A STUDY ON PROTECTION OF ENTERPRISE HUMAN RESOURCE DATA ON WEB PLATFORMS THROUGH ANONYMITY ALGORITHMS

Jing CHEN[1]

The recruitment process on online platforms poses a significant security threat to the confidentiality of enterprise human resource data, necessitating robust protection measures. This paper studied the protection of enterprise human resource data during recruitment on online platforms. This paper presents a comprehensive introduction to various anonymity algorithms, followed by the design of an (α, l) anonymity model based on sensitivity. Furthermore, experiments were conducted on the Adult dataset with occupation as a sensitive attribute. It was found that with the increase of the values of k, l, and the number of quasi-identifier attributes, the information loss of the designed (α, l) anonymity model increased, but it was smaller than that of the l-diversity model. For example, when k = 200, l = 6, and the number of quasi-identifier attributes = 5, the information loss experienced by the l-diversity model exhibited a rise of 21.8% compared to that encountered by the designed model. Then, the running time of the designed model was found to be higher than that of the l-diversity model, while exhibiting superior stability. In addition, when combined with sensitivity, the (α, l) anonymity model effectively safeguarded sensitive attributes in enterprise human resource data processing on actual network platforms. These results demonstrate that the sensitivity-combined (α, l) anonymity model offers the advantages of minimal information loss and greater stability in safeguarding data, making it applicable in practice for better protection of enterprise human resource data.

Keywords: anonymity algorithm, human resource data, protection, sensitive attribute

1. Introduction

Under the influence of the increase in college graduates, the reduction of hiring by enterprises, and the development of artificial intelligence, both the pressure and difficulty of employment have increased significantly [1]. In the employment process, recruiters and employers typically expend substantial time and resources conducting numerous interviews to identify suitable talents and positions. However, with the development of network technology, a plethora of online recruitment platforms has emerged as a partial remedy for the limitations

[1] Department of Management, Taiyuan University, Taiyuan, Shanxi 030032, China, e-mail: ji56966@163.com

inherent in traditional offline recruitment methods [2]. Through these platforms, users independently provide their information to facilitate the matching of talents and positions, thus improving employment efficiency. However, in this process, enterprises' protection of users' private data is inadequate as various human resources data are directly uploaded without undergoing any processing, thus posing security risks for both individuals and enterprises. The protection of data is very important [3]. Protecting data while embracing the convenience offered by online platforms has garnered significant attention across diverse industries. Kumar et al. [4] designed an optimization algorithm to dynamically determine the optimal k value instead of the standard k-anonymization process and used the improved gray wolf optimization algorithm to achieve the optimization. Through experiments, it was found that their method ensured maximum security while minimizing computational time. Abbasi et al. [5] used the k-means++ algorithm based on clustering process to implement an optimal k-anonymity algorithm and improved the quality of the anonymous data by using a normal distribution function. Experimental results demonstrated a 1.5-fold reduction in information loss and a 3.5-fold decrease in execution time. To protect medical data, Mandala et al. [6] used an adaptive awareness probabilistic-based crow search algorithm to select the optimal security key. Bazai et al. [7] proposed Apache Spark subtree data anonymization technique based on distributed dataset and found that the method had high performance by comparing it with other methods. Belmar-Monterrubio et al. [8] proposed a function called Tukmuky Jinma'any-Function to generate chaos for practical applications and used it as an entropy source for symmetric encryption algorithms. Through experiments on image bit streams, this method showed outstanding results and overcame the drawbacks of the previous chaos-based cryptographic system. Pawar et al. [9] designed and developed an identity authentication and data security mode. Through experiments, it was found that this method had a good effect on privacy protection. Thabit et al. [10] proposed a lightweight homomorphic encryption algorithm that enhances the security of data in cloud computing by utilizing a lightweight encryption algorithm in the first layer and a multiplication homomorphic scheme in the second layer. Experimental results demonstrated that this method exhibited strong security properties. This paper mainly explored the anonymity algorithm. In response to the insufficient consideration of similarity in sensitive attributes in current research, a (α , l) anonymity model incorporating sensitivity was designed to impose constraints on different values of sensitive attributes. The method was assumed to possess the capability of ensuring protection for enterprise human resources data, and it was experimentally validated. This work provides some theoretical references for further research on anonymity algorithms and also contributes to strengthening the protection of sensitive data.

2. Literature review on anonymity algorithm

2.1 Privacy protection technology

The term “privacy” refers to the confidential information that individuals or organizations wish to keep undisclosed to the public, encompassing sensitive data pertaining to an individual's personal life or an organization's operations. In the human resources data of companies on online platforms, privacy information includes personal information provided by job seekers, such as education background, past occupations, salary situation, etc., as well as information uploaded by companies, such as financial status and personnel organization. Due to the openness of the Internet, unprocessed or inadequately protected data can be easily obtained by attackers, leading to privacy breaches [11]. In this situation, companies and individuals may refuse to provide or selectively provide relevant data due to concerns about privacy protection, which can also pose certain obstacles to online recruitment. As online recruitment continues to expand and develop, the protection of human resources data becomes particularly important in the process.

In order to realize the protection of such sensitive data, privacy protection technologies have been developed, including data encryption and restricted release methods [12]. Anonymity algorithms fall under the category of restricted release, which ensures data security by releasing information with reduced accuracy.

Table 1

Original data table T

ID	Name	Age	Gender	ZIP code	Disease
1	Alice	22	F	221125	Flu
2	Bob	31	M	221365	Heart disease
3	Carlo	36	M	221452	cancer
4	Daisy	26	F	221458	HIV
5	Elsa	41	F	221154	Flu

Note: ZIP: zone improvement program

Assuming there is an original data table T (Table 1) (Tables 1-6 are all hypothetical examples), the attributes in it can be written as follows.

(1) Identifier (ID) attribute

A unique person can be identified by an ID, such as name and identity card number. The "name" in Table 1 belongs to ID.

(2) Quasi-identifier (QI) attribute

The combination of multiple QI attributes can significantly enhance the likelihood of identifying a specific individual, whereas a single QI attribute, such as age, gender, nationality, is insufficient for recognizing individual identities. In Table 1, "age", "gender", and "ZIP code" are QI attributes.

(3) Sensitive attribute (SA)

SAs are attributes that contain personal privacy data, such as disease, address, income, occupation, etc. In Table 1, "disease" is classified as an SA.

(4) Non-sensitive attribute (NSA)

NSAs are common data, with the exception of the three categories of attributes listed above, and they can be published directly.

Commonly used anonymization methods include generalization and concealment.

(1) Generalization: it can be divided into the following two types.

① Value generalization: a specific value is replaced with a more general value. Taking the "age" item in Table 1 as an example, the treatment process is shown in Fig. 1.

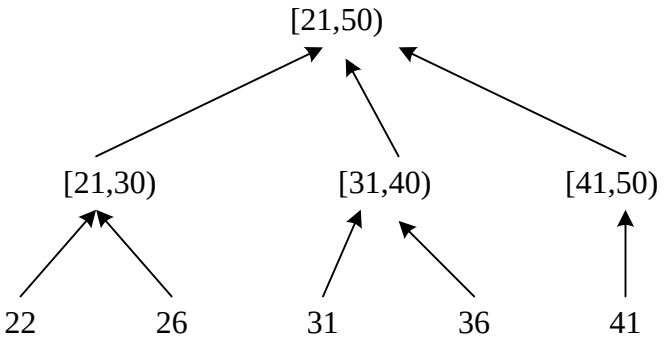

Fig. 1 Value generalization

② Domain generalization: replace the domain of an attribute with a more abstract domain, such as "ZIP code" in Table 1, the process is as follows:

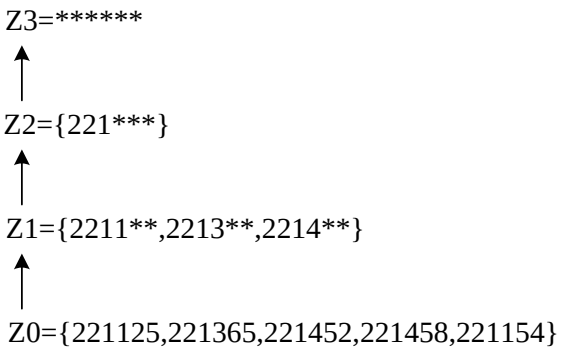

Fig. 2 Domain generalization

(2) Concealment: it means special generalization operations, i.e., direct deletion of the original data, or replacement of the original data with a particular symbol.

2.2 Anonymity algorithm

Currently, the commonly used anonymity algorithms are as follows.

(1) K-anonymity model

If any single record t is identical with at least k-1 records in terms of QI in a data table T', then table T' satisfies the k-anonymity rule. The 2-anonymization processing is performed on Table 1, and the results are shown in Table 2.

Table 2

2-anonymous data table T'

EC ID	Age	Gender	ZIP code	Disease
1	[21,30]	F	221***	Flu
	[21,30]	F	221***	HIV
2	[31,40]	M	221***	Heart disease
	[31,40]	M	221***	cancer
3	[41,50]	F	221***	Flu

In Table 2, the term “EC” refers to the equivalence class, which represents a set of data records sharing the same QI values. For example, according to Tables 1 and 2, if the attacker finds that the age, gender, and ZIP code of Alice are [21,30), female, and 221125, then according to Table 2, two records match these QI values. Consequently, the attacker will be unable to ascertain Alice's disease-related information.

(2) l-diversity model

To expand the amount of data, suppose there is a 2-anonymous table containing eight pieces of data (Table 3).

Table 3

2-anonymous data table

ID	Age	Gender	ZIP code	Disease
1	[21,30]	Person	2213**	Flu
2	[21,30]	Person	2213**	Flu
3	[21,30]	Person	2213**	cancer
4	[31,40]	Person	2214**	Flu
5	[31,40]	Person	2214**	cancer
6	[31,40]	Person	2214**	Heart disease
7	[41,50]	Person	221***	HIV
8	[41,50]	Person	221***	HIV

Suppose an attacker obtains an external information table (Table 4).

Table 4

External information

Name	Age	Zip code
A	33	221456
B	45	221457

The analysis of Tables 3 and 4 reveals that within the last equivalence class, i.e., the data group where IDs 7 and 8 are located, there is only one sensitive attribute value, "HIV". If the attacker knows the other QI values of the user, it becomes possible to deduce that the user is afflicted with HIV. Consequently, this results in a breach of the user's private data, highlighting certain deficiencies in k-anonymization.

Based on k-anonymity, l-diversity requires that each EC must contain at least one sensitive attribute value, which further reduces the possibility of an attacker obtaining sensitive user data. A data table that satisfies 2-diversity is shown in Table 5.

Table 5

2-diversity data table

ID	Age	Gender	ZIP code	Disease
1	[21,30]	Person	2213**	Flu
2	[21,30]	Person	2213**	Flu
3	[21,30]	Person	2213**	Cancer
4	[31,40]	Person	2214**	Flu
5	[31,40]	Person	2214**	Cancer
6	[31,40]	Person	2214**	Heart disease
7	[41,50]	Person	221***	HIV
8	[41,50]	Person	221***	Pneumonia

According to Tables 4 and 5, it can be seen that the QI values within each equivalence class are the same after l-diversity processing. Furthermore, an equivalence class will not contain only one sensitive attribute, thereby enhancing its resilience against potential attacks.

(3) (α, k)-anonymity model

The (α, k)-anonymity model is an extension of the k-anonymity model. On the basis of k-anonymity, a sensitive attribute is considered to satisfy (α, k)-anonymity if the occurrence frequency of each attribute value is less than α within each equivalence class. α in Table 6 is set to 0.5.

Table 6

(0.5,2) anonymous table

Age	Gender	ZIP code	Disease
[21,30]	F	221***	Flu
[21,30]	F	221***	HIV
[31,40]	M	221***	Flu
[31,40]	M	221***	Cancer

The above anonymity models were analyzed (Table 7).

Table 7

Analysis of anonymity algorithms

	K-anonymity model [13]	l-diversity model [14]	(α, k)-anonymity model [15]
Merits	Good performance in resisting connection attacks	Ensuring diversity of sensitive attributes within each equivalence class	Consideration for sensitive attributes is added, enhancing the level of privacy protection.
Demerits	Diversity of sensitive attributes not taken into account	Difficult to resist similarity attacks	Only applicable to data tables with uniformly distributed sensitive attribute values

The anonymity algorithms in Table 7 have been widely applied in various scenarios such as e-commerce and product development. It can be observed that these anonymization algorithms can achieve a certain level of data protection. However, with the continuous improvement of practical requirements, these algorithms are still insufficient in meeting the actual needs for data protection. The l-diversity model and the (α, k)-anonymity model partially overlook the similarity of sensitive attributes during the process of anonymization. The presence of HIV and cancer in the same equivalence class poses a high risk to privacy, as these attributes are both sensitive. In such cases, an attacker can make educated guesses about the user's serious medical condition, leading to potential privacy breaches. Conversely, less critical attributes like colds or toothaches have lower protection requirements and would not significantly impact privacy even if an attacker speculates their approximate range. The imposition of distinct constraints on various values of sensitive attributes is imperative from this perspective.

3. Research methods

From the current perspective of anonymity algorithms, there are still some shortcomings in terms of data protection effectiveness, which cannot fully meet the needs of enterprise human resource data protection on actual network platforms. Therefore, in order to obtain a more reliable method for data protection, this paper improves the existing l-diversity model and proposes an (α,l)-anonymity model that incorporates sensitivity.

If anonymous table T' satisfies l-diversity, and in each EC, every equal sensitivity group SD_i satisfies:

$$\frac{|SD_i|}{|E|} \leq \alpha_i, \tag{1}$$

$$\alpha_i = 1 - D_i, \tag{2}$$

then it satisfies the (α, l) anonymity combined with sensitivity.

In the equation, $|E|$ refers to the number of tuples in the EC, and $0 < \alpha_i < 1$. For each sensitive attribute value s_i, a sensitivity level is set, and the value obtained after quantifying this level is denoted as D_i, $0 < D_i < 1$.

The specific procedure of the (α, l) anonymity algorithm combined with sensitivity is shown below.

(1) For data table T, where the number of sensitive attribute values is n, the values are divided into multiple tuple sets (t_i), and each tuple set has the same sensitivity.

(2) The set of tuples is ranked in descending order according to D_i: $t_1, t_2, \cdots, t_n$.

(3) If $t_i (i = 1,2, \cdots, n)$ is a non-empty set, any record (r) is selected from t_i to create a cluster, quasi-cluster $e' = \{r\}$. Then, one record is selected from each of the $1 \sim l$ sets according to the sensitivity and added to e' to make e' satisfy (α, l) anonymity combined with sensitivity and minimize the increased information loss.

(4) Step (3) is repeated. Then, quasi-cluster e' is added to the cluster set as cluster e, and the extracted records are removed from the original set until the remaining set of non-empty tuples is less than one.

(5) Any record (r) is selected from the remaining tuples. Any cluster (e) is taken out from the cluster set. After r is added to the cluster, it should still satisfy (α, l) anonymity combined with sensitivity, and the increased information loss should be smallest; otherwise, r is concealed.

(6) Return to the set of clusters and generalize the QI values of each cluster to form eligible anonymous table T'.

4. Analysis of results

4.1 Experimental setup

The experimental environment was Windows 1 operating system, Intel(R) Core(TM) i7-9750 CPU@2.6 GHz 2.59GHz, and 8GB memory. Python 3.6 was employed as the programming language.

The sensitive data that require protection in enterprise human resource records include salary, occupation, and other relevant information. In order to facilitate comparative analysis with existing methods, this study conducted experiments using the Adult dataset from UCI (data source: http://archive.ics.uci.edu/ml/datasets/Adult), which has a certain degree of correlation with enterprise human resources data. The Adult dataset consists of U.S. Census data with 14 attribute variables such as age, gender, and race. Occupation was taken as the sensitive attribute, and five other attributes were selected for the experiments. A total of 40,000 tuples were arbitrarily selected from the dataset to serve as experimental data. Based on the prevalence of different occupations in real

society, their sensitive levels were divided, and the degree of sensitivity was determined. The results are shown in Table 8.

Table 8

Sensitivity values for occupation

Serial number	Occupation	Degree of sensitivity
1	Adm-clerical	0.1
2	Sales	0.2
3	Other-service	0.2
4	Tech-support	0.3
5	Craft-repair	0.3
6	Farming-fishing	0.4
7	Machine-op-inspect	0.4
8	Transport-moving	0.4
9	Exec-managerial	0.6
10	Prof-specialty	0.6
11	Protective-serv	0.7
12	Handlers-cleaners	0.7
13	Priv-house-serv	0.7
14	Armed-forces	0.7

The generalization treatment of the other attributes is as follows.

(1) Age: there are a total of 74 attribute values and four generalization levels.

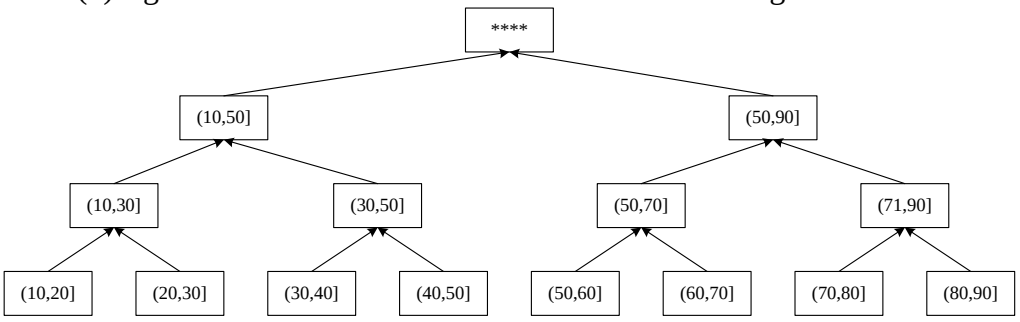

Fig. 3 Hierarchical tree of "age" generalization

(2) Gender: there are a total of two attribute values and three generalization levels.

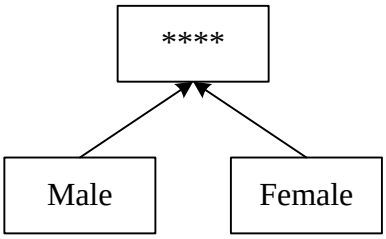

Fig. 4 Hierarchical tree of "gender" generalization

(3) Race: there are five attribute values and two generalization levels.

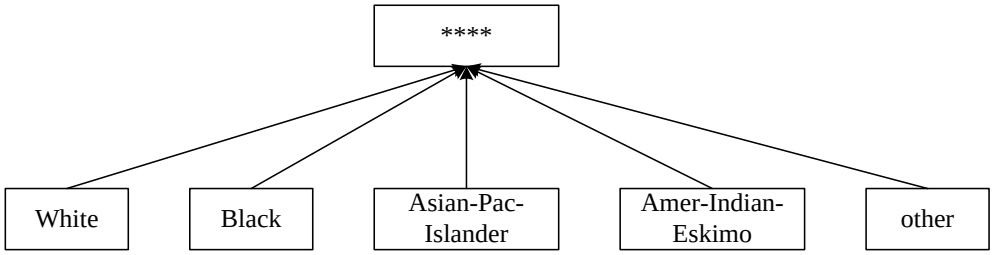

Fig. 5 The "race" generalized hierarchical tree

(4) Marital-status: there are seven attribute values and three generalization levels.

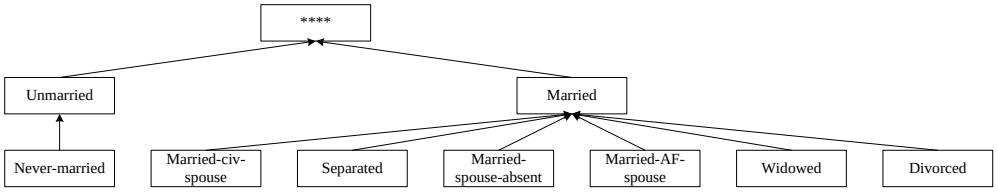

Fig. 6 Hierarchical tree of "marital-status" generalization

(5) Relationship: there are six attribute values and three generalization levels.

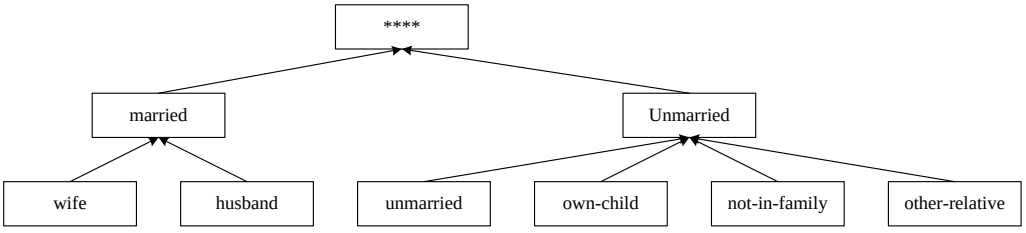

Fig. 7 Hierarchical tree of "relationship" generalization

The performance of the algorithm was evaluated by measuring the loss of information, which refers to the deficiency or distortion in data quality between the anonymous table and the original data table. The calculation of different QIs is as follows.

(1) Numerical attributes: for original attribute A, the QI value is$(x_1, x_2, \cdots, x_n)$, the anonymization process yields $([a_1, b_1], [a_2, b_2], \cdots [a_n, b_n])$. Then, the information loss of this QI is:

$$TIL_{(A)}(x) = \frac{b_i - a_i}{|A_i|}, \tag{3}$$

where $|A_i|$ is the value field of attribute A.

(2) Classification attributes: for original data A, the QI value is $(x_1, x_2, \cdots, x_n)$, and the anonymization process yields $(c, c_2, \cdots, c_n)$. Then, the information loss of this QI is:

$$TIL_{(A)}(x) = \frac{h(c_i)}{|A_i|}, \tag{4}$$

where $h(c_i)$ is the set of all leaf nodes in the subtree with c_i as the root node and $|A_i|$ is the number of all nodes in attribute A.

Ultimately, the information loss of tuple t and data table T is:

$$TIL(t) = \sum_{i=1}^{n} TIL_{(A_i)}, \tag{5}$$

$$TIL(T) = \sum_{t \in T} TIL(t). \tag{6}$$

3.2 Analysis of results

Firstly, the information loss rate of the (α, l) anonymity model combined with sensitivity was analyzed. The variation in the amount of information loss of the algorithm under different values of k and l is presented in Fig. 1.

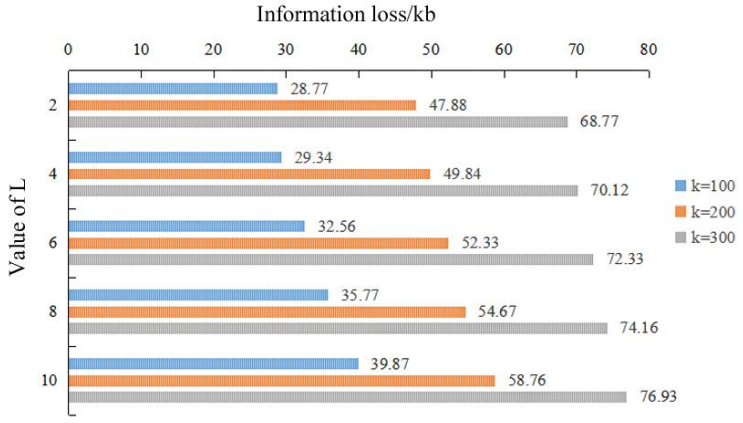

Fig. 8 Effect of k and l values on the amount of information loss

According to Fig. 8, it can be observed that, under the condition of the fixed k value, the information loss of the (α, l) anonymity model combined with sensitivity increased with the increase of l value. Taking k = 100 as an example, the information loss of the algorithm at l = 2 was 28.77 kb; when the value of l was increased to 4, 6, 8, and 10, the information losses were 29.34 kb, 32.56 kb, 35.77 kb, and 39.87 kb, which were increased by 1.98%, 13.17%, 24.33%, and 38.58%, respectively, as compared to that at l = 2.

Then, under the condition of the fixed l value, the amount of information loss of the model combined with sensitivity increased with the increase of k value. Taking l = 2 as an example, the information loss of the algorithm at k = 100 was 28.77 kb. When the value of k was increased to 200 and 300, the information losses were 47.88 kb and 68.77 kb, which were increased by 19.11% and 139.03%, respectively, compared to that at k = 100. It was concluded from Fig. 8 that as the values of k and l increased, the degree of data anonymization improved, and the information loss became larger.

Then, the effect of the number of QIs on the amount of information loss was analyzed while keeping the values of k and l constant. The model combined with sensitivity was compared with the l-diversity algorithm under conditions where k = 200, l = 6, and the number of QIs ranged from 2 to 5. The results are shown in Fig. 9.

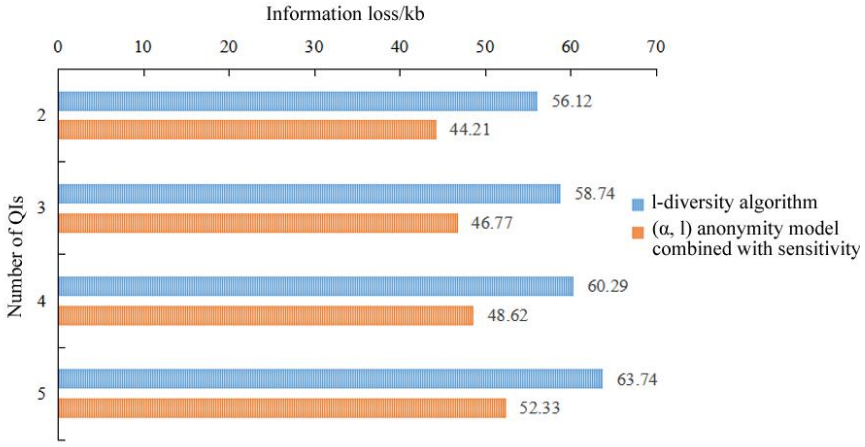

Fig. 9 Effect of the number of QIs on the amount of information loss

As shown in Fig. 9, the amount of information loss gradually increased as the number of QIs increased, both in the l-diversity algorithm and in the model combined with sensitivity. Taking the model combined with sensitivity as an example, the amount of information loss was measured at 44.21 kb when the number of QIs was set to 2. This value increased to 46.77 kb, 48.62 kb, and 52.33 kb when the number of QIs was set to 3, 4, and 5 respectively. Notably, compared

to QI = 2, there was an increase in information loss by approximately 18.37% at QI = 5. These results showed that as the number of QIs increased, the amount of data to be anonymized increased, and therefore, the amount of information loss in the algorithm also increased. The comparison between the two methods showed that the amount of information loss of the l-diversity algorithm was significantly higher than that of the model combined with sensitivity. For example, when the number of QIs was 5, the amount of information loss of the l-diversity algorithm was higher than that of the model combined with sensitivity, which indicated that the model combined with sensitivity was more effective in safeguarding data.

The running time of the algorithm was analyzed by performing experiments under different data volumes. When k = 100, l = 2, QI = 5, the comparison of the running time is shown in Fig. 10.

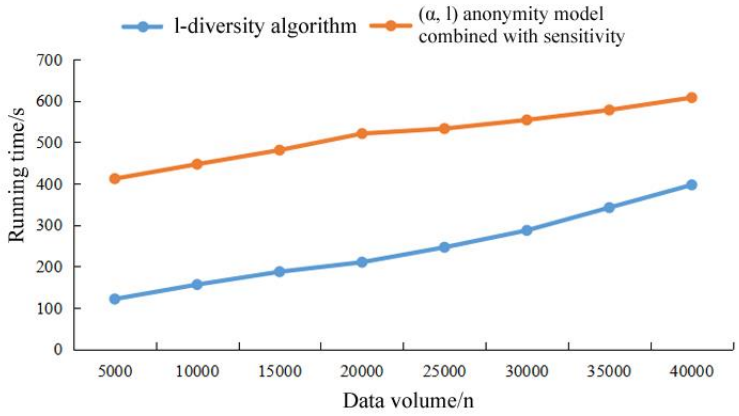

Fig. 10 Comparison of running time under different data volumes

As shown in Fig. 10, the running time of both methods continued to increase with the increase of data volume. In comparison, firstly, the running time of the model combined with sensitivity was higher than that of the l-diversity algorithm. This can be attributed to the fact that during anonymization, the model intensified data anonymization to enhance data protection. Then, the comparison of the two methods showed that the l-diversity algorithm always maintained a large increase in the running time with the increase of the amount of data, whereas the model combined with sensitivity demonstrated a relatively gentler increase in running time when influenced by larger amounts of data. The results showed that the time overhead in the model combined with sensitivity increased to some extent, in order to strengthen the protection of the enterprise's human resource data; however, this model demonstrated improved stability when confronted with expanding data volumes.

Then, data were collected from real online recruitment platforms in order to verify the applicability of the proposed method in protecting human resources data

on actual online platform enterprises. Enterprise human resource data were crawled from the network recruitment platforms such as Zhaopin (https://www.zhaopin.com/) and 51job (https://tel.51job.com/) through crawler technology between October 2022 and December 2022. After eliminating duplicates and incomplete values, 11,254 pieces of data were obtained, and part of the data is shown in Table 9.

Table 9

Enterprise human resources data

Serial number	Gender	Age	Education background	Salary	Position
1	Male	45	Doctor's degree	20,000	Front-end developer
2	Female	28	Master's degree	15,000	Graphic designer
3	Female	31	Master's degree	18,000	Sales manager
4	Male	35	Master's degree	19,000	Electronics engineer
5	Male	27	High school degree	6,500	Business merchandiser

The collected data were anonymized using the (α, l) anonymity model combined with sensitivity. Taking "job" as the sensitive attribute, when k = 2 and l = 2, part of the anonymous data is presented in Table 10.

Table 10

Anonymous table of enterprise human resources data

Serial number	Gender	Age	Education background	Salary	Position
1	Person	(40,50]	Higher education	2****	Front-end developer
2	Person	(20,30]	Higher education	1****	Graphic designer
3	Person	(30,40]	Higher education	1****	Sales manager
4	Person	(30,40]	Higher education	1****	Electronics engineer
5	Person	(20,30]	Secondary education	6***	Nurse

According to Table 10, after anonymization, the leakage risk of enterprise human resource data was significantly reduced. For example, when examining the 3rd and 4th datasets, it was observed that the two datasets were consistent in quasi-identifier attributes such as gender and age. If the attacker obtains the user's quasi-

identifier information by using the external information, he still cannot determine whether the user's position is a sales manager or an electronics engineer. This proved the reliability of the model combined with sensitivity in protecting enterprise human resource data.

4. Conclusion

In this paper, an anonymity algorithm was applied to protect enterprise human resource data on the network platform. Additionally, it employed the (α, l) anonymity model combined with sensitivity to strengthen the protection of private human resource data. Through the experiments with "occupation" as the sensitive attribute, it was found that the information loss of the model combined with sensitivity increased with the increase of the values of k and l and the number of QIs. However, when compared to the l-diversity algorithm, the model combined with sensitivity exhibited lower information loss and greater stability. As a result, it achieved superior anonymization effects for enterprise human resource data on actual network platforms. Therefore, the (α, l) anonymity model combined with sensitivity can be applied to actual network platforms to protect enterprise human resource data.

The study has yielded certain findings and demonstrated the efficacy of the proposed method in safeguarding enterprise human resource data. However, there are still some limitations to consider:

(1) it solely focuses on scenarios with a single sensitive attribute and does not explore cases involving multiple sensitive attributes;

(2) it primarily concentrates on protecting static data without considering dynamic data;

(3) the experimental scope is relatively limited, leaving the feasibility of this method in safeguarding data across other domains uncertain.

In future work, based on the method described in this paper, analysis can be conducted on multiple sensitive attributes and dynamic data scenarios to expand the application domains of the proposed approach and further enrich research on anonymity algorithms.

R E F E R E N C E S

[1]. W. Ren, “A Study on College Graduates' Employment Problem in The Context of Big Data Based on the Event of COVID-19”, ICBDE'21: 2021 4th International Conference on Big Data and Education, 2021, pp. 88-91.

[2]. C. Brando, R. Silva, and J. V. dos Santos, “Online recruitment in Portugal: theories and candidate profiles”, Elsevier, **vol. 94**, no. Jan., 2019, pp. 273-279.

[3]. X. Lili, “New anonymous attribute encryption algorithm in cloud computing environment”, IPPTA, **vol. 30**, no. 7, 2018, pp. 536-540.

[4]. S. S. Kumar, A. R. Reddy, B. S. Krishna, J. N. Rao, and A. Kiran, "Privacy Preserving with Modified Grey Wolf Optimization Over Big Data Using Optimal K Anonymization Approach", J. Interconnect. Netw., **vol. 22**, no. Supp01, 2022.
[5]. A. Abbasi, and B. Mohammadi, "A clustering-based anonymization approach for privacy-preserving in the healthcare cloud", Concurr. Comp. Pract. E., **vol. 34**, no. 1, pp. 1-18, 2021.
[6]. J. Mandala, and M. V. P. C. S. Rao, "Privacy preservation of data using crow search with adaptive awareness probability", J. Inf. Secur. Appl., **vol. 44**, no. FEB., pp. 157-169, 2019.
[7]. S. U. Bazai, J. Jang-Jaccard, and H. Alavizadeh, "Scalable, High-Performance, and Generalized Subtree Data Anonymization Approach for Apache Spark", Electronics, **vol. 10**, no. 5, pp. 1-28, 2021.
[8]. R. Belmar-Monterrubio, J. E. Quiroz-Ibarra, and F. Cervantes-Sodi, "A versatile mathematical function for generating stable and chaotic systems: A data encryption application", Chaos Soliton. Fract., **vol. 167**, pp. 1-5, 2023.
[9]. A. B. Pawar, S. U. Ghumbre, and R. M. Jogdand, "Privacy preserving model-based authentication and data security in cloud computing", Int. J. Pervasive Comp., **vol. 19**, no. 2, pp. 173-190, 2023.
[10]. F. Thabit, O. Can, S. Alhomdy, G. H. Al-Gaphari, and S. Jagtap, "A Novel Effective Lightweight Homomorphic Cryptographic Algorithm for data security in cloud computing", Int. J. Intell. Netw., **vol. 3**, pp. 16-30, 2023.
[11]. A. Rodríguez-Hoyos, J. Estrada-Jiménez, D. Rebollo-Monedero, A. Mohamad-Mezher, J. Parra-Arnau, and J. Forné, "The Fast Maximum Distance to Average Vector (F-MDAV): An algorithm for k-anonymous microaggregation in big data", Eng. Appl. Artif. Intell., **vol. 90**, no. C, pp. 1-20, 2020.
[12]. A. H. Aljammal, H. Bani-Salameh, A. Qawasmeh, A. Alsarhan, and A. F. Otoom, "A new technique for data encryption based on third party encryption server to maintain the privacy preserving in the cloud environment", Int. J. Bus. Inform. Syst., **vol. 28**, no. 4, pp. 393-403, 2018.
[13]. A. Rodríguez-Hoyos, J. Estrada-Jiménez, D. Rebollo-Monedero, A. Mohamad-Mezher, J. Parra-Arnau, and J. Forné, "The Fast Maximum Distance to Average Vector (F-MDAV): An algorithm for k-anonymous microaggregation in big data", Eng. Appl. Artif. Intell., **vol. 90**, no. C, pp. 1-20, 2020.
[14]. A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity", ACM T. Knowl. Discov. D., **vol. 1**, no. 1, pp. 1-24, 2007.
[15]. R. Wong, J. Li, A. Fu, and K. Wang, "(α, k)-anonymous data publishing", J. Intell. Inf. Syst., **vol. 33**, no. 2, pp. 209-234, 2009.