

Data Segmentation Scheme for Decentralized Social Networks

Haoyue Zhang

Abstract—Social network is one of the most promising Internet application services, while the traditional social network is completely centralized. While the centralized social network provides users with rich personalized services, there are also data security problems caused by a large amount of data transmission. Therefore, it is urgent to establish a secure and efficient decentralized social network scheme. Based on the security problem of information transmission in traditional social networks, this paper divides social network information into node information and edge transmission information to improve its transmission efficiency. In addition, the traditional data segmentation in file transmission is further optimized, and the randomization processing is carried out to eliminate the correlation between data blocks after traditional data segmentation, and to protect the confidentiality of the original data block information.

Index Terms—Decentralization, social networks, privacy protection, data segmentation

I. INTRODUCTION

With the advent of the Internet era, social network has become one of the most promising Internet application services. By June 2021, the scale of domestic social network users has reached 983 million^[1], and the data magnitude of information has developed from TB level to PB level. The era of big data has come^[2]. Traditional social networks are completely centralized, with social platforms setting rules, storing data and sending content. User's information and the communication between the user and dynamic needs to be released by a central server, the platform can be through a central server to achieve the purpose of access to user data, such as precision advertising push to profit^[3], the purpose of in recent years, due to data mining and the development of cloud computing and other technical popularization, While the centralized social network provides users with rich personalized services, there are also data security problems caused by a large amount of data storage. The personal information and privacy security of users are facing unprecedented challenges. Therefore, aiming at the problem of information transmission security, designing an efficient and secure data transmission scheme is one of the key research points in the field of information security.

In traditional social networks, all data are classified into one category and transmitted after the same type of encryption and segmentation. If the data type is not distinguished, the corresponding computational power will be wasted. Therefore, the corresponding classification of these data can increase the network utilization. The traditional data

segmentation technology also fails to differentiate the processing of different social network data, only to guarantee the system performance, ignoring the difference of transmission modes of different data types, and does not deal with the data security of the segmented file sub-blocks [4]. How to make the segmented data sub-block store the original data information completely, and ensure that any single data sub-block cannot leak the information of the original data block, is a new challenge to the traditional data segmentation technology.

This article, from the perspective of the social network data transmission, to social network data types are classified, based on the former data characteristics using the improved file segmentation technology to break up, for users to share files data provides a more efficient and safer way, make the file after segmentation block to block of mutual authentication between, achieve the goal of tamper proof.

II. RELATED WORK

A. Classification of social network information

According to the structure of social network, privacy issues can be divided into three categories, which are individual identity disclosure, connection disclosure and content disclosure^[5]. Individual identity disclosure refers to the individual's attribute information, structure location information, label information and so on are obtained by the attacker and identified, that is, a user node is associated with a specific individual; Connection leakage refers to the data leakage problem caused by external attacks in the process of data interaction, that is, the privacy information of edges between user nodes is attacked by attackers. For decentralized social networks, we divide them into node information and edge transmission information according to the data transmission characteristics of social networks, as shown in Figure 1. Nodes are individuals in the network and participants in the social network, which can represent a specific person, group or event^[6]. Node information is the basic user information entered during user registration. Such as name, age, cell phone number and other information. The essence of edge transmission is the information transmission between nodes. In social networks, nodes are connected to each other to form edges when they transmit information such as publication, audio or video. The information transmission in edges is called edge transmission.

Manuscript received July 26, 2022

Haoyue Zhang, School of Computer Science and Technology, TianGong University, Tianjin, China

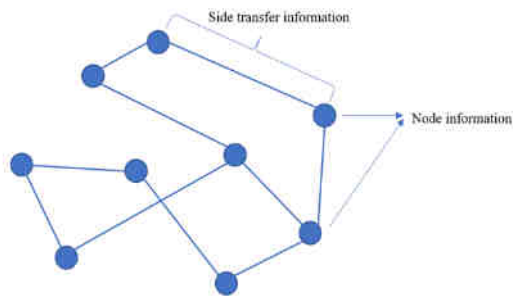


Figure 1 Social network infographic

B. Traditional file segmentation

In the traditional file segmentation, large file data are linearly cut into small data subblocks in order [7]. The principle is to use the input stream to read the file, after reading the specified size of the stream and then output to the specified file, until the end of the file read. When the segmented files need to be merged, all the sub-files need to be read into an input stream in order, and then all the files in the input stream are output to the same file output stream, so that the segmented files can be merged into the original file [8]. The specific segmentation process is shown in Figure 2.

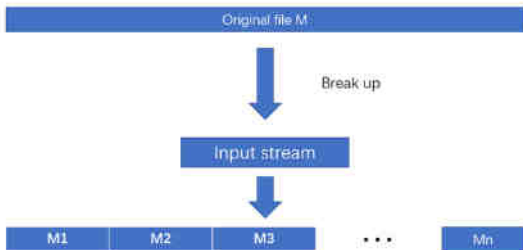


Figure 2 Flowchart of file segmentation

III. THE OPTIMIZED DATA SEGMENTATION SCHEME

The information stored in the node is user attribute information and local cache data. Attribute information is the basic information of user registration, such as name, height, weight, etc. [9]. Because the information itself is meaningless, the key factor in privacy disclosure is to correspond it to the specific user. This type of data stored in the local device after data at the time of the partitioning, relative to the transmission of data is more secure, combined with the consideration of the social network information storage efficiency, only need to be physically stored separately after the traditional linear segmentation, can hide the information protection can achieve very objective.

Edge transmission involves the data exchange between nodes [10], that is, there is an information transmitter and an information receiver, including the signature to verify identity information. Edge transmission faces more security problems in network switching than the local data of nodes, so file segmentation in edge transmission needs to adopt a more secure file segmentation scheme than ordinary linear segmentation.

Because transmission of social network information of different sizes, if we put all the documents are divided into the same size it is not practical, suitable for small file segmentation method when applied to large files, may be

integral number too much, so we will all documents are divided into the same number of a file, set the number of file segmentation of fit. The method of file splitting can also ensure the confidentiality of large files. Block cap on the number affected by equipment number and size of the storage space, suppose that we will file segmentation rules for 10 number, each file that we will be the average into 10 groups, such as 100 m, each size is 10 MB, if the file size is 1000 m, each size is 100 m, but if the file is too large, even got the ten copies, each will also be very big, It will cause a lot of negative impacts on transmission, so we need to set an upper limit on the size of a single file to segment it more reasonably. But at the same time, we also need to set an upper limit on the number of blocks in order to prevent too many blocks from causing too much subsequent encryption work. Therefore, the improved file segmentation follows the formula: As $N=xX*P$ for segmentation (the final number of blocks is N, the file size is X, the upper limit of files is X, and the upper limit of blocks is P), each segmented file block will have index information to achieve the purpose of mutual verification between blocks. The file segmentation flowchart is shown in FIG. 3:

- (1) Generate segmentation number: According to the file segmentation function in the previous section, generate N.
- (2) Map file blocks: Map file blocks separately, such as 1-AN.
- (3) Generate random sequence: Generate 1-N random sequence by random function.
- (4) File block rearrangement: Rearrange data block a1-an according to the generated random sequence
- (5) File transfer: Data is transferred according to the rearranged file order.

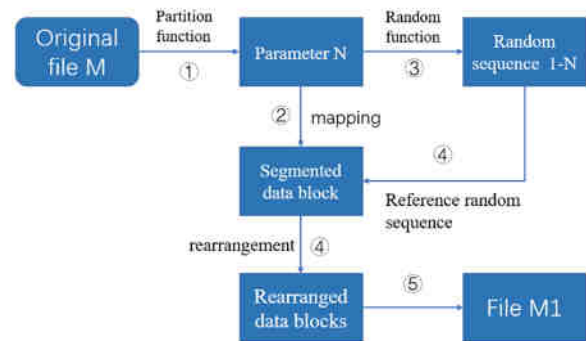


Figure 3 Flowchart of file segmentation

IV. CONCLUSION

With the continuous development of media technology, the issue of privacy protection in social networks has attracted extensive attention. People pay more and more attention to the security of data transmission when exchanging information with others. Based on the characteristics of social network data, this paper divides it into two types: node information and edge transmission information. And on the basis of classification, different types of social data file segmentation methods are designed. In the node information, the common linear segmentation method combined with decentralized storage is used to achieve the purpose of information hiding. In the edge transmission information, the improved file

segmentation method is used to optimize the segmentation process, such as randomization, to eliminate the correlation between data blocks after traditional data segmentation, realize the mutual verification between blocks, and achieve the purpose of protecting the confidentiality of the original data block information. In addition, the improved file segmentation technology combined with encryption algorithm and other technologies are applied to decentralized social networks, which will further improve the security of data transmission in social networks and provide a new research direction for privacy protection in decentralized social networks.

REFERENCES

- [1] China Internet Network Information Center. The 49th Statistical Report on the Development of Internet in China [R/OL]. (2022-02-25).
- [2] JIANG C S. Research on privacy and security issues in social networks: A case study of Facebook [J].2015(6):1-1. (in Chinese).
- [3] Lian Jie. Research on Social Network Data Mining Based on User Characteristics: [D]. Beijing Jiaotong University, 2014.
- [4] Liu Guiyun, Shao Zhichao. Research on Content Storage Strategy Based on Slicing technology in IPTV [J]. Information and Communication, 2009,(02).
- [5] Yao Ruixin, Li Hui, et al. Privacy protection in social networks [J]. Journal of network and information security,2016,2(4) : 33-43.
- [6] Xu Fei, Yang Guangwen, Ju Dapeng. Design of distributed storage system based on Peer to Peer [J]. Journal of software, 2004,2(15):268-277.
- [7] YanYin shot. Implementation of Text File Segmentation Method [J]. Computer programming skills and maintenance,12-16 (5) : 2015
- [8] Ueno Y, Miyaho N, Suzuki S. Disaster Recovery Mechanism using Widely Distributed Networking and Secure Metadata Handling Technology [C].Proceedings of the 4th edition of the UPGRADE CN workshop, Session II:Networking, 2009.
- [9] Yongjun Li ,Zhaoting su,et al.Exploiting similarities of user friendship networks across social networks for user identification[J].Information Sciences,2019,8(22):78-98.
- [10] Junchi Funasaka, Nozomi Nakawaki. A parallel download method of coping with variable bandwidth[C].Proceedings of the 23rd International Conference on Distributed Computing System Workshops. May 2003.