Personal Information Leakage Risk Analysis and Prevention Research in the Era of Big Data

Weichao Li [1,2,*], Junyao Tan [3], Qihang Jiao [4], Qin Li [1]

[1] Zhengzhou University of Aeronautics, Zhengzhou 450046, Henan, China
[2] Collaborative Innovation Center for Aviation Economy Development, Zhengzhou450046, Henan, China
[3] Beijing Normal University, Beijing 100875, China
[4] University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: liweichao@zua.edu.cn

Abstract: In the era of big data, massive data brings convenience to people's life, but also faces the threat of personal information leakage. Then how to prevent personal information leakage has become a research hotspot. This paper analyzes the problems of personal information leakage based on the big data environment, and puts forward the preventive measures for personal information leakage, which has certain reference value and significance for the prevention of personal information leakage threats in the era of big data.

Keywords: big data era; personal information; information disclosure; risk analysis

1. Introduction

With the advent of the era of big data and the development of data mining technology, personal information has great commercial value, and network security risks and personal information leakage problems are increasingly prominent [1]. In 2017, Chinese netizens encountered increasingly prominent network security problems. According to CNNIC's 42nd Statistical Report on Internet Development in China, the number of Internet users in China reached 802 million, the Internet penetration rate was 57.7%, and the highest personal information leakage problem occurred, reaching 27.1% [2], a large number of information intensified, has become a more serious social problem. The disclosure of personal information will not only bring serious losses to information subjects, but also seriously threaten the healthy development of big data industry [3].

2. Big Data Environment and Features

2.1. Big Data Environment Analysis

According to the definition of Wikipedia, big data is a collection of data that cannot be captured, managed, and processed in an affordable time frame with conventional software tools. Zhang yanxin believes that big data is the data with huge scale and complicated form that are processed by colleges and universities with standard database technology [4]. In short, big data is dependent on the development of information technology. In the massive information resource data, through the use of information technology to analyze the content of big data and explore its inherent value for our use it.

2.2. Characteristics of Big Data

Different scholars have different definitions of the characteristics of big data. Some scholars believe that big data has three characteristics, which are large number, various types and fast speed. Some scholars believe that big data has four characteristics: large number, various types, fast speed and high value. This paper divides the characteristics of big data into the following four aspects:

(1) Rich data information. The data in big data is counted in units of tens of billions of units. The data search in big data needs to be carried out in the amount of information in this trillions of levels, reflecting the richness of data information.

(2) Diversity of data types and forms, including structured data, semi-structured data and unstructured data.

(3) Convenient and fast data processing. Because data information has timeliness, in order to prevent the loss of effectiveness of data information, big data is processed in the first time in seconds, which is convenient and fast.

(4) Low value density of data. Each data in the big data is of great value, but the value density of each data is very low, which requires deep excavation and research to obtain the value.

3. Personal Information Leakage

3.1. Personal Information

The content of personal information in this paper is mainly divided into the following three aspects: On the first hand, basic personal information, such as name, gender, age, id number, address, telephone number, email address, etc. Second, the device information used by the individual mainly refers to the IP address and MAC address of the computer PC, such as mobile positioning information, wifi information that is available on our mobile phones. Third, personal account information including personal communication information, SMS information, micro blog account, QQ account, WeChat account and other chat software information.

3.2. The Way and the Danger of Personal Information Leaks in the Big Data Age

3.2.1. The ways of personal information leakage in the era of big data

(1) Personal information leakage is caused by personal online behavior. The majority of Chinese Internet users are young and middle-aged, and the proportion of people aged 40 and above is increasing, as shown in Figure 1. With the change of network penetration rate, the penetration rate of Internet users is still rising. With the rapid development of the Internet today, people use various apps or small programs in shopping, work, travel, entertainment and other aspects to obtain our account nickname, personal profile picture, geographical location and other information, leaving our personal information inadvertently.

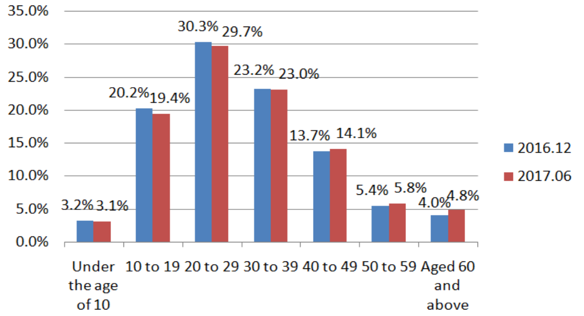

Figure 1. Distribution of Chinese Internet users (data source: public data collection).

Countries, enterprises and other organizations through the collection, sorting and use of personal information, to provide personal services become more convenient and quick. There is no shortage of elderly or teenagers due to insufficient identification of the Internet, and it is impossible to identify phishing websites or scan irregularities. The personal wrong behavior such as the QR code causes the personal information to be collected, organized, and used twice by the criminals. According to the China Consumers Association, pair 100 APP personal information collection and the privacy policy evaluation, according to data, 10 types of APP are widespread suspicion of too much personal information collection, 59 APP allegedly excessive collection “location”, 28 APP allegedly excessive collection “directory information”, 23 APP allegedly excessive collection “identity”, 22 APP allegedly excessive collection “phone number”, etc.; Some mobile phones even have abnormal phenomena such as automatic sending of short messages [5]. Usually, personal information is a part of personal privacy, which is not allowed to be obtained without personal consent. However, the APP is excessively collecting personal information, which makes many users disclose their personal information without their knowledge.

(2) Using software technology to steal personal information. In today’s big data environment, the common application of computers not only facilitates people’s information exchange, but also brings challenges and threats to the security of personal information. In Table 1, ten personal information leakage incidents at home and abroad in 2017 are sorted out. Personal information leakage is caused by internal system vulnerabilities, configuration errors or internal staff; some are also attacked by external hackers and virus software. No matter what causes the information leakage, it will cause huge economic loss and negative social impact.

Table 1. Personal information leakage events at home and abroad in 2018.

An enterprise name	Event release date	What profile	Range
Facebook	2018.3 & 2018.9	Cambridge analytica used Facebook users’ personal information without their permission, and hackers took control of 400,000 accounts	Cambridge analytica improperly used the private information of 87 million unauthorized users; The hackers used 400,000 accounts under their control to gain access to information about 30 million Facebook user accounts.
Ac Fun	2018.6	User data was stolen by hackers	Ac Fun’s 9 million user data were leaked and sold on the dark web for 400,000 yuan.
HZH owns several chain hotels	2018.8	Hackers attack website vulnerabilities	240 million check-in records leaked. It involved more than 3700 hotels in 370 cities.
Marriott starwood	2018.11	Using the wrong security system	About 500 million users’ information has been leaked in the past four years.
Yuantong express	2018.6	Yto employees sell express data internally	A billion users’ data was openly sold on the dark web.
SF Express	2018.8	Sf express users personal information flow to the market, the source of the leak is not officially specified	300 million user data is sold on the dark web.
51 job	2018.6	Some user accounts are hacked	1.95 million resumes leaked.
Under Armour	2018.3	The data breach was hacked	The 150 million user data breach included user names, emails and passwords, not including credit CARDS. More private information, like id CARDS.

MyHeritage	2018.6	The website server was attacked	MyHeritage, a website for family genetic and DNA testing, has 92 million users whose information has been compromised with potentially serious consequences.
Panera Bread	2018.4	Site loopholes	More than 37 million users have had their information compromised for more than eight months.

Source: according to the Internet open data collation.

(3) Personal information leakage is caused by the interests of Internet black and gray industry. Network black ash industry, it is the behavior that USES network to begin illegal and criminal activity such as fraud of telecommunication, phishing website, Trojan virus, Hacker blackmailing. In a slightly different way, “black production” refers to cyber crimes that directly violate national laws, while “gray production” is a controversial act that walks on the edge of the law and often provides assistance for “black production”. There are four types of black ash industry: false account registration and other sources of black ash industry; a platform for illegal transactions and exchanges; Trojan implantation, phishing websites, all kinds of malware, etc. Most of the black accounts on the Internet are realized in the form of malicious registration, false authentication and identity theft [6].

According to the “2018 research report on the governance of network black ash industry” released by nandu big data research institute and other institutions, it is estimated that the scale of China’s network security industry in 2017 is more than 45 billion yuan, while the scale of black ash industry has reached nearly 100 billion yuan. The huge profits of the black ash industry drive employees to obtain personal information through various means, and the black ash industry on the Internet has become one of the important sources of personal information leakage.

3.2.2. Harm of personal information disclosure

At the present stage, the commercial value and legal significance of personal information are increasing day by day, and personal information leakage also exists in a normalized and multiple form. The individuals who suffer from information leakage will be troubled by harassing phone calls, junk messages, telecom fraud, account theft, property damage and so on. According to statistics, in 2016, Chinese netizens suffered a total economic loss of about 91.5 billion yuan due to junk information, fraudulent information and personal information leakage, and the average loss was about 133 yuan. There may be some degree of growth every year; Personal information disclosure will not only have a bad impact on our personal lives, but also bring us huge economic losses, and may even threaten our lives. When our personal information is used by illegal branches, they can carry out more criminal activities in a more targeted manner. Citizens are panicked by various security threats, which affects the stability of society as a whole [7].

4. The Causes of Personal Information Leakage in the Era of Big Data

4.1. Weak Awareness of Personal Safety

In the context of big data, most of the disclosure of personal information is caused by insufficient legal knowledge of information security and weak security awareness. When using the Internet, some citizens fill in their personal information arbitrarily and carry out real-name authentication without any defense. Others voluntarily fill in their personal information for a small profit. Little do they know that in the Internet age, really effective personal information has important value. Individual safety consciousness is weaking, bring about illegal element to be able to take advantage of.

4.2. Technical Precautions are Lacking

The common use of computers greatly facilitates people’s information exchange, but also brings challenges and threats to personal information security. Driven by personal interests, lawless elements develop various types of viruses and implant Trojan horses and various phishing websites to invade users’ computers, mobile clients and steal personal information. In addition, the existence of the computer itself is also a problem of personal information leakage. Surveys show that about 70% of China’s information products and technologies are imported from abroad, and there may be security traps in hardware facilities and operating systems [8]. In dealing with these problems, the imperfect preventive measures and the lack of information technology will increase the risk of personal information leakage.

4.3. Laws and Regulations are not Sound

Personal information disclosure is so rampant, a very important factor is that China’s laws and regulations are not sound, criminals of low cost of crime, less responsibility. Although there are nearly 40 laws, more than 30 regulations and 200 rules and regulations in China, such as the General Principles of Civil Law, the Law on the Protection of Consumer Rights and Interests, the Law on Electronic Commerce and the Decision of the Standing Committee of the National People’s Congress on Strengthening the Protection of Network Information. Rules and regulations, all related to the protection of personal information [9], but overall, the relevant laws and regulations are still relatively scattered, and don’t have a specific and unified personal information protection law, though the enterprises to protect the personal information of legal obligation, not specific provisions enterprise information security system, as it should be also lack specific inspection standard, Moreover, in the face of the emerging new problems in the information age, there are still many loopholes [10].

4.4. Insufficient Supervision

Our country not only lack of complete regulations complete personal information protection law, but also the lack of solely responsible for user information management and supervision organization, many of the rules and regulations need every industry self-regulation and supervision between each other, and in the aspect of administrative supervision in our country is relatively backward, and there is no unified administrative supervision mechanism, so there is no way for enterprises to supervise, so also does not reach the designated position for citizens' personal information protection measures. In the context of information network, there is still a lack of user information security protection in China. It is mainly reflected in the following aspects: the lack of user information protection regulations, unable to provide guidance and systematic legal provisions for the protection of user information.

5. Measures to Prevent Personal Information Leakage in the Era of Big Data

5.1. Popularize Personal Information Security Knowledge and Improve Personal Information Protection Awareness

The main body of cyber security threats is individual citizens. No matter how perfect the network security technology and legal system is, it is impossible to completely eliminate the threat of information leakage [11]. We need to enrich our own security knowledge and improve our own information security literacy. When facing the risk of information disclosure, we will consciously increase our vigilance and protect our information security. In the context of big data, citizens should be clearly aware of the threat that information leakage may bring. They can take the initiative to participate in the information security education activities organized by relevant companies and enterprises, and develop good habits of surfing the Internet. Not to connect to free WiFi at will, not to click on the Connection at will ,nor to believe in winning prizes easily, not to browse the web without trace. You can install effective anti-virus software on your own work or home computer or mobile client to strengthen the protection of computer firewall; finally, we should supervise the companies or enterprises that have provided information, strengthen education supervision and management for the personnel who have access to confidential data, and prevent them from reselling information. As the subject of information, we should always understand the relevant laws and regulations on the protection of our information. In case of violation, we will be held accountable.

5.2. Strengthen the Technical Prevention System of Information Leakage Protection

According to data analysis, 65% of products related to information security protection measures such as firewall system and encryption mechanism in China are purchased by the state, which shows that China has a big gap with foreign countries in the field of information technology [12]. In the era of big data, in order to prevent the disclosure of personal information, it is not only necessary to improve the legal treaties and strict internal supervision system of enterprises or companies, but also need to have a strong sense of personal security protection and a strong technical prevention system and high-end technology development and application. Its prevention and protection system can be established from the perspectives of system security, network security, storage security and operation security [13]. Under the strict protection of the prevention system, not only personal information security can be protected, but also the problem can be dealt with strictly before the occurrence, so as to avoid greater losses, so as to protect personal information security more effectively.

5.3. Improve Relevant Laws and Regulations

Law is the foundation of the country, and improving the legal system is an important factor to ensure social stability and development. With the increasing phenomenon of citizens' personal information infringement and disclosure, it is extremely urgent to formulate relevant laws and regulations. Mandatory laws can regulate some bad behaviors on the Internet. Foreign countries have issued relevant laws and regulations on the issue of personal information disclosure. In February 2015, the United States issued the consumer privacy rights case (draft), which mainly regulates the handling of personal information in the business environment and provides a programmatic basic guarantee for the protection of personal information [14]. In May 2018, the EU implemented the EU general personal information protection regulation, which established new requirements for personal information protection in the era of big data and realized a new situation of "one continent, one law" personal information protection system [15]. China is also making continuous efforts to protect personal information security. In September 2018, the standing committee of the 13th National People's Congress included the personal information protection law and data security law into the first category of the legislative program. If the "personal information protection law" and "data security law" is issued, this will be a targeted personal information security laws and regulations in our country, can change the current situation of relevant laws and regulations and the lack of unified and strong enforcement agency, is expected to form a standard, systematic legal rules and regulations.

5.4. Establish a Unified Regulatory Mechanism

Personal information due to the insufficient supervision and management, information leakage is serious, Government regulatory authorities need to strengthen the protection of personal information security at this stage, monitor and manage the contact, collection and circulation of personal information, urge enterprises to continuously input costs and fulfill relevant obligations, and also increase leakage of user data from the source. Strike strength and punishment [9]. At present, although our country has a small number of regulatory agencies, it is not unified management, so not only management difficulty, but also cannot effectively prevent the disclosure of personal information. The establishment of a unified supervision mechanism can not only reduce the work pressure of relevant staff, but also effectively prevent the disclosure of information in business, administration, education and other network information leakage problems. In this way, it can not only make up for the lack of supervision, but also reduce the threat and loss to users caused by information disclosure.

6. Conclusion

The development of big data not only brings convenience to people's life, but also faces the threat of information leakage. Although big data is rich in information, diverse in data and convenient in data processing, it is applied in various fields of society, and the value of data information is constantly increasing, which leads to the continuous leakage of citizens' personal information. The state and citizens should pay attention to the harm of personal information disclosure, the state should improve relevant laws and regulations, citizens should improve the awareness of personal information protection, jointly maintain the network information security order, effectively resist the risk of personal information disclosure.

Acknowledgment

Foundation support: Science and technology innovation talents in Colleges and universities in Henan province "Research on network public opinion big data mining and intelligent early warning service system" (18HASTIT034), the higher education in Henan province study and teaching reform project "Research and practice on the reform of the training mode of information management personnel based on E-learning" (2017SJGLX403), Henan Education Science 13th Five-Year plan "Research on the construction of lifelong education system in Henan Province under the Internet+ environment" ([2018]-JKGHYB-0168), Research on Information Co-construction and Sharing System in Henan Cross-border Electronic Commerce Comprehensive Test Zone. Henan Soft Science Research Project (192400 410222).

Reference

[1] C. Lovelock, J. Wirz. *Services Marketing: People, Technology, Strategy*. Pearson Prentice Hall: USA, 2006.

[2] CNNIC China Internet information center. Statistical report on Internet development in China. Available online: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201808/t20180820_70488.htm.

[3] Zhan Xin, He Yuyu. Illegal disclosure of personal information in the network environment and its countermeasures. *Chongqing and the world (academic edition)*, **2016**, 12: 99-102.

[4] Zhang Yanxin, Kang Xuran. Research on personal information security of social networks in the era of big data. *Lantai world*, **2014**, 5: 24-25.

[5] Anonymous. APP should not be a channel to reveal personal privacy. *China business news*, 2018-11-29 (A02).

[6] The China youth daily. The Internet network black ash industry nearly billions of personal information leakage is source. Available online: http://tech.qq.com/a/20181009/001578.htm.

[7] Fu Xiaoqiong, Shi Enlin, Wang Wei. A brief analysis of personal information security in the era of big data. *China public security (academic edition)*, **2018**, 3: 89-93.

[8] Xu Yi, Shen Jianfeng. Prevention and control of personal information leakage in the context of big data -- based on the investigation of xu yuyu's death after being cheated. *Industry and technology BBS*, **2017**, 4: 207-209.

[9] Chen Jing. Protection of personal information cannot be achieved without strong supervision. *Economic daily*, 2018-11-28(008).

[10] Tong big column, Discussion on personal information security and protection in network environment. *Network security technology and application*, **2012**, 8: 19-21.

[11] Zhang Liping, Zhang Ying. Personal information security protection in the context of big data: A study on legal prevention. *Research on securities policy and commercial law*, **2018**, 16: 126-127.

[12] Bai Xia. Problems and countermeasures of personal network information security management. *Shandong University*, **2016**.

[13] Wang Huixia, Yao Yao. Survey on current situation of personal information protection of college students in the era of cloud computing and analysis of influencing factors. *Heilongjiang science and technology information*, **2016**, 2: 176-177.

[14] Jing Lijia. The legal interests of crimes against citizens' personal information in the context of big data should be transferred to. *Law review*, **2018**, 2: 116-127.

[15] Liu Yun. Development process and reform and innovation of European personal information protection law. *Journal of Jinan University*, **2017**, 2: 72-84.

Weichao Li, PhD, professor, specializes in digital information resource management.
Phone: 13838518007
E-mail: liweichao@zua.edu.cn.