International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-03, Issue-08, Nov 2017

A Survey on Security Issues in Big data

[1]Mr. J. Sagar Babu, [2]Mr. K. Yakhoob

[1,2]Assistant Professor, Department of CSE, Princeton College of Engineering & Tech, Hyderabad, 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.

Keywords – Security Issue, Big Data, privacy issue, analysis.

I. INTRODUCTION

Data generation and collection quickly surpass the bounds in the digital universe of today. The data has been doubling every 2 years since 2011 [1]. It is predicted that the data will increase 300 times, from 130 exabytes in 2005 to 40,000 exabytes in 2020 [2]. As a result of this technological revolution, the big data is becoming increasingly an important issue in the sciences, governments, and enterprises. Big Data is a data set, which is difficult to capture, store, filter, share, analyze and visualize on it with current technologies [3]. Despite such difficulties, if you can cope with big data, it provides you with generating revenue, executive efficiency, strategic decisions, better services, defining needs, identifying new trends, and developing new products, all of which is covered in the data science [3]. In addition, data science studies parallel and distributed processing, similarity search, graph analysis, clustering, stream processing, search ranking, association analysis, dimensionality reduction and machine learning algorithms [4]. However, in this complex computation environment, traditional security and privacy mechanisms are insufficient to analyze big data. These challenges in big data consist of computation in distributed and non-relational environments, cryptography algorithms, data provenance, validation and filtering, secure data storage, granular access control, and real time monitoring [5].

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

II. 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

III. DEFINITION AND CHARACTERISTICS OF BIG DATA

Big data refers to large and complex datasets that typical software is inadequate for managing [2]. There are various Explanations of big data via Vs. 5Vs are typically used to characterize of Big Data as volume, velocity, variety, veracity and value (Fig. 1) [3,6,7]. Volume is the size of data; velocity is the high speed of data; variety indicates heterogeneous data types and sources; veracity describes consistency and trustworthy of data; and value provides outputs for gains from large data sets.

DOI : 10.18231/2454-9150.2017.0033

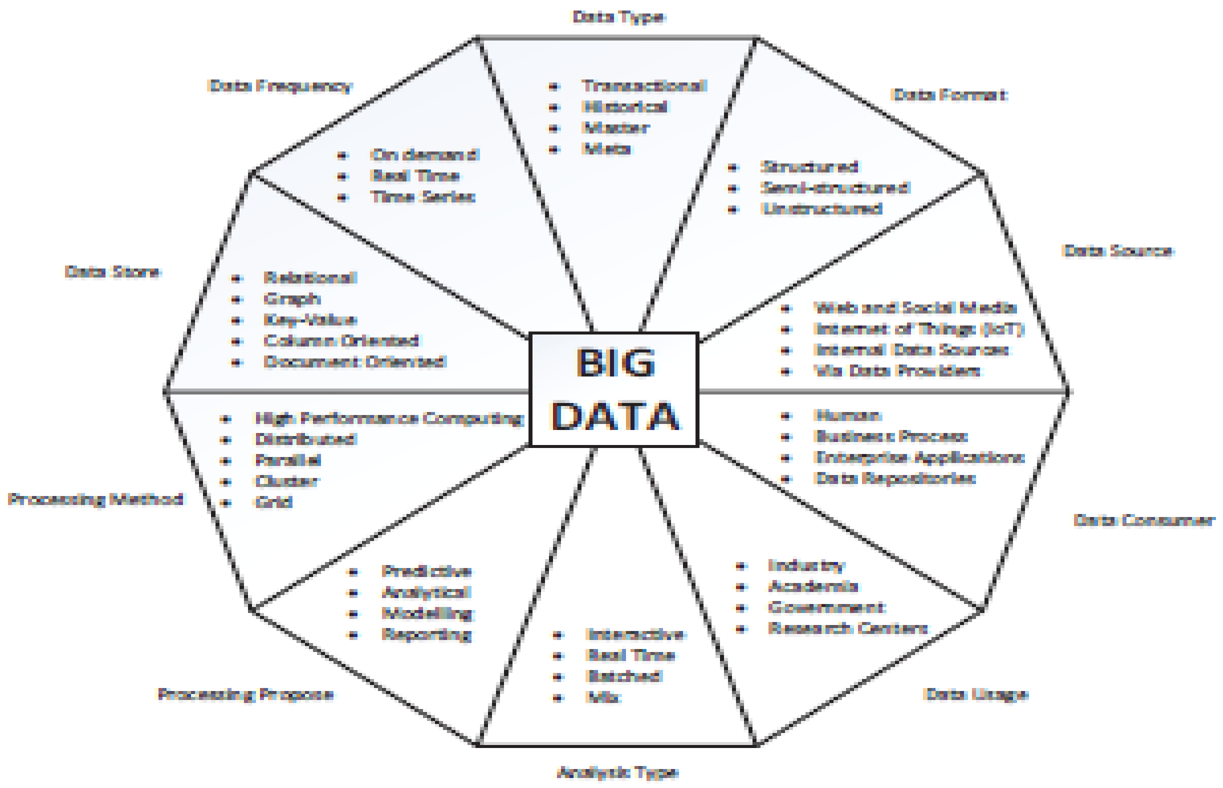

Fig 1: Bigdata Classification

IV. 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.

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.

V. 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.

DOI : 10.18231/2454-9150.2017.0033

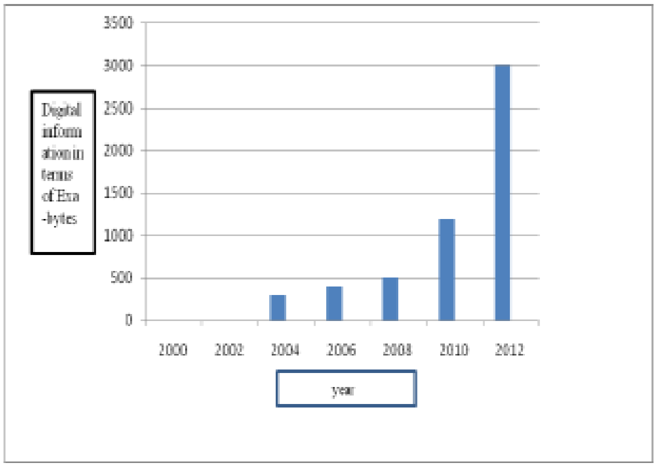

Fig 1 shows the rapid growth in digital information

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.

VI. HOW BIG DATA CAUSES PRIVACY VIOLATION IN VARIOUS APPLICATIONS

6.1 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.

6.2 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.

VII. CONCLUSION

Big data needs extra requirements for security and privacy in data gathering, storing, analysing, and transferring. In this paper, we examined studies on big data security and privacy, comparatively. According to the literature, network traffic should be encrypted with suitable standards; access to devices should be checked; employees should be authorized to access systems; analysis should be done on anonymised data; communication should be made for the secure channel to prevent leakage, and network should be monitored for threats. Big data privacy, safety and security are the biggest issues to be discussed more in the future, so new techniques, technologies and solutions need to be developed in terms of human-computer interactions or existing technologies should be improved for accurate results. It is hoped that this study would help understand the big data and its ecosystem better and develop better systems, tools, structures and solutions not only for today but also for the future.

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] Vinayak D. Shinde, "Fear of Data Privacy and Security in Cloud Computing Technology", International Journal for Research in Engineering Application & Management (IJREAM), ISSN:2494-9150 Vol-01, I-09, pp 01-09, 2015.

[7]. “Big Data Analytics” ericsson White paper,284 23-3211 Uen, August 2013

DOI : 10.18231/2454-9150.2017.0033