Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 7, No. 1, 2024

ES Data Aggregation Scheme based on Personalized Local Differential Privacy

Qiong Liu [a, *], Xueyan Liu [b], Jia Wang [c] and Hao Sun [d]

College of Computer Science and Engineering, Northwest Normal University, Lanzhou, Gansu, China
[a, *] m13893678075@163.com, [b] liuxy@nwnu.edu.cn, [c] wj15719323392@163.com, [d] sh_ttdb0728@163.com
*** Corresponding author:** Qiong Liu (Email: m13893678075@163.com)

Abstract: In modern society, with the rapid development of technology, research on Epidemiological Survey Data (ESD) has become increasingly crucial. As various diseases continue to evolve and spread, there is a growing need for a profound understanding of the patterns and trends of disease transmission. Epidemiological investigation, as a key method, provides fundamental data support for devising effective prevention and control strategies by tracking cases, contacts, and potential infectees. However, accompanying this progress is the issue of dealing with a large amount of privacy-sensitive data. Once these data are compromised, it may pose severe harm to individuals and society. To address this challenge, we propose a uOUE-based epidemiological survey data aggregation scheme for the collection and processing of ESD, aiming to enhance the efficiency and accuracy of data coding. We ensure the secure, efficient, and accurate aggregation processing of ESD. Through rigorous validation and comparative analysis, our algorithm complies with the requirements of local differential privacy and unbiased estimation. It demonstrates good practicality and accuracy in ESD collection, providing users with a reliable privacy protection effect.

Keywords: Privacy Protection; Data Encryption Sharing; Personalized Local Differential Privacy.

1. Introduction

In recent years, the world has witnessed the significant impact of various infectious diseases such as SARS, swine flu, Ebola, novel coronavirus epidemics, and influenza virus [1][2]. In response to these outbreaks, Epidemiological investigation (EI) has emerged as a critical measure to control their spread. EI typically involves the meticulous tracking of patients, close contacts, and potential contacts conducted by health departments or other relevant organizations. This investigative process aims to identify potential close contacts based on the patients' fundamental information and behavioral patterns, followed by comprehensive tracking, screening, and necessary interventions in various settings like hospitals, communities, and workplaces [3]. Consequently, the importance of ensuring timely and accurate EI cannot be overstated, with a particular emphasis on preserving personal privacy and information security.

When it comes to analyzing the transmission of infectious diseases, the data used necessitate a higher level of privacy protection compared to general health data. These data encompass the identities, health statuses, and detailed information about close contacts of Epidemiological Survey Objects (ESO), which are extremely sensitive from the perspective of data owners. Emergencies escalate the urgency and scale of epidemic investigations, often leading to physical recording, thereby increasing the risks of information exposure. Information leaks can have severe consequences, ranging from repeated privacy breaches to the dissemination of misguided public opinions, social panic, and potential harm. Moreover, inadequate or excessive desensitization of patient information by authorities may result in the disclosure of irrelevant data, further complicating the challenge of data privacy protection. Therefore, safeguarding citizens' privacy is of paramount importance during epidemic investigations, necessitating the implementation of effective encryption and security measures to prevent the leakage of sensitive information.

1.1. Related Work

In epidemiological studies, there have been many research results on the encryption protection of ESD. Blumenberg C et al. [4] used the REDCap system to explore the advantages and limitations of the electronic data collection environment and solved data inconsistency through real-time reporting and field verification, which is expected to play a role in time-saving and data quality improvement. However, there are shortcomings in its method of detail description, limitation discussion, replicability, and data quality analysis. Dong E et al. [5] introduced the COVID-19 dashboard, which provides real-time epidemic data for the world, and described the data collection process in detail. However, the inconsistency and time delay of the data affects the reliability of the data and the limitations of the application. Sperber A D et al. [6] deeply compared the two data collection methods of face-to-face interviews and internet surveys, but the preferences and participation of different audiences may affect the bias of the method and did not provide a detailed data collection method design and description. In summary, most studies only encrypt the ESD itself, ignoring the privacy protection of non-medical data, and may not be able to fully protect the user's privacy information, their methods cannot be replicated, the limitations are large, and the data cannot be applied. For data involving correlations, such as influenza status, sending data separately may lead to increased errors and reduce data accuracy and availability. In addition, the correlation between non-sensitive attributes and sensitive attributes may result in the disclosure of personally sensitive information. Therefore, when dealing with the ESO, we must consider the relevance and take privacy protection measures to ensure data security and privacy.

DP [7] is a privacy protection method with a strict formal

security model. It is characterized by high efficiency and low cost and has attracted much attention in the research field. In 2013, LDP [8] was proposed as a variant of DP, which inherits its advantages and abandons the dependence on trusted third parties, thus improving the practicality of the model. LDP can add noise to all data and effectively protect data privacy. It has a wide range of applications, including machine learning, network services, data statistics, and optimization. LDP has been widely used in the industry. For example, Apple uses it to protect users' mobile phone usage data, and Google uses LDP-based components (such as randomized aggregatable privacy-preserving ordinal response, RAPPOR) to collect user behavior data. In 2017, Wang et al. [9] proposed a framework to incorporate most LDP protocols into the pure LDP protocols framework to optimize and generalize existing protocols and compared the accuracy and communication cost of different LDP protocols. They also introduced the OUE protocol with higher accuracy. In addition, the study compares a variety of encoding methods and provides suggestions for selecting protocols. Histogram encoding (HE) and unary encoding (UE) require $O(d)$ communication costs, and direct encoding (DE) and local hashing (LH) require $O(\log d)$ or $O(\log n)$ communication costs (n is the number of users). All protocols except DE estimate the computational cost $O(n \cdot d)$ of the frequency of all values. When the number of values that users may input is large, UE is an effective coding method, and its communication cost $O(d)$ is equivalent to that of HE. Therefore, when the user may input the number of values $d > 3e^{\varepsilon} + 2$ and $d < n$, to avoid the high computational cost of DE and LH, the OUE coding mechanism offers improved accuracy and unbiased estimation results. Furthermore, the UE, known for its simplicity and intuitiveness, is easy to implement and offers superior performance in terms of computational and communication costs. This makes it a more practical choice for adoption in various applications. However, many existing LDP mechanisms often overlook the varying privacy protection requirements of different data in practical scenarios. Consequently, they may increase estimation errors by over-protecting non-sensitive data. In 2019, the ULDP model [10] provided a feasible method to reduce the protection of non-sensitive data according to the privacy requirements of different data, thereby improving data utility and protecting the privacy of sensitive data. In 2022, He et al. [11] proposed the uOLH protocol for big data domains, which aims to achieve low communication costs and high data utility. Nonetheless, it focuses on the big data and does not apply to all scenarios, and complex data hashing increases the complexity and cost of implementation. Additionally, Cao et al. [12] devised a frequency estimation mechanism conforming to the set-valued data ULDP model, offering a privacy protection solution for set data. However, its applicability may be limited when dealing with different data types, thereby increasing the complexity and cost of practical implementations. These studies collectively aim to address privacy leakage issues and enhance data practicality and utility.

Based on the particularity of data in EI, in the process of LDP perturbation, it is necessary to ensure that ESD does not lose information and that ESD is fully protected by privacy. The OUE protocol scheme based on the ULDP protocol is used to transmit the frequency estimation information of large-scale data sets, which can achieve low communication costs and high data utility in the big data domain.

Therefore, we improved the utility-optimized unary encoding (uOUE) mechanism based on the OUE mechanism [10] to address the personalized privacy protection requirements for ESD. Moreover, we demonstrated that uOUE complies with the ULDP model. This scheme effectively handles situations where sensitive and non-sensitive ESD are mixed and ensures that the protection effect of sensitive data is not affected.

Our main contributions are as follows.

(1) For users' personalized privacy protection requirements, we improve the uOUE protocol that conforms to the ULDP model based on the OUE mechanism.

(2) By proving that the uOUE protocol satisfies the ULDP model and calculating the theoretical variance of the frequency estimation results, the proposed protocol has been deemed to have low communication cost and high data utility.

1.2. Organization

The rest of our scheme is organized as follows. Section 2 introduces some preliminary knowledge; Section 3 gives the specific content of the uOUE mechanism; Section 4 gives the theoretical proof and comparative analysis of the scheme. Section 5 Summary.

2. Preliminary Knowledge

2.1. Utility Optimization LDP

ULDP divides the original data set X into sensitive data set X_S and non-sensitive data set X_N and divides the output set into protected data set Y_S and reversible data set Y_N. Its formal definition is as follows.

Define 1. $(X_S, Y_S, \varepsilon) - ULDP$, $\varepsilon \geq 0$, for the perturbation mechanism with input domain X and output domain Y, $M: X \to Y$, if and only if the perturbation mechanism M satisfies the following properties, satisfies $(X_S, Y_S, \varepsilon) - ULDP$.

(1) For any $y \in Y_N$, here is and only one:

$$\Pr[M(x_1) = y] > 0 \tag{1}$$

And for any $x_1 \neq x_2$, satisfy the following:

$$\Pr[M(x_2) = y] = 0 \tag{2}$$

(2) For any input $x_1, x_2 \in X$, get any output $y \in Y_S$, satisfy the following:

$$\Pr[M(x_1) = y] \leq e^{\varepsilon} \Pr[M(x_2) = y] \tag{3}$$

2.2. Utility Evaluation

The mean square error (MSE) is used to evaluate the effectiveness of the protocol and experiment. The formal definition of mean square error is shown in(4):

$$MSE(\hat{F}) = E\left[\sum_{i=1}^{n}\left(F_x - \hat{F}_x^2\right)\right] \tag{4}$$

F_x represents the real frequency, and F_x represents the estimated frequency.

3. uOUE Agreement

In the realm of ESD studies[13], sensitivity issues are examined, encompassing both sensitive and non-sensitive

attributes. For instance, when assessing Epidemiological Investigation (EI) data, an ESO is assumed to have an EI record (i.e., attribute set) comprising attributes such as gender, age, symptoms, and more. Here, gender is identified as a sensitive attribute, while attributes like allergic drugs and chronic diseases are considered non-sensitive. Sensitive attributes consist of both sensitive candidate values and non-sensitive candidate values. For example, when scrutinizing regional attributes, the travel area of an ESO (i.e., regional attribute candidate value set) may include values such as Beijing, Shanghai, Guangxi, and Hubei. In this scenario, Beijing and Shanghai are regarded as sensitive candidate values, while the remaining regional attributes are considered non-sensitive. Addressing the intricate sensitivity of ESO privacy data, this paper introduces an ESD aggregation scheme based on the uOUE mechanism. This scheme enhances data utility and ensures data transmission privacy by minimizing the protection of non-sensitive data. Diverging from the traditional OUE protocol, the uOUE protocol takes into account privacy budget considerations and incorporates the ULDP mechanism to reduce communication costs, effectively mitigating the risk of information leakage while preserving data accuracy.

3.1. Utility Evaluation

The uOUE scheme encodes each ESD value as a binary vector, and the dimension is equal to the number of candidate values, so that each data item can be expressed as a vector, where each dimension indicates that the data item belongs to a specific ESD candidate value. The privacy perturbation of ESD is calculated by vector operation to maintain the attributes of original data items. The protocol distinguishes sensitive sets and non-sensitive sets and uses different privacy protection methods. Sensitive data is disturbed by random response mechanism, and non-sensitive data is disturbed by individual candidate values, and then frequency estimation is carried out to improve accuracy and utility. This processing method reduces the protection strength of non-sensitive data and improves the accuracy of frequency estimation results, while still protecting privacy and improving data utility.

3.2. Utility Evaluation

The participants of the uOUE protocol include three parties: ESO, Epidemiological data control center (EDCC) and Epidemiological data users (EDU). ESO holds the original data, encodes it and sends it to EDCC after perturbation. EDCC aggregates and statistically analyzes the perturbation data of all users, estimates the frequency distribution results of all the original data, and finally sends the results to the corresponding EDU.

The original data set is recorded as $X=\{x_1,x_2,\cdots,x_{|X|}\}$, the dimension size is $d=|X|$. Among them, the original data set is divided into two parts: sensitive data set X_S and non-sensitive data set X_N. The two do not intersect, that is, $|X|=|X_S|+|X_N|$.

The uOUE scheme is divided into three steps: encoding, perturbation, and aggregation. The specific steps are as follows:

3.2.1. Encoding

In uOUE, the data is first UE encoded, that is, the classification data in the user's hands is encoded as a d-bit vector v, each bit corresponds to data in the original data domain. If the user data contains data k, let the k bit of v be 1.

Assume that there are n users, and each user holds raw data x. To reduce the subsequent communication cost, UE is used to encode it, and the encoding result $v=Encode(x)$ is obtained. Suppose $\{x_1,x_2,\cdots,x_{|X_S|}\}$ is sensitive data X_S and $\{x_{|X_S|+1},x_{|X_S|+2},\cdots,x_{|X|}\}$ is non-sensitive data X_N, then the sensitive output is $Y_P=\{(y_1,y_2,\cdots,y_{|X|})|y_1,y_2,\cdots,y_{|X|}\in\{0,1\}\}$ and the reversible output is $Y_N=X_N$. When x is sensitive data, it will be encoded as data in $\{x_1,x_2,\cdots,x_{|X_S|}\}$, and when x is non-sensitive data, its encoding result is data in $\{x_{|X_S|+1},x_{|X_S|+2},\cdots,x_{|X|}\}$.

3.2.2. Perturbing

Users encode and perturb their original data locally to generate perturbation data v'. According to the sensitivity x, $Perturb(v)$ is processed by different perturbation methods. The specific method is to perturb each v_k in v to obtain v_k', as shown in(5). The processed perturbation data v_k' will be sent to the server for aggregation and analysis.

$$\Pr[v_k'|v_k]=\begin{cases}\alpha & v_k=1,v_k'=1;v_k\in X_S\\ 1-\alpha & v_k=1,v_k'=0;v_k\in X_S\\ \beta & v_k=0,v_k'=1;v_k\in X_S\\ 1-\beta & v_k=0,v_k'=0;v_k\in X_S\\ \gamma & v_k=1,v_k'=1;v_k\in X_N\\ 1-\gamma & v_k=1,v_k'=0;v_k\in X_N\\ 1 & v_k=0,v_k'=0;v_k\in X_N\\ 0 & v_k=0,v_k'=1;v_k\in X_N\end{cases} \tag{5}$$

For sensitive data x, some probabilities remain unchanged $\alpha=\frac{1}{2}$, $\beta=\frac{1}{1+e^{\varepsilon}}$ and probability deflection occurs, as shown in Figure 1 below; for non-sensitive data x, the probability of having $\gamma=\frac{e^{\varepsilon}-1}{2e^{\varepsilon}}$ remained unchanged. If $v_k\in X_N$ and $v_k'=1$, then the reversible data v_k is output, that is, we can use v_k' to represent $(y_1,y_2,\cdots,y_t,t\le|X|)$ directly. The perturbation examples are shown in Figure 1 and Figure 2. After the disturbance is completed, the disturbance data will be sent to the server.

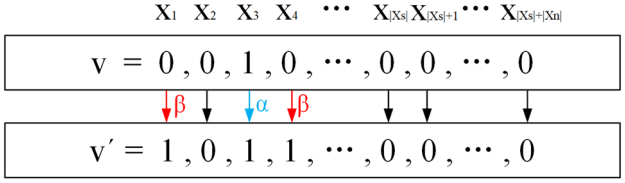

Figure 1. Example of uOUE sensitive data perturbation

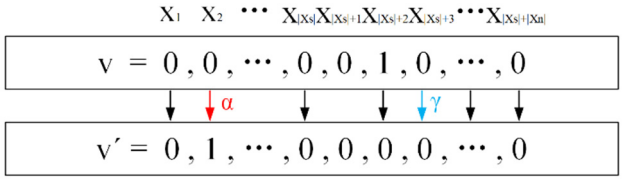

Figure 2. Example of uOUE non-sensitive data perturbation

3.2.3. Aggregation

After the server receives the disturbance data sent by the user, the server will count whether each bit in v' is 1. Suppose that the number of occurrences of 1 in the k-th bit is $\hat{F}_x$, by counting the number of occurrences of 1 in each bit, the server can estimate the probability of occurrence of x in each original data, and the statistical analysis results $\hat{F}_x$ close to the frequency distribution of the original data are as follows:

$$\hat{F}_x = \begin{cases} \dfrac{F_x / n - \beta}{\alpha - \beta}, if\ x \in X_S \\ \dfrac{F_x}{n\gamma}, if\ x \in X_N \end{cases} \tag{6}$$

4. Scheme Analysis

This section provides a theoretical and comparative analysis of the uOUE protocol, demonstrating its low communication cost and high data utility. It also includes a security proof and comparative analysis of the ESD aggregation scheme based on uOUE.

4.1. Theoretical Analysis of uOUE Protocol

This section will introduce some related properties of the uOUE protocol and give the corresponding theoretical proof.

Theorem 1. The perturbation process of uOUE conforms to the ULDP model.

Proof of Theorem 1. For any $v_1, v_2 \in X_m$, the probability of outputting the same result A satisfies(7):

$$\begin{aligned} \frac{\Pr[A|v_1]}{\Pr[A|v_2]} &= \frac{\prod_{k\in[m]} \Pr[A[k]|v_1]}{\prod_{k\in[d]} \Pr[A[k]|v_2]} \\ &\le \frac{\Pr[A[v_1]=1|v_1] * \Pr[A[v_2]=0\|v_1]}{\Pr[A[v_1]=1|v_2] * \Pr[A[v_2]=0|v_2]} \\ &= \prod_{k\in d} \frac{\alpha}{\beta} \cdot \frac{1}{1-\gamma} = e^{\varepsilon} \end{aligned} \tag{7}$$

The above (7) satisfies the second nature of (3).

In uOUE, for any output $A \in Y_N$, there is only one original data item $x \in X_N$ that can be perturbed into the reversible data, that is, if and only if $x \in X$, the reversible data is output with probability γ. Therefore, uOUE satisfies the properties of (1) and (2) in the ULDP model definition.

In summary, the perturbation process of uOUE conforms to the ULDP model.

Theorem 2. The result of uOUE frequency estimation is an unbiased estimation.

Proof of Theorem 2. For the original data x, its estimated frequency is denoted by $\hat{F}_x$.

If x is sensitive data, it can be known from the disturbance process that:

$$\begin{aligned} E(\hat{F}_x) &= E\left(\frac{F_x / n - \beta}{\alpha - \beta}\right) \\ &= \frac{\alpha F_x + \beta(1 - F_x) - \beta}{\alpha - \beta} = F_x \end{aligned} \tag{8}$$

If x is non-sensitive data, it can be known from the disturbance process that:

$$E(\hat{F}_x) = E\left(\frac{F_x}{n\gamma}\right) = \frac{n\gamma F_x}{n\gamma} = F_x \tag{9}$$

In summary, the frequency estimation result of uOUE is unbiased.

Theorem 3. In the uOUE protocol, the mean square error of the estimated frequency $\hat{F}_x$ is shown in *(10)*:

$$MSE[\hat{F}_x] = \begin{cases} \dfrac{4e^{\varepsilon}}{n(e^{\varepsilon}-1)^2} + \dfrac{F_x}{n}, x \in X_S \\ \dfrac{e^{\varepsilon}+1}{n(e^{\varepsilon}-1)} F(x), x \in X_N \end{cases} \tag{10}$$

Proof of Theorem 3. From Theorem 2, Equation (4.2) is an unbiased estimation, so MSE is equal to the variance of $\hat{F}_x$.

x is sensitive:

$$\begin{aligned} MSE\left[\hat{F}_x\right] = Var\left[\hat{F}_x\right] &= Var\left[\frac{F_x / n - \beta}{\alpha - \beta}\right] \\ &= \frac{nF(x)\alpha(1-\alpha) + n(1-F(x))\beta(1-\beta)}{n^2(\alpha-\beta)^2} \\ &= \frac{4e^{\varepsilon}}{n(e^{\varepsilon}-1)^2} + \frac{F_x}{n} \end{aligned}$$

x is non-sensitive:

$$\begin{aligned} MSE\left[\hat{F}_x\right] = Var\left[\hat{F}_x\right] &= Var\left[\frac{F(x)}{n\gamma}\right] \\ &= \frac{nF(x)\gamma(1-\gamma)}{n^2\gamma^2} = \frac{1-\gamma}{n\gamma} F(x) \\ &= \frac{e^{\varepsilon}+1}{n(e^{\varepsilon}-1)} F(x) \end{aligned}$$

4.2. Comparative Analysis of uOUE Protocol

We evaluate the utility of six mechanisms, including GRR, RAPPOR, uGRR, uRAP, uOLH, and uOUE. Since $MSE\left[\hat{F}\right] = \sum_{x\in X_S} MSE\left[\hat{F}_x\right] + \sum_{x\in X_N} MSE\left[\hat{F}_x\right]$ and Theorem 3, we compute the MSE for both the proposed and existing LDP mechanisms when $\varepsilon = O(1)$. The results are displayed in Table 1, which F_{X_N} represents the actual total frequency of non-sensitive data.

In practical applications, most of the MSE comes from sensitive data, but sensitive data usually only accounts for a part of the entire data set. Therefore, by optimizing the utility of the personalized component privacy mechanism, its MSE is significantly smaller than the non-utility mechanism. Our improved mechanism is easy to implement in practical applications and shows better performance in terms of computational cost and communication costs. This mechanism can protect sensitive data more accurately and

reduce its impact on the overall error.

Table 1. Comparison of MSE when $\varepsilon = O(1)$

Mechanism	MSE	Mechanism	MSE
GRR	$O\left(\frac{d^2}{n\varepsilon^2}\right)$	GRR	$O\left(\frac{\lvert X_S\rvert}{n\varepsilon^2}+\frac{F_{X_N}}{n\varepsilon}\right)$
RAPPOR	$O\left(\frac{d}{n\varepsilon^2}\right)$	RAPPOR	$O\left(\frac{\lvert X_S\rvert}{n\varepsilon^2}+\frac{F_{X_N}}{n}\right)$
uGRR	$O\left(\frac{\lvert X_S\rvert^2}{n\varepsilon^2}+\frac{\lvert X_N\rvert F_{X_N}}{n\varepsilon}\right)$	uGRR	$O\left(\frac{\lvert X_S\rvert}{n\varepsilon^2}+\frac{F_{X_N}}{n\varepsilon}\right)$

In addition to data utility, communication cost is also an important criterion to evaluate whether a mechanism is good or not. We summarize the communication cost of existing ULDP protocols, which can be seen in Table 2.

As the privacy budget increases, the error of the six protocols gradually decreases. However, in terms of data utility, uGRR is significantly behind the other two protocols. This difference is mainly due to the use of large data sets in the original data domain in the experiment, which poses a challenge to the adaptability of uGRR. For a wide range of raw data domains, uOLH shows superior communication cost performance. In practical scenarios, especially when the original data domain is moderate, the uOUE protocol is superior in communication overhead. Considering the unique application scenarios and data characteristics of ESD, the uOUE protocol effectively meets the needs of actual scenarios while ensuring privacy protection.

Table 2. Communication cost comparison

Mechanism	MSE	Mechanism	MSE
GRR	$O(\log d)$	uRAP	$O(\lvert X_S\rvert+\lvert X_N\rvert)$
RAPPOR	$O(d)$	uOLH	$O(\log\lvert H\rvert+\log(g+\lvert X_N\rvert))$
uGRR	$O(\log\lvert X_S\rvert+\lvert X_N\rvert)$	uOUE	$O(\lvert X_S\rvert+\lvert X_N\rvert)$

5. Using THIS TEMPLATE AND ITS Automatic Formatting

In the context of ESD aggregation, we have improved and designed an epidemiological survey based on the uOUE protocol, ensuring the enhanced accuracy of ESD frequency estimation while safeguarding sensitive data. Our goal is to alleviate privacy risks associated with the transmission of patient raw data across channels and servers, ensuring the integrity and security of data during processing. These enhancements make ESD aggregation more secure, efficient, and accurate, enhancing the precision and reliability of results while effectively preventing data tampering and forgery.

In the future, our research will focus on improving the utility of data within personalized models and designing multi-level, multi-dimensional privacy ESD aggregation schemes. In personalized models, we will explore adaptive data processing methods to meet the diverse needs of different users and data requirements. In the design of multi-level privacy-level ESD aggregation schemes, we will comprehensively consider varying data sensitivities and privacy requirements, optimizing the utilization of data information while preserving privacy. These studies will drive advancements in the field, providing comprehensive and optimized solutions for data aggregation and privacy protection.

References

[1] Giabicani, M.; Le Terrier, C.; Poncet, A.; Guidet, B.; Jean-Philippe, B. Limitation of life-sustaining therapies in critically ill patients with COVID-19: a descriptive epidemiological investigation from the COVID-ICU study. Critical Care, 2023. [CrossRef].

[2] Song, Q. X.; New Coronavirus Pneumonia Epidemic-related Rumors and Its Mechanism of Generation and Dissemination - Discussion on the Cooperative Principle of Emergency Information Release. Language Planning Research, 2021, pp. 57-66.

[3] Feng, B.; Chao, L. Analysis of epidemic prevention and control behavior and influencing factors of employees in public places in the normalized prevention and control stage of COVID-19. Anhui Journal of Preventive Medicine, 2022. [CrossRef].

[4] Blumenberg, C.; JD, A. Electronic data collection in epidemiological research. Applied clinical informatics 2016, Volume. 7, pp. 672-681. [CrossRef].

[5] Boneh, D.; Lynn, B.; Shacham, H. Short signatures from the weil pairing. Journal of cryptology: the journal of the International Association for Cryptologic Research 2004, Volume 17, pp. 297–319. [CrossRef].

[6] Sperber, A. D.; Bor, S.; Fang. X. Face-to-face interviews versus Internet surveys: Comparison of two data collection methods in the Rome foundation global epidemiology study: Implications for population-based research. Neurogastroenterology & Motility 2023. [CrossRef].

[7] Dwork, C. Differential privacy: A survey of results. In International conference on theory and applications of models of computation, Berlin, 2008. [CrossRef].

[8] Duchi, J. C.; Jordan, M. I.; J, M. Wainwright. Local privacy and statistical minimax rates. In 2013 IEEE 54th Annual Symposium on Foundations of Computer Science. IEEE, 2013. [CrossRef].

[9] Wang, T.; Blocki, J.; Li, N. H. Locally differentially private protocols for frequency estimation. In 26th USENIX security symposium (USENIX Security 17), 2017.

[10] Murakami, T.; Kawamoto, Y. Utility-optimized local differential privacy mechanisms for distribution estimation., In 28th USENIX Security Symposium (USENIX Security 19), 2019.

[11] He, X. Y.; Zhu, Y. W.; Zhang, Y. Utility optimization of local differential privacy mechanism based on OLH. Journal of Cryptography 2022, Volume. 9, pp. 820-833. [CrossRef].

[12] Cao, Y. R.; Zhu, Y. W.; He, X. Y; Zhang, Y. Utility-optimized local differential privacy set data frequency estimation mechanism. Computer research and development 2022.

[13] Dong, E.; Ratcliff, J.; Goyea, T. D.; MS, A. K. The Johns Hopkins University Center for systems science and engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. lancet infectious diseases 2022, pp. e370-e376. [CrossRef].