Scientific Research Publishing

Open Journal of Social Sciences, 2017, 5, 139-145
http://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Research on Personal Privacy Protection of China in the Era of Big Data

Hui Zhao, Haoxin Dong

Graduate School of Chinese Academy of Social Sciences (GSCASS), Beijing, China
Email: zhaohui_cool88@163.com

How to cite this paper: Zhao, H. and Dong, H.X. (2017) Research on Personal Privacy Protection of China in the Era of Big Data. *Open Journal of Social Sciences*, **5**, 139-145.
https://doi.org/10.4236/jss.2017.56012

Received: April 24, 2017
Accepted: June 16, 2017
Published: June 19, 2017

Abstract

The purpose of this essay is to investigate the privacy concerns of Chinese, and to develop relevant protective measures. The groups are divided into two parts by gender and six parts by ages to analyze the different gender and different age groups of privacy concerns. The significance of this study is protecting personal data property. The data of personal information after finishing processing have economic value. These data once disclosed, will be not reversible, so it is important to study the personal privacy in the era of big data and to initiate and enforce legal and regulatory protection measures. Results show that Chinese's privacy in public places for Internet records, friends dynamic and age's awareness is insufficient; most people especially female lack privacy protection skills. Educators need to improve the relevant laws and regulations, promote privacy protection skills and strengthen the conception of privacy.

Keywords

Personal Data, Big Data, Privacy, Survey

1. Introduction

The notion of privacy was first put forward in 1890, which included ideas, views and emotions (Warren & Brands, 1979) [1]. Privacy content changes with the times background and regional culture. Through the summary of more than 200 cases, it was pointed out that privacy mainly includes: portrait, name, private space and misunderstood information by others (Prosser, 1960) [2]. With the development of the Internet, especially the popularity of e-commerce and social networks, the content of privacy has become a hot topic discussed by experts and scholars. The EU has a relatively wide range of privacy content, such as personal age, salary, credit status, property status, physical status, employment

status, family status, hobbies, network comments and online shopping records (Schmidt, 2014) [3]. Privacy range is relatively narrow in US and Japan. It mainly includes the following factors: personal age, home address, property status and other content, such as online shopping records and online comments (Newell, B. C., 2013) [4]. According to Cavoukian, A. & Jonas, J., "the right of habitation, property status, social relations, living and reluctance to publicly available information belong to privacy" (2012) [5]. Juels claims that illegal access to personal information, illegal providing personal information to others and personal decision interference, all belong to the infringement of personal Privacy Act (2008) [6]. Obviously, personal information may not have research value. However, in the era of big data, the property attributes of personal data are becoming increasingly prominent. Much personal information such as family members, work units, marriage status, online shopping records, Internet records and so on in real life are not necessarily practical significance, but in large data background, it has economic value. For example, Decide is a company which forecasts commodity price and provides suggestions about purchase time for consumers (Lazar, N., 2015) [7]. Through the analysis of data in Amazon, it shows that there are one billion data integration in a page to facilitate consumer comparison, predict the price trend of products, and help the user to identify the best time to buy (Brunozzi, S., 2012) [8]. Through the association of data mining, it is easy to find a lot of privacy information in the surface of mass data, which brings a new problem to the information security and privacy protection. About 80% organizations have their own database, such as Baidu and Tencent, and these data are mostly in the form of digital, and it is easy to spread out. Once spread out, it is irreversible. Necessarily, it is significant to study privacy in the era of the big data. According to Madden, S., the existing literature about the China's personal privacy lacks of specialized research (2012) [9]. To this end, this essay is designed to determine the main content of personal privacy through questionnaires and put forward privacy protection measures.

It can be argued that in the field of business through the analysis of sales data, companies can understand customers' shopping habits and know which products are suitable for the sale of collocation in order to provide personalized services. However, it is my contention that in fact personal information is completely exposed and used by the third party without informing us, so as to determine our identity. In addition, When internet users become accustomed to the web and visit social networking sites, the users may not have noticed that personal activities is being monitored. According to Victor, Amazon surveils customers' shopping habits, Google watches people's web browsing habits and Facebook knows user's social relations network, which involve 1/5 population of the world (2013) [10]. Even though database passwords use the protection measures, the probability of hackers restoring the password is about 70% - 80% (Glasgow, K., 2015) [11]. I request that it is necessary to pay much attention to privacy in the era of big data.

2. Literature Review

The strengthening of legislation to regulate the use of personal information under the network environment is a general way, and explore some of the new data usage patterns, in order to ensure the legitimate rights and interests of individuals to dig up the premise of the use of information, innovative value-added services (Tene, 2012) [12]. Weber (2012) [13] believes that the relevant laws and regulations in terms of information protection play a crucial role, and legislation cannot be overanxious for quick results. From the perspective of access to the big data privacy protection, combing the current big data security and the relevant key technologies, Feng believes that big data has brought new security issues, but there is also an important way to solve the problem, through technical means, related policies and regulations to address the big data security and privacy issues (2014) [14]. According to professor Liu, "it is not enough only by laws and technologies to protect personal information, and it is necessary to strengthen the comprehensive of data literacy from the perspective of ethics" (2015) [15]. Peter Leonard claims that Data analysis has economic value, while our personal information may be collected and used by the third party without informing us, so as to determine our identity (2013) [16]. Finally, Leonard focuses on the use of anonymous and de-identified measures to ease the re-identification of risk and these measures are a combination of technology, business and contract security measures (2013) [16]. Smith claims that "when data about customers is collected and mined by companies such as Facebook, Google, mobile phone companies, retail chains and governments, it is a sociological problem to surveil the use of personal data" (2012) [17]. It is feasible to look at this issue from a new perspective, namely how can the user gain awareness of the personally relevant part Big Data that is publicly available in the social web. The amount of user-generated media uploaded to the web is expanding rapidly and it is beyond the capabilities of individuals to sift through them to see which media impacts our privacy. Based on the emerging trends of privacy implications and geo-tagged social media, users can stay informed about which parts of the social Big Data deluge is relevant to them (Smith, 2012) [17]. According to Wang Zhong, the privacy protection in big data environment not only related to ethics, law, industry, technology and other fields, but also involves a large number of individuals, groups, enterprises and institutions, and thus put forward to build the standard system, product information registration system, traceability system, traceability information reward system (2014) [18]. These systems are regarded as the main content of personal data privacy disclosure mechanism, to attract the relevant stakeholders to provide a solution for the rational use of personal data.

The most difficult part of this assignment will be the collection of foreign literature since there is little literature studying Chinese privacy trust. In addition, big data privacy issues is a new field and the current domestic research in this area is less, so finding sources concerning privacy will also be difficult.

3. Methodology

The essay adopts questionnaires to collect data and analyze people's attitude. In order to guarantee the wide range of the data, this survey chooses different education level, different sex, different ages and different occupations of people as the object of investigation. About 90% of the questions are single and closed response, so it is better to analyze the questionnaire and initiate a clear legal and regulatory protection measures. The survey adopts the Likert Scaling approach, which can be more clear understanding of consumer attitudes towards privacy. This questionnaire has been disseminated through online and offline methods. In order to carry out online survey, the questionnaire is hung in the famous professional survey website "questionnaire star". The investigation lasted for 15 days. A total of 300 copies have been sent and received a total of 209 questionnaires, of which 123 online questionnaires, 86 copies of offline, including 98 males and 111 females. Details are shown in Table 1.

Ranking from high to low according to the probability of individual privacy are: ID number (73.52%), telephone number (53.18%), IP address (45.26%), Internet records (37.69%), friends dynamic (33.45%), ages (28.12%) and full name (22.37%).

4. Results

The vast majority of citizens have realized that telephone numbers are privacy. Telephone numbers have become the most common way to get personal information, and customers are often subject to the merchant's telephone harassment. In addition, We Chat uses phone numbers to bind and log in. As long as users have received the verification code, it is possible to ignore previous password and log in directly, which leads to the huge risk.

Table 1. Descriptive statistics of the study sample.

	Features	Proportion		Features	Proportion
Gender	Male	49.12%	Occupation	Full time students	14.73%
	Female	50.88%		Professionals	6.29%
Age Group	Below 18	2.75%		Marketing	16.11%
	18 - 30	57.56%		Finance/audit staff	16.7%
	31 - 40	29.08%		Service personnel	7.07%
	41 - 50	8.06%		R & D personnel	7.86%
	51 - 60	2.16%		Teacher	22.2%
	60-	0.39%		Others	9.03%
Education Level	Junior high	3.14%			
	School	3.93%			
	High school	10.41%			
	Specialist	45.2%			
	Undergraduate	16.7%			
	Master degree	20.62%			

Data source: all the data comes from the questionnaire.

Friends dynamic and ages are not paid much attention. In fact, these two are very important personal information. In foreign countries, ages are the personal privacy. However, in China, ages are not regarded as personal privacy. The reason for this phenomenon may be due to the difference between Chinese and foreigners (Glasgow, K., 2015) [11]. Criminals through the acquisition of other people's dynamic, quickly understand the preferences, activities, and other information of users, and make criminal planning. In order to reduce the disclosure of the personal information, it is necessary to remind citizens to pay attention to the social networking site's friend dynamic and set the browse permissions.

Females believe that the proportion of personal information belonging to the privacy are higher than males, indicating that females are more sensitive to personal information, as shown in Table 2. But female's personal privacy leak ratio is higher than male, and the ratio of taking the initiative measures to protect the privacy of personal information is lower than male. Obviously, females are conscious of privacy but lack of prevention technology. Most of them don't know encryption technology, firewall technology, security authentication and digital hiding technology. when the personal information is infringed, the probability of female choice silently endured 31.66% is more than 22.80%, which shows that women are less flexible than men in taking specific measures, as shown in Table 3.

5. Discussion

This essay is designed to study the contents of personal privacy in the era of big data, and propose the strategy to protect the privacy of individuals, which requires large amount of data. The previous studies haven't use questionnaire to

Table 2. Cross analysis on the cognition of different gender for the content of privacy.

	Male	Female
ID number	47.51%	52.49%
Telephone number	44.88%	55.12%
IP address	45.00%	55.00%
Internet records	45.10%	54.90%
Friends dynamic	47.77%	52.23%
Ages	46.67%	53.33%
Full name	45.99%	54.01%

Data source: all the data comes from the questionnaire.

Table 3. Cross analysis of privacy protection measures in different gender.

	Silently Endure	Report	Lawsuit	Network posting	Don't know
Male	22.80%	17.20%	14.00%	6.34%	39.66%
Female	31.66%	10.97%	12.76%	1.54%	53.07%

Data source: all the data comes from the questionnaire.

study privacy, while this essay use questionnaire to study it. Because questionnaire is efficiency and availability, the essay adopts questionnaire to collect data and analyze people's attitude. In order to guarantee the wide range of the data, this survey chooses different education level, different sex, different ages and different occupations of people as the object of investigation.

6. Conclusion and Recommendation

The research indicates that data has economic value, while most of Chinese are not aware of the importance of personal information, and our information may be collected and used by the third party without informing us, so as to determine our identity. The fact that citizens are aware of the privacy of telephone numbers is due to the merchant's telephone harassment. It is also indicated that ages and friends dynamic also belong to the content of privacy, while most people don't pay much attention to them. On the issue of personal information, males are generally "careless" and "lazy", while females are more "careful" and "active", while female's technical ability is relatively poor. Both of them don't know how to protect their privacy, so it is necessary to learn from the overseas related legislation and combine with China's actual situation to accelerate the formulation of laws and regulations on the protection of the privacy of citizens and make the violation of privacy clear, as a result, let the citizens abide by the law. For example, in Britain, it is required that business or government should inform the subjects when personal information is collected and used by them; in Germany, if citizens receive spam messages, it is judged to be a violation of personal privacy (Wilson, S., 2014) [19]. In addition, it is necessary to strengthen the publicity and education of the concept of privacy. There are two ways to promote this concept: the first is through public service ads or lectures; the other is through school education. At last, privacy protection skills training should be promoted and privacy protection technology innovation should be supported. Through the national science and technology plan or industrial development fund, it is available to support enterprises or research institutions to innovate privacy protection technology.

References

[1] Warren, S.D. and Brandeis, L.D. (1973) The Right to Privacy. *Harvard Law Review*, **4**, 193-220.

[2] Prosser, W. (1960) Privacy. *California Law Review*, **48**, 383-423. https://doi.org/10.2307/3478805

[3] Schmidt, R., Möhring, M., Maier, S., Pietsch, J. and Härting, R.C. (2014) Big Data as Strategic Enabler—Insights from Central European Enterprises. *Lecture Notes in Business Information Processing*, **176**, 50-60. https://doi.org/10.1007/978-3-319-06695-0_5

[4] Newell, B.C. (2013) Local Law Enforcement Jumps on the Big Data Bandwagon: Automated License Plate Recognition Systems, Information Privacy, and Access to Government Information. Elsevier, Amsterdam.

[5] Cavoukian, A. and Jonas, J. (2012) Privacy by Design in the Age of Big Data. *Euro-*

control Int, 1-17.

[6] Juels, A. (2008) Targeted Delivery of Informational Content with Privacy Protection. US Patent No. 7472093 B2.

[7] Lazar, N. (2015) The Big Picture: Big Data and Privacy. *Chance*, **28**, 39-42. https://doi.org/10.1080/09332480.2015.1016848

[8] Brunozzi, S. (2012) Big Data and NoSQL with Amazon DynamoDB. *Proceedings of the* 2012 *Workshop on Management of Big Data Systems*, ACM, New York, 41-42. https://doi.org/10.1145/2378356.2378369

[9] Madden, S. (2012) From Databases to Big Data. *IEEE Internet Computing*, **16**, 4-6. https://doi.org/10.1109/MIC.2012.50

[10] Victor, J.M. (2013) The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy. *Social Science Electronic Publishing*, **123**, 513-528.

[11] Glasgow, K. (2015) Chapter 4—Big Data and Law Enforcement: Advances, Implications, and Lessons from an Active Shooter Case Study. Application of Big Data for National Security, 39-54.

[12] Tene, O. (2012) Privacy in the Age of Big Data: A Time for Big Decisions. Stanford Law Review Online, 64.

[13] Weber, S. (2012) Big Data Privacy and Security Challenges. In: *Proceedings of the* 2012 *ACM Workshop on Building Analysis Datasets and Gathering Experience Returns for Security*, ACM, New York, 1-2. https://doi.org/10.1145/2382416.2382418

[14] Feng, D.G., Zhang, M. and Hao, L.I. (2014) Big Data Security and Privacy Protection. *Chinese Journal of Computers*, 17-23.

[15] Liu, Y. (2015) Privacy Protection Method in the Era of Cloud Computing and Big Data. MATEC Web of Conferences. EDP Sciences, 162-168. https://doi.org/10.1051/matecconf/20152201041

[16] Leonard, P. (2013) Customer Data Analytics: Privacy Settings for "Big Data" Business. *International Data Privacy Law*, **4**, 53-68. https://doi.org/10.1093/idpl/ipt032

[17] Smith, M., Szongott, C., Henne, B. and Von Voigt, G. (2012) Big Data Privacy Issues in Public Social Media. Digital Ecosystems Technologies (DEST). 6*th IEEE International Conference on Digital Ecosystems Technologies*, Campione d'Italia, 18-20 June 2012, 1-6. https://doi.org/10.1109/DEST.2012.6227909

[18] Wang, Z. (2014) Research on Concern for Privacy of Personal Data in the Era of Big Data. Information Studies Theory & Application, 15-19.

[19] Wilson, S. (2014) The Collision between Big Data and Privacy Law. Social Science Electronic Publishing, 2. https://doi.org/10.7790/ajtde.v2n3.54