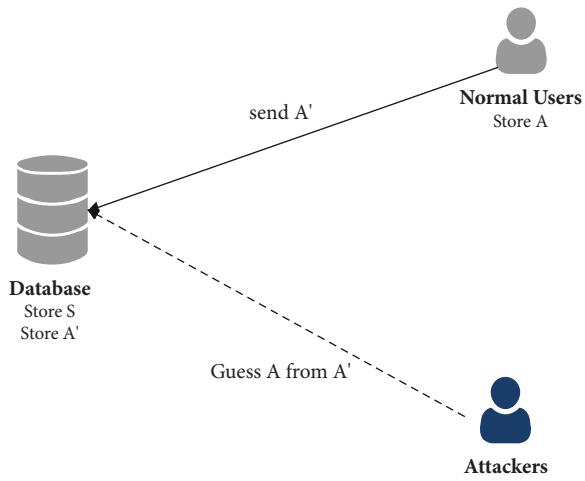

Figure 1: The relation of S, A, and A'.

I work

(i) Less than 3 hours a day

(ii) Between 3 and 6 hours a day

(iii) More than 6 hours a day

In a negative survey, I do not work

(i) Less than 3 hours a day

(ii) Between 3 and 6 hours a day

(iii) More than 6 hours a day

If the positive version of the survey is being answered by an individual working less than 3 hours a day, the first option must be chosen. If the same person is answering the negative version of the survey, one of the last two options should be selected. After selecting a series of answers, the negative information is stored in the database for every individual which is a unique user profile. Once the negative information is exposed to the attackers, they will guess the positive answers from users by brute force attack. The problem thus falls to one point: the probability of revealing or guessing elements of A after viewing on a set that is out of A. As the selection of negative information could be repeated multiple times, the problem will become subtle.

In this paper, we formally analyze above probability. The main contributions of the paper are as follows:

(1) We prove the privacy strength in terms of probability analysis and some key bounds are provided

(2) We also analyze the possible methods and probability of recovery of original information from negative information

(3) We test the probability of original information recovery from negative information with different bound values under a brute force attack

The rest of the paper is organized as follows: Related work is reviewed in Section 2. Section 3 presents the problem formation and analysis. We discuss some bounds in further advanced discussions in Section 4. Section 5 shows the experimental results on the bound of the negative information, and we conclude the paper in Section 6.

2. Related Work

Negative surveys (NSs) are designed to get more accurate data from the negative answers given by interviewees, which are generated by a randomized response model. NS mainly focuses on two folds: data accuracy and data privacy [7, 8]. For data accuracy, users tend to avoid sensitive questions and give implicit or fake answers, thus generating inaccurate datasets. Therefore, NS asks the respondents to choose between t possible answers to a single question; that is, other t-1 answers are eliminated by users one by one. For data privacy, attackers can get user profiles and infer the user identities from the negative answers they give [9]. As a result, there is a need to protect original information recovery from negative information. For example, Bao et al. [10] designed a greedy algorithm to calculate the smallest confidence areas. With the dependable level of the negative survey, the reasonable range of the positive survey results could be estimated, which is analyzed by studying the confidence coefficient of the negative survey. However, there is currently no research on the bounds of negative information selection [11].

A recent study on negative surveys focuses on improving accuracy by requiring each respondent to select multiple false answers of the same category [12]. In 2016, Esponda et al. [13] introduced a statistical approach for collecting sensitive data, which allows each participant to customize the amount of information that she is willing to reveal, as each respondent has a different criterion in terms of the sensitivity of a specific topic. To improve the accuracy of estimation, Liu et al. [14] proposed a multiple negative survey (MNS), which collects each user's multiple different negative categories to get more accurate results. Two crucial scientific problems (accuracy and confidence level) are analyzed, and the anonymity vote model is then introduced. Jiang et al. [15] indicated that the typical type of NS fails to achieve satisfactory privacy preservation. As a result, they proposed two novel negative survey models that use negative combined categories (NCCs), namely, NCC-I and NCC-II. The experimental results show that the proposed methods can achieve excellent privacy preservation in only two categories. Xu et al. [16] proposed to retain aggregate scores in negative surveys and designed an algorithm to exploit the aggregate scores to enhance the accuracy during result reconstruction. Experimental results show that the proposed approach could outperform existing algorithms since it considers the questionnaire which has multiple questions and aggregate scores from them to have global results.

When it comes to data privacy protection in NS, we found that NS can be leveraged for both user and object privacy protection. For example, Aoki and Sezaki [17] argued that it was difficult to enable a complicated security system on resource-constrained mobile phones. Therefore, they proposed a method of combining the NS with randomized response techniques for privacy preservation. By using this method, the participatory sensing

applications can ensure the data integrity while protecting data privacy with low computation complexity. Similarly, Jiang et al. [18] introduced a privacy-preserving aggregation scheme based on NS for smart meters. However, they do not have a discussion on the bounds of the NS information selection. In 2017, Luo et al. [19] applied NS in the location- and trace-privacy protection of the moving object. They analyze the effectiveness of both the single-selection NS and the multiple-selection NS for location- and trace-privacy protection and theoretically prove that the single-selection NS is a more effective method in this scenario. Yang et al. [20] introduced a privacy-preserving aggregation scheme based on NS for vehicle fuel consumption data. In this paper, they found that although the individual real-time fuel consumption data is meaningless, continuous real-time fuel consumption data may reveal the user's privacy. Therefore, they proposed an anonymous algorithm on the user side and an estimation algorithm on the server side that are able to prevent data collection by attackers. However, no research is focused on the bounds of negative information selection, which makes it important to find the balance between the data accuracy and data privacy.

3. Problem Formulation and Analysis

Let $A = \{a_1, a_2, \ldots, a_m\}$ is a set. S is a full domain. a is included by A.

$\overline{A} = S - A$. $A' \subset \overline{A}$ is a set.

Definition 1. negative information selection mapping $f\colon A \longrightarrow A'$). It that takes a set A as input and output a new set A′. That is, $A' \Longleftarrow f(A)$.

Remark 2

(1) f can be conducted by, in each run, selecting $a \in \overline{A}$, and then let $A' = A' \cup \{a\}$

(2) f is randomly mapped instead of a function, as A' could be different for the same A

(3) $f(A) = A' \subset \overline{A}$

Next, we will analyze some bounds for the security of f. Roughly speaking, the security is quantitative measurement of the information leakage on A from A'.

Let $|A|/|S| = t_1 \in (0, 1)$, where $|X|$ returns the number of elements in a set X. Let $|A|/|A'| = t_2 \in \mathbb{R}^+$.

Remark 3

(1) If $t_2 > 1$, then $E(A') < E(A)$, where $E(\cdot)$ is an entropy function that represents the entropy of a set.

(2) If $t_2 \in (0, 1)$, then $E(A') > E(A)$.

(3) Because $|A'| < |S|$, $|A'|/|S| = (|A|/t_2)/(|A|/t_1) = t_1/t_2 < 1$. Thus, $t_1 < t_2$, $t_1/t_2 \in (0, 1)$.

Suppose S is public. A' is public. A is private. We afterward intend to explore the privacy leakage of A when given A'. Or, we want to quantitatively measure the privacy leakage in terms of main threshold t_1 and t_2.

Definition 4 (privacy of A given A'). It is defined as a conditional probability $\Pr(A \mid A')$ that denotes the probability of successfully guessing the set of A upon given A'.

Let $C(m, n)$ be the combination counts of selecting m from n.

Proposition 5. $\Pr(A) = C(|A|, |A|)/C(|A|, |S|) = 1/C(t_1|S|, |S|) = 1/C(t_1 n, n) = 1 * 2 * \ldots * t_1 n/n * (n-1) * \ldots * (n - t_1 n + 1)$, *where* $|S| = n$.

Proof (straightforward). The combination counts for selecting $|A|$ from A are $C(|A|, |A|)$, and the combination counts for selecting $|A|$ from S are $C(|A|, |S|)$. □

Proposition 6. $\Pr(A \mid A') = C(|A|, |A|)/C(|A|, |S| - |A'|) = 1/C(|A|, |S| - |A'|) = 1/C(t_1 n, n - n * t_1/t_2) = 1/C\ (t_1 n, (1 - t_1/t_2)n)$.

Proof (straightforward). The combination counts for selecting $|A|$ from A are $C(|A|, |A|)$, and the combination counts for selecting $|A|$ from $S - A'$ are $C(|A|, |S| - |A'|)$. □

Proposition 7. *The privacy leakage of A upon given A' is denoted as $\delta(A \mid A')$ and*

$$\delta(A \mid A') = \Pr(A \mid A') - \Pr(A) = \frac{1}{C(t_1 n, (1 - t_1/t_2)n)} - \frac{1}{C(t_1 n, n)}. \quad (1)$$

Proof. It is due to Propositions 5 and 6, and the gap between them is the leakage. □

Remark 8

(1) $\delta(A \mid A')$ grows with the decreasing t_2. Recall Remark 3.

(3) $t_2 > t_1$. Therefore, it is better when $|A'|/|A|$ is larger.

(2) Whether $\delta(A \mid A')$ grows or not with the increasing of t_1 depends on the decreasing gaps between $1/C(t_1 n, (1 - t_1/t_2)n)$ and $1/C(t_1 n, n)$. Recall Remark 3 (1), $t_1 \in (0, 1)$.

Proposition 9. $\delta(A \mid A') > 0$. *That is, privacy leakage cannot be averted.*

Proof. As $\Pr(A \neq A') = 1$, $\Pr(A \mid A') \neq \Pr(A)$. $\delta(A \mid A') = \Pr(A \mid A') - \Pr(A) > 0$. □

4. Advanced Discussion

4.1. Simplified Estimation. Next, we want to give a simplified estimation for the leakage.

Suppose $|A|/|S| = t_1 \in (0, 1)$. $|A|/|A'| = 1$.

Proposition 10. $\Pr(a \in A) = t_1$, *where* $\Pr(a \in A)$ *is the probability that an adversary successfully guesses an element of A.*

Proof (straightforward). The number of elements in A over the number of S is t_1. □

Proposition 11. $\Pr(a \in A \mid A') = t_1/1 - t_1$, *where* $\Pr(a \in A \mid A')$ *is the probability that an adversary successfully guesses an element of A after viewing* A'.

Proof. $\Pr(a \in A \mid A') = |A|/|S| - |A'| = 1/1/t_1 - 1 = t_1/1 - t_1$. □

Proposition 12. $\delta(A \mid A') = t_1 t_1/1 - t_1$.

Proof. $\delta(A \mid A') = \Pr(a \in A \mid A') - \Pr(a \in A) = t_1/1 - t_1 - t_1 = t_1(1/1 - t_1 - 1) = t_1 t_1/1 - t_1$. □

4.2. Interleave Impacts of Multiple Negative Information. Next, we will discuss the impact of multiple execution of f.

If $A_i' = f_i(A)$, $i = 1, 2, \ldots, j$, $f_i = f$.

Proposition 13. $\Pr(a \in A \mid \cup A_i') \le t_1/1 - j * t_1$, *where* $\Pr(a \in A \mid \cup A_i')$ *is the probability that an adversary successfully guesses an element of A after viewing all* A_i' $(i = 1, 2, \ldots, j)$.

Proof. $\Pr(a \in A \mid \cup A_i') = |A|/|S| - |\cup A_i'| \le 1/1/t_1 - j = t_1/1 - j * t_1$. □

Remark 14

(1) As $\cap A_i' \neq \varnothing$, the upper-bound of $|\cup A_i'|$ is $j * |A|$

(2) The worst case is usually the concern, $\Pr(a \in A \mid \cup A_i') = t_1/1 - j * t_1$

(3) In the worst case, the probability grows with the increasing of j

(4) For the viewpoint of information recovery, the private collection of A_i' can recover A with expected probability

4.3. Correlations in Mixing Selection. In this section, we will discuss the implicit impact of other negative information for designated information.

Let $B = \{b_1, b_2, \ldots, b_n\}$. $\overline{B} = S - B$. $B' \subset \overline{B}$.

Proposition 15. $A \cap B \subset \overline{A' \cup B'}$

Proof. $A' \subset \overline{A}$ and $B' \subset \overline{B}$. Thus, $A' \cup B' \subset \overline{A} \cup \overline{B} = \overline{A \cap B}$. Thus, $A \cap B \subset \overline{A' \cup B'} \subset \overline{A'}$.

It means that the openness of B' will damage the privacy of A.

Similarly, we have the following result. □

Proposition 16. $A \cap \bigcap_{i=1}^{j} B_i \subset \overline{A' \cup \bigcup_{i=1}^{j} B_i'}$.

Proof. $A' \subset \overline{A}$, $B_i' \subset \overline{B_i}, i = 1, 2, \ldots, j$. Thus, $A' \cup \bigcup_{i=1}^{j} B_i' \subset \overline{A} \cup \bigcup_{i=1}^{j} \overline{B_i} = \overline{A \cap \bigcap_{i=1}^{j} B_i}$. Thus, $A \cap \bigcap_{i=1}^{j} B_i \subset \overline{A' \cup \bigcup_{i=1}^{j} B_i'}$. □

Remark 17

(1) The addition of the openness of B_i' increases the probability of the leakage on an element in A

(2) Sufficient number of B_i' can recover an element of A

5. Experiment Evaluation

5.1. Experiment Setup. In this section, we do the attacker simulation experiment under different parameters. We assume that the attacker can launch brute force cracking attacks, which is more powerful to guess the right set a in A. As shown in Figure 2, attackers can both guess from the whole set S or the set S–A'. In the first situation, attackers do not know A' and can only guess a from S, while in the second situation, attackers can guess a from S–A'. Attackers can make a query of a to check whether their guesses are right. If a-guess equals to real a, it means a success attack.

The experiments are done by a Python project in Windows 10 and the main function is random sample function supported for combination calculation. The sizes of S are 30, 48, and 60, which is flexible and customized. As we need to calculate C_S^A, setting |S| as 60 in our experiment could already support a large number of combinations of possible subsets chosen from S.

5.2. Simulation Time. We calculate the time consumption when an attacker cracks the information successfully under different parameters. When S is 30, 48, and 60, brute force attacks take hours or days to crack the information, while after A' leakage, we can see a clear decrease in the time consumption in information guessing. For example, when no data are leaked, it may take around 4 days to crack the user information. When half the information is leaked when S is 48, the average cracking time is 17.3 s, 62.4 s, and 174.2 s for different values of the ratio of a to A. The short time of cracking information indicates the low security of the system to deal with the leakage.

5.3. Guess Times Comparison. We first calculate the attacker's guess times under two sets, namely, S and S-A', and then calculate the drop percentage. Besides, we vary the ratio of A to S and the ratio of A to A', so as to see the effect of the parameters t_1 and t_2 on the drop percentage.

The results of the drop percentage of brute force attack times under S and S–A' are shown in Figures 3–5. The drop percentage is the difference between the times guessed under S and S–A' minus the times guessed under S. To notice, the large drop percentage indicates low security, which means that once A' is exposed to the attackers, the set a is dangerous.

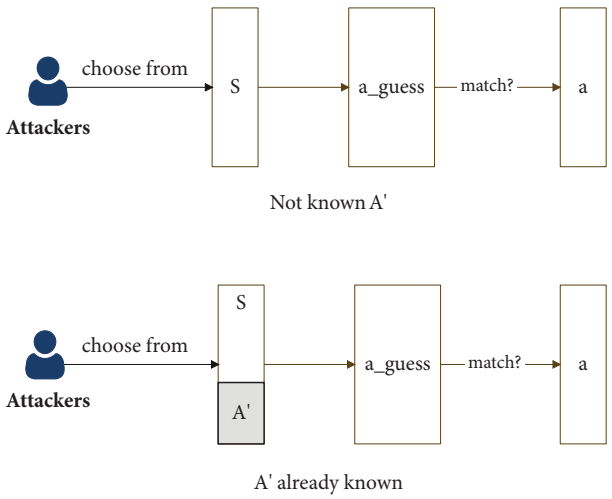

Figure 2: The attacker's simulation experiment overview.

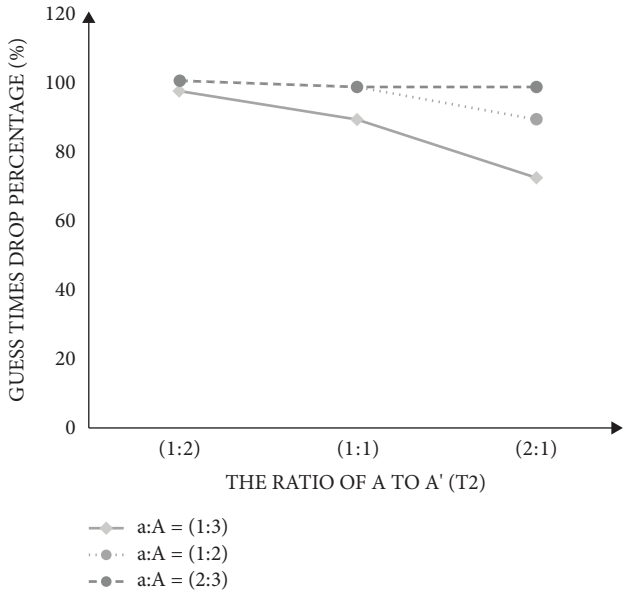

Figure 3: The guess times drop percentage when $|S| = 30$.

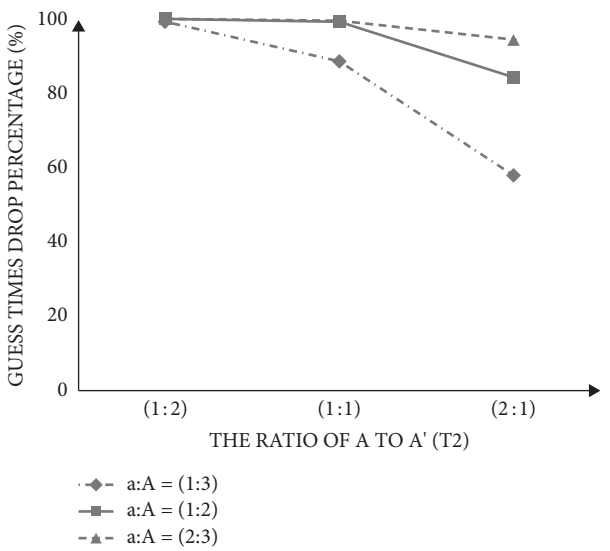

Figure 4: The guess times drop percentage when $|S| = 48$.

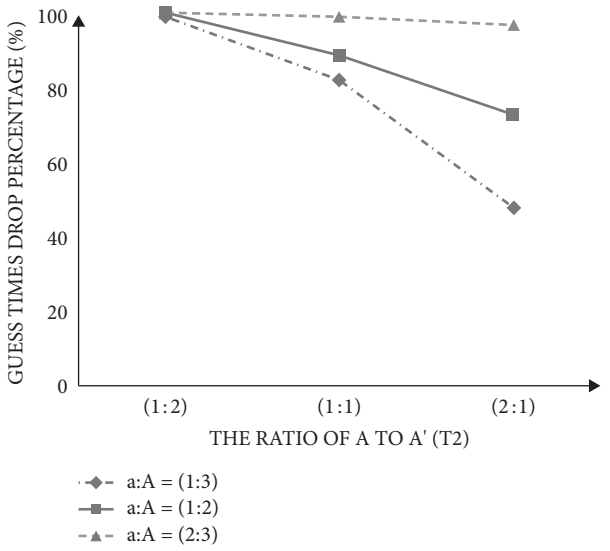

Figure 5: The guess times drop percentage when $|S| = 60$.

In Figure 3, we can see that the drop percentage decreases when the ratio of A to A' (t_2) increases. Besides, the drop percentage is highest when the ratio of a to A is the highest as well. When t_2 is 1 : 2, the value of drop percentages is 97%, 99.8%, and 99%, respectively. When t_2 is 1 : 1, the value of drop percentages is 88.7%, 97.8%, and 98%. When t_2 is 2 : 1, the value of drop percentages is 71.9%, 89%, and 97.8%. These values are still too high, over 30% exposure of set S is dangerous; therefore, we cannot expose much A' information and also we need to mix set a in a large set A.

In Figure 4, we can draw the same conclusion that with the increase of the value of t_2, the drop percentage decreases, since A' counts less part in S. When t_2 is 1 : 2, the value of drop percentages is all around 99%. When t_2 is 1 : 1, the value of drop percentages is 88.2%, 98.8%, and 99%. When t_2 is 2 : 1, the value of drop percentages is 58.1%, 84%, and 94.3%. To notice, when t_2 is 2 : 1 and the ratio of a to A is the smallest, the drop percentage decreases to 58%, which means we can find a boundary to make sure the set a is hard to guess. As long as the ratio of A to A' is high, as the ratio of a to A (S) is small at certain degree, we can protect the set better.

In Figure 5, when the ratio of A to A' is 1 : 2, the drop percentages are all around 99% under different ratios of a to A. When t_2 is 1 : 1, the value of drop percentages is 82.2%, 88.9%, and 99%. When t_2 is 2 : 1, the value of drop percentages is 47.9%, 73%, and 97%. In conclusion, the successful guessing percentage drops when few elements in A' set are exposed to attackers, and we could protect the information by increasing the value t_2 and decreasing the value t_1. Besides, the possibility of guessing the exact a in set A is lower with larger $|S|$, as the attackers should conduct more combination operations to guess the

original information. Therefore, the final suggestion is that to keep a large database and guarantee the lowest ratio of data exposure.

6. Conclusion

In this paper, we formally analyze the bounds of negative information selection, which is an important approach for privacy protection by using negative information to replace original information. We prove the bounds in privacy leakage, probability estimation, multiple negative information implication, and correlation implication. The experimental findings reveal that when data leakage exceeds 30%, it leads to a dangerous situation. Also, the leakage possibility shows minimal correlation with the number of elements in full set, but it is influenced by the ratio of the leaked information. Further works focus on the system application of the proposed method.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was financially supported by the Guangdong Basic and Applied Basic Research Foundation "Research on Key Technologies of Intelligent Interconnection of Industrial Internet" (2022B1515120054), the State Key Laboratory of Geo-Information Engineering and Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology of MNR, CASM (2023-04-04), the Key Laboratory of Data Protection and Intelligent Management, Ministry of Education, Sichuan University, and also the Fundamental Research Funds for the Central Universities (SCU2023D008).

References

[1] Y. Bao, W. Luo, and X. Zhang, "Estimating positive surveys from negative surveys," *Statistics and Probability Letters*, vol. 83, no. 2, pp. 551–558, 2013.

[2] F. Maritsch, I. Cil, C. McKinnon et al., "Data privacy protection in scientific publications: process implementation at a pharmaceutical company," *BMC Medical Ethics*, vol. 23, no. 1, p. 65, 2022.

[3] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated learning for Internet of things: a comprehensive survey," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 3, pp. 1622–1658, 2021.

[4] S. Alneyadi, E. Sithirasenan, and V. Muthukkumarasamy, "A survey on data leakage prevention systems," *Journal of Network and Computer Applications*, vol. 62, pp. 137–152, 2016.

[5] X. Hu, L. Lu, D. Zhao et al., "Privacy-preserving k-means clustering upon negative databases," in *Proceedings of the International Conference on Neural Information Processing*, Springer, Montreal, Canada, December 2018.

[6] F. Esponda, "Negative surveys," 2006, https://arxiv.org/archive/math.

[7] R. Liu and S. Tang, "Negative survey-based privacy protection of cloud data," in *Proceedings of the International Conference in Swarm Intelligence*, Springer, Beijing, China, June 2015.

[8] L. Yu, Y. Fu, J. Oakley, O. Hambolu, and R. Brooks, "On accuracy and anonymity of privacy-preserving negative survey (NS) algorithms," *Computers and Security*, vol. 105, no. 2021, Article ID 102206, 2021.

[9] S. Aoki, M. Iwai, and K. Sezaki, "Limited negative surveys: privacy-preserving participatory sensing," in *Proceedings of the 2012 IEEE 1st International Conference on Cloud Networking (CLOUDNET)*, pp. 158–160, Paris, France, November 2012.

[10] Y. Bao, W. Luo, and Y. Lu, "On the dependable level of the negative survey," *Statistics and Probability Letters*, vol. 89, pp. 31–40, 2014.

[11] C. R. Gjestvang and S. Singh, "A new randomized response model," *Journal of the Royal Statistical Society- Series B: Statistical Methodology*, vol. 68, no. 3, pp. 523–530, 2006.

[12] H. James, M. M. Groat, S. Forrest, and F. Esponda, "Anonymous data collection in sensor networks," in *Proceedings of the 2007 Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services (MobiQuitous)*, IEEE, Philadelphia, PA, USA, August 2007.

[13] F. Esponda, K. Huerta, and V. M. Guerrero, "A statistical approach to provide individualized privacy for surveys," *PLoS One*, vol. 11, Article ID e0147314, 2016.

[14] R. Liu, J. Peng, and S. Tang, "Multiple-negative survey method for enhancing the accuracy of negative survey-based cloud data privacy: applications and extensions," *Engineering Applications of Artificial Intelligence*, vol. 62, pp. 350–358, 2017.

[15] H. Jiang, W. Luo, B. Duan, and C. Wu, "Enhancing the privacy of negative surveys using negative combined categories," *Applied Soft Computing*, vol. 96, Article ID 106578, 2020.

[16] Z. Xu, D. Zhao, F. Li, and J. Zhou, "Negative survey with aggregate scores from multiple questions," in *Proceedings of the 2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 1257–1262, Hangzhou, China, May 2022.

[17] S. Aoki and K. Sezaki, "Negative surveys with randomized response techniques for privacy-aware participatory sensing," *IEICE- Transactions on Communications*, vol. 97, no. 4, pp. 721–729, 2014.

[18] H. Jiang, W. Luo, and Z. Zhang, "A privacy-preserving aggregation scheme based on immunological negative surveys for smart meters," *Applied Soft Computing*, vol. 85, Article ID 105821, 2019.

[19] W. Luo, Y. Lu, D. Zhao, and H. Jiang, "On location and trace privacy of the moving object using the negative survey," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, no. 2, pp. 125–134, 2017.

[20] W. Yang, X. Chen, Z. Xiong, Z. Xu, G. Liu, and X. Zhang, "A privacy-preserving aggregation scheme based on negative survey for vehicle fuel consumption data," *Information Sciences*, vol. 570, no. 2021, pp. 526–544, 2021.