Who Moved My Data? Information Privacy Concerns in the Big Data Era

Lingjun Xie
School of media and communication, The University of Sydney, Australia.
lxie9393@uni.sydney.edu.au

Keywords: big data, information privacy, data breach, privacy protection, digital.

Abstract. Big data can bring public many benefits and opportunities. However, it also causes many critical issues. Privacy is one of the most important ethical issues in the information age, which has aroused public concerns. Users' personal data will be collected and analyzed through big data technology, but in the process of data monitor and information collection, lots of security risks may threaten public's data privacy. Some people may not fully aware that their information privacy is violated in many digital areas, so the essay will start from a case study of PRISM, which will help readers understand information privacy is a serious issue and the necessity of protecting public's data privacy in contemporary. The next part is explaining causes of data breaches and illustrating the existing challenges in the era of big data. Finally, from the perspective of technology, legislation, and individuals, the essay introduces methods of protecting data privacy in big data era.

1. Introduction

Big data is a term for "massive data sets having large [1], more varied and complex structure with the difficulties of storing, analyzing and visualizing for further processes or results" [2]. User's footprints on the internet, like browsing records, personal information, online shopping preferences, reviews on social media and other cookies form datasets, and these datasets will be integrated through the technology of big data. For example, the user's research and purchase records will be recorded when they are shopping online, and the shopping website will recommend related products according to user preferences. Some companies will conduct targeted advertising or precise marketing based on the user's search records. According to this, the wide application of big data can bring individuals benefits and organizations opportunities.

Big data will lead to critical problems while creating values, personal information privacy is a serious issue today. Personal information privacy refers to "the ability of the individual to personally control information about oneself" [3]. This information can be regarded as data property of people themselves, and they have right control their own information or data. Individuals have right to control data and decide whether their data can be available to others or not [4]. Individuals' will and right should be respected.

Big data is a technology while privacy is belonging to human value, and the relationship between big data and privacy is like the relationship between technology and human value. This is a paradox: big data technology analyzes massive data through collecting user information. However, for individuals, they have the demand of protecting personal information. Users are producers of these data and they have right to preserve and control data. Normally, users use the password to protect their account information, but their information is visible to web developers or data analysts. If the user data is a library, the big data technology is the key that can open the library, and data analysts are readers who can browse all the books in the library. However, user data is not always safe; their information privacy is often violated because of data breaches or companies using data without their permission. Invading personal information privacy is a common phenomenon at present. However, there are many existing challenges hampering the protection of public information privacy. How to protect information privacy becomes a critical issue in big data era.

2. A Case Study of PRIS

In the book 1984, the author George Orwell describes a society with ubiquitous government surveillance and people who live in a society without privacy. Reality has similarities with the book. Prism is a surveillance program promoted by United States National Security Agency (NSA). In June 2013, Snowden disclosed the secret documents of the PRISM to the Guardian and the Washington Post. PRISM is a top-secret surveillance program which allows official organizations to collect public's digital information on the internet, including users' search history, email, files, and chats [5]. It is found that PRISM is a data-collection program that promoted by the government of the U.S. and without individual warrants. The original intention of PRISM is monitoring online communication of terrorists, and ensures the safety of citizens, but now the program is far more invasive [6]. Today's PRISM monitors people's behaviors on the internet and causes psychological panic and people don't realize how their data will be used. Even more disturbing is that the wide range of data is provided by Internet giants, like Microsoft, Google, Yahoo!, Facebook, YouTube, Skype and Apple [7]. Users' data can be monitored and analyzed without informed, what's more, some internet companies even provide the data source for government, which is an obvious violation of users' information privacy. People use these websites in their daily life, but they may not consider that their data is being monitored when they are surfing the internet. 57% of Americans believe that the monitoring of public's information is unacceptable [8].

There is an emerging information privacy concern among the public after Snowden exposed the NSA's online surveillance program, but what makes public disappointed is that even they are aware of their data is monitoring by government, they cannot stop the NSA spying. In addition to the NSA, other organizations or companies may also monitor users' data all the time. If you use the Internet, search on Google, social with others on Facebook, chat through Skype, your data may be collected and personal information may be exposed to "invisible eyes". But it does not mean that the future of information privacy is pessimistic. Disseminating knowledge related to information privacy is necessary because it can arouse public's awareness of information privacy, help users better understand why data were leaked, why it is difficult to protect information privacy and methods of protecting their privacy. If more and more users are aware of the seriousness of the privacy issue, it is possible that more measures will be taken to protect information privacy in the big data era.

3. Cause and Challenges

Public needs to understand how their information privacy is violated. Information privacy will be violated when data breaches occur and data breach is the main reason for invading users' information privacy. Most breaches are about money and attackers are motivated by greed and miscellaneous errors are one of the top patterns of data breaches [9]. It is demonstrated that because of a technical glitch, Facebook has inadvertently exposed 6 million users' personal data, including phone numbers and email addresses to unauthorized viewers [10]. As one of the most widely used social networking websites in the world, there are hundreds of millions of users in Facebook. Users' information is accessible on the social networking site, their information privacy will be challenged if data breaches occur. The incident caused panic among users who worried that data breaches could have an impact on their privacy.

In addition, crime ware or hacking is also the main reason for data breaches. It is essential to maintain the privacy of confidential customer information such as personal information, credit card, and password [11]. In 2017, a giant data breach happens in Equifax, which owns the credit history data and personal information of 800 million people around the world, confirmed 143 million people's personal data has been hacked [12]. Another example is Yahoo, a world well-known portal web, has been attacked by hackers and more than 1bn user accounts were compromised, which illustrates that even large internet company can be attacked by hacks [13]. Hackers seek improper benefits by getting access to users' personal information, such as credit card number, password, transaction records, etc.

Furthermore, it is demonstrated that enterprises may share customers' information with other parties without their permission. Many retailers are using complex ways to track consumers online, and the giant retailer, Walmart is gathering big data or massive information of online customers, then analyze these data to shape business decisions [14]. Customers' data are important for decision makers. However, consumers produce data, including their footprints in the website, their location information and so on, these data are the property of consumers. However, customers do not understand how the retailer uses their online data. If the company use consumers' data for commercial purpose and share them with other parties without customers' permission, it will violate customers' information privacy.

Data can be attacked by hacks and shared by enterprises without permission. Public's sensitive information, property security may be threatened by hacks or sold by organizations in many areas. But there are many challenges hampering information protection and the difficulty of privacy management is increasing in recent years. "Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources", which means big data is dynamic and diverse. In technology dimensions, the process of analyzing big data can be complex, some unexpected errors may happen during data processing [15]. Even large internet companies, like Facebook, can't completely avoid miscellaneous errors. These errors during data processing will result in data breaches and increase the difficulty of information protection.

In legal dimensions, big data ratchet up the significance of data protection standards but the current data protection laws are not consistent or completed [16]. The definition of "invade individual privacy "is obscure, which makes it hard to determine whether the behavior constitute crimes. In addition, the scope of privacy right is also changing with the development of era, which makes it difficult to determine sensitive information and personal privacy. The inherent contradiction between enterprise service and user privacy is another challenge for privacy protection in big data era. Big data is an essential method to extract various internal and external data sources, and by researching these data, business strategies and marketing decisions can be made [17]. This means big data is the key to business development and enterprises need to collect users' data by applying big data technology. However, customers are not in the same position with companies, they are worrying that their privacy can be violated in the process of integrating data. More seriously, sometimes their data can be sold by websites to make profits.

4. How to Protect Information Privacy?

Although there are many existing challenges in protecting personal information privacy, public's information privacy can be preserved from different perspectives. Organizations and companies should be responsible for their customer's information privacy, governments should enact policy to protect personal information privacy. As the owner of data, individuals also need to concern data security and protect personal information in daily life.

Some internet companies are a victim of hack attack, to avoid malicious attacks, they need to build a strong firewall to detecting and thwarting cyber-attacks. The existing technical methods to protect data privacy are "passwords, controlled access, a two-factor authentication, which requires a user to submit two of three authentication factors before gaining access to a resource or service."[18] There are advanced technological solutions to protect data privacy, including cryptography, encryption, and virtual barriers and monitoring software. If these technical methods can be applied to practices, then some cyber-attacks can be monitored. One of the noticeable solutions is that controlling the possible data breaches by applying explicit authentication mechanisms [19]. Take social networking as an example, if authentication mechanism is applied in the area, the possibility of being attacked by malware can be reduced to a certain extent.

Governments have the responsibility to protect public's information privacy by enacting regulations. Related rules and regulations that protect privacy need to be updated with the development of big data. The European Union released General Data Protection Regulation to protect public's data privacy [20]. It is illegal to share personal data with other parties or use these data for a

purpose without users' consent [21]. However, there are no clear regulations that protect information privacy in many areas. If data can move to these unregulated areas, users' information will be invaded without restrictions. In some areas, many existing regulations cannot keep pace with era development [22]. As a matter of fact, rules that aimed at protecting personal privacy and responding to data crimes need to be promoted. There is no comprehensive data protection legislation and no clear definition of sensitive data [23]. One of the issues is defining the concept of information privacy protection in the era of big data and standardizing the process of personal data collection and sharing mechanism. Governments need to strengthen cooperation and negotiation with others to promote regulations.

Individuals also need to improve privacy awareness and take measures to protect their sensitive information. For example, when a user is asked to fill out some forms containing personal information on the website, they need to think twice before filling the form and do not submit sensitive information. Users create data via different methods, like websites, social networking, and many mobile apps. They share their personal information in social networking, use many apps to track their everyday life, and they are those who create or share data while organizations are collecting these data to gain insights [24]. According to this, if users want to protect their privacy, they should avoid disclosing sensitive data to a third party because the information may be lost or embezzled. And delete cookies on internet regularly is also important to protect their data privacy. Users also need to install antivirus software and firewall to strengthen the security of computer system, which can help to resist hack attacks. Backup important data and do not leak sensitive information are also needed.

5. Summary

Big data bring public and organizations many opportunities but also cause many issues on privacy protection. For individuals, they are in the demand of preserving their personal information and sensitive data, but big data is a technology that collects, manage and analyze these data. In the development process of big data, privacy breaches are happening frequently in the digital field, which makes public increasingly concern privacy protection issue. Data breaches happen in many areas, including e-commerce, social networking, finance, business, almost every aspect of personal life. Currently, there are many challenges, which involves technical challenges, legal challenges, social and ethical challenges, these challenges make privacy protection more difficult in big data era. Governments and organizations urgently need to develop regulations to standardize the industry, and against data breaches. Individuals should take their responsibility to protect their sensitive information. The digital world is evolving while big data is developing, we believe that big data can improve the well-being of the public, and privacy can be well protected in the future.

References

[1]. Smith H J, Milberg S J, Burke S J. Information privacy: measuring individuals' concerns about organizational practices [J]. MIS quarterly, 1996: 167-196.

[2]. Sagiroglu S, Sinn D. Big data: A review[C]//Collaboration Technologies and Systems (CTS), 2013 International Conference on. IEEE, 2013: 42-47.

[3]. Stone E F, Guitar H G, Gardner D G, et al. A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations [J]. Journal of applied psychology, 1983, 68(3): 459.

[4]. Chen K, Rea Jar and I. Protecting personal information online: A survey of user privacy concerns and control techniques [J]. Journal of Computer Information Systems, 2004, 44(4): 85-92.

[5]. Greenwald G, MacAskill E. NSA Prism program taps in to user data of Apple, Google and others [J]. The Guardian, 2013, 7(6): 1-43.

[6]. Stoycheff E. Under surveillance: Examining Facebook's spiral of silence effects in the wake of NSA Internet monitoring [J]. Journalism & Mass Communication Quarterly, 2016, 93(2): 296-311.

[7]. Sports P O L. NSA slides explain the PRISM data-collection program [J]. 2013.

[8]. Madden M, Rainie L. Americans' attitudes about privacy, security and surveillance [M]. Pew Research Center, 2015.

[9]. Solutions V E. Data Breach Investigations Report [J]. Verizon, Report, 2016.

[10]. Shih G. Facebook admits year-long data breach exposed 6 million users [J]. Reuters (June 21). http://www. Reuters. Com/article/2013/06/21/uk-facebook-security-idUKBRE95K19120130621, 2013.

[11]. Webster M. Data protection in the financial services industry [M]. Rutledge, 2017...

[12]. Smiley, S. Equifax: Australians' sensitive financial information at risk in data breach of US company [J]. Retrieved from http://www.abc.net.au/news/2017-09-08/smiley-credit-check-australians-financial-information-at-risk/8887198,2017.

[13]. Thiemann S. Yahoo hack: 1bn accounts compromised by biggest data breach in history [J]. The Guardian, 2016, 15.

[14]. Davey I. Technologies," Consumers, Big Data, and Online Tracking in the Retail Industry: A CASE STUDY OF WALMART," 10 August 2014[J].

[15]. Wu X, Zhu X, Wu G Q, et al. Data mining with big data [J]. IEEE transactions on knowledge and data engineering, 2014, 26(1): 97-107...

[16]. Konner C, Cate F H, Millard C, et al. The challenge of 'big data 'for data protection [J]. 2012.

[17]. From H S. Big Data: Opportunities and privacy challenges [J]. Arrive preprint arXiv: 1502.00823, 2015.

[18]. Schmitt C, Shaffer M, Owen P, et al. Security and privacy in the era of big data[J]. White Paper. ARENCI/National Consortium for Data Science. ARENCI White Paper Series, 2013.

[19]. Cuzzocrea A. Privacy and security of big data: current challenges and future research perspectives[C]//Proceedings of the First International Workshop on Privacy and Security of Big Data. ACM, 2014: 45-47.

[20]. Victor J M. The EU general data protection regulation: Toward a property regime for protecting data privacy [J]. Yale LJ, 2013, 123: 513.

[21]. Littman J. Information privacy/information property [J]. Stanford Law Review, 2000: 1283-1313.

[22]. Yu W E, CISM C. Data Privacy and Big Data—Compliance Issues and Considerations [J]. 2014.

[23]. King N J, Raja V T. Protecting the privacy and security of sensitive customer data in the cloud [J]. Computer Law & Security Review, 2012, 28(3): 308-319.

[24]. Pereira C, Raman R, Wang L, et al. Big data privacy in the internet of things era [J]. IT Professional, 2015, 17(3): 32-39.