Security and Privacy Data Protection Methods for Online Social Networks in the Era of Big Data

Lei Ma(✉) and Ying-jian Kang

Telecommunication Engineering Institute, Beijing Polytechnic, Beijing, China
malei235@tom.com

Abstract. In order to improve the security of online social network security and privacy data and shorten the delay of privacy data protection, this paper proposes a method of online social network security and privacy data protection in the era of big data. In order to enhance the security of online social network security and privacy data, a network security and privacy data architecture is constructed. In view of the risk of online social network security and privacy data loss, this paper proposes a set of security and privacy data backup processing scheme. Complete the formulation of online social network security and privacy data protection scheme; Combined with the current attack methods commonly used by attackers, the shortcomings of the traditional privacy protection algorithm in protecting the security of online social network privacy data security are concluded. Design online social network security privacy data privacy protection algorithm; Finally, complete homomorphic encryption of online social network security and privacy data is adopted to realize online social network security and privacy data protection in the era of big data. Experimental results show that the proposed privacy data protection method has a shorter delay than the traditional one.

Keywords: Big data era · Online social · Network security · Private data · Protection scheme

1 Introduction

With the rapid growth and evolution of these social networking sites, many platforms have evolved into a complex network, and social networking data has been applied to various fields [1]. In the traditional user login, the user is regarded as an independent individual. With the emergence and wide application of third-party login, for example, the user binds a microblog account in toutiao, and after logging in toutiao, he can get the information of the user's friend relationship and friend dynamics in the microblog account. Therefore, the security of social network data needs to be paid close attention to. Privacy data in social networks mainly include published user information (such as name, age, geographical location, graduate school, etc.) and information of friends among users [2]. The source of the social network privacy problem is that when the Owner of social network service provider and user Data DO (the Data Owner is not in the same trust domain, at this point in the social network Data privacy will be out of the control range of the DO, then the user Data including all the information provided by the

Y.-D. Zhang et al. (Eds.): ICMTEL 2020, LNICST 327, pp. 53–66, 2020.
https://doi.org/10.1007/978-3-030-51103-6_5

SNSP will face the danger of privacy Data leaked and security threats [3]. In addition, in the era of data sharing and in the field of scientific research, data in social networks need to be released and collected in large quantities. In the process of such data dissemination, it is inevitable that privacy information will be misappropriated or abused. For example, some famous SNS websites such as Facebook and Twitter have been reported to leak or lose users' private data, which has also led to serious consequences.

Considering the social network is a keep in touch, and share information with others for the purpose of network of communication media platform virtualization, obviously serious privacy protection problems, and based on its vast number of users, the traditional method on the protection of the encryption technology is far from alone meet user requirements for data protection and sharing. Therefore, it is necessary to balance the validity of data and the protection intensity of user privacy data to ensure the safe release of user privacy data. In addition, in the project of new media broadcasting, the privacy protection work under ELGG social platform is only done in the node encryption, and the protection mechanism of edges and graphs is not perfect. Therefore, it is necessary to study the privacy protection technology combining nodes, edges and graphs under the social network platform to achieve the overall privacy protection function under the social network.

2 Design of Online Social Network Security and Privacy Data Protection Methods

2.1 Develop Online Social Network Security Privacy Data Protection Program

In order to enhance the security and availability of online social network security and privacy data and shorten the delay of privacy data protection, a network security and privacy data architecture diagram is constructed in the era of big data, as shown in Fig. 1.

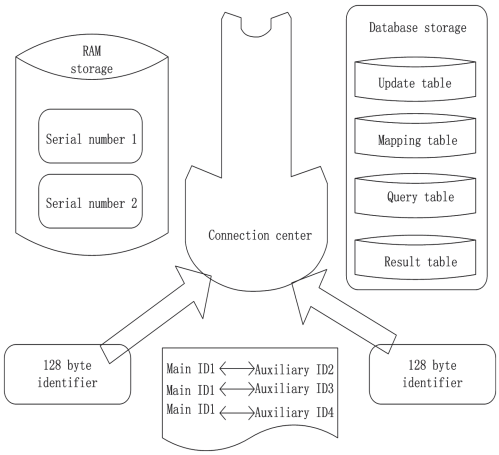

Fig. 1. Network security and privacy data architecture in the era of big data

Each user is given a special ID, and each primary ID can identify itself as well as multiple secondary ids. The primary ID is used when the owner of the secure private data of an online social network USES the secure private data of an online social network. When not the owner of online social network security and privacy data USES the data, he/she needs to log in with auxiliary ID. All ids are one master ID, but only some are auxiliary ids, and each subsidiary ID can only assist one master ID, but multiple auxiliary ids can be selected [4].

According to the network security and privacy data architecture in the era of big data designed above, in order to make the online social network security and privacy data safe and reliable, real-time backup of the online social network security and privacy data must be conducted. In the era of big data, online social networks have a high concentration of secure and private data. Unlike the previous online social networks, the secure and private data are backed up by the data owners themselves. In the era of big data, the centralized storage of online social network security and privacy data increases the risk of loss of online social network security and privacy data. Once the server is paralyzed, the loss is immeasurable. This section proposes a backup processing method that can backup private data at any time and enhance the security of private data [5].

The proposed backup solution means that the online social network security and privacy data owner determines a fully trusted person who is responsible for the online backup of the online social network security and privacy data and synchronizes with the data owner. In other words, let the user make sure that a fully trusted third party is hosting his copy, so that the online social network security privacy data has a guarantee, will not lose data due to the failure of their own server. When a third party loses data, it is more obvious to the user who is responsible. The master ID sends a duplicate message to the third party agent, who updates the mapping table to record the mapping of the master ID. Can modify the mapping relationship between main ID at any time, but when the main ID modify the mapping table, that is changed corresponding to the third party agent, a copy of the third party agent before storage is not stored in the new third party agent under, but if the primary ID want to restore all copies, you can restore all haven't delete all copies of [6].

In order to better understand the program flow, the program flow chart is analyzed, as shown in Fig. 2.

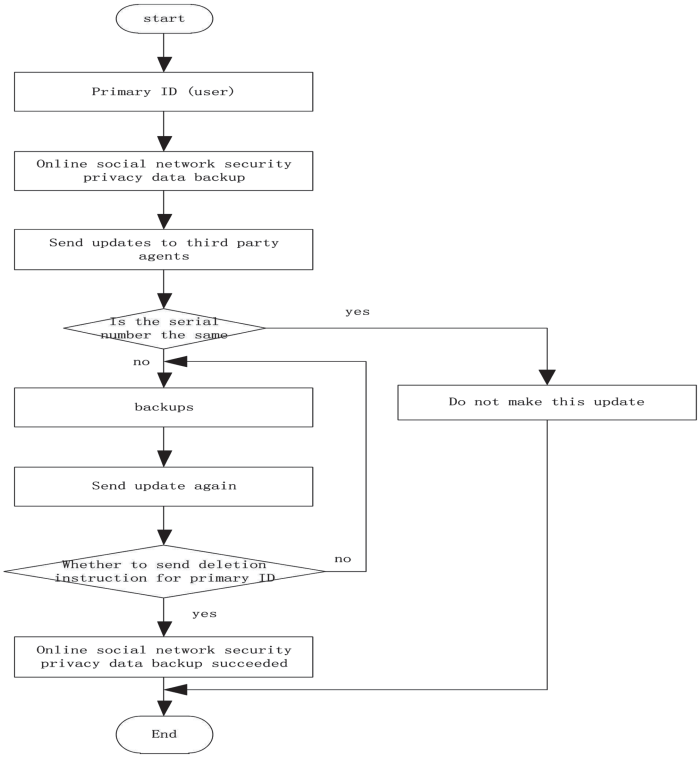

Fig. 2. Flow chart of online social network security and privacy data backup processing

The newly inserted online social network security and privacy data is stored as temporary data in the third-party agent. The primary ID periodically tracks to obtain its latest copy sequence, and the primary ID obtains the latest copy sequence to compare with the previous sequence. If the serial number is found to be the same, the master ID sends a deletion request to the third-party agent, who will delete all the updates [7].

The third-party agent is responsible for storing online social network security and privacy data information, and keeps it updated by sending messages to the master ID [8]. The third party receives the update request from the master ID and records the update in the update table. If the sequence number sent by the master ID is less than or equal to the last sequence number, the third party will not insert the update information into the update table. Conversely, if the sequence number is greater than the last sequence number, it is updated to the update table.

If the online social network security and privacy data is not successfully backed up, the serial number of the copy of the primary ID to the agent and the secondary ID are both on the agent and the secondary ID, which means that the serial number will not be lost easily. After each backup to the third party agent, the master ID will automatically send the update information to the third party again until the message sent by the master ID is a delete instruction.

The third party agent only assists in temporary storage during the whole process, and the version consistency is completely determined by the master ID.

In the era of big data, data service providers like bank store of money people store online social network security data privacy, and we have put forward the third party backup agent as high safety in the bank safe keeping money, backup the online social network security data privacy, such a backup plan to ensure that the integrity of the online social network security data privacy and security [9].

In the era of big data, users can query personal online social network security and privacy data anytime and anywhere, which brings a great impact on online social network security and privacy data protection in the era of big data. In view of such access mode anytime and anywhere, a query processing scheme controlled by access rights is designed [10].

Firstly, the query process is briefly described. When querying a user's online social network security and privacy data, it can be obtained by the personal master ID or any auxiliary ID. In this scheme, an embedded built-in relational database is adopted to process the query request. The query statements that apply to this scenario are given below.

[Q1:] SELECT COUNT(*) FROM her WHERE diagnosis = "friend" AND doctorName = "liu"
[Q2:] SELECT DATE FROM her WHERE diagnosis = "Friend" AND doctorName = "liu"

Two main roles in the query process are represented by the requester and the respondent. When a requester requests a query, the respondent may be a primary ID or a secondary ID. the query process is described as shown in Fig. 3.

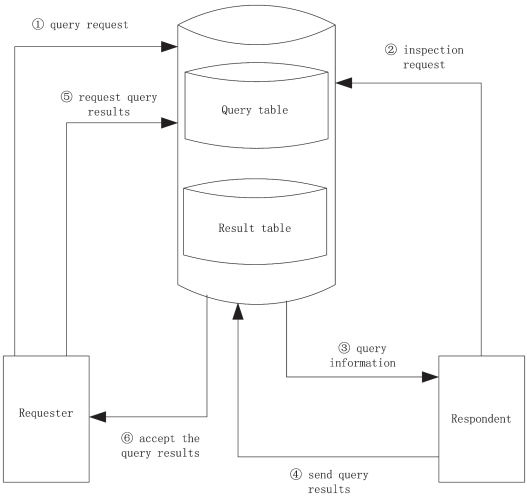

Fig. 3. Online social network security privacy data query process

Step 1: the requester initiates the request, and the request information is stored in the request form.
Step 2: the respondent accepts the request information and checks the rationality of the request.
Step 3: the respondent receives the information from the requester.
Step 4: the respondent feeds back the query content of the requester.
Step 5: send the query response and check whether the respondent has made information feedback.
Step 6: the requester receives the feedback from the respondent.

In order to increase the security of privacy data of online social network, the authority control method is used here, as shown in Table 1.

Table 1. Permission settings for categories of online social network security privacy data

Online social network security privacy data	Class A	Class B	Class C
Data 1	Whole	Only user primary ID is allowed	Allow secondary ID
Data 2	Whole	Only user primary ID is allowed	Allow secondary ID
Data 3	Whole	Only user primary ID is allowed	Allow secondary ID
Data 4	Whole	Only user primary ID is allowed	Allow secondary ID
Data 5	Whole	Only user primary ID is allowed	Allow secondary ID
Data 6	Whole	Only user primary ID is allowed	Allow secondary ID
Data 7	Whole	Only user primary ID is allowed	Allow secondary ID

According to the requirements of permission, relevant managers classify the specific online social network security privacy data, and users can set permission in their own online social network security privacy data interface. It can be seen from the above table that the content of the user's online social network security privacy data includes three main parts of data. Type a data is the online social network security privacy data that the user does not want anyone to see except himself. Type B data is the data that the user wants specific users to see, but does not want other users to see. Type C data is the data that the user specifies some or some users can see.

Set an identification table in the data online social network security privacy database. When the user checks the check box in the table, the online social network security privacy data in the identification table changes from 0 to 1, and only one check box in the column set in the foreground can be checked.

When anyone logs in to the online social network security privacy data system, the system will determine the identity of the login. When ordinary users log in, the built-in query statement is:select Online social network security privacy data from General table where Summary table. Data 1 = Identification table data. Identification table A; If a secondary ID is logged in, the built-in query statement is:select Online social network security privacy data from General table where General table. Data 1 = Identification table data.Identification table A;

Similarly, this is only for the rights control of viewers, and other rights control can be done, such as the rights control on the query location.

In this paper, we put forward some protection schemes of online social network security privacy data based on the era of big data. First, we put forward an auxiliary system construction. Then, on the basis of this construction, we put forward a proxy backup processing method to increase security, and then we put forward a query processing scheme, and analyze the use of authority control in query. In the end, the author suggests that our country should make laws on online social network security and privacy data security. Next, through online social network security privacy data privacy protection algorithm, to improve the security of online social network security privacy data.

2.2 Online Social Network Security Privacy Data Privacy Protection Algorithm Design

Based on the online social network security privacy data protection scheme developed in the era of big data, combined with the common attack means of attackers at present, the shortcomings of the traditional privacy protection algorithm in protecting the online social network security privacy data security are obtained, and the privacy data protection delay is shortened. The traditional anonymization algorithm can achieve anonymity and diversification, but it can not meet the setting problem of different users for different online social network security privacy data attributes. At the same time, if the online social network security privacy data attributes are set with unified anonymization operation, it will also increase the efficiency of the algorithm.

Because the perturbation method based on spectrum constraint is relatively fixed, it will maintain a high availability of online social network security privacy data after being processed by this method. However, because the perturbation process of this method is relatively simple, the degree of online social network security privacy data protection is very low.

The availability of s-spectrum switch method is similar to that of spectrum constraint based perturbation method in social networks. Due to the randomization of the edge perturbation algorithm, its privacy protection is improved. However, if the map features are unchanged, the significance of the map features may change greatly. Therefore, the measurement standard of online social network security privacy data restricted by spectrum radius is deceptive. At the same time, the feature significance of graph can be used as the measurement theory of graph structure.

Based on the deceptive problem of spectral radius, a perturbation method is proposed, which combines the spectral radius constraint of graph with the feature significance of graph. In this method, the balance between the availability and security of the online social network security privacy data after randomization is guaranteed by adjusting the significant changes of the characteristics of the online social network security privacy data graph, the harmonic mean shortest distance of the graph and the value of the spectral radius of the graph. However, there are obvious disadvantages in the implementation process of this method: high calculation cost, because every random edge modification operation will need to do many times of graph eigenvalue calculation, however, in the graph scrambling algorithm, the edge information must be modified repeatedly, and the proposed algorithm has no high application value in the use of the project.

Based on the previous requirement analysis and data mining related knowledge, this paper proposes the fine-grained attribute anonymity algorithm, group based node division method and group based edge randomization algorithm in online social network security privacy data to achieve the online social network security privacy data protection.

In this part, from the user's point of view, the attributes of the nodes in the online social network security privacy data are protected. In the past, the operations of anonymity have been refined to the records in the data table at the data table level, so as to meet the different needs of different users.

In this paper, we protect the privacy attributes of users, fully consider the different requirements of different users for anonymity, and ensure the k-anonymity model of privacy attributes. The method adopted is the combination of personalized anonymity model and k-anonymity model, that is, fine-grained attribute anonymity algorithm. The method used is the combination of concealment and generalization in data mining. Concealment is to remove all values or single attribute values of a tuple. Generalization is to replace the original property value with a larger range of values. Concealment can reduce the amount of data generalization, thus reducing the loss of data.

In this algorithm, P is defined as the user's attribute, K as the degree of anonymity, and ε as the level of attribute generalization.

Algorithm idea: first, we need to traverse the leaf nodes of each attribute. We can use the depth search algorithm to determine which layer of parent node to replace the current leaf node value according to the preset value set by the program. If a special user sets a higher value for the generalization degree of an attribute value, the corresponding higher-level parent node is replaced in the generalization tree of the attribute. In addition, in order to reduce the information loss of data anonymity, we can adjust the result set to meet the l-diversity model at last, using clustering algorithm and Datafly algorithm. The specific algorithm steps are as follows:

Input: Data table T in database;
Output: Table T' after fine-grained attribute generalization;
Steps are as follows:
Step 1: Select the database, connect to the T data table in the database, and select some of its fields as the property collection of privacy protection.
Step 2: According to the attribute set P set by the user and the generalization degree ε corresponding to the generalization attribute, perform the corresponding generalization operation on the corresponding generalization attribute, and save the results of the generalization operation in the temporary data table t_temp table.
Step 3: Continue k-anonymization for the temporary data table in step 2, and complete the anonymization privacy protection. If l-diversity protection is needed, the final data table T' can meet the l-diversity model by setting the l-value, clustering algorithm and Datafly algorithm.

In the fine-grained attribute generalization algorithm adopted in this paper, by introducing the personal privacy constraint value PC_i, C to set the degree of generalization for the user on this attribute, so through the setting of parameter K and parameter C, the K anonymity and generalization function of the user's privacy data

can be realized. In order to further resist privacy attacks, we can also combine l-diversity algorithm to achieve the privacy protection function of user attributes.

Next, the nodes of online social network security privacy data are classified. There are three types of nodes in online social network security privacy data: free point, neutral point and conservative point. Next, the nodes of online social network security privacy data are classified. There are three types of nodes in online social network security privacy data: free point, neutral point and conservative point. By dividing the nodes, not only the complexity of the algorithm can be reduced, but also the availability of the online social network security privacy data graph can be improved after the partition because of the better response to the disturbance of the graph under the constraint of the characteristic significance of the graph. The specific classification method is as follows:

Input: Adjacency matrix M of SNS;
Output: SNS conservative point group and free point group;
Steps are as follows:
Step 1: According to the mathematical formula, the adjacency matrix M of online social network security privacy data is calculated to get the α_u of each node, that is, the coordinates in k-dimensional spectrum space.
Step 2: According to the first k eigenvectors, all nodes are divided into groups: if the components satisfying the k-th eigenvector are smaller than the given threshold, then the nodes are divided into free point groups, if larger than the given threshold, then the nodes are divided into conservative point groups.

The flow chart of the algorithm is shown in Fig. 4.

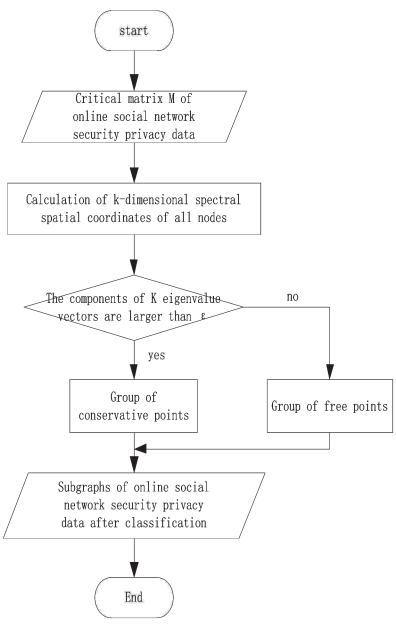

Fig. 4. Flow chart of online social network security privacy data privacy protection algorithm

Through the operation of the node division algorithm in the online social network security privacy data graph, all nodes in the online social network security privacy data graph are divided into the conservative point group and the free point group; through the setting of different K values, the different division basis of nodes can be realized, so as to form more groups.

2.3 Full Homomorphic Encryption of Online Social Network Security Privacy Data

In the era of big data, the online social network security privacy data is relatively complex, and the protection delay of the security privacy data is long. Through the full homomorphic encryption of the online social network security privacy data, the privacy data protection delay is shortened, so as to realize the online social network security privacy data protection in the era of big data. The so-called homomorphic encryption is to do any operation on the online social network security privacy data, and the result of the operation after decryption is the result of the same operation on the plaintext. The scheme is ε and the security parameter is λ. To realize homomorphic encryption, we need four algorithms: key generation algorithm (keygen), encryption algorithm (encrypt), online social network security privacy data evaluation algorithm (evaluate), decryption algorithm (decrypt). The overall working structure is shown in Fig. 5.

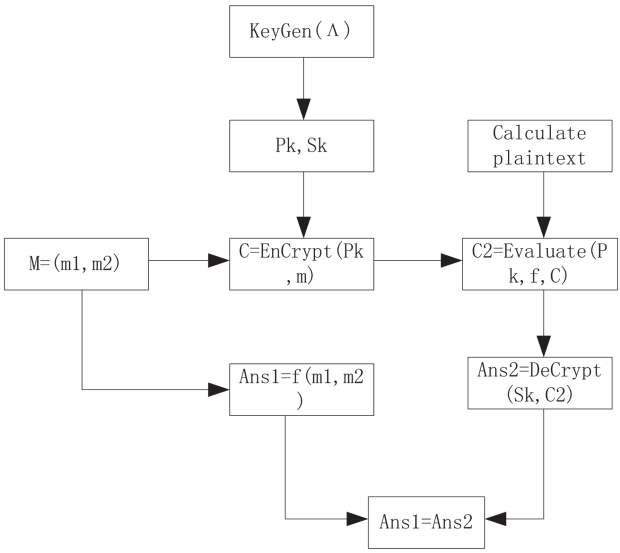

Fig. 5. Structure of homomorphic encryption

(1) the keygen algorithm generates the public key (PK) and private key (SK) according to the security parameter λ. The public key is used for encryption algorithm and evaluation algorithm, and the private key is used for decryption

algorithm. In the era of big data, the public key PK needs to be shared to the server for evaluation.

(2) encrypt the plaintext vector m, and use the encryption algorithm encrypt (PK, m) to encrypt m to get the online social network security privacy data C. In the era of big data, when the requester is allowed, the sensor data is encrypted to get online social network security privacy data C, which is sent to the requester.

(3) the evaluation function processes online social network security privacy data. When the requester gets the online social network security privacy data, it needs to process the online social network security privacy data accordingly. The evaluation algorithm can be used directly. Its input parameters include public key PK, online social network security privacy data C and corresponding operation F. Thus, the computing result C2 of online social network security privacy data is obtained.

(4) decrypt the online social network security privacy data C2 to get the calculation results. In the era of big data, it is the service provider that sends the processed online social network security privacy data to the user, and then the user uses his own private key to decrypt the online social network security privacy data C2 in step 3 to get ans2. At this time, the result of ans2 is consistent with that of the corresponding operation f for plaintext M. The whole process of homomorphic encryption is over.

(5) The method of full homomorphic encryption protects the security and privacy data of online social networks in the era of big data. It not only ensures the security of data in the transmission process, but also ensures the security of data in the storage and processing process of the server. It can meet the user's all-round protection of privacy data and realize the protection of online social network security privacy data in the era of big data.

3 Comparative Experiment

3.1 Experimental Background and Evaluation Index

In order to ensure the validity and reliability of the experiment, we use the online social network security privacy data set produced in the era of big data to do relevant experiments. This data set is composed of sensor type, user's daily behavior, sensor's perception data and time stamp to generate the perception data. In addition, a large number of real data are downloaded from CASAS (Center for Advanced Studies in Adaptive Systems) as sample data for clustering analysis to generate FDR parameters, as well as learning samples.

In the windows 10 environment, we use PHP language to process the data, standardize the data in the format of txt and dat, and store it in the unified MySQL database, so as to facilitate the reading and operation of the data in the experiment.

In order to measure the effect of online social network security privacy data protection, the average clustering accuracy parameters are set. Because in the process of bypass attack, the most important step is to cluster the sensor information, that is, to classify the RF signals emitted by different sensors, and simultaneous interpreting the

RF signals emitted by the same type of sensors. If the sensor nodes are wrongly identified and classified, the attacker will not recognize the real behavior of the user, or recognize the wrong behavior. In view of this, ACA is set as a parameter to measure the effect of privacy protection. The value range of ACA is [0, 1]. If ACA is close to 1, the attacker can analyze the distribution of wireless sensors by monitoring the radio frequency signal, which is the same as the distribution of real wireless sensors. On the contrary, if the ACA value is close to 0, the attacker can analyze the distribution of wireless sensors by monitoring the radio frequency signal, which is totally different from the distribution of real wireless sensors. In other words, the lower the ACA value, the better. Through experimental analysis, when the ACA value is maintained at [0.1, 0.4], the privacy protection effect is the best.

In order to compare with the traditional protection method, FVR parameters are set. FVR is the ratio of noise data to real sensor sensing data in unit time. Under different conditions, it is difficult to compare the advantages and disadvantages of the two methods. By changing the size of FVR, it can be used to unify the conditions of the two methods, and then compare the energy consumption, privacy protection effect and delay. If FVR increases, it means the number of noise packets added to the wireless sensor network increases; if FVR decreases, it means the number of noise packets added to the wireless sensor network decreases. When the two methods are in the same FVR condition, the better the privacy protection effect is, and the algorithm with lower delay has more advantages.

In order to compare the data protection of the two methods, we compare the size of FVR of the two methods under the condition of the same privacy protection effect, i.e. the same ACA.

3.2 Experiment Implementation

During the experiment, limited by the existing experimental environment, the two protection methods can not be deployed to the real reproduction social network environment. Only with the open real data set, combined with the algorithm idea, the algorithm can be simulated to achieve the overall privacy protection plus noise method. The specific experimental steps are as follows:

STEP1: The FDR parameters are obtained by analyzing the sample data set. In the process of analyzing and obtaining sample data sets, because the format of each open data set is different, we need to normalize different data sets, and then call the analysis to get FDR parameters.
STEP2: FDR parameters are used in the process of supervised learning to get the learning parameters.
STEP3: According to the real transmission data of the wireless sensor and the learning parameters in step 2, a random noise packet is generated and added to the transmission sequence of the wireless sensor.
STEP4: Take the execution result of step 3 as the fingerprint information of the wireless sensor monitored by the bypass attack, execute the bypass attack algorithm, and calculate the privacy data protection delay.

For the traditional privacy data protection method and the online social network security privacy data protection method in the era of big data, two algorithms are implemented through the same experimental data set, and then the noise data is added to the experimental data set according to the privacy data protection delay and FVR, and finally the data is statistically compared.

3.3 Analysis of Experimental Results

Using the above experimental background and implementation scheme, the following experimental results are obtained, as shown in Fig. 6.

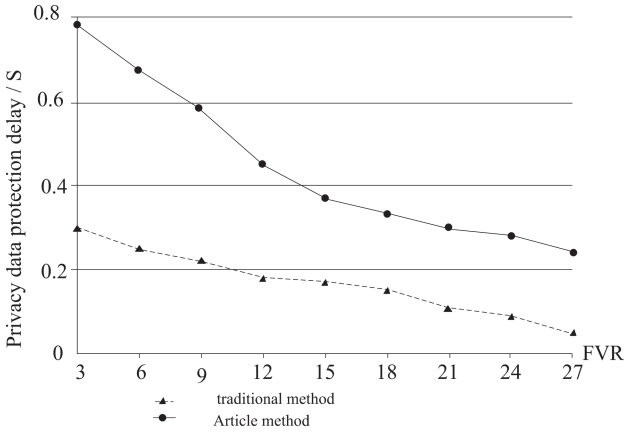

Fig. 6. Experimental results

As shown in Fig. 6, according to the privacy data protection delay comparison of the two methods, the method is to compare the amount of noise data added when the privacy protection effect is the same [illegible]. When the delay value of privacy data protection is lower than 0.4, the effect of privacy protection is the best. Compared with the traditional privacy data protection method, the delay value of privacy data protection method in the era of big data is lower. Without considering the delay, it is obvious that the energy consumption of online social network security privacy data protection method is the smallest in the era of big data. For each privacy data protection delay is the same, the online social network security privacy data protection method is shorter than the traditional privacy data protection method in the era of big data.

4 Conclusion

In this paper, we propose a secure privacy data protection method for online social networks in the era of big data. First of all, combined with the characteristics of wireless sensor networks, this paper analyzes the problems that need to be considered

in the design of noise model, that is, the design principles of privacy data protection. Then, from the construction of data protection and specific implementation steps, the specific implementation steps of privacy data protection methods are introduced in detail. Through experimental comparison and analysis, this paper compares the protection method of this paper with the traditional method from the perspective of privacy data protection delay value. The results show that the privacy data protection delay value of online social network security privacy data protection method is low in the era of big data, which has the advantage of good privacy protection effect.

References

1. Huang, W., Huang, J., Li, Y.: An analysis of the relationship between anti-terrorist intelligence work and the protection of personal privacy information in the era of big data. Library Inf. **182**(4), 43–50 (2018)
2. Lina, Z.: Personal information privacy protection in the informatization construction of university personnel files in the era of big data. Arch. Shanxi **238**(2), 70–72 (2018)
3. Zhang, C., Liu, C., Guo, Q.: Optimization of user information security protection in large data. Comput. Simul. **34**(7), 154–157 (2017)
4. Wang, T., Liu, Y., Jin, X.: Research on k-anonymity-based privacy protection in crowd sensing. J. Commun. **39**(S1), 176–184 (2018)
5. Huang, R.: Simulation research of network user privacy information protection. Comput. Simul. **11**, 319–322 (2017)
6. Huang, R., Long, L.: The problems and countermeasures of individual privacy protection under open government data in China. Library **10**, 1–5 (2017)
7. Lin, Y., Duan, X.: Digital information encryption technology for network privacy protection. Modern Electron. Tech. **41**(9), 45–48 (2018)
8. Li, X., Luo, X.: Simulation of user information security protection for communication transmission under big data. Comput. Simul. **35**(05), 178–182 (2018)
9. Wei, D., Ma, H.: Research on the influencing factors of personal data storage security and privacy protection under network environment. Library Theory Pract. **1**, 89–95 (2018)
10. Wang, X.: Computer network data security encryption technology in internet environment. Mobile Commun. **3**, 49–53 (2019)