IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Research On Personal Information Security On Social Network In Big Data Era

Ms. Ritu Tailor
Assistant Professor
Geetanjali Institute of technical Studies,
Dabok, Udaipur (Rajasthan)

Dr. Priyanka Sisodia
Associate Professor
Geetanjali Institute of technical Studies,
Dabok, Udaipur (Rajasthan)

Abstract

The soaring development of network information technology and mobile terminals has changed our work and life dramatically. An increasing number of people like to post their activity information on the internet. At the meantime, the upcoming of social network and social big data science in big data era makes personal information security issues more obvious. The paper bases on big data background, researches the current situation of privacy protection on social network and analyzes the reasons of privacy infringement on social network, and then proposes corresponding countermeasures from the perspectives of users information security literacy, In big data science the process of current situation of security on sites for analysis data security protection technology as well as laws and regulations, looking forward to guaranteeing personal information security on social network.

Keywords-Big data, Social network, Personal information security.

INTRODUCTION- **Today**, the development of informatization and networking lead to explosive growth of data. According to statistics, 2 million users are using Google's search engine in every second, Face book users share 5 billion resources every day, Twitter process 340 million tweets every day, At the same time, the large amount of data are produced continuously in scientific calculation, medical services, finance, retailing. 8ZB data will be generated in 2015.

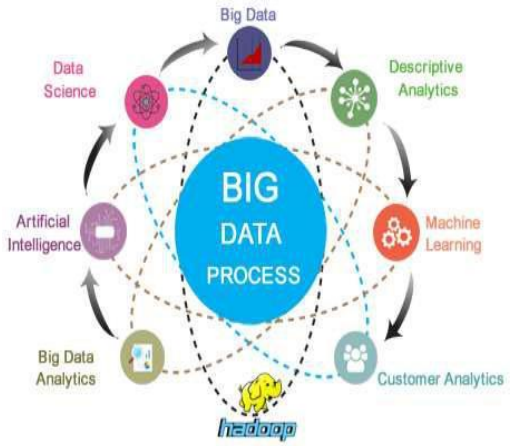
Social network comes around us with lightening speed. We would like to share our dynamic conditions on the network, interact with friends and thumb up for them. However, each operation of us on computer and mobile phone will be recorded by the server. Our life track even position coordinates can be detected from the information. If someone

spitefully collect or abuse information data on social network, huge hidden danger will be brought for our life. Personal information security even personal safety will be threatened. Therefore, it is a problem in urgent need of solution in big data era about how to guarantee personal information security. At present, the development of big data still faces many problems, Security and privacy issues is one of the key issues that people recognized widely, currently, people's every word and action on the internet are recorded by businesses, including shopping habits, friends contact situation, reading habits, searching habits, etc.Number of cases show that even after a large number of harmless data is collected, personal privacy will be exposed. In fact, the security implications of big data are more widely, the threat people faced, is not limited to leak of personal privacy, like other information, big data is facing many security risks during storage, processing, transmission, etc. and it needs data security and privacy protection. But data security and privacy protection in big data era is more difficult than in the past (such as data security in cloud computing, etc.). In the cloud computing, the service providers can control the storage and operation of data.

1. BIG DATA ERA AND SOCIAL NETWORK-

A. Big Data Era

Big data refer to the collection of data set in a certain period. The data are huge and complicated, so it is very hard to dispose by using database management tools grasped at present or traditional application program for data process. The challenges include collection, preservation, storage, search, sharing, transfer and analysis as well as visualization. Generally speaking, 5V are used to define and embody characteristics of big data, namely, Velocity, Volume, Value, Variety, and Veracity.



Velocity- Obviously, velocity refers to the speed at which vast amounts of data are being generated, collected and analyzed. Every day the number of emails, twitter messages, photos, video clips, etc. increases at lighting speeds around the world.

Volume- Volume refers to the incredible amounts of data generated each second from social media, cell phones, cars, credit cards, M2M sensors, photographs, video, etc. The vast amounts of data have become so large in fact that we can no longer store and analyze data using traditional database technology.

Value-When we talk about value, we're referring to the worth of the data being extracted. Having endless amounts of data is one thing, but unless it can be turned into value it is useless.

Variety-

Variety is defined as the different types of data we can now use. Data today looks very different than data from the past. We no longer just have structured data (name, phone number, address, financials, etc) that fits nice and neatly into a data table.

Veracity-

Veracity is the quality or trustworthiness of the data. Just how accurate is all this data? For example, think about all the Twitter posts with hash tags, abbreviations, typos, etc., and the reliability and accuracy of all that content.

B. Social network in big era data

Recent social networking websites such as Twitter, Facebook, LinkedIn, YouTube, and Wikipedia have not only connected large user populations but have also captured exabytes of information associated with their daily interactions. Social networking has its beginnings in the work of social scientists in the context of human social networks, mathematicians and physicists in the context of complex network theory, and, most recently, computer scientists in the examination of information or Internet-enabled social networks. In recent years, "cyber manhunt" can be heard without end. Every time when big events or events attracting high public attention happen, all information of parties involved is made public very fast. It owes to the big data.

2. Why does Big Data Threat Personal Privacy

3.1 Connectivity of Social Network

Data has become available for not only legitimate uses but also for abuses. Big data has the ability to change our lives. This wealth of big data can allow social scientists to study social interactions on a scale and at a level of detail that has never before been possible. Our goal is to evaluate the value of big data in various social.

There is a possibility of malicious use, there are security and privacy threats to the big data that you must be concerned about especially if you are the who spends more time on the internet.

In our social activities, there is often the case : Social networking sites recommend some people you may know to you. Why is there such a situation? As our society has connectivity. Because many social interactions currently take place in online networks, social scientists have access to unprecedented amounts of information about social interaction. Computer has a massive user information by analyzing any common social networks of two users, or by reading the phone contacts to determine whether acquaintance between two users.

3.2 Size of the Big Data Problems

There are sheer scales of people who are involved in big data security incidents, the stakes have grown in number. The professional development system at Arkansas University got breached in 2014 and 50,000 people were affected. This is a huge amount of people however if in comparison to 145 million people whose birth dates, home, and email addresses and other data were stolen in a data breach at eBay in the same year.

3.3 Need for Public Power Connectivity

In order to meet the needs of law enforcement, many countries in the world usually require network or telecom operators to store certain user datain a certain period of time, and provide the raw data and the results when the government need. This requirement is certainly legitimate, and does not pose a great threat to personal privacy in the era of the small data. However, in the era of big data, information communication capacity of the network increases rapidly, the data can reflect the personal background, characteristics, habits, behavior, becomes more and more specific, once this information is abused by public authority in the absence of supervision, it does exist the possibility that personal information has security risk.

4. Data resource disclosure, sharing and privacy protection are contradictory

In the age of big data, information disclosure and sharing are beneficial to business development and government administration. Because personal information can bring commercial value, it stimulates illegal merchants and personnel, illegally acquires and USES personal information to gain profit for business opportunities. In addition, some governments have demanded access to personal privacy in order to safeguard national 620 cyber security and security, thus making national security and personal privacy at odds.

5. Current situation of privacy protection on social network

A. Personal Information Leakage

With the emergence of new Internet technologies, due to the users in the use of the registered account in the process of personal information is recorded, so the user experience and enjoy the convenient way at the same time, also feel system all the time in the "monitoring" of his own actions. According to China Internet network information center, the number of Internet users in China reached 731 million by the end of 2016, up by 0.43 billion from 688 million in 2015. Along with the further promotion and application of the Internet, personal information will be recorded and stored by the various systems and platforms, which to a certain extent, improve the chances of the personal information was leaked, personal information security has become an important social problem. We also begin to expose more information on the network. When receiving fraud short messages or phones, we forget where our information is disclosed. However, most of us have “been accustomed to it” in new environment. The registration needs personal information to verify. The security of e-commerce payment is worrying, but we cannot help shopping online. The information sent out cannot be taken back. It is also difficult to stop.

B. Self-defect of database

Database operates on the basis of operation system, so database security depends on that of the operation system. The most common is to inject SQL by using the bug of application system and database. The whole table or huge amount of data can be obtained directly after successful injection of SQL.

C. Data access methods

There are several ways for third parties to access user information. Flickr is an example of a social media website that provides geo tagged photos that allows users to view the exact location of where a person is visiting or staying. Geo tagged photos make it easy for third party users to see where an individual is located or traveling to. There is also growing use of phishing, which reveals sensitive information through secretive links and downloads through email, messages, and other communications. Social media has opened up an entirely new realm for hackers to get information from normal posts and messages.

D.*Share it with third parties*

Nearly all of the most popular applications on Facebook—including Farmville, Causes, and Quiz Planet—have been sharing users' information with advertising and tracking companies. Even though Facebook's privacy policy says they can provide "any of the non-personally identifiable attributes we have collected" to advertisers, they violate this policy. If a user clicked a specific ad in a page, Facebook will send the address of this page to advertisers, which will directly lead to a profile page. In this case, it is easy to identify users' names.

6. The Main Principles of Privacy Protection

In era of big data, the focus on privacy issues shifted to the users .Only to regulate users ’behavior , their actions are consistent with the professional norms of the big data industry practitioners ,the protection of personal privacy is possible.

6.1 The Principle of theCertain UsingScope of Data –

The goal of handling of personal information must be specific, clear, reasonable, does not expand Gang ZengInt. Journal of Engineering Research and This principle is more difficult to do, but we can use "negative list", we stipulate what kind of behavior is not allowed, at the time of collection and use of data, Because these action encroach on personal privacy, as long as not to touch the place, other behavior of data using are acceptable.

6.2 MAIN PRINCIPLES OF PERSONAL DATA PROCESSING

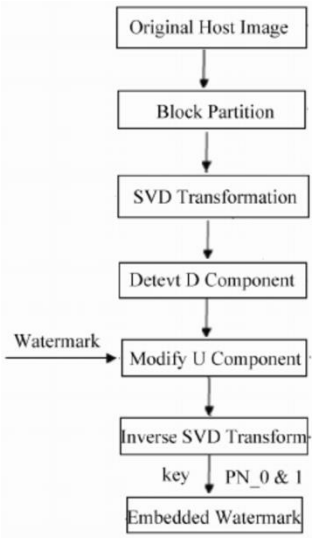
The term "personal data" determines the information that permits to identify a person directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. Any operation or set of operations with personal data (using automatic or not-automatic means) is called "processing of personal data". The main principles of personal data processing require strong rules for personal data protection (PDP).

6.3 The principle of individual participation Individuals have the right to decide whether their data is collected, knowing what data is collected, to confirm the data can be collected, modifiedanddeleted.Personal information is divided into two types: general information and sensitive information. Sensitive personal information may include ID number, phone number, race, political views, religious beliefs, genes, fingerprints, etc.By default, general information can be collected, but before the collection and use of sensitive personal information,firstly,the user must obtain the consent.

7. Key Technologies of Privacy Protection

7.1 Digital watermarking

The identification information is embedded into the data carrier through watermark and other technology, so as to protect the private data without affecting the use of the data carrier. In addition, fingerprinting is also embedded in the digital Watermarking, which enables quick identification of information owners and helps to track the burglars. Embedding fragile watermarks into the database also allows for timely detecting modification to data (Jia and Guo, 2015). The text watermark is generated using multi methods. The text watermark mainly includes the document structure fine-tuning watermark, natural language watermark and text content watermark. Some watermark may verify the data. For these characteristics, the digital Watermarking technology has great application value and development potential



7.2 AnonymityData Protection Technology

In the big data environment, anonymity protection is necessary to protect the data. For example, in social networks, anonymity protection can be divided into user identity anonymity, attributes anonymity and relationshipanonymity(known as edgeanonymity).Theinformation of user identification and user attributemust be hidden when published, the relationshipanonymity is to hide the relationship between users when data is released. At present, the relationship anonymity is a hotspot of research, many scholars have studied multiple methods for the relationship anonymity.Through other public information,an attacker may beinfer anonymous users, especially relationship between the users.

7.3 Anonymous protection of data publishing

Revista de la Facultad de Ingeniería U.C.V., Vol. 32, N°14, pp. 790-794, 2017 793 In the relational data, data publishing anonymity technology is a key means to protect the privacy data. During data publishing, repeated publications often appears. With data publishing anonymity technology, lawless analysis of repeated publishing of privacy data can be avoided, so as to prevent data anonymity being compromised (Wei, 2015). Criminals have multiple channels for obtaining data. Data publishing anonymous protection needs further study.

7.4 Data Provence Technology

Due to the diversification of data sources,it is necessary to record the origin and the process of dissemination, to provide additional support for the latter mining and decision. Before the emergence of the concept of big data, Data provence technology has been widely studied in database fields.Its purpose is to help people determine the source of the data in the data warehouse. The method of data provenceis labeled method， through the label, we can know which data in the table is the source, and can easily checking the correctness of the result, or updatethe datawith a minimum price. In the future data provence technology will play an important role in the field of information security. But Data provence technology for big data security and privacy protection also need to solve the following two questions: 1, The balance between privacy protection and data provence;2,to protect the security of data provence technology itself.

7.5 Access Control Technology

7.5.1Role Mining

Role-based access control (RBAC) is an access control model used widely. By assigning roles to users, roles related to permissions set, to achieve user authorization, to simplify rights management, in order to achieve privacy protection.In the early, RBAC rights management applied "top-down" mode: According to the enterprise's position to establish roles. When applied to big data scene,the researchers began to focus on "bottom-up" mode, that is based on the existing "Users - Object" authorization, design algorithms automatically extract and optimization of roles, called role mining. In the big data scene, using role mining techniques， roles can be automatically generated based on the user's access records，efficiently provide personalized data services for mass users. It can also be used to detect potentially dangerous that user’sbehavior deviates from the daily behavior. But role mining technology are based on the exact, closed data set, when applied to big data scene, we need to solvethe special problems: the dynamic changes and the quality of the data set is not higher.

7.5.2Risk Adaptive Access Control

In the big data scene, the security administrator may lack sufficient expertise,Unable to accurately specify the data which users can access, risk adaptive access control is anaccess control method for this scenario.By using statistical methods and information theory, define Quantization algorithm, to achieve a risk-based access control.At the same time, in the big data environment, to define and quantify the risk are more difficult.

CONCLUSION

This paper first introduces the security problems faced by big data, discusses the reasons of privacy problems, then, discusses the principles to address privacy issues, finally, and from four aspects discusses the technology to solve the problem of privacy protection. At present, although there have been some methods to solve the problem of privacy protection, but research is not enough, only combination of the technical and legal means can solve the problem better.

REFERENCES

[1] Bu Ying-Yi,Fu Ada WaiChee,Wong Raymond Chi Wing,et al. Privacy preserving serial data publishing by role com-position//Proceedings of the 34th International Conference on Very Large Data Bases(VLDB’2008).Auckland, New Zealand, 2008:845-856

[2] Feng Deng-Guo, Zhang Min, Li Hao. Big Data Security and Privacy Protection[J]. Chinese Journal of Computers,2014,14(1):246-258.

[3] Chen ChangFen, yuxin. Privacy Protection in the Era of Big Data[J], News and Writing, 2014,6:44-46.

[4] Liu Yahui, Zhang Tieying, JinXiaolong,ChengXueqi.Personal Privacy Gang ZengInt. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 5,

Issue 5, (Part -6) May 2015, pp.46-50 www.ijera.com 50|P a g e Protection in the Era of Big Data[J]. Journal of Computer Research and Development, 2015,15(1):229-247.

[5] Li Fenghai, Li Shuang, Zhang Bailong, etc. Research on Anti APT Attack Plan of High Grade Safe Net [J], Information Network Safety, 2014, (8):109-114

[6] Viktor Mayer-Schonberger, Kenneth Cukier. Big Data: A Revolution that Will Transform How We Live , Work and Think. Boston: Houghton Mifflin Harcourt, 2013.

[7] Ying X, Wu X. Randomizing social networks: A spectrum preserving approach//Proceedings of the SIAM International Conference on Data Mining (SDM'08).Georgia, USA, 2008:739-750

[8] Zhang Yanxin, Tang Xuran. Research on Personal Information Security on Social Network in