PERSONAL INFORMATION SECURITY CRISIS IN THE ERA OF BIG DATA

Li Keshu[1] and Shi Meixia[2]

[1]Postdoctoral in Economics, School of Economics and Management, Beijing Jiaotong University

[2]Professor, School of Economics and Management, Beijing Jiaotong University

Abstract: The arrival of the big data era has facilitated all aspects of our lives, meanwhile the rapid development of big data technology has provided channels for the illegal collection, dissemination and leakage of personal information. The personal information security crisis has arrived. However, the big data protection technology in China is not complete, definition of personal information in domestic legislation is not clear, and there is no systematic and perfect legal system in protecting personal information. The author will study the protection of personal information in the big data era from the legislative level, as well as analyze the causes of the personal information security crisis.

Keywords: Big Data Personal Information Security Crisis

1. THE FORMATION, CHARACTERISTICS AND DEVELOPMENT TREND OF BIG DATA

1.1 The formation of the big data era

What is big data? There is no unified definition of big data in today's academia. Gartner Company first used the concept of "Big Data" in 2001. The world-renowned consulting firm McKinsey Company first defined "Big Data" and clearly pointed out the arriving of the big data era. The definition given by McKinsey & Company is: " Data collection whose size is so large that is beyond the ability of traditional database software tools in aspect of capture, store, manage, and analyze." In May 2011, Professor Victor Mayer-Schoenberg clearly affirmed the importance of big data by stating "The main reason to name an era by big data is not because experts and scholars are research big data, but big data is affecting everything in society". Hereafter, all sectors of society began to gradually realize the value of big data and began to conduct a lot of research on it. Among them, the business community paid great attention to big data.

Business entities are also aware of the deeper value of personal information - providing direct reference for corporate strategy. Therefore, data analysis has become common in various business entities. Huge amounts of personal information are collected and stored in the database by them, which can be used to analyze consumer demand and draw users' profile.

1.2 The characteristics of the big data era

Many domestic and foreign scholars have analyzed the characteristics of the big data. The author has summarized the characteristics of the big data on this basis, that is volume, variety and velocity.

1.2.1 Volume, which means the total amount of data is huge. According to the recent report by International Data Corporation (IDC), the volume of global data amount will expand 50 times by 2020.

1.2.2 Variety, which means there are many types of data. Data can be classified into infinite categories in conformity with different classification methods.

1.2.3 Velocity, which means that generating and changing speed of data is fast . In the internet era, new data generates while old data is stored in every moment of every day.

Infinite amount of data has been in a state of constant developing and changing.

1.2.4 High value, which means the value of data is high. One general development trend arise after infinite amount of data is collected, analyzed and summarized, which can be used to predict people's consumption preferences, and this prediction function will produce infinite value.

1.3 The development trend of the era of big data

After the rapid development of big data in recent years, the author believes that the development trend can be summarized as follows:

1.3.1 Data utilization

Data utilization means that data is a resource collection that has great value and can be applied to all aspects of enterprise to achieve the purpose of predicting certain trend or using the data as the transaction target.

1.3.2 Deep integration with cloud computing technology

According to the 2019 Big Data Landscape released by foreign big data landscape production company Matt Turck on July 16, 2019 ,it clearly indicate that the close integration of big data technology and cloud computing technology is the general trend, it also predict that the mutual relationship will be closer in the future.

1.3.3 Breakthrough in scientific theory

With the more and more in-depth analysis and research on big data by various subjects in society, the value of big data on the technical level will gradually be explored, and this value may bring the breakthrough in scientific theory on the whole social level.

1.3.4 Forming data science

When data and society interact to a certain extent, data will form a complete system as mathematics and physics in the foreseeable future, and become an independent and interconnected science with other subjects.

2. OVERVIEW OF PERSONAL INFORMATION IN THE ERA OF BIG DATA

2.1 Different doctrines of citizen’s personal information definition

In the process of domestic scholars' research on personal information, they have been trying to define personal information from different angles. Some scholars believe that personal information refers to the identity card, date of birth, genetic characteristics, fingerprints, and other information of natural person that can identify the individual. The scholar used enumeration to define the scope of personal information, defining personal information as information that can identify the individual including but not limited to the above-listed scope. In addition, there are also generalized models. Personal information refers to information that can distinguish from others and identify a person, either directly or indirectly combined with other information.

2.2 Classification of personal information

Comprehensive consideration of the attribute and type characteristics of personal information in the era of big data, the personal information in the era of big data can be divided into three categories: user identity and authentication information, user data and service content information, and user service related information:

2.2.1 User identity and authentication information

User identity and authentication information is virtual identity information that can be used alone or in combination with other information to identify the user's natural person identity, or to replace the user's natural person identity attributes, and also include authentication-related information used to verify identity. The user identity and authentication information includes two subcategories, one is the user's natural person identity and identification information, and the other is the user's virtual identity and authentication information. Specific description is listed in Table 1.

Table 1.Subcategory and scope of user identity and authentication information

Subcategory	Scope(including without limitation)	Take specific information for example
A1:User's natural person identity and identification information	A1-1:User Profile	Name, ID type and number,age,gender,occupation, work unit, address, religion,ethnicity, nationality,ect.
	A1-2:Certificate of identification	Photocopies of ID cards, officer IDs, passports, driver's licenses, and social security cards,ect.
	A1-3:Physiological identification	Fingerprint, voiceprint, iris, Facebook, etc.
A2:User's virtual identity and authentication information.	A2-1: Identity and authentication information in general service.	Passwords, words of command, and password-protected answers , etc related to phone number, account number, nickname, IP addresses, email address, personal digital certificates in general service.
	A2-2: Identity and authentication information in trading service.	Various account numbers and corresponding passwords, password-protected answers, etc in trading service.

2.2.2 User data and service content information

User data and service content information is data and content information with user privacy attributes that may be collected by others in the era of big data. The user data and service content information include two subcategories,one is user service content and profile data, and the other is user social content information. Specific description is listed in Table 2.

Table 2. Subcategory and scope of user data and service content information

Subcategory	Scope(including without limitation)	Take specific information for example
A1:User's natural person identity and identification information	B1-1: Service content information	Telecom network service content information, such as call content, SMS, MMS, etc . Internet service content information, such as instant messaging content, data file, email content, online purchase orders, logistics information, etc related to personal information transmitted over the Internet.
	B1-2:Contact Information	User profile data such as address book, friends list, group list, Wechat circle of friends list, and list of followers, etc.
	B1-3:User private data	Material data information such as user text, multimedia and other information stored in user cloud storage, terminal, SD card,etc.
	B2-1: Private social content	Social information posted to specific user groups, such as posting content in groups, setting permissions to Weibo content, etc.

2.2.3 User service related information

User service related information is service usage and service related auxiliary information that may be collected by others in the era of big data. The user service related information includes service usage information and device information. Specific description is listed in Table 3.

Table 3. Subcategory and scope of user service related information

Subcategory	Scope(including without limitation)	Take specific information for example
C1:User service usage information	C1-1:Business subscription, subscription relationship	Service order information, service registration time, modification, cancellation status information, etc.
	C1-2:Service records and logs	Service details: such as voice, SMS, MMS and other telecommunications business service details, which may include calling number, calling location, called number, communication start time, duration, traffic information, etc; Internet or mobile Internet business usage details, such as cookie content, service access records, such as URLs, business logs, etc., and online shopping records, etc.
	C1-3:Spending information and bills	Power-on,power-off, network access time, online time, points, pre-deposit, credit rating, credit limit, payment situation, payment method, etc; Bills: such as fixed expenses paid, communication fee, data fee, collecting charge, balance, etc.
	C1-4:Location information	Latitude and longitude of location, area code, community code, base station number, etc of the user.
C2:User equipment information	C2-1:Equipment information	Hardware model, unique device identification code IMEI, device MAC address, SIM card IMSI information, etc.

3. PERSONAL INFORMATION SECURITY CRISIS IN THE ERA OF BIG DATA

3.1 Reasons for the personal information security crisis

3.1.1 Drive of Interests

Personal information is known as one of the most valuable resources in the 21st century, and it contains huge business value, intelligence value and public management value. It can not only provide the basis for government decision-making, generate efficiency and benefits in public management ,but can also be used as intelligence to predict consumer consumption habits, generating user profile which brings business profits.

The beginning of the recognition of the value of personal information is meanwhile the beginning of robbery and violation. The first malicious use of personal information in human history occurred in Nazi Germany. The Nazis in Germany used computers produced by IBM to automatically process and classify personal information about Jews from the census, especially sensitive information such as race and religion, in order to confine Jews to concentration camps accurately and rapidly. During the same period, the U.S. government illegally used household registration information to track and investigate Japanese-Americans , that is the reason personal information protection legislation first occurs in Europe and the U.S.

3.1.2 The absence of laws

Although many domestic existing laws include provisions about the protection of citizen's personal information, the domestic existing legal system for the protection of citizen's Personal Information is relatively fragmented and non-systematic. During the rapid development of the Internet, the cases of internet infringement of citizens`personal information emerge in an endless stream, while the existing laws are lagging which is not enough to meet the demands of society.

3.1.3 Improper management by personal information holders

The improper management of the internet personal information is in the most cases the leakage of citizen's personal information due to problems in protection measures

at the aspects of technology, awareness, and management taken by the personal information holders. After the personal information is collected, the administrator shall take reasonable measures to avoid the risk the personal information is abused, distorted or leaked. Otherwise, the leakage of personal information may not only involve business secrets of commercial subjects and infringe on the interests of the commercial subjects itself, but also cause inconvenience to every citizen, even cause citizens to suffer property damage and mental loss. At present, internet environment in China is still immature, the sense about effective management of personal information of personal information holders is relatively weak, which affects the reputation and credibility of information holders, there's a great possibility of causing greater economic losses to information owners.

3.2 Specific manifestations of the violation to personal information security

The value of personal information lies it will create greater value. Driven by economic interests, there are a large number of criminals, the possibility of personal information being violated is increasing, and the possibility of personal information being unreasonably used, collected, distorted, deleted, copied, stolen, spread multiplies is increasing.

3.2.1 Improper Collection of Personal Information

3.2.1.1 Failure to fulfill notification obligations and collect amounts of personal information

Many organizations dig and collect personal information for business purposes, using various forms of surveys to know and record users' personal information. Users are required to provide a large amount of personal information sufficient to identify them as a prerequisite for being provided services. However, these commercial organizations does not tell users items as what's the purpose of collecting these information and what measures were taken to ensure the user's information security .

3.2.1.2 Collection of personal information without users` permission

The internet can make the collection of personal information more hidden, and the citizens doesn't know, many users 'personal information is collected without personal consent, for example: collecting user's usage records and other content through IP

addresses; using Cookies software which has tracking function to measure and track users` actions on the website; using virus programs to infiltrate computer systems to steal users` personal information, and other acts of collecting personal information without users` consent.

3.2.2 Improper Dissemination of Personal Information

3.2.2.1 Improper Leakage

Due to reasons such as lack of relevant security measures, the user's personal information is leaked after government agencies and website service providers collects personal information, which is an infringement on personal rights. In addition, many websites leak obtained personal information to advertisers for economic benefit.

3.2.2.2 Malicious Spread

The website spreads personal privacy to increase the users` click-through rate to obtain higher economic benefits.

3.3 Improper use of Personal Information

The website conducts secondary development and utilization of personal information for business purposes. For example, search engine companies collect information about users` cookies and IP addresses, search terms entered, and search results that were clicked etc. The information is gathered together, then through analysis and digging the user's personality, interests, purchasing behavior, and online activities are obtained, they use these results for advertising or other commercial purposes.

4. CONCLUSION

Against this background of the era of big data, the situation that internet infringes personal information occurs frequently. Based on the domestic current legislation condition of personal information protection, and drawing on the foreign experience of protecting personal information, the author puts forward his personal opinions. In order to handle cases of infringement of personal information in the context of big data, the author believes that it is necessary to start with the nature of infringement of personal information, and to clarify the ways of Internet infringement and the corresponding measures. As China does not yet formulate perfect "Personal Information Protection Law", as to the purpose of this article, the author puts forward personal suggestions on the protection of personal information in order to contribute to the protection of personal information.

References

[1] Zhao Jin, Zhou Yang. May 2016. "Analysis on the Criminal Law Protection of Personal Information Security in the Big Data Era [J]" Technology and Market.

[2] Zhang Shiyun, Yu Hongen, Yin Weijie. May 2016. "Study on personal information infringement in the era of big data——thinking about the strategy of governing the country according to law [J]". Legal System and Economy.

[3] Zhang Li'an, Han Xuzhi. Mar 2016. "Private law attribute of personal information right in the era of big data [J] ". Law Forum.

[4] Yang Zhen, Xu Lei. Feb 2016. "Research on China's Personal Information Protection Legislation in the Big Data Era [J]". Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition).

[5] Sun Ping. Law. April 2016 . "Systematic construction of basic rights model for personal information protection legislation [J]".

[6] Qian Li, Ou Lijia. Jan.2016. "Conflict and balance between personal privacy and public right to know in the era of big data [J] ". Journal of Pingxiang University.

[7] Hou Zhiying. Dec.2015. "Personal information protection in the era of big data from the perspective of legislation [J] ". Productivity Research.

[8] Qi Aimin, Wang Jiyan. Nov. 2015. "The application of personal information protection law in the era of big data and its extraterritorial effect [J] ". Social Scientist.

[9] Wang Ran. May 2015. "E-forensics from the perspective of personal information protection in the era of big data: from the perspective of network platform [J] ". Journal of Shandong Police College.

[10] Zhou Dong. Apr. 2015 . "Personal information and privacy in the era of big data: a comparative study based on extraterritorial law [J] ". Graduate Law.

[11] Hou Fuqiang. Jul. 2015 . "Crisis and Governance of Personal Information on the Internet in the Big Data Era [J] ". Theoretical Perspective.

[12] Zhang Maoyue. Jul.2015. "New threats and protection of personal information data security in the era of big data [J] ". China Science and Technology Forum.

[13] Di Hanyu, Zheng Yunfan, Lu Lin. Dec. 2015. "Research on Personal Information Security Issues in New Media Dissemination in the Big Data Era——Taking "The First Case of Guangdong Human Flesh Search" as an Example [J] ". Western Radio and Television.

[14] Kuang Wenbo. Jun. 2015 . "Personal privacy protection in the era of big data [J] ". China Broadcasting.

[15] Zhang Maoyue. Jun. 2015. "Risks and Countermeasures of Citizen Personal Information Data in the Big Data Era [J] ". Information Theory and Practice.

[16] Yang Xiu. May 2015. "Personal Information Protection in Targeted Advertising in the Big Data Era: Analysis of 'China Internet Targeted Advertising User Information Protection Industry Framework Standard' [J] ". International Press.

[17] Qi Tong, Shi Xiaoyan. Mar. 2015 . "Personal Information 'Right to Be Forgotten'in the Era of Big Data: Comment on Gonzales v. Google [J]. ". Law of Finance and Economics.

[18] Mo Xiaochun. Mar. 2015. "Discussion on the protection of personal information of Chinese citizens in the era of big data [J] ". Innovation.

[19] Li Wenlong. Feb. 2015. "Privacy protection and right to be forgotten in the era of big data [J] ". Graduate Law.

[20] Wang Xuehui, Zhao Xin. May 2015. "Explore the Integration and Protection of the Public and Private Law of the Right to Privacy——From the Perspective of Personal Information Privacy in the 'Big Data Era' [J] ". Hebei Law Science.

[21] Ni Bin. Feb. 2015. " The protection of personal privacy by the judicial system in the context of the era of big data [J] ". Journal of Ezhou University.

[22] Yang Lixin, Han Yan. Feb.2015. " Chinese Localization of Right to Be Forgotten and Application of Laws [J] ". Law Application.

[23] Wang Yonghong. Feb.2015 . "Civil Law Protection of Citizens' Right to Privacy in the Age of Big Data [J] ". Legal System and Economy.

[24] Fan Xiaofeng, Wang Hao. Jun. 2014. "Legal protection of privacy of personal data in the era of big data [J] ". Journal of Hebei Radio and TV University.

[25] Liu Yahui, Zhang Tieying, Jin Xiaolong, Cheng Xueqi. Jan. 2015 . "Personal privacy protection in the era of big data [J] ". Computer Research and Development.

[26] Liu Yukun. Feb. 2014. "Civil law protection of personal information in the era of big data [J] ". Chinese Lawyer.

[27] Chen Xuan, Li Yan. Dec. 2013. "The definition of 'personal electronic information'in the era of big data: a comparative law-derived study of rights [J] ". International Press.

[28] Shi Weimin. Dec. 2013. "The Predicament and Path Choice of Personal Information Protection in the Big Data Era [J] ". Information Magazine.

[29] Xu Xuguang. May 2003 . " On Personal Data and Protection in the Network Age [J] ". Library and Information Service.