Research on Personal Privacy Protection Technology in the Age of Big Data

Yanlin Yin, Rui Han

Heilongjiang University of Technology, Jixi, Heilongjiang, 158100, China

Keywords: Personal Privacy Protection, Big Data, Protection Method

Abstract: Big data is of great value, and the era of big data has brought changes to our lives, work, and thinking, and it has made it difficult to hide personal privacy. How to protect personal privacy has become a concern of the whole society. The paper analyzes the definition of personal privacy, the threat of big data to personal privacy and the way to disclose privacy, and proposes suggestions such as improving laws and regulations, strengthening corporate self-discipline, improving personal privacy protection technology, cultivating personal information protection awareness and data sales licensing system.

1. Introduction

Big data is about a wide variety of sustainable data, using new tools, new models, and new systems to gain insightful and new value. Big data contains great value and is a valuable asset of the company. The arrival of the era of big data is changing our lives, work and thinking, as well as threatening our personal privacy. Personal privacy leaks occurred frequently at home and abroad, such as the most shocking "prism" incident in foreign countries; 20 million user information on Russian dating sites were leaked; the second largest medical insurance company in the United States was attacked by hackers, resulting in 80 million user data. Impact; domestic 360 browser infringes on user privacy incidents, these incidents have seriously violated the privacy of customers and damaged the legitimate rights and interests of customers. It can be seen that in the era of big data, how to protect personal privacy will become a serious problem for every citizen and even the whole society.

2. Threats to personal privacy in the era of big data

Various information about individuals in the network is everywhere. People can use the information analysis tool software to collect, store and convert people's identity information, network browsing traces and geographical location, and then summarize and statistically analyze them. The data between the various systems can be mutually verified and explained to each other, and a variety of databases will emerge. According to a recent study published by Harvard University, as long as there is a person whose age, gender and zip code, it is possible to search the public data for about 87% of the person's personal information.

In the big data application environment, data presents dynamic characteristics, data types and data sources are diversified, and a large number of structured and unstructured data are continuously generated. Many storage systems cannot meet the needs of big data applications. At present, most privacy protection technologies and algorithms are aimed at traditional relational data, and cannot be directly transplanted into big data applications. Therefore, storage system architecture and security protection for big data are facing challenges.

In the era of big data, the free flow and sharing of information brings benefits to business development, technological innovation, and government management. Because personal information can bring commercial value, it stimulates unscrupulous merchants and personnel to illegally acquire, utilize, and store personal information to gain benefits for business opportunities. In addition, some governments require personal privacy in order to maintain national cyber security and security stability, resulting in national security and personal privacy. According to the "Prism" incident, the US National Security Agency has implemented a top-secret electronic monitoring

  DOI: 10.25236/csam.2019.005

program since 2007; in February 2016, the US FBI required Apple's mobile phone to open the back door to obtain personal encrypted information in the name of national security, and obtained California. The court's support, this incident also triggered a discussion of national security and personal privacy.

3. The challenges of personal privacy in the era of big data

The network of big data generated by the interaction, integration and integration of the “human, machine and material“ ternary world brings great opportunities, and brings many scientific problems and extremes to the existing IT architecture, machine processing and computing power. Big challenge. In addition, big data has the characteristics of large data volume, various data types, fast data generation and low value density. In addition, the characteristics of personal privacy changing dynamically with many factors make it more difficult to protect personal privacy in the era of big data. The following six challenges and research questions are addressed for the privacy protection of big data. 1) The scope of personal privacy protection is difficult to determine. According to the above description of the concept of personal privacy, the concept of privacy changes with the development of information technology, and also considers the characteristics and background of different people. Therefore, it is difficult to define which sensitive data is protected by privacy. 2) Infringement of personal privacy is difficult to identify. The form of infringement of personal privacy is complex and diverse, and it is impossible to judge whether it constitutes an infringement according to the current law. Users often use pseudonyms on the web. This anonymity makes it difficult for victims to collect evidence and find real infringers. Even if the victim obtains evidence through webpage backup, etc., the webpage is always constantly updated, and it is difficult to exert the validity of the evidence as long as the infringer does not recognize it. Therefore, how to determine who violated personal privacy is facing great challenges.

As information and communication technologies become more common, managing personal privacy information becomes more difficult. Management of personal privacy information includes the collection, storage, use, and distribution of personally identifiable information. 1 When collecting personal information, how to ensure that the collected information maintains its integrity during transmission; 2When storing personal information, what technology is used to ensure that information is not stolen or illegally accessed; 3 for the use of personal information, should How to set a strict access control policy so that different people can see data of different access levels without increasing the amount of management workload; 4When publishing information, control what information needs to be released and who can access the published on the network. Information has become an issue of increasing concern to enterprises. For the data to be released, how to ensure that the data will not reveal personal privacy information, while ensuring the effectiveness of the data, but not to hide all the data in order to protect privacy, this does not reflect the value of the data. Corporate managers are increasingly aware of the importance of protecting personal privacy data because it is directly related to the company's interests. However, how to manage data, that is, to ensure the use of data while protecting personal privacy, is one of the great challenges faced by enterprises in the era of big data.

4. Big data personal privacy protection technology

Data encryption technology has a long history. After entering the digital age, it is still a reliable method for computer systems to protect sensitive information. The role of data encryption is to prevent intruders from stealing or tampering with important data. According to the encrypted key algorithm, data encryption can be divided into symmetric encryption algorithm and asymmetric encryption algorithm. 1) Symmetric encryption algorithm uses the same key for encryption and decryption, mainly used to ensure the confidentiality of data. The most representative algorithm is the DES (dataencryptionstandard) algorithm proposed by IBM in the 1970s. Based on this, many improved algorithms of DES, such as triple DES (tripleDES), randomized DES (RDES), and IDEA (Internationaldataencryptionalgorithm), generalized DES (generalizedDES), NewDES, Blowfish,

FEAL, and RC5. In 2001, the American National Institute of Standards and Technology published advanced densification standards (AES) instead of DES, becoming one of the most popular algorithms for symmetric key encryption. The advantage of the symmetric encryption algorithm is that the computational overhead is small, the encryption speed is fast, and it is suiTable for the encryption of a small amount or massive data. It is the main algorithm currently used for information encryption. The disadvantage is that the two parties use the same key, it is difficult to ensure the security of the two keys; when the amount of key data increases, the key management will impose a burden on the user; in addition, it is only suiTable for encrypting and decrypting data. Provides the confidentiality of data, it is not suiTable for use in distributed network systems, key management is difficult, and the cost is high.

Asymmetric encryption algorithm is also called public key algorithm, and its encryption and decryption are relatively independent, using different keys. It is mainly used in the field of information exchange such as identity authentication and digital signature. The most famous representative of the public key cryptosystem algorithm is RSA, in addition to the back-packet cipher, DSA, McEliece cipher, Diffe_Hellman, Rabin, zero-knowledge proof, elliptic curve, EIGal algorithm. The advantage of the asymmetric encryption algorithm is that it can adapt to the openness requirements of the network, and the key management problem is also relatively simple, which can easily realize digital signature and verification. The disadvantage is that the algorithm is complex and the rate of encrypted data is low. However, both symmetric and asymmetric encryption algorithms have the risk of key leakage. Therefore, Rivest developed the MD2 algorithm 1 in 1989, which does not require a key, and triggers the study of a hash algorithm (also called a hash function), which converts an arbitrarily long input message string into a fixed-length output string without the need for a secret. Key, and the process is one-way, irreversible. The more popular algorithms are MD5, sha-1, RIPEMD and Haval. The hash algorithm does not have the problem of key storage and distribution. It is very suiTable for use on distributed network systems. However, due to the complexity of encryption calculation, it is usually only used when the amount of data is limited. For example, password encryption and software widely used in registration systems. Use period encryption, etc. Data encryption technology can guarantee the accuracy and security of the final data, but the calculation and sales ratio is larger. Encryption can not prevent the data from flowing to the outside. Therefore, encryption itself cannot completely solve the problem of protecting data privacy. Data encryption algorithm is a key technology of privacy protection. The research focus in the era of big data will focus on the improvement of existing algorithms; the combination of symmetric encryption algorithm and asymmetric encryption algorithm. With the advent of new technologies, new encryption algorithms that conform to the development of new technologies will be developed.

The database is still the main body of the information system, such as a large amount of personal and family information stored in the government database; personal financial information stored in the financial database; personal medical history information stored in the medical database, online banking, mail information and personal registration information used on the network. Wait. In the era of big data, although MapReduce technology is widely used in related data analysis to become a database competitor, MapReduce cannot completely replace the database, and they can learn from each other and integrate into a new ecosystem. The database not only faces the threat of intruders, but also faces the threat of insiders, including unauthorized data viewing, incorrect data modification, and data unavailability. To ensure database security, consider four levels: physical security, operating system security, DBMS security, and database encryption. The first three layers are not enough to ensure the confidentiality of data. Database encryption can ensure that sensitive information exists in the form of ciphertext and is protected. In order to protect sensitive data in the database, a dual mechanism of data encryption and access control is adopted. Since the research work on data encryption and access control is relatively mature, only the matters needing attention when using encryption and access control are described here.

5. Conclusion

"Without trust, there is no big data." The era of big data requires a new mechanism to protect personal privacy. Through research, this paper proposes practical and personal privacy protection measures from legislation, industry self-discipline, self-protection, privacy protection technology and sales license mechanism of public data to assist relevant management departments in making decisions. Personal data privacy protection needs constant exploration and research, which can alleviate people's concerns about personal privacy leakage and contribute to the sustainable development of the big data industry itself.

References

[1] Luo Bingmei. Online personal privacy information protection strategy [J]. Modern Intelligence, 2003, 23 (11): 25-27.

[2] Gao Wenbiao. On Privacy and Its Protection [J]. Journal of Xinjiang Public Security and Judicial Administration Cadre College, 2000(4): 32-34.

[3] Yang Lihua. Exploration and Research on Data Privacy Protection Strategy in China [J]. Journal of Information, 2009, 28(B06): 300-302.

[4] Zhuge Yudan. The Famous Name of Property Service Contract [J]. Journal of Hunan Public Security College, 2009, 21(4): 97-101.

[5] Shan Yunjuan. The company improperly blocked the unnamed shareholders from being named for compensation liability [J]. People's Justice, 2015, 0 (16): 61-64.