Study of Personal Information Security and Privacy Protection based on Big Data

Xuan Huang

School of Software and Internet of things Engineering, Jiangxi University of Finance and Economics, Nanchang, Jiangxi, 330000

Keywords: big data, privacy protection, information security

Abstract: With the development of information technology, emerging services such as blogs, Weibo, and social networks based on Web2.0 technology and the unprecedented growth rate of the Internet of Things have produced a wide variety of data, and cloud computing provides the basis for data storage. Platform, all this has created the official arrival of the era of big data. Big data contains great value and is a valuable asset of the company. But big data also brings huge challenges, and personal privacy protection is one of them. The rapidly evolving Internet has become an indispensable part of people's lives. People have left a lot of data footprints on the network. These data footprints are cumulative and relevant. When multiple data footprints are gathered together, you can discover individual ones, privacy information. The use of this information by malicious elements for fraud and other acts has brought many troubles or economic losses to the individual's life. Therefore, the personal privacy of big data has attracted wide attention from industry and academia. Firstly, it introduces the related concepts of personal privacy protection in the era of big data, and discusses the challenges and research problems faced by personal privacy protection.

1. Introduction

In the Internet environment, the "information explosion" has produced a huge amount of data, which has spawned the arrival of the era of big data. According to statistics, Facebook generates more than 300TB of log data per day; the data generated on the Internet can be engraved with 168 million DVDs per day; Twitter processes more than 340 million Twitters per day; and so on. These data are enough to show that we are already in a big data environment. In the era of big data, big data and deeply affect people's work, study and life, people's behavior on the Internet will be recorded by the system, including personal basic information, purchase records, daily operating habits. For merchants, this data can in-depthly explore the needs of users, thereby improving the corresponding products and services, while also achieving precise marketing, resulting in more economic benefits. However, these data also expose the privacy of individuals, and there is a great risk in the security of personal information. Therefore, this article stands in the context of big data to discuss the key issues of personal information security, and analyzes the information security risks, and proposes specific solutions and solutions in combination with the corresponding technologies.

2. Big data related concepts

Big data is a relatively abstract concept, and there is currently no uniform definition in academia. According to Wikipedia's definition, big data is a collection of data that cannot be captured, managed, and processed by conventional software tools within an affordable time frame. McKinsey pointed out that big data refers to a collection of data that cannot be collected, stored, managed, and analyzed by traditional database software tools over a period of time. According to Gaudner, a big data research organization, big data is a massive, high-growth, and diverse information asset that requires a new processing model to have greater decision-making, insight, and process optimization capabilities. From this, no matter which definition, big data does not refer to a new product or

  DOI: 10.25236/systca.18.017

technology, it is only a phenomenon in the digital age. In the ever-changing era of information explosion, academics, industry, and education all have different opinions on big data, but it is generally accepted that big data has evolved from the previous definition of 3 "V" to the definition of 4 "V". The 3 "V" features of big data refer to Volume, Variety, and Velocity, but the fourth V feature of big data is still controversial. Some scholars believe that it is Value, that is, value, and emphasizes that it is the most critical point in the characteristics of big data. Other scholars believe that the fourth V should be Veracity, that is, authenticity, which means that the data has higher authenticity.

3. The concept of personal privacy and the challenges it faces in big data

The introduction of privacy goes back to Warren et al.'s "Privacy" published in 1890, which became a pioneering work on traditional American law. Warren and Brandeis propose that personal privacy is a unique right that should be protected from the unfounded release of others who want to keep secret details in their lives. The concept of privacy has been studied in all areas of social science (such as philosophy, psychology, sociology) for more than 100 years, but there is no clear definition that meets both the needs of the times and the practice test. The definition of privacy is mainly divided into two categories: value-based, considering privacy as a human right, part of the social moral value system, a commodity, the value of people and society (such as users worrying about privacy issues while surfing the Internet) In many cases, in order to meet some of their own needs, they still submit their personal information); based on homology, the privacy relationship to the individual's thoughts, perceptions and perceptions is regarded as a state (including 4 seed status: anonymous, Concealment, retention, and privacy. A type of control that represents transactional control between individuals and others. The ultimate goal is to enhance autonomy or reduce leakage. Control-based privacy definitions used to be the mainstream of privacy research, but there are also studies that use control as an element of privacy. Two studies have become one of the focuses of academic debate. In a specific situation, for different things, different people, privacy refers to information that users think is sensitive and unwilling to disclose. Banisar et al. classify personal privacy into four categories: 1) Information privacy, ie the management and use of personal data, including identity card numbers, bank account numbers, income and property status, marriage and family members, medical records, consumption and demand information (eg Shopping, buying a house, car, insurance), traces of network activities (such as IP address, browsing trails, activity content), etc.; 2) communication privacy, that is, personal use of various communication methods and communication with other people, including telephone, QQ, E-mail , WeChat, etc.; 3) spatial privacy, that is, a specific space or area for personal access, including home address, work unit, and public places for personal access; 4 physical privacy, that is, to protect the integrity of the individual's body, to prevent invasive operations, such as drug testing Wait. The personal privacy referred to in this article is personal information that is unwilling to be disclosed or known to others in the personal life of the citizen, such as the user's identity, trajectory, location and other sensitive information. The scope of privacy includes private information, private events and private spaces.

The Internet has become a part of our lives, leaving us with a data footprint for visiting major websites. In the big data environment, this makes our privacy leaks easier. We are always exposed to the “third eye”. Major shopping websites such as Taobao, Amazon, and Jingdong are monitoring our shopping habits; Baidu, Bing, Google, etc. monitor our query records; QQ, Weibo, phone records, etc. eavesdropped on our social network; the surveillance system monitors our E-mail, chat history, Internet records, etc.; Flashcookies leaked our Some information such as usage habits or location, advertisers track our information and push related ads. Our daily activities are also monitored, such as smart phones monitoring our location; work units, major event venues, shops, communities, etc. to monitor our access behavior. The development of digital sensor technology allows us to collect new data in our daily situations, such as radio frequency identification (RFID)-based automatic payment systems and license plate recognition systems, implantable sensors to monitor patient health, and surveillance systems. Senior man at home waiting. As sensor technology continues to mature, various types of sensors will be widely used in our individuals or

organizations. The hallmark of these systems is that interactions are becoming increasingly blurred, so new mechanisms are needed to manage the risks posed by personal information and privacy.

4. Big data personal privacy protection technology

The data in the communication can use the SSL protocol to ensure the security of the data. Therefore, the data protection of the data layer mainly refers to the protection of the storage and management of the data. Ensuring the security of personal information at the data layer is fundamental to all other data-based applications, including ensuring the confidentiality, integrity, and availability of data. This section mainly describes the related research on protecting personal privacy data from the aspects of data encryption and access control. 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.

The symmetric encryption algorithm uses the same key for encryption and decryption, and is mainly used to ensure the confidentiality of data. The most representative algorithm is the DES algorithm proposed by IBM in the 1970s. Based on this, many improved algorithms of DES, such as triple DES, randomized DES (RDES), and IDEA, generalized DES New DES, Blowfish, FEAL, and RC5. In 2001, the National Standards and Technology Institute issued advanced densification standards (AES) to replace DES, which became one of the most popular algorithms in 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.

The asymmetric encryption algorithm is also called the 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.

5. Conclusion

The privacy protection of big data is still in its infancy. Although privacy protection is an important issue for users, companies are reluctant to implement privacy protection and not make full use of user information or provide better services to users. Limit the development of the company or its competitiveness in the market. Based on the analysis of personal privacy protection issues in this paper, it is expected that there will be a complete and understandable security solution to meet the needs of personal privacy protection in the future. For the majority of users to improve the self-protection awareness of personal information through the implementation of education and technical prevention. Of course, in the information age, as long as we use the network, it is unrealistic to completely protect personal privacy. At the same time, the user's data is spread across the Internet, and it is very difficult to ensure that all companies publish consistent information. Therefore, legislation and industry norms should be integrated into the process of technology implementation and corporate behavior, and they should be synchronized to maximize the use of data and minimize the leakage of privacy to meet current needs and solve more challenges. The article answers the solutions to the challenges of protecting personal privacy in the era of big data from different perspectives of technology and legislation and industry norms, hoping to provide some reference for subsequent research.

Acknowledgment

Key research and demonstration of key research and development projects and digital copyright trading platforms in Jiangxi Province. No.: 2018ACE50033

References

[1] Wu Zhenqiang. Research on network technology security and network defense under the information generation [J]. Network Security Technology and Applications, 2014(08): 140-141.

[2] Fu Jizhang. Research on Computer Network Technology Security and Network Defense [J]. Digital World, 2016 (06): 5.

[3] Zhang Lei. Research on Computer Network Technology Security and Network Defense [J]. Computer Optical Disk Software and Applications, 2013(16): 154+298.

[4] Su Bin. Application and research of power information communication technology in the era of smart grid [J]. North China Electric Power University, 2015.

[5] Li Chao. Analysis of the application of computer network technology and its development in power information communication [J]. China Science and Technology Investment, 2012 (33): 21.