Big Data Security and Privacy Protection

Qiang Mei

Jiangxi University of Engineering, Xinyu, Jiangxi, 338000, China

Keywords: Big Data, Security Technology, Privacy Protection

Abstract: The rapid development of the network has brought about the era of big data. Big data has a subtle influence on people's daily life and production economics. It is a hot spot of concern in all walks of life. At present, the collection and comprehensive application technology of big data is not mature enough. When using big data, it also faces a series of security problems: the authenticity of information is not guaranteed, and user privacy is leaked. Based on the basic overview of big data and the security challenges faced by the current development of big data, the paper discusses key technologies for big data security and privacy protection.

1. Introduction

With the extensive use of data by various industries, big data has become a major symbol in the field of information technology after the mobile Internet, cloud computing, and the Internet of Things. Due to the formation of such a large and complex data system, people's analysis and in-depth study of data information has become less convenient. To handle and manage such complex data systems requires more comprehensive security and privacy protection technologies, but now people are facing the growing information security and privacy issues of big data. This is a major challenge that requires the entire information technology industry to focus on and actively seek solutions.

2. Big data Overview

Big data refers to a large and complex collection of data that is difficult to process using current database management tools or data processing methods. Data sources can be divided into: various data information such as pictures, texts, and audio that people voluntarily issue on the Internet; various types of logs and files generated by machines and stored in computers. database, media materials, etc.; item attribute class, device record data, such as various product information recorded in the warehouse, data calculated in astronomical glasses, and so on. Characteristics of big data: Scale – As mentioned above, big data is large and complex. According to statistics, the total amount of information in the world in 2012 has been 2.7ZB, and it is expected to increase to 8ZB in 2015; 2) Diversity—In the past, in order to facilitate storage and viewing, data is mostly structured data based on text. Now, due to the diversification of information carriers, non-structures containing information such as pictures and audio are made. More and more data; value - through the analysis and statistics of the overall data, extracting valuable parts for users to use is also one of the basic characteristics of big data; high speed - In the era of information explosion, there is a growing need for efficient processing of information and the provision of real-time information.

Big data analysis is mostly used in different fields such as science, medicine, and commerce, and its uses are very different, but their goal of analyzing data is no more than three. In order to obtain valuable information, the original data containing a large amount of information is analyzed and integrated at different angles, and finally the more excellent information is summarized to help people understand the essence of things, grasp the development and operation of things, and then The next step in the development of things and things to make a predictive response. For example, in the field of fashion sales, the marketing department can understand the consumer's consumption trend and demand by analyzing the consumer's consumption data, and can produce more market-oriented products in advance to satisfy consumers. By analyzing the accumulated data in

  DOI: 10.25236/eeems.2018.033

multiple dimensions, not only can people grasp the general group characteristics, but also can specifically describe the differences between different individuals. Enterprises can use this data to introduce more humanized services to customers. For example, Amazon analyzes behaviors (ie, search, browse, join shopping carts, and purchases) before users purchase items, and can understand the user's purchase goals and psychological activities at the time of purchase, and implement effective recommendations. Under the condition that information can be quickly transmitted through the network, it is more necessary to analyze the authenticity of the data information. The information provided by the erroneous data may cause the user to make an incorrect decision, and sometimes cause irreparable errors. Therefore, it is necessary to carry out a detailed and in-depth analysis of the data, to take its brilliance and to ruin it. For example, filtering spam in a mailbox can also use big data analysis technology to protect users from interference.

3. Security test for big data

With the continuous development of science and technology, the era of big data has come, which brings us opportunities and values, but also brings new security challenges. In recent years, the security and privacy issues of big data have been widely concerned and worried, and the exposure of the “Prism Gate” has highlighted this problem. Different from the traditional security issues, in the era of big data, the security tests faced by data mainly have the following aspects. It turns out that if big data is not properly handled, it will pose a great threat to users' privacy. According to the protected objects, privacy protection can be divided into three categories, namely, location protection, connection relationship protection, and identifier protection. In the era of big data, the threat to user privacy is not just the privacy leak of individuals, but also the analysis and prediction of big data on its status and behavior. Nowadays, many companies believe that as long as the information is anonymized and the information without the user identifier is published, the privacy of the user can be protected. However, the protection effect obtained by this method is not satisfactory. In general, when collecting, storing, using, and managing user data, there is a lack of standards, regulations, and regulations, and they are overconfident and dependent on corporate self-discipline. In addition, users are not told where their private information is being used.

At present, it is generally believed that the data in front of us is a fact, which can fully prove everything. However, data is deceptive, and if it cannot be selected, it can easily be deceived by the illusion of data. The deception of big data is mainly reflected in two aspects, one is forged data and the other is distorted data. In order to achieve some effect, some people may create illusions by falsifying data, and then induce data analysts. Due to the scale and diversity of data, true and false information is often difficult to discern, resulting in erroneous conclusions. In addition, due to errors in the process of data collection, storage, etc., it is easy to cause data distortion, which will have a certain impact on its analysis results.

4. Big Data Security and Privacy Protection Technology

The technology was widely adopted before the big data concept was formed. It was designed to help users determine the source of the data and then verify that the results were correct or updated. The notation method is the basic method of the technology, and gradually evolves into the two forms of Why and Where in the course of practice. The focus is on the calculation method and the source. This technology plays a huge role in the traceability and recovery of files and can be applied to cloud storage scenarios. In 2009, data traceability technology was listed as one of the three important technologies to ensure national security by related reports, and it still has a lot of room for development in the future data and information security field.

RBAC (access control based on related roles) adopted a top-down management model in the early stage - the role division based on the role of the enterprise, and the bottom-up management mode was selected in the latter stage - the role was automatically implemented according to the existing role. Optimization and extraction, the latter is role mining. Usually, the technology can be used to automatically generate characters according to the user's click status, not only to complete

personalized services in time, but also to discover potential dangers against abnormal behavior of users. By collecting and analyzing the behavior data of users and their devices, the technology acquires the behavior characteristics of users and their devices, and then can verify the identity of operators and their devices by using the acquired feature information. The use of identity authentication technology increases the difficulty of hacker attacks, reduces the burden on users, and effectively unifies the authentication mechanisms of different systems.

Watermarking technology refers to embedding identifiable information into a data carrier in some unobtrusive manner without affecting the content of the data and the use of the data. Generally used in media copyright protection, there are also some database and text files that apply watermarking technology. However, the application of watermarking technology on the multimedia carrier and the database or text document is very different. The characteristics of the data based on the disorder and dynamics of the two are not consistent. Data watermarking technology can be divided into strong watermarks, which are used to prove the origin of data and protect the original author's creative rights. Fragile watermarks can be used to prove the authenticity of data. But watermarking technology is not suitable for big data that is now mass-produced quickly. This is something that needs improvement. Research on data traceability technology began in the database field and is now being introduced into big data security and privacy protection. Marking the source data can shorten the time when the user judges the authenticity of the information, or help the user to verify whether the analysis result is correct or not. The marking method is the most basic method in the data traceability technology, mainly the calculation method (Why) and the data source (Where) of the recorded data. Data traceability technology also plays a significant role in the traceability and recovery of files.

In terms of structured data, to effectively implement user data security and privacy protection, data publishing anonymous protection technology is a key point, but this technology needs to be continuously explored and improved. Most of the existing data publishes the basic theory of anonymous protection technology. The setting environment is mostly that users publish data once and statically. If the identifiers are grouped by tuple generalization and suppression processing, the collection of common attributes is anonymized using the k anonymous mode, but it is easy to miss a particular attribute. But in general, the reality is changeable, and data is generally published continuously and repeatedly. In the complex environment of big data, it is more difficult to implement data distribution anonymous protection technology. An attacker can obtain various types of information from different publishing points and different channels to help them determine a user's information. This also requires researchers in the information field to invest more energy and research.

5. Conclusion

When the era of big data came, it brought opportunities for technological development, but also brought new problems and challenges. Big data security and privacy protection are just some of the problems that need to be solved. Through the specific research and technology mining of the status quo of big data security and privacy protection, this paper discusses the key technologies that can solve the existing information security and privacy protection problems, such as anonymous technology, watermark technology and traceability technology. Of course, these are not just these. Realize big data security and privacy protection. At the same time, we must master some national policies to provide a good environment for the development and application of related technologies, so that big data can better promote the development of human society information technology.

References

[1] Feng Dengguo, Zhang Min, Li Wei. Big Data Security and Privacy Protection [J]. Computing Machine, 2014(1).

[2] Ge Yueying. Information Security in the Age of Big Data and Protection of Citizens' Personal

Privacy [J]. China Information Industry, 2014(1).

[3] Xie Bangchang, Jiang Yefei. How to protect privacy in the era of big data [J]. China Statistics, 2013 (6)

[4] Zheng Chenyang. Research on Network Security Strategy for Big Data [J]. Digital Library Forum, 2014(2).

[5] Li Wei, Wu Xuefang, Zhang Chaoliang. Research on related technologies of privacy protection data mining in information security [J]. Electronic Production, 2013 (24).