SURVEY ON BIG DATA AND RELATED PRIVACY ISSUES

P.Kamakshi[1]

[1]*Professor, Dept. of Information Technology, Kakatiya Institute of Technology and Science, Warangal, India*

Abstract

With recent development in technology, networking and cost reduction in storage devices, today we are flooded with huge amount of data . The data is collected from heterogeneous sources and wide range of application areas. Analysis was performed on these data by means of methodically developed models. Big data is a conventional term used to describe the exponential increase and accessibility of structured and unstructured data. In future big data will be essential to business as well as society like internet facility. Resolutions that were previously build on estimation or on conceptual models of reality can now be done based on the collected and stored data itself. Big Data analysis is now used in almost every phase of our society, communication services, marketing, banking and research. The big data phase has shown the ways for huge opportunities in science, health care system, economic decision, educational system and novel forms of public interaction and entertainment. But these opportunities also result in challenges in the area of privacy and security. Big data utilizes huge quantity of data that may be available in the cloud and it may require data processing distributed across numerous servers. It is found that the development progress of big data also intensified the threats to information security. This paper focus on the big data and various privacy issues related to it.

Keywords: *Big data, Data privacy, Privacy issues*

---***---

1. INTRODUCTION

The buzzword big data is a catchword used to illustrate a great volume of structured as well as unstructured data. As the data size is very huge, it is difficult to use traditional database and software techniques to process it. In many organizations either the data is too large or it moves at extremely high-speed or it goes beyond existing processing capability. Big data is likely to facilitate business in improving their operations and help in making faster and more[5] intelligent decisions. Though the term big data look as if representing huge volume of data, but that is not always the situation. The retailer refers big data as technology which contains tools and techniques required by an organization to deal with huge quantity of data and storage amenities.

It is believed the term big data started with companies handling web search applications and looked-for queries on large distributed collection of data. The range of big data may be petabytes or exabytes of data consisting of huge number of records of millions of people related to sales, health care system, mobile information etc. Generally such data is un- structured data and is commonly unfinished and unapproachable. It is not necessary that Big Data only refers to extremely huge data and tools and measures used to process and investigate them, but it also gives directions for new ideas and challenges [4] in research area.

1.1 Few Major Challenges of Big Data are as below:

A: Short of efficient tools and techniques for safely organizing large-scale data and distributed data sets
B: Security and privacy issues while sharing data and susceptible ever growing public databases
C: Deliberate or malicious leakage of data

2. APPLICATIONS OF BIG DATA

Though the term ‘Big Data’ simply looks like a great buzzword today, In the long run, every phase of our lives will be influenced by big data. The applications of big data can be categorized as below:
i. Customer analysis: Big data helps companies in analyzing the customer purchase patterns and predict the future requirement to companies by means of various models.
ii. Optimize business processes
iii. Improvement in personal performance optimization
iv. Improvement in public healthcare system
v. Growth in Science and Research
vi. Enhancement in laws of protection and security
vii. Improving and Optimizing Cities and Countries
viii. Economic improvement
ix. Fraud analysis
x. Analysis of social media access

3. FEATURES OF BIG DATA

“Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”. With respect to Gartner definition, big data is often described in terms of the ‘three Vs’: volume, variety and velocity.

Volume: Big data uses huge datasets[10] which include data internet searches, online purchases and transactions, social media interactions, mobile information, data from sensors in vehicles and other devices. The amount of big data may be petabytes or exabytes. It is also possible to hold very large datasets, due to the reducing price of storage and the accessibility of cloud-based services. As these datasets are very large, they cannot be analyzed using conventional techniques like spreadsheets or SQL queries. New tools like NoSQL and open source software Hadoop have been developed to analyse big data.

Variety: Big data often necessitate collection of data from heterogeneous sources. Presently it seems that big data analytics primarily employs structured data like tables with defined fields as well as unstructured data. For example, the data is collected from various sources like social media source like twitter, on line products purchases and the comments related to products etc. merging data from diverse sources in this way presents various challenges with respect to IT perspective. Practitioners analyzed and suggested that of the 'three Vs', variety is the most significant characteristic of big data. This view propose that, when a company is analyzing its own customer database which is very large, may not essentially publish any innovative ideas in terms of either analytics or data protection. On the other hand, when it joins its own information with the data extracted from various sources, then it will give results that are qualitatively different.

Velocity: In some situation like in real time it is essential to analyze data as fast as possible. Big data analysis can be employed to analyze the static data like database of a store as well as the data which is time varying and continuously created or documented like online purchases and credit card payments.

4. PRIVACY ISSUES IN BIG DATA

With rapid growth in technology, networking and cost reduction in storage devices, data revolution has taken place naming it as Big data. Big data is referred as one in which enormous quantity of data can be collected, stored and analyzed at reasonably low price.

The graphical representation of increase in flow of new digital data is show in Fig.1below.

Such collection of huge data provide benefits to health care, government services, fraud protection, retailing, manufacturing and other sectors.

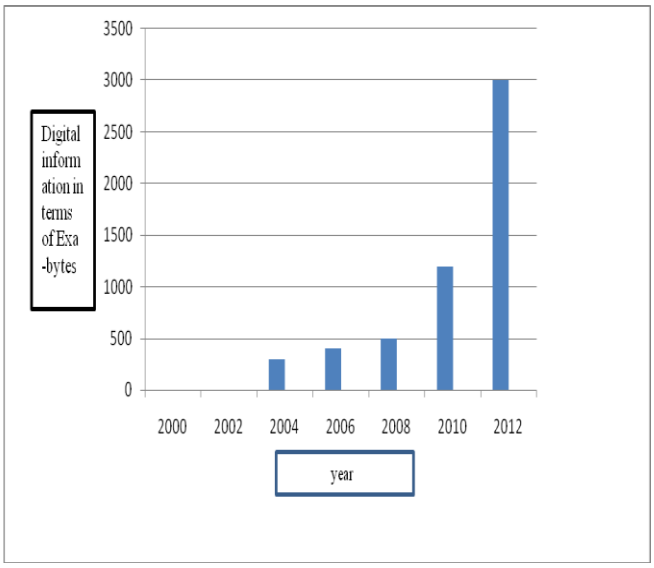

Fig 1 shows the rapid growth in digital information

Today big businesses are dependent on peoples or customer data. Every day the data about health, on line purchases, shopping and entertainment preferences in the form of digital information is collected, which in turn helps the business in decision making and achieving their goals. Freely available data encourages many businesses for analyzing the data and aim at their consumers exclusively, but sometimes in illegal, fashion.

Many organizations like mobile companies, insurance companies, banks offering Loans etc. collect the personal information of an individual like phone numbers, addresses, email information etc. from various sources and utilize it for their personal usages, causing problems to the people and customers. The customers even don't know how his/her information is distributed to[1] different organizations without their knowledge.

The online privacy management services provider survey reveals that in 2014, 92% of US internet users were worried about their online privacy . Only 55% of US internet users said they trust most businesses with their personal information online and 89% of consumers revealed that they avoided doing business with companies as their online privacy information is not protected.

Ken Parnham, TRUSTe's European Managing Director, mentioned that it is predictable that consumer online trust has dropped to its lowest point, and only 55% of internet users prepared to trust companies with personal data online. It is an alarm for businesses that collection and sharing of commercial data increased online privacy concerns..Lack of consumer confidence can starve the businesses of valuable data, product and sales, restricting the livelihood of the digital market as people are less liable to access and share advertisements, use apps, or enable location tracking on their mobile phones. These results show that accomplishment is no longer just about modernization or technology development, It is necessary that co must take decisive action to deal with online privacy concerns to continue competition, risk reduction and build online trust in future.

A UK-based survey conducted by the Global Research Business Network (GRBN) reported that 40% of respondents in the UK and 45% of respondents in the US were highly concerned about the safety of their personal data. Data that consumers considered to be personal were national insurance numbers, healthcare data , and geographical location data. Only 27% of respondents trusted law enforcement with their personal data, followed by banks and retail.

This is an indication of threat and an alarm to every company's big data strategy under the category of risk management. The big question is that how do big data can be used productively and profitably without jeopardizing consumers trust and business patronage.

Though the emergence of big data provides enormous benefits, but has also triggered privacy concerns. To a great extent the anxiety is associated with misuse of data which collected and used by governments for national security, healthcare purposes, insurance etc.

Today with advancement of technology and internet, people are able to realize that the information which they really think as private to them is no more private, but became public. Today this is the major concern in people day to day lives.

5. HOW BIG DATA CAUSES PRIVACY VIOLATION IN VARIOUS APPLICATIONS:

5.1 Health Care System

Because of the tremendous advantages in protecting the health of patients, big data is highly supported by health care[2] system**.** Big data information is used to recognize[6] people with a high risk of certain medical conditions at early stage and providing improved quality care and lowering the increase cost of health care. Although there are tremendous benefits, new studies are revealing that big data may be riskier than initially thought. As per survey it is found that, though the health care data is personal, it is easily accessible. It is important to be conscious about security and privacy implications tapping into big data.

5.2 Predictions can cause Discrimination

Big data allows the prediction of quite a bit of other information about people. The information big data can predict is increasingly developing the potential to be used as a way of discriminating against people in[7] a variety of demographics. A study shows that when observation of status i.e. like information from face book was analyzed, it gave accurate information to discriminate men depending on race , alcohol consumption, gender etc. It is very much concerned by many people that organizations, employers, education system may use such models and start discriminating people based on many human oriented parameters.

5.3 Product Sales

One of the major applications of big data is marketing where the marketers try to place their products and services in front of highly targeted customers. But, when the customer is categorized into one category based on their behaviors, there is possibility for harm. In spite of the possibility for harm, marketers still use big data to aim at people on social media platforms like search engines and email. Forceful entry into personal area by providing advertisements based on friends, likes and email content is causing anxiety among consumers.

6. CONCLUSION

In today's world many companies and organizations like banking, educational system, credit card companies, insurance companies etc. are using big data for analysis purpose. Huge amount of digital information collected from various sources like credit card companies, government institutes, banking, health care systems are used regularly for analysis purpose. Inspite of several applications [8]and advantages, big data has raised new challenges in the area of privacy and security. The main reason behind the privacy problem is that today huge amount of personal information is freely available directly or indirectly [11]in the form of digital information. Many organizations are utilizing Big data for their personal benefits ,profit and to achieve their goals by using the personal information of customers. Misusing of such information is causing loss[3] of trust and faith of customers in organizations. In this paper the strength and applications of big data as well as various privacy issues are discussed

REFERENCES

[1]. Boyd, Danah and Kate Crawford, "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon."Information, Communication, & Society 15:5, p. 662-679(2012).
[2]. "Big Data is the Future of Healthcare", Cognizant 20-20 insights, September 2012
[3]. Thomas M. Lenard and Paul H. Rubin, "The Big Data Revolution: Privacy Considerations", December 2013
[4]. Big Data Analytics for Security Intelligence,, CLOUD SECURITY ALLIANCE , September 2013
[5]. Agrawal R.,Srikant R., ``Privacy Preserving Data Mining.,"In the Proceedings of the ACM SIGMOD Conference.2000.
[6]. "Big Data Analytics" ericsson White paper,284 23-3211 Uen, August 2013
[7]. VINT research report on " Privacy, technology and the law Big Data for everyone through good design"
[8]. Ira S. Rubinstein, ' Big Data: The End of Privacy or a New Beginning?", International Data Privacy Law Advance Access published January 25, 201
[9]. P. Russom, " Big Data Analytics", Best Practices Report, Fourth Quarter, The DataWarehouse Institute , Renton, WA, September 18 2011
[10]. IDC, Digital data to double every 18 months, worldwide marketplace model and forecast, Framingham, MA. available at www.idc.com May 2009
[11]. R. Agrawal, R. Srikant, "Privacy-preserving data mining", In: Proceedings of the 2000ACM-SIGMOD on management of data, Dallas, TX, USA, May 15-18, 2000