# Privacy Security and Protection Strategies in the Big Data Era

**Menglu Zeng[1, 2, a], Meng Wang[3], Mengjie Zeng[4, b, *]**

[1]Henan University of Animal Husbandry and Economy Zhengzhou, China

[2]School of Physics & Information Technology Shaanxi Normal University Xi'an, China

[3]College of Intelligent Manufacturing and Automation Henan University of Animal Husbandry and Economy Zhengzhou, China

[4]College of Agricultural and Forestry Economics and Management Henan University of Animal Husbandry and Economy Zhengzhou, China

[a]mlmk1039@163.com, [b]Zengmengjie008@163.com

*Corresponding author

**Abstract:** Internet has become a part of people's lives. While people browse information online, they also leave many data traces. The collection of data traces mentioned above is the big data described in this paper. Big data, as the hot spot of social tracking, brings convenience to people's lives, but also brings great challenges. Privacy disclosure is one of them. Starting from the connotation of big data and privacy, and on the basis of analyzing the current situation of big data privacy protection, this paper puts forward some strategies of big data privacy protection from the aspects of individuals, enterprises and governments and so on.

## 1. Introduction

Internet has become a necessary part of people's life. While people browse information on the Internet, they also leave many footprints of access. In the era of communication, people's daily behavior is constantly monitored by "the third eye", such as Taobao, Jingdong and so on record our shopping habits; Baidu, Google and so on monitor our browsing records; QQ, Weibo and so on eavesdrop on our social network. People's daily activities are also monitored, such as smart phones to monitor our real-time location; work units, residential areas to monitor our access behavior, we are constantly exposing ours ID card numbers, bank accounts, social accounts and so on in various ways, even travel habits, shopping habits, living habits, family income and so on are unavoidable. Privacy disclosures become difficult to defend against in big data environments.

The essence of data is production data and assets. An important function of big data is to analyze and predict users' behavior. For example, through analysis of shopping records, e-commerce may know more about wives' preferences than husbands; through leaked medical information, users' privacy disorders can be known; through analysis of social networks, users' interests and hobbies can be captured; and even through information mining, users' political tendencies are likely to be disclosed. The threat of privacy disclosure is not only confined to the exposure of personal information, but also lies in the consequences of such information disclosure, which may threaten personal and property security, social stability and even national security.

Big data is like a double-edged sword. While opening up a new era of human society, it also brings great challenges to personal privacy. How to enjoy the convenience of big data while protecting personal privacy and preventing sensitive information disclosure has become a focus problem [1-2].

## 2. The Overview of Big Data

### 2.1 What is Big Data?

Big data, in theory, refers to data sets that cannot be captured, managed and processed by conventional software tools in a certain time range. It is a huge, high growth rate and diversified information asset [3]. Popularly speaking, many of our daily activities are transformed into invisible data traces. With the accumulation of time, these traces become more and more huge, so that traditional technology can not be timely analyzed and processed, thus forming so-called "big data".

### 2.2 The Characteristics of Big Data

At present, a more unified understanding is that big data has so-called "four V" characteristics [4], as shown in Fig. 1 below.
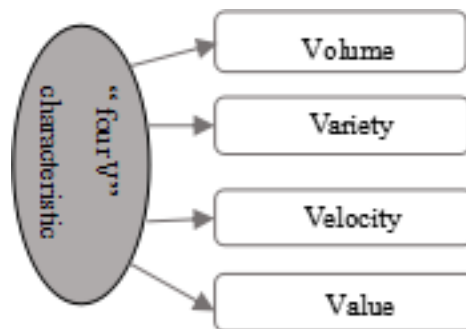


Figure 1. The Characteristics of Big Data.

Volume refers to the large scale of data. The volume of big data is amazing and continues to grow at an explosive rate. It is predicted that by 2025, the amount of data on the Internet will exceed the total brain capacity of all human beings.

Variety refers to the variety of data. The sources of data are wide, the types of data are various and the correlation between data is strong.

Velocity refers to the speed at which data is generated. The generation of big data is very fast, and the data generation, acquisition, processing and flow rate increase exponentially.

Value refers to the low value density of data, that is, the low gold content of data, which is a typical feature of large data. Valuable data accounts for a small proportion in big data world. Data mining is similar to gold panning in the sand, mining sparse but precious information from massive data, which is also the core feature of big data.

### 2.3 The Advantages of Big Data

Personal information in big data era has distinct value and can bring considerable economic benefits. Many countries pay more and more attention to big data and regard it as the golden resource of the new era, and gradually elevate it to the height of national strategy. The U.S. government regards big data as "oil of the future". In 2012, the U.S government launched the Program of Big Data Research and Development, which aims to enhance the ability of the United States to extract truth from massive data and help strengthen U.S homeland security; in the same year, the British government announced the establishment of the world's first open data institute, which aims to encourage people to seek innovative opportunities from open big data; the United Nations has also launched "the Global Pulse project", hoping to use "big data" to predict unemployment or disease outbreaks in some areas in order to provide timely assistance. Big data itself is not valuable, what we excavate is the development trend behind big data, which is reshaping our daily behavior and thinking mode. At present, the application of big data analysis in science, medicine, commerce and other fields is of great value [5], which can be summarized as follows:

*1) Acquiring knowledge and predicting trends:* Big data analysis can effectively get knowledge through phenomena. Based on the acquired knowledge, people can predict the future development

trend of real things more accurately. For example, people can predict the stock market according to Twitter information [6].

*2) Analyzing and Mastering Personalized Data Characteristics:* Individuals have distinct personality characteristics while satisfying some group characteristics. By analyzing the accumulated long-term and multi-level big data, enterprises can capture users' behavior rules, acquire users' behavior characteristics, and provide users with more precise personalized products and services. For example, Google uses big data to analyze users' habits and preferences, and then helps advertisers to deliver advertisements efficiently and accurately in order to maximize their benefits [7].

*3) Identify the truth through analysis:* Because the dissemination of information in the network is more convenient, the harm caused by false information in the network is also greater. At present, people begin to try to use big data to filter false information. For example, Yelp, the largest social comment website in the United States, filters false comments by using big data analysis to provide more authentic comments for users [8]. Yahoo and Thinkmail use big data analysis to filter spam e-mails such as advertisements and sales promotion.

## 3. The Overview of Personal Privacy

Privacy refers to the secret that citizens are unwilling to disclose or know for others. Popularly speaking, privacy refers to the unwillingness to tell other people's personal information. In the era of big data, our privacy space has been compressed, and it's hard for us to have so-called secrets anymore.

Before the advent of the digital age, people's activities were hard to record and share. Today, most behaviors can be captured digitally and used to predict the next trend of people's behavior. With the continuous upgrading of monitoring technology, image processing technology and positioning technology, complete personal privacy is almost impossible for ordinary people. For example, when we use location navigation, our behavior privacy has been broken; when we browse the web, our information privacy is also being violated. Applications that we take for granted in our lives are actually exposing our personal privacy from time to time.

## 4. The Current Situation of Privacy Protection in Big Data Era

### 4.1 Most netizens do not pay attention to privacy protection, and their awareness of privacy protection is weak

The Internet has broken the privacy rules of the traditional era, and "public privacy" has emerged [9]. The increasing transparency of information also facilitates the disclosure of privacy. Individual privacy protection awareness is the main reason for personal information disclosure. There are many ways to disclose personal privacy: logistics sheets on express delivery, real-name train tickets, social networking software such as Wechat, QQ on the website, consumption records of Internet users and other information [10]. The random abandonment of these objects and the wanton theft of data are the main reasons for the disclosure of personal information.

### 4.2 Public regulatory capacity is lacking

In the era of big data, the amount of information contained in citizens' network communication has increased dramatically, and the information about personal background, characteristics and behavior habits has become more and more specific. If this information is abused without supervision or with weak supervision, it will cause serious harm to personal life and work.

### 4.3 Policy legislation is not clear

Because privacy protection law of our country is not perfect enough, and its provisions are simple and scattered, only limited protection of data privacy, which makes our enterprises in the European Union, North America and other regions have been in the "unfair competition" position, the

economic interests have been seriously damaged, so our country urgently needs to introduce a sound privacy law with clear rewards and penalties.

## 4.4 Privacy protection technology is immature

Privacy protection technology is relatively weak. In big data era, the development speed of personal privacy protection technology and big data is not synchronous and coordinated. Data processing technology has not yet fully provided a solid technical support for personal information protection, there is still much space for improvement.

## 4.5 The social responsibility of enterprise is not strong.

Most enterprises take profit maximization as the principle, and constantly collect and utilize public privacy to seize benefits. In addition to Internet companies, mobile companies also have a large amount of data about users' real-time location, calls and Internet access. Although it is stipulated that enterprises are not allowed to push marketing advertisements to users, some companies are still trying to bypass "social responsibility" in a new way, they use a large amount of data to analyze the potential needs of users, and then push advertisements in a relatively vague group. This practice only delays the time of advertising push, but also "intentionally" expands the scope of users, whose behavior is more serious, which is closely related to the lack of enterprise social responsibility.


## 5. The Privacy Protection Strategies in Big Data Era

Personal information is regarded as the "new oil" in the contemporary economy. In the age when Internet technology has not been popularized, the protection methods of personal information are relatively simple. As long as the relevant laws and regulations are set up and the necessary physical and personnel control is carried out, information leakage can be prevented to a great extent. However, with the development of information technology and driven by commercial interests, the situation of personal information security is worrying. Faced with this situation, we should strengthen the protection of privacy information from all levels in order to promote personal development, maintain social stability and consolidate national security. Here are some strategies, as follows:

## 5.1 Strengthening Quality Education and Enhancing Personal Data Privacy Protection Consciousness

First of all, we should enhance the awareness of personal privacy protection and not leave personal information at will; secondly, we should be strict with ourselves and not violate the law; finally, when personal privacy is infringed, we can use legal weapons to protect ourselves.

In short, in order to protect personal privacy, the first task is to improve personal privacy protection awareness.

## 5.2 Breaking through the Core Technology of Data Security

It is very necessary to use technical means to protect our privacy data. Privacy data is protected by encryption devices and algorithms, so that even if the data is lost, the original data can not be obtained by the theft. But at present, the most effective protection measures still adopt the data encryption method combining software and hardware.

The technical framework of large data processing is shown in Fig. 2. We should start with data acquisition, data analysis and data presentation, break through key technologies, develop more efficient and optimized processing mode, and protect our privacy data from all aspects.

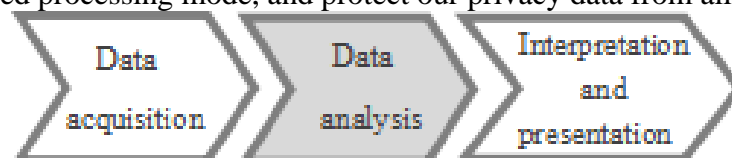Data acquisition → Data analysis → Interpretation and presentation

Figure 2. The Framework of Big Data Technology.

In data collection, analysis and presentation, a solid fortress is established, which does not give illegal elements the opportunity to protect personal privacy from the technical level, so as to ensure that they can get more secure security protection.

## 5.3 Strengthening public supervision

In recent years, privacy leaks have emerged in endlessly. For example, HSBC lost 370,000 users' insurance information; Aliyun suffered the largest DDoS attack in the world, an Asian network operator's data center suffered 334,000 BPS spam data flow attack in early 2015; Sony suffered ATP attack in 2014, and a large number of employees' information was leaked. Although our country has formulated supervision standards for personal information protection, which only play a guiding role, lacking necessary mandatory. Public supervision is too formal, some supervision departments are good at post-processing, lack of necessary pre-defensive measures. Based on this, our country should fully consider the national conditions, drawing lessons from foreign experience, setting up administrative supervision institutions, formulating clear standards and scope of personal information security supervision, and further protecting the security of data privacy from the perspective of public power supervision.

## 5.4 Perfecting legislation

Since the requirements of data privacy in all walks of life may have their own industry characteristics, before formulating the protection law, we can first formulate the principles of privacy protection in different walks of life, and establish the protection law covering different walks of life as far as possible. From a broader perspective, we can establish a general privacy protection law among countries, and improve the general type of privacy protection law, so as to make it more effective in solving the problem of personal privacy security.

## 5.5 Enhancing the sense of social responsibility of enterprises and cultivating the self-discipline ability of industries

Customers are the source of enterprise's interests. While abiding by relevant laws, enterprises should also bear in mind their social responsibility and cultivate their self-discipline ability in order to attract more potential customers. At present, our information industry self-discipline mechanism is still in the initial stage of development, there are many defects in the protection of personal information security. Firstly, we should learn from foreign mature self-regulation mechanism and establish industry self-regulation mechanism suitable for China's national conditions; Secondly, a special third-party certification and rating agency should be set up to guide and encourage enterprises to conduct self-regulation and self-restraint; Finally, the management and punishment of enterprises should be strengthened to promote the healthy and harmonious development of the information society.

## 6. Conclusion

In the era of big data, the protection of privacy information has a unique significance. The traditional theory and technology of privacy protection can not keep pace with the development of big data. It is necessary to rethink and locate the problem of privacy protection of big data. This paper puts forward some countermeasures for the protection of personal privacy in the era of big data from different perspectives, such as individuals, enterprises, governments and society. In conclusion, it is the law first, the technology guarantee, the standard leading, the strengthening of supervision and personal participation.

In a word, it is still an arduous task to construct a set of privacy protection strategies suitable for China's national conditions, which requires the joint efforts of the whole society.

**References**

[1] Mengxiao, Feng. Big data management: Concepts, techniques and challenges [J], Journal of Computer Research and Development, vol.1, pp. 146 - 169, 2013.

[2] Viktor Mayer. Big Data: A Revolution that Will Transform How We live, Work and Think[J], Boston, Hought on Mifflin Harcourt, 2013.

[3] http://en.wikipedia.org/wiki/Big_data.

[4] Liguo, Jie. Research status and scientific thingking of big data[J], Bulletin of Chinese Academy of Science, vol. 6, pp. 647 - 657, 2012.

[5] Fengdeng, Guo. Big Data Security and Privacy Protection [J]. CHINESE JOURNAL OF COMPUTERS, vol. 37, pp. 246 - 258, 2014.

[6] Goels, Hofman JM.Predicting consumer behavior with Web search [J]. National Academy of Sciences, vol. 41, pp. 17486 - 17490, 2010.

[7] http://server.yesky.com/datacenter/172/34705172.shtml.

[8] https://adage.com/article/digital/fake-reviews-rise-yelp-crack-fraudsters/237486/.

[9] Yuan Ma. Research on Personal Information Protection in Big Data Age [J], COMPUTER ENGINEERING & SOFTWARE, vol. 12, pp. 249 - 253, 2017.

[10] Zhi Yao. Research on Personal Privacy Security Protection in the Era of Big Data [J]. Innovation Science and Technology, vol. 18, pp. 75 - 78, 2018