A Study on Privacy Techniques in Big Data Technology

P. Sowndarya[1] Dr. V. Kathiresan[2]
[1]MCA Student [2]Head of Department
[1,2]Department of Computer Applications
[1,2]Dr.SNS Rajalakshmi College of Arts and Science, Coimbatore, Tamil Nadu, India - 641 049.

Abstract— Privacy issues are showy by the velocity, volume, and variety of Big Data, such as major cloud infrastructures, diversity of data sources and formats, streaming nature of data acquisition and high volume inter-cloud relocation. Therefore, traditional security mechanisms, which are tailored to securing small-scale, static (as opposed to streaming) data, are inadequate. We emphasize the top ten Big Data security and privacy challenges. Importance the challenges will motivate increased focus on fortifying Big Data infrastructures.

Key words: Big Data, Social Media, Public Power, Security, Quality Assurance, Variety

I. INTRODUCTION

During in the recent years, data production rate has been growing exponentially [1, 11]. Many organizations demand adept solutions to store and analyze these big amount data that are preliminary generated from various sources such as high throughput instruments, sensors or connected devices. Big Data is a very difficult concept to define accurately, since the very notion of Big in terms of volume of data varies from one area to another. Today,the development of information and networking lead to volatile growth of data. According to statistics, 2 million users are using Google's search engine in every second, Facebook users share 4 billion resources every day, Twitter process 340 million tweets every day, At the same time, the large amount of data are produced happening in scientific calculation, medical services, finance, retailing. 8ZB data will be generated in 2015.

At present, the development of big data still faces many problems, Security and privacy issues is one of the key issues that people acknowledged 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 shows 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. Currently, many organizations have realized the security problems of big data, and take action to focus on big data security issues. In 2012, the cloud security alliance (CSA) formed a big data working group, aimed at finding solutions for data center security and privacy issues. It is based on carding the research situation of big data, analyzes the security challenges to big data, discusses the key technology of the current big data security and privacy protection.

Table 1 summarizes the big data technologies from batch processing in 2000 to present with most significant stages and products.

Stage/Year	Characteristics	Examples
Batch Processing *2003-2008*	Big amount of data is collected, entered, processed and then the batch results are produced. Distributed file systems (DFS) are used for fault-tolerant and scalability. Parallel programming models such as MR are used for efficient processing of data.	GFS, MR, HDFS, Apache Hadoop
Ad-hoc (NoSQL) 2005 – 2010	Support random read/write access to overcome shortcomings of DFS that are appropriate for sequential data access. NoSQL databases solve this issue by offering column based or key-value stores, in addition to support for storage of large unstructured datasets such as documents or graphs.	CoachDB, Redis, Amazon DynamoDB, Google Big Table, HBase, Cassandra, MongoDB
SQL-like *2008 - 2010*	Simple programming interfaces to query and access the datastores. This approach provides functionalities similar to the traditional data warehousing mechanisms.	Apache Hive/Pig, PrestoDB, HStore, Google Planner
Stream Processing *2010 - 2013*	Data are pushed continuously as streams to servers for processing before storing them. Streaming data usually have unpredictable incoming patterns. Such data streams are processed using fast, fault-tolerant, and high availability solutions.	Hadoop Streaming, Google Big Query, Google Dremel, Apache Drill, Samza Apache Flume/Hbase, Apache Kafka/Storm
Real-time Analytical Processing *2010 - 2015*	Automated decision making for streams that are generated from the machine-to-machine applications or other live channels. This architecture helps to apply real-times rules for the incoming events and existing events within a domain.	Apache Spark, Amazon Kinesis, Google Dataflow

Table 1: Evolution of the Big Data from batch to real-time analytics processing. [1]

All rights reserved by www.ijsrd.com

II. BIG DATA THREAT PERSONAL PRIVACY

A. Social Network

In our social tricks, 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. Although individual users can set to turn off the reading function of social networking sites, as long as users use social networking sites, he will leave marks - logs, status, messages, and even point admire, connection between user and the whole social network is well-known, there is the possibility of associated with other users in the network.

B. Commercial Interests

Many operators record the user's scene and behavior for a long time, and label the user characteristics analyze the possible behavior habits and needs, and then push advertising information in a range of relatively obscure user groups.

C. Need for Public Power

In order to meet the needs of law enforcement, many countries in the world usually require network or telecom operators to store certain user data in a certain period of time, and provide the raw data and the results when the government need. This requirement is certainly justifiable, and it does not a great threat to personal privacy in the era of the small data.

III. NEED OF BIG DATA

Many of the research organizations, uses big data, but may not have the efficient mechanism, particularly from a security perspective if a security problem occurs to big data, it causes more serious legal consequences and reputational damage than at present. The present era, many organizations are using traditional security mechanisms which are used to secure small scale static data, are inadequate. To provide security for big data, techniques such as encryption, logging sugar pot detection must be essential. Security and privacy issues are magnified by volume; variety and velocity of big data, such as large scale, cloud infrastructure, multiple sources and different formats, streaming nature of data acquisition and high volume inter cloud migration.

IV. MAIN PRINCIPLES OF PRIVACY

A. Principle of the Certain Using Scope of Data

The goal of handling of personal information must be specific, clear, reasonable, does not expand range of use, not change the purpose of the use of personal information when the owner of the information do not know. This standard 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.

B. Quality Assurance

Information managers must ensure that the processing of personal information is confidential, complete, available and up to date. and need to ascertain internal control mechanism to protect personal information, and regularly detect security, protection and the implementation of information systems, measured by themselves or an independent evaluation agency, to develop plans for loss, damage, tampering, improper use and other events during processing;

V. KEY TECHNOLOGIES OF PRIVACY

Privacy protection technology is mainly studied from the following perspectives: user privacy protection, data content verifiable, and access control.

A. Anonymity Data Protection Technology

The big data environment, inscrutability protection is necessary to protect the data. For example, in social networks, inscrutability protection can be divided into user identity anonymity, attributes anonymity and relationship obscurity. At here, 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 be inferring anonymous users, especially relationship between the users.

B. Data Watermarking Technology

Digital watermarking refers to the identification information is embedded imperceptible within the data carrier and does not affect the method of its use, usually used for copyright protection of multimedia data, there is also a watermarking scheme for databases and text files. Due to the characteristics of arbitrariness and dynamic data, watermarking methods are very different on the marked database, document and multimedia files. Stout Watermark can be used to prove the starting point of big data. Delicate Watermark can be used to prove the dependability of big data.

C. Data Provence Technology

The diversification of data sources, it's indispensable 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 point is to help people establish the source of the data in the data warehouse. The method of data province's 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 update the data with a minimum price.

D. Access Control Technology

1) Role Mining

Role-based access control (RBAC) is an access control model used usually. By passing on 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.

All rights reserved by www.ijsrd.com

2) 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 an access 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 situation, to define and measure the risk are more difficult.

VI. CONCLUSION

In this paper first introduce the security a problem faced by big data, discusses the reasons of privacy problems then, discusses the principles to address privacy issues, finally, from four aspects discusses the technology to solve the problem of privacy protection. At here, even if 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.

REFERENCE

[1] Ali Gholami and Erwin Laure, "Big Data Security And Privacy Issues In The Cloud", International Journal of Network Security, January 2016.
[2] C.Yosepu, P Srinivasulu, Bathala Subbarayudu "A Study On Security And Privacy In Big Data Processing", International Journal of Innovative Research in Computer and Communication Engineering.
[3] European Data protection Supervisor, "Meeting The Challenges Of Big Data", European Data Protection Supervisor is an independent institution of the EU,19 November 2015.
[4] Gang Zeng, "Privacy Protection In Big Data Environment", Journal of Engineering Research and Applications, May 2015.
[5] Sreeranga Rajan, Fujitsu, "Big Data Security And Privacy Challenges", cloud security alliance , April 2013.
[6] Omer Tene Jules Polonetsky "Big Data For All: Privacy And User Control In Theage Of Analytics" Northwestern Journal of Technology and Intellectual Property, April 2013.
[7] Azzeddine RIAHI Sara RIAHI, "The Big Data Revolution, Issues And Applications" International Journal of Advanced Research in Computer Science and Software Engineering, August 2015.

All rights reserved by *www.ijsrd.com*