Protection of Personal Information Security in the Age of Big Data

Hui Zou
Sch. of Marxism, Wuhan Univ of Technol.
WuHan, China
e-mail：donna1314520@163.com

***Abstract*—As a production factor, big data is being intensively integrated with the development of various industries. The consequent security issues concerning personal information are becoming increasingly severe. This article aims to analyze the major issues which arise with big data, including the breach of personal information, the potential security risks and the users' reduced personal information control rights. It is followed by the analysis of their causes: users' weak awareness of information security, the under-developed laws and regulations concerning the security of personal information, the laggard technology for security maintenance of big data, and motivations for profits. In order to solve these problems, firstly, the awareness of information security and users' ability of security maintenance should be strengthened. In addition, the security management system of big data should be improved with deepening reform of security technology. Furthermore, legislation and regulation system should be reinforced. Lastly, standardized codes of practice and self-discipline pact of the industry should be reinforced to ensure the security of personal information in the age of big data.**

Keywords-The age of big data;Security of personal information;Information breach;Information security threat;Protection countermeasures

Ⅰ. INTRODUCTION

With the prevalence of mobile internet, the internet of things, smart terminals, network transmission and social networks, big data is generated in every passing second. As is indicated by IDC, with the development of information age, the amount of data worldwide doubles every two years and is predicted to reach 40ZB in the year of 2020. As a production factor, data has been integrated with the development of all walks of life, including telecommunications, energy, medical care, transportation, business, finance and education, and has provided scientific instruction and precise basis to government and enterprises in the aspects of management, decision-making and services.

As for the definition of big data, it can be interpreted from various perspectives. McKinsey Global Institute released a report contended that big data refers to the data sets whose abilities of collection, storage, management and analysis exceed typical database software. Wikipedia defines big data as the data integration which is difficult to be processed by current database management tools, with the features of massiveness and complexity. Ma Yun, the founder of Alibaba, defines big data as a certain kind of service. At present, the most widely-accepted definition was put forward by Douglas Laney, an analyst from US Gartner Group. He emphasized the 3V features of big data, namely, volume, variety and velocity[1]. With an increasingly in-depth understanding of big data, the 3V features has developed to 4V (added by veracity) or 6V (added by value, vender and veracity). Some common features of big data can be summarized from the various definitions mentioned above. Firstly, big data is the massive large-scale data sets including structuralized and non-structuralized data. Secondly, the essence of big data lies in the increasing data value by constant processing and utilization. Thirdly, the "One Second Rule" of big data processing indicates that the result of data analysis should be presented within limited seconds. If it takes too much time, the data value will gradually lose. Fourthly, the prospect of big data industry is promising, which will be intensively integrated with the development of various industries.

However, with the concentration and growth of volume, data, especially the core data, is gradually controlled by the minority, which gives rise to the phenomenon of "data monopoly" to some extent. Furthermore, due to the economic and political profits of data, big data has become the target of cyber-attack. When users are enjoying the convenience of surfing the internet, someone ill-intended is analyzing their needs according to the data of browsing history in order to sell things. What's worse, user's private information will be acquired without permission or notice. Traditional information security measures can never meet the requirements of the new age. Urgent issues have arisen, including how to ensure the personal information security, how to prevent personal information from illegal collection and utilization, and how to strengthen users' control of their information.

Ⅱ. MAIN PROBLEMS ABOUT PERSONAL INFORMATION SECURITY IN THE AGE OF BIG DATA

A. Information Breach

Gemalto, a research company of digital security, released a report of Breach Level Index for the year of 2015. It pointed out that 2015 witnessed serious incidents of data breach. During the 12 months, security staff of the company collected and categorized 1673 cases of data breach, resulting in the breach of 707 million data records. The major source is the hostile outside attacker (Figure 1). The major forms include the attack of personal information and identity theft (Figure 2). According to the report about *the Protection of the Rights of Chinese Netizens 2015* released by Internet Society of China, the personal identity information of nearly 80% netizens had been breached. The information about personal online activities of over 60% netizens had been breached. The

978-1-5090-4840-3/16 $31.00 © 2016 IEEE
DOI 10.1109/CIS.2016.141

breach of personal information, spam and fraud information led to a loss of 80.5 billion yuan. From the perspective of the security of personal information, internet users from android system are one of the main targets of hackers. The use of wireless Wifi also increases the risks of data breach. The most typical type of personal data breach is the disclosure of personal private information. In recent years, the incidents of bank card fraud have been more prevalent. Criminals steal personal information about the card by sending phishing messages by simulated base-stations, free Wifi, or refitted POS machines to steal the assets of users. Personal information breach does harm to normal lives and interests of citizens, either to their personal or property safety.

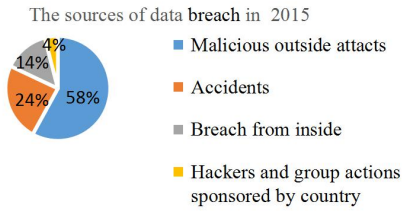

Figure.1 The sources of data breach in 2015 (released by Gemalto)

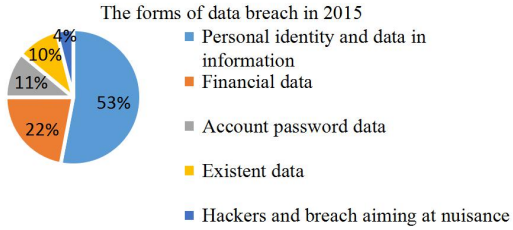

Figure.2 The forms of data breach in 2015 (released by Gemalto)

B. Security Risks

The existing non-big data security solutions are not designed to handle the scale, speed, variety and complexity of big data. Most organizations lack systematic approaches for ensuring appropriate data access mechanisms[2]. The issues of security risks are expressed in the following aspects. Firstly, the big data itself is the tool for hackers to attack users' information. Motivated by material benefits, hackers will use social platform, electronic commerce, e-mail and self-media to collect personal information about phone number, home address, family income and recent demands. The information is then sold to merchants for marketing. Some merchants even simulate mobile banking, website and operators to conduct fraud online shopping and lottery fraud. Secondly, big data becomes the main target of online attack. The feature of nonlinear structuring of big data means that huge amounts of data is as well as the interweaving complex data. In addition to its own value, the data also has potential value, which can be used by attackers to make more profits. Thirdly, big data becomes the high-level carrier which is vulnerable to continuous attack. The attack techniques of APT are highly targeted with long attacking period, which has become the major threat to the security of big data. It is a continuous and complex process, which is not subjected to real-time detection. In addition, the value density of big data is very low. The current security protection tools are unable to focus on certain value point. As a result, virus and trojan have many opportunities to be hidden in big data.

C. Users' reduced Personal Information Control Rights

Personal information rights are frequently called informational self-determination rights, which embody the protection of personality interests, including self-determination[3]. Everyone should have the right to control his or her personal information. Every person's information rights should be protected by the law. Compared to the traditional environment, in the age of big data, users' personal information control rights are remarkably reduced. On the social platform, personal information is easy to be browsed, collected and transmitted. Through integrated analysis of personal information on different social media, it is easy to establish an information system, including preference, social circle and belief of the target person. In the age of big data, data collection and usage are easier compared with traditional information transmission, which lead to the changing position of original data owners from active to passive. To some extent, it even involves the infringement of property rights, which imposes challenges on personal information security.

III. CAUSES FOR SECURITY ISSUES OF PERSONAL INFORMATION IN THE AGE OF BIG DATA

A. Users' weak Awareness of Information Security

For those with weak awareness of information protection, their security issues will be worse. The misbehaviors of netizens include the followings. Firstly, the account password is rarely changed. According to *the Survey of Security Awareness of Netizens in China (2015)*, 81.64% netizens in China didn't change password regularly. 75.93% netizens used the same password for different accounts for convenience. 44.42% netizens chose their birthday, phone number or name spelling as passwords. These problems also bother netizens in many other countries. Secondly, related rules about information protection are usually ignored. When signing in a website or downloading software, users will be informed of related rules concerning users' information protection and responsibility. If they choose to accept, they are allowed to download, install and use the software. Otherwise, they will be prohibited to do so. A majority of users never read the terms, and click "Accept" immediately, which might mean authorizing related operators to access their personal information. Thirdly, many people link to Wifi or scan two-dimension code without much care. The risks behind free Wifi and two-dimension code are frequently neglected. Once a user is linked to phishing Wifi and opens the consumer app on the phone, the order and expense calendar will be fetched completely. Criminals might pant two-dimension codes into virus program, and tempt victims to scan it by fraud or pretending to be

coupon. Important information of victims will be easily disclosed. Then, the password can be tampered by short message identification, and the property in the account is removed easily.

B.The Under-developed Laws and Regulations Concerning the Security of Personal Information

At present, China is in lack of a complete legislation system for the protection of personal information, and the country hasn't attached enough importance to the protection of customers' information. The existing laws and regulations are unilateral in content with limited range of information protection. They are in lack of concrete and feasible terms, as well as effective implementation measures.The causes of the threat can be attributed to three factors. Firstly, the subject of law enforcement is not clearly identified. At present, Ministry of Public Security, Industry and Commerce Authorities and Ministry of Industry and Information Technology, all have rights to supervise the misconduct on the internet. However, as for the implementation of specific cases, the responsible party should be further identified distinctly. Secondly, the companies which are trading users' data illegally are seldom punished. Lastly, several companies are not aware of related laws, so they may never notice their illegal conducts. Moreover, the loss caused by personal information breach is not easy to define, which results in difficult evidence acquisition and legal rights protection.

C. The Laggard Technology for Security Maintenance of Big Data

At present, many websites suffer from security flaws, including network protocol flaws, software flaws, and security management flaws. Because of these flaws, attackers can visit or destroy the system without authorization, which severely threatens the security of the system. In the new age, the super-massive volume of data and the existence of half-structuralized and non-structuralized data have exceeded the maintenance ability of traditional database. The laggard technology for security maintenance of big data cannot provide a reliable protection. In terms of the processing technologies, they are still at the stage of development, without mature solution or corresponding equipment. Moreover, distributed system deployment entails stricter requirements for techniques, which is unable to ensure security for multiple users. As for data mining technique, it can act as an intrusion detection technique to spot unknown intrusion behaviors in internet attack. The methods of data mining include association rules mining, decision-making tree, neural network, rough bayesian networks and visualization[4].However, as big data is various, complex, inconstant, and discrete, data mining technique cannot be used directly, which imposes more challenge. As a result, valuable data might be flooded by massive data, which is also an impediment to data security protection.

D. Motivations for Profits

In the information age, data can bring about abundant economic benefits. The transaction of personal information is motivated by the profit chain behind it. Many criminals attempt to acquire personal information of other people to seek their own benefits. After accessing personal information, hackers can intrude the account and email of users to acquire more information, or even sell it. The buyers will use it for internet promotion, telecom pam or electronic junk mail. Hackers often have more than millions of user information, which can be sold to different people for various times with abundant profits. Both the sellers and buyers are motivated by profits. Buyers can secure a dominant place in market competition through a third-party company to make their marketing more targeted. And the sellers can directly acquire profits by selling personal information. This is a market of massive demands, which is difficult to supervise.

Ⅳ. COUNTERMEASURES OF PERSONAL INFORMATION PROTECTION IN THE AGE OF BIG DATA

A.The Awareness of Information Security and Users' ability of Security Maintenance Should Be Strengthened

Personal information security is an important part of national security. From the perspective of national security strategies, the public awareness of information security should be strengthened comprehensively. As for individuals, when they use information resources, they should attach great importance to the protection of personal information. They should be cautious, when browsing the web pages, logging in websites, using social platforms and electric commerce service platforms. The awareness of setting hierarchic password should be strengthened. The password to account involving privacy should be made more complex. Users had better not use free Wifi in public places, or scan unknown two-dimension codes without care. It is reasonable to reduce the use of online payment software in public places. The calls and messages from strangers should never be trusted. Government departments should include information security into their routine work as an essential part. The top-level design of national information security system should be enhanced. Big data security development plans should be settled down. Internet users should be frequently cultivated about the knowledge of internet security and risks in various ways. Moreover, the monitoring and management of core data and sensitive data should be reinforced. Proper privacy education should be strengthened to remove narrow-minded consciousness of privacy protection and help people reasonably share personal data. Thus they can then choose the scope of the data to be released as well as the data sensitivity according to the various situations[5].

B.The Security Management System of Big Data Should be Improved with Deepening Reform of Security Technology

The financial investment should be put in big data

security management system, in order to make breakthroughs of technological research and development. The protection of the security of phone terminals, and the construction of security infrastructure should be enhanced. Talents in information security techniques should be cultivated to increase the capability and level of the whole team. Threshold should be set concerning companies' collection of users' personal information. The scope, extent and range of data usage should also be strictly managed. In terms of information security protection, a unified standard system of network identification should be established to standardize data of various forms. In addition to the use of traditional anti-virus software and setting of firewall, big data can be utilized to clearly classify different risks according to their features, in order to identify malicious calls, harassing information and phishing sites, which could provide reliable supports for users to make distinction. As for the monitoring and tracking of big data, emphasis should be laid on the tracking of internet attacks based on big data. The attack models should be handled with big data technology to establish models of data monitoring and tracking and to establish security tools. As for data management, the range of key database should be determined, in order to make thorough rules about database management and security operation, as well as to monitor key database.

C.Legislation and Regulation System Should be Reinforced

The establishment and implementation of legal system by legislative, judicial and supervision departments should adapt to the development of big data. Based on the current legal system, personal information privacy and public information security should be clearly defined to specify the sentencing criterion about information security crime. Through legislation, the responsibility and obligation of different subjects should be specified to normalize the legal usage of data. A set of laws, regulations, rules and administration should be carried out more rapidly. A stricter system of big data service, technology and products should be set up, to prescribe the operation of big data service providers in a legal way. Systematic prevention mechanism should be improved, including equipment purchase mechanism, information security mechanism, safety precaution mechanism, emergency response mechanism, staff management and professional team cultivation mechanism, in order to construct internal security compulsorily and to improve the ability of big data security protection[6]. The advanced experience of the developed countries can also be learned. For instance, the US has a sound regulation system to supervise the internet effectively in terms of strategy, tactics, technology and management. In 1995, the European Union issued *the Guidance to Protection of Personal Data*, which protected personal information during the whole process before, during and after the event.

D. Standardized Codes of Practice and Self-discipline Pact of the Industry Should be Reinforced

Supervision should be strengthened from the industry. Companies should improve self-regulation, and lay emphasis on the software and hardware construction of their information security and the construction of regulations concerning inside information security. As for the dishonest websites with frequent flaws, the whole industry should impose more supervision and warming. For the severely illegal websites, which refuse to mend their ways, the supervision departments of the industry will publish the Hit List and impose strict punishment. Self-discipline pact of the industry should be carried out about users (personal information owner), data service provider, and data customer, in order to ensure the validity and legality of data usage and sharing. The usage period and forms of data should be specified, with a clear definition of punishment to offense to codes of practice, in order to create a safe and sound environment for data usage.

Ⅴ. CONCLUSION

As the product of information age, big data aims to exploit the potential values between data in order to serve the economic and social development through the storage and analysis of massive data. However, with the further development of the industry, the issues arising thereby have attracted increasingly public attention. The leading companies around the world are caught in the scandals of user information disclosure, which imposes severe threats to the information security of citizens. In the age of big data, the initiative in the market is possessed by those who own the maximum data. As a result, the requirement for data, including personal information, is expanding, which further raises the importance of protecting personal information in the age of big data. In this context, this article introduces the definition, features and current development of big data, on the basis of which it analyzes the major issues of personal information security and their causes, and ends with the countermeasures of protection. The author hopes to add some inspiration to the topic through the study.

REFERENCES

[1]Laney, D.,2001.3D Data management: controlling data volume,velocity,andvariety[EB/OL].http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

[2] Nir Kshetri. Big Data's Impact on Privacy, Security and Consumer Welfare [J]. Telecommunications Policy. Volume 38, Issue 11, December 2014, Pages 1134–1145.

[3]Margaret C. Jasper. Privacy and the Internet: Your Expectations and Rights under the Law [M]. New York: Oxford University Press,2009:52.

[4] Xiangxiang Cong. A Study on the Space-time Simulation Mining Algorithm in the Age of Big Data. [D]. Shang Hai: East China University of Science and Technology,2012.(in Chinese)

[5]Zhong Wang, Qian Yu. Privacy Trust Crisis of Personal Data in China in the Era of Big Data: the Survey and Countermeasures[J]. Computer Law & Security ReviewVolume 31, Issue 6, December 2015, Pages 782–792.(in Chinese)

[6] Zuoning Chen, Guangyi Wang. Big Data Security and Autonomous Controllability. Chinese Science Bulletin. Volume 60, Issue 5 to 6, February, 2015. (in Chinese)