Research on Network Privacy Protection Mechanism Based on Trust Agent in Big Data Environment

Hong Wang, Na Lei and Liumei Zhang

School of Computer, Xi'an Shiyou University, Xian,Shaanxi,China
Email: wanghong@xsyu.edu.cn, swhlzln2045@163.com, rikrun@gmail.com

Abstract. The problem of personal information and behavioral privacy disclosure in the big data environment has aroused great concern in the whole society. This paper analyzes the user privacy leakage problem in the big data environment supported by cloud computing, proposes a privacy protection mechanism based on trust agent model, and studies the specific method and implementation path of using distributed trust routing and private cloud to realize privacy protection, and its core is to prevent users from leaking real private information through the localization and virtualization of privacy data. It also discusses the problem of user privacy information being corrupted by active privacy data pollution.

1. Introduction

The rapid convergence of massive data and the continuous advancement of cloud computing have led to the rapid development of big data analytics technology and the era of efficient data utilization.

On the one hand, cloud computing and big data are inextricably linked. This is because that cloud computing is a solution platform for achieving higher capabilities through technology and products, while big data focuses on the deep mining of business data itself. However, big data analysis is increasingly relying on the power of cloud computing [4]. It can be said that the big data analysis supported by cloud computing has achieved great success in data mining, intelligence gathering, network marketing, monitoring and tracking, and sociological research, bringing great value to users, and strongly and profoundly changing people's lives.

On the other hand, privacy of personal information and behavior has been greatly threatened as the explosive growth of Internet user data and the continuous progress of data mining technology [6]. In the era of simulation and small data, the institutions that can control the personal data of citizens in large numbers can only be government agencies with public power, but now many Internet enterprises can also have huge amounts of data, and even exceed government agencies in some respects, so that the privacy protection posed a serious challenge [5]. The overwhelming precision advertising and frequent privacy leaks continue to sound alarm: the value of big data is unquestionable, and the privacy risks behind it are even more important.

At present, privacy protection mainly relies on social governance and technical means [14]. The former includes privacy protection legislation [10] and corporate and industry self-regulation, while the latter mainly has two types of methods based on data encryption and data anonymity and distortion. The methods based on symmetric encryption, public key encryption and heuristic encryption play a certain role in the privacy data, but some limitations have shown that the application scenarios of the data are greatly restricted [11]. The main purpose of the anonymous method is to make it difficult for an attacker to identify which of the N eligible candidates is the real information of the target user, thereby reducing the probability of privacy leakage to 1/N. The idea based on the distortion method is to randomly modify the user's privacy data to make an attack. Thus, the attackers cannot accurately

guess the true raw data and thus protect it. Data-based anonymity and distortion are key technologies for current privacy protection.

2. Analysis of the characteristics of big data analysis mining

Data mining in the era of big data shows some distinct features compared to traditional data mining.

a. It provides integrated analytical and processing capabilities for a wide variety of structured and unstructured data sources. This is an important feature of big data analytics mining.

b. Replace the sampled data with the full amount of data as the object of analysis and mining. In the big data analysis environment supported by cloud computing, the efficiency of full-scale data analysis may be higher than the analysis of sampled data.

c. Correlation analysis is more than causation hypothesis. The main point of big data analysis is to use the machine learning algorithm to find the correlation between the various factors and the linkage trend in the massive data, and to achieve the prediction through regression analysis.

A real life user in an information society, under the big data environment of full data collection, integrated processing and pan-association analysis [3], and privacy data not only has the possibility of unintentional disclosure, but more risks come from encountering a full-scale excavation for a specific purpose. Taking the "user portrait" technology as an example, it can characterize the user's characteristic attributes through various dimensions, and abstract a user's information after analyzing and extracting these feature data [12]. From a good point of view, the "user portrait" can help businesses achieve precision marketing, user research, personalized services and business decisions, from a malicious point of view, it is very worrying!

In view of this, this paper proposes a method and approach for user privacy protection in a big data environment based on trust agent and private cloud.

3. Network privacy protection mechanism based on trust agent

In the context of big data analytics, there are two basic orientations of privacy protection. There are as few and irregular privacy data as possible in the big data source for analysis, and the possibility of big data association analysis is reduced by means of data isolation to reduce the utilization value [15]. Therefore, the basic order should be followed. First, through anonymous access, followed by hiding private information when it is not anonymous, then masquerading private data when it cannot be hidden, and finally obstructing it when it is difficult to disguise breaking the relationship analysis of multi-source data, fragmenting [1], contradicting and polluting private data, reducing the threat of data analysis means to private data, so as to achieve different degrees of protection of network privacy.

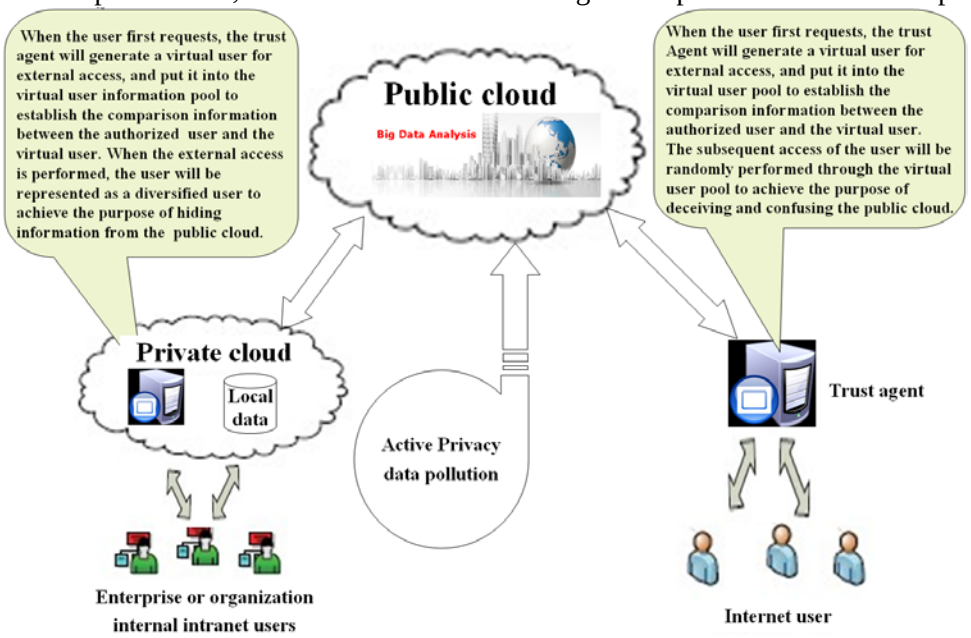

Figure 1 schematic diagram of network Privacy protection system based on trust agent

The core idea of network privacy protection based on trust agent is to transfer personal privacy information to the trust agent or private cloud for processing. When the user accesses, the virtual connection between the private cloud and the public cloud and trust agent and the public cloud can be utilized to ensure that the scope of user privacy information limited to a specific range [7][8][9], and through the isolation and blocking methods, personal information and behavioral privacy collected in the public cloud big data environment are fragmented, thus preventing larger scale big data analysis and reducing the opportunity and channel of privacy leakage. At the same time, through the use of confusion [2], deception [13], the active privacy data pollution of the big data used for analysis further reduces the accuracy of big data analysis of private information. Figure 1 shows a schematic of this core idea.

4. Trust agent construction

The so-called trust agent means that the user enters and borrows the access path of a trusted third party (for example, the government) to implement network access. The core of the trust agent is trust routing. The basic structure is shown in Figure 2.

In Figure 2, when an authorized user accesses the network, the trust agent generates a corresponding virtual user according to a specific algorithm, and the virtual user replaces the authorized user for network access. After the authorized user exits the access, the virtual user does not continue to exist. As the number of user access increases, the virtual users in the trust agent will form a large virtual user pool. When a new authorized user accesses, a matching pair is established between the authorized user and an existing virtual user under the trust routing assignment to access the network. At this time, the big data source on the public cloud exposes only the relevant information of the virtual user, which can hide the authorized user information, and their correspondence is not fixed, but can changed constantly. Every time the user's private information exposed to the public cloud will be different, thus showing a diversity and confusing characteristic. This will bring difficulties to the analysis of big data, thus protecting the privacy data.

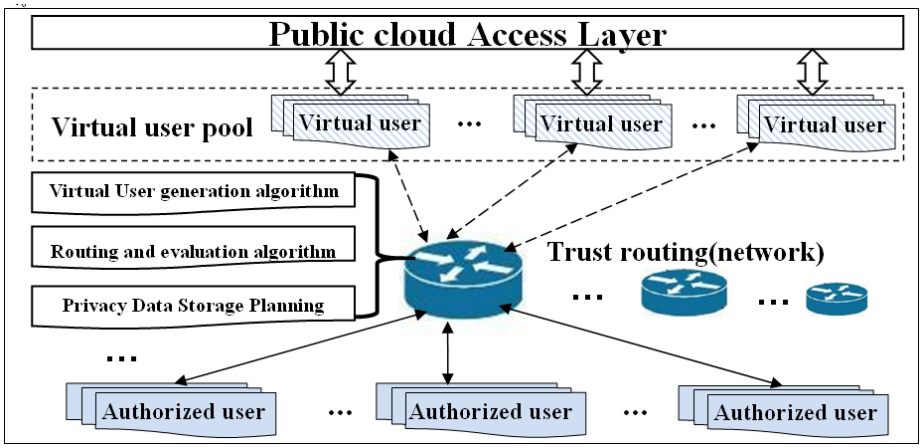

Figure 2 Trust agent basic structure diagram

The main functions of trust routing implementation include virtual user generation, virtual user pool resource management, routing and evaluation, and privacy data storage planning. At the same time, multiple trust routes can be established and form a distributed and controllable routing processing network. Each trust routing in the network has its own independent security control mechanism, and also has a protocol coordinated with the trust routing to achieve controllable data correlation, thus leaving the possibility of tracking related links for special departments (such as public security, national security). When an authorized user accesses a network application, he can automatically or manually select a trust route that establishes a local trust relationship with himself to perform various network operations.

5. Active privacy Data pollution

The basic principle of active privacy data pollution is shown in Figure 3.

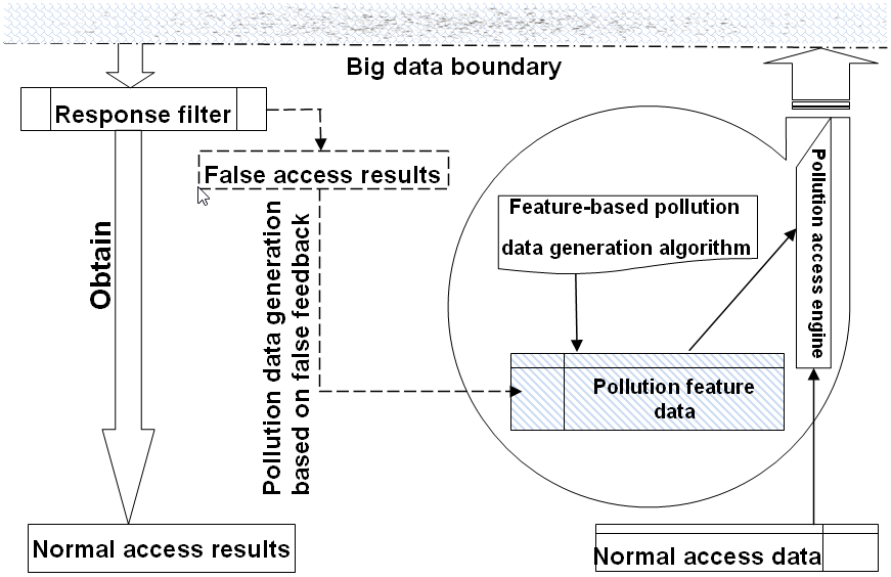

Figure 3 Active privacy data pollution schematic

In Figure 3, a pollution feature data set is generated by a feature-based pollution data generation algorithm or an empirical method. When the network access is performed, the normal access data and the pollution feature data are all sent to the public cloud through the pollution access engine. After the cloud platform responds to the request, the returned result can be filtered by a special response filter, and the normal access result is obtained therefrom and the false access result formed by the pollution data is filtered out. After such active pollution access, data within the big data boundary that may contain behavioral habits and access trajectories are “contaminated” (pollution data is shown in the upper part of Figure 3). Thus, the purpose of interfering with the credibility of big data analysis conclusions from the source is to protect the user's network privacy. In Figure 3, the false results of feedback can also be returned to the pollution feature data set, and further pollution operations can be performed through these data to form a closed-loop system to achieve a "virtuous cycle" of data pollution, greatly improving misleading analysis of big data, and increasing the level of protection of private data. A typical example of this is the use of search keyword pollution to counter the behavior of search engines to obtain user access information.

6. Conclusion

The trust agent access model and hybrid cloud construction method proposed in this paper can partially isolate the private information from the big data analysis application on the public cloud. It restricts the privacy content to the smaller private cloud and private area. Thus, it is able to cut off the data source foundation of the correlation analysis of the personal information aggregation analysis and the behavior privacy. The main features of the proposed mechanism are: cloud access based on trust agent, fine-grained privacy restrictions and active privacy data pollution. There main innovations are: the trust agent model helps users hide data and cheat information for large data analysis by means of virtual users and trust routing; the active privacy data pollution technology can make data source used for big data analysis find confusion and protect user privacy to a certain extent.

7. References

[1] CAO Laicheng,LIU Yufei,DONG Xiaoye,GUO Xian.User privacy-preserving cloud storage scheme on CP-ABE[J].Journal of Tsinghua University(Science and Technology),2018, 58(2):150-156.

[2] ZHANG Shaobo,LIU Qin,WANG Guojun.Trajectory privacy protection method based on location obfuscation[J].Journal on Communications, 2018, 39(7):81-91.
[3] CUI Yi-hui,SONG Wei PENG,Zhi-yong,YANG Xian-di.Mining Method of Association Rules Based on Differential Privacy[J].Computer Science, 2018, 45(6):36-40,56.
[4] FAN Kai,DENG Hai,LI Hui and YANG Yintang.Privacy Protection Smartcard Authentication Scheme in Cloud Computing[J]. Chinese Journal of Electronics, 2018, 27(1):41-45.
[5] LI Zhi-peng,SUN Ming-song,SONG Zeng-lin.The Location Privacy Protection Technology of Mobile Intelligent Terminal[J].Journal of Harbin University of Science and Technology, 2018, 23(2):58-64.
[6] Shen Liyan,Chen Xiaojun,Shi Jinqiao and Hu Lanlan.Survey on Private Preserving Set Intersection Technology[J].Journal of Computer Research and Development, 2017, 50(10):2153-2169.
[7] SHI Yu-Liang,CHEN Yu,SUN Shi-Bin,CUI Li-Zhen.Data Chunks Adjustment Mechanism for Privacy Protection[J].Chinese Journal of Computers,2017, 40(12):2719-2733.
[8] LI Lu-lu,HUA Jia-feng,WAN Sheng,ZHU Hui,LI Feng-hua.Achieving efficient location privacy protection based on cache[J].Journal on Communications,2017, 38(6):148-157.
[9] LIU Xia,FENG Chao-sheng,LUO Wang-ping.Multilevel privacy protection mechanism for cloud backup system[J].Computer Engineering and Design,2017, 38(12):3241-3246.
[10] LIU Ling,LUO Rong.Research on Open Government Data and Personal Privacy Protection in Big Data Perspective[J].Information Science, 2017, 35(2):112-118.
[11] Wang Qianyi Ouyang Rongbin.Application research on privacy preservation for data service platform[J].Journal of Huazhong University of Science and Technology(Natural Science Edition),2016, 44(S1):152-157.
[12] ZHANG Lin,LIU Yan,WANG Ru-chuan.Location publishing technology based on differential privacy-preserving for big data services[J].Journal on Communications,2016, 37(9):46-54.
[13] LEI Kai-yue,LI Xing-hua,LIU Hai,PEI Zhuo-xiong,MA Jian-feng,LI Hui.Dummy trajectory privacy protection scheme for trajectory publishing based on the spatiotemporal correlation[J].Journal on Communications, 2016, 37(12):156-164.
[14] Lu Xuemei,Gu Chunsheng.Analysis on Causes and Protective Strategy of User Privacy Disclosure in the Big Data Environment[J].Journal of Modern Information,2016, 36(11):66-70.
[15] GENG Kui,LI Feng-hua,LI Wei-hao,LI Hui,NIU Ben,XIE Rong-na.Proxy-based privacy-preserving scheme for mobile Internet[J].Journal on Communications,2015, 36(11):25-32.