Advances in Computer Science Research (ACSR), volume 61
7th International Conference on Education, Management, Computer and Society (EMCS 2017)

Based on the Technological Analysis of Big Data Security and Privacy Protection Technology

Lei Wang
Department of Computer Engineering, Suzhou Vocational University, Suzhou 215104, China
wlei@jssvc.edu.cn

Keywords: Big Data; Big Data Security; Privacy Protection; Data Mining; Protection Strategy

Abstract. With the rapid development of network society and intelligent society, and the increase of data acquisition source, the volume of data is on an explosive growth, which has led to the emergence of big data and wide use in all walks of life. Big data also brings a lot of trouble to people at the same time they bring in the convenience. People found that in the process of big data collection, transmission, storage, management, analysis, release, use, destruction and other processes there were security problems and big data security faced more challenges. By analyzing the data security problems and the privacy risks in the big data environment, this paper puts forward the strategies of big data security privacy protection, and provides reference for the construction of big data security and privacy protection system.

Introduction

With the popularity of mobile smart terminals, the rapid development of the Internet and the rise of Internet of Things, every moment will produce a lot of data information as we have entered the era of big data. In the enjoyment that the development of big data brings convenience to people's lives, people also found that the users' personal information was more easily access to others, which would give users a lot of unnecessary trouble and bring great risks to the social security and stability. Therefore, more and more scholars began to study big data security and privacy protection technology.

The Concept and Characteristics of Big Data

Big data refers to a collection of data that cannot be captured, managed, and processed with conventional software tools within an affordable timeframe, requiring a new processing model to have greater decision-making power, insight, and process optimization Of the massive, high growth rate and diversified information assets[1-2]. The concept of "4V + 1C" can be extracted from the concept of big data, namely Variety, Volume, Velocity, Value, and Complex[3]. Types of big data include digital, text, pictures, audio, video, network log, etc. Common big data are generated from the hospital, banking, post and telecommunications, communication systems and other official collection, the data processed after log on the Internet platform, different types of computer systems, and mobile phone system data files. Big data volume is very large, and the general processing volume of data is in the PB level above. Big data processing runs fast, following the "1 Second Law" because it can have access to high-value information from all types of data quickly. With rational use of big data, you can create high value at a low cost. It can be said that modern society, big data and our lives are closely related.

Big Data Security Analysis

Big Data Security Issues. Big data contains a huge commercial value at present, so many industries are doing big data analysis and mining to provide data support for their business decisions. However, in the enjoyment of convenience that big data analysis and mining bring to their production and life, industries should also realize that in the era of big data, information security is facing more and more problems. In the data collection, transmission, storage, management, analysis, distribution, use, and destruction of the whole life cycle, big data faces many security threats[4].

Technical Analysis of Big Data Security. In big data environment, security requirements from all walks of life are changing, the traditional pattern of data security has not completely satisfied the security requirements of big data. With the increase of the data amount, how to protect the data becomes more and more difficult. At the same time, the distribution and open process of data will also increase the risk of data leakage. Big data security technology researches, therefore, involve the data on the big data business chain production, storage, processing, extraction, commercial application value of data security and protection technology from ensuring the key data security technology. And the use of safety information involves big data analysis and application in the field of information security, including the safety of data collecting, sorting, filtering, integration, storage, mining, audit, the application of key technologies[5]. Different industries have different requirements for big data security because of their needs according to the actual situations for analysis.

Big Data Privacy Protection Strategy

The application of big data has produced great data analysis value, but at the same time also brings a major challenge to privacy protection.

The Concept of Privacy. Privacy often refers to the information that individuals, institutions and entities don't want to be exposed to the external world. Privacy is often divided into two classes based on the value. Privacy is regarded as a kind of human rights, a part of social moral value system, a commodity, and the value of people and society. Based on homologous, privacy is related to personal thoughts, awareness and understanding, which is a state including 4 seed states: anonymous, concealing, retention, and secret, and a kind of control that says a transaction between individuals and others. Its ultimate goal is to enhance autonomy or reduce leakage[6].

In information society, privacy is sensitive information that the owner of doesn’t want to be more disclosed of, including the characteristics of sensitive data and the data representation. It changes with the experience of life. It is not only confidential, anonymous, safe and ethical concepts overlap, but also relying on the special situation[7]. In the big data environment, multi-source cross validation is more likely to find the real users behind the anonymous data, resulting in privacy disclosure.

Big Data Privacy Risks. In January 2012, Mr. Obama said in a consumer privacy bill conference, privacy from the beginning has been our democratic system of the heart, and now more than ever need it, more so big data era[8]. In August 2015, the related department of our country issued on big data development outline of action, "much starker choices-and graver consequences-in planning also made it clear that China should expand network economy space, open and share data resources, realize the national strategy of big data, and advance the layout of the next generation Internet. The Chinese central and local governments have Internet privacy regulators specialized in monitoring network information, controlling the spread of the Internet rumors [9].

1. The Risks of Collecting Data. In big data age, you can get through the users’ information in many ways. When you log in most sites to register or install some application software, you will be forced to provide the users’ geographical location, personal information, network logs and other related data, otherwise it cannot be used or you cannot use the important functions of the software. Some network companies collect and use the privacy and information of users in order to pursue the economic benefits of big data through the provision of network services. Do users rarely know what their information is being used for doing? Who is responsible for the risk of their own information? Is your information spread maliciously? Which will pose a risk to the users’ privacy.

2. The User on the Risks of Ego to Protect Consciousness is not Strong. With the rapid development of the network, people are keen on using social tools for daily contact, publishing their own personal information, thus often provide their privacy to others in the case of unconscious will, resulting in their own privacy leakage. When you are infringed, you cannot take the right measures to protect yourself.

3. The Risk of Data Mining. At present, even with the anonymous way of information collection, through the user information, location information, operation information and other data portfolio analysis, each person’s behavior can easily emerge. In addition, data scientists can mine out valuable information from published data analysis which is likely to involve privacy information of users, for

example, fast clustering method based on graphs k - center[10] and k - median[10], correlation clustering methods Co - Cluste[11], etc. These high performance algorithm can not only dig deeper analysis data of the tiny, unrelated pieces of data between each other, but also provide the background knowledge of attack for malicious data mining analysis through the analysis of leaked privacy information of big data. Big data mining analysis may cause leaks or directly lead to some of their privacy protection method losing efficiency.

3.3 Protection Strategy of Big Data Age Privacy. In view of the privacy characteristics and privacy risks of the big data age, the privacy protection strategies are put forward from the following aspects.

1. To Strengthen and Perfect the Legal Protection for the Construction of Privacy Protection Agencies. At present, in our country no specific laws and regulations can be used to regulate the analysis of the data mining and other technical means to obtain personal information management and use of data through the network. Laws and regulations relevant to personal privacy protection are lack of effective connection between industry norms, and even in some ways, there are conflicts and contradictions. Therefore, our country should perfect relevant laws as soon as possible, so that the range of privacy data can be determined, allowing users to appropriately use private data on the basis of data security.

The United States, the European Union, Japan and other developed countries have set up relatively perfect privacy protection agencies, specifically for the protection of privacy, including privacy and a variety of privacy contents. China's current situation is lack of specialized privacy protection functions, so it is difficult to meet people's urgent needs for privacy protection in big data age. Construction of privacy protection agencies should be given priority to in current China.

2. Develop Privacy Protection Technology. To protect data privacy technology is mainly considered how to ensure that big data applications don't leak privacy, in the process of how to be more conducive to the application of big data. The commonly used techniques include big data privacy protection technology, anonymous data encryption storage technology, privacy of data mining technology and data access control technology.

Anonymity refers to the conditional release of data according to specific circumstances. In big data environment to release the identifiers of anonymous data can remove users' information. But if the attacker can contain the user identifier data sets from other sources, and set up a user identifier and the data records of corresponding relations according to the standard identifier connected to multiple data sets, he can still be accurate to the individual information. In addition, the big data environment for information collection, storage and analysis provides a more powerful support, leading to the increase of the attacker's ability, thus anonymous protection becomes more difficult, therefore the researchers need to spend more efforts ensuring the safety of the big data environment anonymity[12,13].

Big data storage needs to ensure the confidentiality and availability of data to guarantee the security of communication, and data encryption and storage technology can meet this requirement. Based on the privacy of data mining researches focused mainly on association rules mining, privacy protection classification mining, clustering mining and privacy protection, such as mining patterns and so on. Second, the big data mining security technology also needs to strengthen the third party excavator identity authentication and access management to ensure that in the process of data mining third parties are not implanted in malicious programs, do not steal system data to ensure the safety of big data [3]. Access control refers to the users' access to different access rights to the visitors, restricting their access to key resources, and preventing illegal users from entering the system and legitimate users for illegal use of resources. It is also an effective means to protect privacy by the users in the big data age. The common access control modes include autonomous access control, mandatory access control and role-based access control.

3. Strengthen the Privacy Safety Awareness. Users in the privacy protection are more concerned about the protection of personal information. It is urgent to improve their own security awareness of understanding the information will bring the serious consequences of leakage. Some companies collect information for the purpose of improving the service of the users, but it is necessary to strengthen the data management and prevent the loss of information after the collection of information on the personal and corporate interests of the harm.

Conclusion

With the rapid development of network technology, the amount of data to further increase the value of big data will continue to appear, the analysis of big data not only changed people's lives, people have been aware that the traditional data security mechanism cannot meet the security needs of big data, so the security of big data and privacy protection are particularly important. By analyzing the data security and privacy risk in big data environment, this paper puts forward the strategies of big data security and privacy protection, and provides reference for the construction of big data security and privacy protection system.

References

[1] Viktor Mayer-Schonberger, Kenneth Cukier: Big Data: A Revolution That Will Transform How We Live, Work and Think (Hodder Export , Boston 2013).

[2] Cukier K, Mayer-Schoenberger V: The Rise of Big Data: How it is Changing the Way We Think About the World. The Foreign Affairs, Vol. 92 (2013), p.28.

[3] Ni Zhang, Y.Y. Zhang: Big Data Security Technology and Application. (Posts and Telecom Press, Bei Jing 2014). In Chinese.

[4] Xin Lv, X.L Han: The Big Data Security System. Journal of Information Security Research, Vol. 1(3) (2015), p.211-216. In Chinese.

[5] Y.M Nie: Big Data and Its Safety Research. Information Security and Communications Privacy, Vol. 5 (2013), p.15-16. In Chinese.

[6] Y.H. Liu, T.Y. Zhang, X.L. Jin and X.Q. Cheng: The Era of Big Data Privacy Protection. Journal of Computer Research and Development, Vol. 52(1) (2015), p.229-247. In Chinese.

[7] Banscal G, Zahedi F, Gefen D: The moderating influence of privacy concern on the efficacy of privacy assurance mechanisms for building trust: A multiple-context investigation, Proc of the IntConf on Information System (Australian: AIS, 2008) p.14-17.

[8] http://www.whitehouse.gov/sites/default/flies/docs/big_data_privacy_report_may_1_2014.pdf

[9] H.D. Sun: Research on Personal Privacy Protection in Big Data Age. Computer Knowledge and Technology, Vol. 13(2017) No 1, p.19-34. In Chinese.

[10] Alina E, SungjinIm,Moseley B: Fast clustering using MapReduce, Proc of the 17th ACM SIGKDD IntConfon Knowledge Discovery and Data Mining(KDD 2011)(NewYork: ACM, 2011), p.681-689.

[11] FlavioC,NileshD,Ravi K: Correlation Clustering in MapReduce, Proc of the 20th ACM SIGKDD IntConfon Knowledge Discovery and Data Mining(KDD 2014)(NewYork: ACM, 2014), p.641-650.

[12] Sedayao J, Bhardwajr, Goraden: Making big data, privacy, andanonymization work together in theenterprise: experiences and issues, Proceedings of the 3rd InternationalCongress on Big Data(Anchorage, Alaska, USA, June 27-July 2, 2014,) 2014.

[13] Sun G Z, Wei S, Xie X: Deanonymization technology and applications in the age of big data. Information&Communications Technologies, (2013) No 6, p.52-57.