Big Data and its privacy and security concerns

Mr. Srikant Singh,
Assistant Professor, Department of Computer science and Engineering, Kalinga University, City- Naya Raipur, State- Chhattisgarh, Country- India.

Abstract:*In the current age the amount of data gathered around us form a group of both structured and unstructured data. The unstructured data around us is about 80% of the total data which leads to a problem of data analysis. As this data is not kept in a proper way so it needs to be analysed for the purpose of research but this leads to the security and privacy concern. This paper deals with the issues related to the privacy and security of Big Data. It focuses on the various areas in which the data can be accessed and that can be used for the purpose of data analysis.*

Date of Submission: 21-07-2018 Date of acceptance: 6-08-2018

I. Introduction

Database systems technology has advanced a great deal during the past four decades from the legacy systems based on network and hierarchical models to relational and object database systems. Due to the explosion of web-based services, unstructured data management and social media and mobile computing, the amount of data to be handled has increased from terabytes (TB) to petabytes (PB) and zetabytes (ZB) in just two decades. Such vast amounts of complex data have come to be known as Big Data(Workshop Report: Big Data Security And Privacy Sponsored by the National Science Foundation, , The University of Texas at Dallas, September 16-17,

2014). „Big Data‟ refers to novel ways in which organizations, including government and businesses, combine diverse digital datasets and then use statistics and other data mining techniques to extract from them both hidden information and surprising correlation.

II. Issues in Big data

When we come across the term Bigdata the first thing that comes in mind is about the vast or the enormous amount of data present around us in unstructured form. This enormous amount of data is needed for further analysis which is termed as data analytics. When we pick such kind of data then we have security and privacy concern related to data. Security means the state of being free from danger or threat while Privacy means a state in which one is not observed or disturbed by other people and the state of being free from public attention. In this paper I am presenting a problem related to data analytics. We can come across various domains to review the privacy issues with Big data. Some of the domains are: Social Network, Social Media, Healthcare and IOT.

Most of us are using social media or social network for different purposes like chatting, reading, entertainment etc. We share our things easily to other persons or group but what about the safety of the data? Whether we are concerned about data safety or privacy? The answer is that only some of us are concerned while others are not aware of data safety and data privacy. They do not know about data safety even they do not know the ways in which their private data can be manipulated and presented to others. So for this companies must undergo the ways to prevent the data during data analysis.

III. Big Data: Group of unstructured data

Almost all the large firms like – Apple, Google, Amazon, Flipkart are engaged in Big Data handling in some form. One of the best example is Google as it relies on the availability of the data it collects from its own services not only to fund its operations (by determining and delivering relevant search ads) but also to train its search algorithms and develop new data-intensive services such as voice recognition, translation, and location-based services.

But if we talk only about these companies‟ data then it will be injustice to Big Data as the field of Big Data id Stretched from a single firm to any government or any company. Basically these datasets are used for the statistical algorithms and data analysis purpose. Data can be included from:

a. Media Houses: A group which has a large extent of data that can be given for research purpose or for the analysis of data. The media houses are collecting enormous amount of data of rich content.

b. Healthcare Services: A large group which focuses on the amount of data gathered from different hospitals, Pathology departments and that can be further used for the study or to find a solution or the cure for a disease. The healthcare firms are maintaining the records in electronic form and also maintain the scanned images for a particular person, which can be used for short-term/long-term health monitoring and epidemiological research programs.

c. Sensor networks: The data from sensor networks includes everything like finger prints or face sensors or image sensors but they have a very wide range in which the data can be used. If the data in this field is not kept secretly then there may be the problem of the data theft which can lead to the issue of security related to the data. Sensor data is circulated at enormous rate from various devices like GPS, RFID, remote sensors etc.

d. Data from social media or network: Various social networks like whatsapp, facebook, Viber, snapchat, instagram are generating data at a very fast rate and people used to share or store information on these platforms. All these are free services so people use it instantly to share their information.

All the above four categories are generating data at a very high speed and every network has a serious concern about the privacy and security of the data.

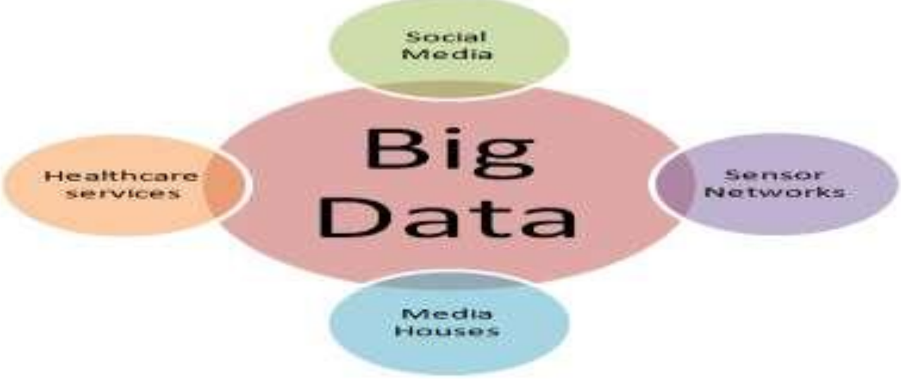

Fig. 1.1

IV. Big Data: Security and Privacy issues

Big Data has a serious privacy concern during transmission of data from one end to another for mining purposes of for data analytics. In big data we have differenent kind of data brought together from different ends and so it raises the security issues in different domains. Big Data includes data logs that may be used for various purposes, which may lead to the security and privacy of an individual to risk. Let us focus on some of the areas which can be used for data analytics :

1. **Social Insurance:** The social insurance industry bridles the intensity of enormous information, security and protection Issues are at the point of convergence as developing dangers and vulnerabilities keep on growing. In human services, a few variables give the essential power to bridle the intensity of huge information. Tackling the intensity of enormous information investigation and genomic examine with continuous access to persistent records could enable specialists to settle on educated choices on medications. As of late, mechanical achievements have assumed a critical part in engaging proactive social insurance. For example, ongoing remote checking of essential signs through installed sensors (connected to patients) permits social insurance suppliers to be alarmed if there should be an occurrence of any issue or troublesome circumstance.

a. **Security and Privacy issues in HealthCare/Social Insurance**: today the growth of healthcare industry leads to the enormous growth in the big data. It also leads to the security and privacy concerns with it. We can focus on following information when we discuss about security and privacy issues related to it:

i. As we know that the amount of unstructured data around us is in very huge amount and must be analyzed for further research and use. The healthcare industry participate a lot when we deal with the big data. The data in the healthcare data centers are certified but most of the certifications does not guarantee safety of patient‟s records but they are focused on the security policies and procedures than implementing them.

ii. Different studies showed that most of the hospitals had at least one security breach in the recent years. In most cases, the attacks were from an insider rather than external [10].

2. **Big Data Future in HealthCare:**

The more extensive acknowledgment and utilization of Big Data over the world has given numerous measurements to the continuous observing and it has offered chances to go anyplace, whenever, with nearly anything in not so distant future. Beneath specify data must be considered for the eventual fate of Big Data in human services:

a. Huge Data has turned into a rising power for the development of IOT. Gartner gauges 26 billion IOT gadgets will be practical by 2020 and the measure of activity created by such gadgets will be sufficiently huge to put it in the classification of huge information (P. Middleton, P. Kjeldsen and J. Tully, "Forecast: The Internet of Things, Worldwide," Gartner, 2013.).

b. Moreover, with the presentation of BSN (Body Sensor Networks) and their immediate application to human services industry, mind suppliers will have the capacity to screen indispensable parameters, medicine viability, and anticipate a plague. Body sensors create huge measure of information, and connecting such medicinal services information from dissimilar asset obliged systems will be significant for driving social insurance investigation. Subsequently, medicinal services suppliers have colossal chances to alter social insurance by saddling the intensity of Big Data (M. Hanson, H. Powell, A. Barth, K. Ringgenberg, B. Calhoun, J. Aylor and J. Lach, "Body Area Sensor Networks: Challenges and Opportunities," Computer, pp. 58-65, 2009.).

2.1. Internet based life: Social media is the media (content) that we transfer or download – whether that is a blog, video, sound, slideshows, web recording, pamphlet or an eBook and so on. Open online networking are for general society and individuals share their data with each other. This data could be as content, pictures, sounds, and recordings.

2.2. **Security and Privacy issues in Social Media:** Public web based life has offered ascend to the huge measure of information. This information is shareable and is in different structures. This has expanded the security and protection worries with such immense measure of information. Following data gives an outline of security and protection issues engaged with Big Data condition:

a. The measure of web based life as far as content, pictures, sounds and recordings, and so on, is becoming too quick and the way it is expanding, it appears that there is no conclusion to this continuous pattern because of which it will be hard to anchor the individual information and security of the individual information will be excessively troublesome, making it impossible to deal with.

b. When taking a gander at protection issues in online networking in Big Data, there is a need to separate which of the numerous Big Data applications areas are being talked about. As the customary Big Data applications, for example, stargazing and other e-sciences as a rule work on non-individual data and all things considered don't have security issues since this sort of data isn't of individual pertinence (. Boyd and K. Crawford. Six Provocations for Big Data. SSRN eLibrary, 2011.) The protection basic Big Data applications exist in the new spaces of social web among which datasets of web-based social networking is critical from the security and protection perspective.

2.3 Internet of Things (IOT) Era: The IOT is an ongoing correspondence worldview that imagines a not so distant future, in which the objects of regular daily existence will be outfitted with microcontrollers, handsets for advanced correspondence, and appropriate convention stacks that will make them ready to speak with each other and with the clients, turning into an essential piece of the Internet (Andrea Zanella et.al. “Internet of Things for Smart Cities” IEEE Internet Of Things Journal, Vol. 1, No. 1, February, 2014).

2.3 Security and Privacy issues in IOT: IOT is a system of systems in which countless, sensor gadgets are associated through the ICT (Information Communications Technology) framework to offer some benefit included administrations. The Internet of Things interfaces individuals and things whenever, wherever with anything and anybody; in a perfect world utilizing any way or system and administration (C. Perera et al., “Context Aware Computing for the Internet of Things: A Survey,” IEEE Comm. Surveys & Tutorials, vol. 16, no. 1, 2013, pp. 414–454). The security and protection issues in IOT can be characterized and comprehended as takes after:

a. Online administration purchasers know that when they utilize free online administrations, (for example, messages, person to person communication sites, newsfeeds, and so on.), they consequently move toward becoming information hotspots for the organizations, which can break down their information to enhance consumer loyalty. What's more, the more terrible thing is that this information of the online purchasers can be sold to the any outsider for facilitate investigation without the worry of the buyers of these online administrations (C.Perera et.al. “Big Data Privacy in the Internet of Things Era”, IT Pro May/June 2015).

b. In the IOT time, the measure of client information that can be gathered will be essentially higher than previously. For instance, ongoing wearable advances, for example, Google Glass, Apple iWatch, Google Fit, Apple Health Kit, and Apple Home Kit can gather touchy data about clients extending from their wellbeing conditions to money related status by watching or recording their every day exercises

2.4 Social Networks: Social systems administration has risen as another time of correspondence, regardless of whether close or far, anybody can be associated through informal organizations to anybody they need to. Individuals share pictures recordings, sounds, content, and so on nearly on regular routine and this causes the requirement for Big Data in informal organizations as the information produced by the long range interpersonal communication destinations and other interpersonal organizations can reach up to numerous gigabytes in volume every day and this information must be taken care of with outrageous care. Facebook, Twitter, Google, Mobile Phone Companies, Retail Chains, and Government offices are the best cases of web-based social networking where individuals share their data.

2.5. Security and Privacy issues in Social Networks: Social systems administration destinations give the specialist to its clients to utilize the protection settings so a client can set protection according to his prerequisites from the protection settings that are given by the specific long range informal communication site the client is utilizing. For instance, Facebook gives such a large number of alternatives to security settings like 'Who can see my stuff?', 'Who can get in touch with me', 'How would I prevent somebody from annoying me?' and so forth., that a client can use to make his profile private and secure according to his needs. Following issues are of significant concern:

a. Security worry here is that these protection settings are at the client's end, shouldn't something be said about the opposite end where the person to person communication destinations are dealt with and created? Does a client get protection at that opposite end too?

b. An interpersonal organization client transfers pictures, recordings, sounds, content, and so on with those whom he needs to get associated. Is it really safe from the point of view of security in such a vast datasets of Big Data? As informal organizations are getting more well known step by step, it has offered ascend to new classes of security and protection worries in the Big Data time.

To face and treat the new difficulties in Big Data that has been produced and getting created in not so distant future, new methodologies that must include the present and future parts of social and mechanical arrangements will be required. Different hypotheses in inquire about have been mulled over for watching protection related data that is accessible on the web. Underneath specified are the work that has been received for security and protection concerns:

To speak to private circle of the clients and consequently bind the entrance, security bubbles were utilized as the first limit between clients to share the online pictures. To secure the protection in informal communication destinations security measure is utilized as defensive conduct and precursors were utilized to empower protection, for example, self' viability, sexual orientation, saw defenselessness.

V. Conclusion

In this paper we have looked into security and protection issues in various space of Big Data. It has likewise been said that what are the distinctive wellsprings of datasets that constitutes the Big Data. Security and protection issues identified with human services, internet based life, IOT period and informal community has been thought about for survey. In future we will survey protection and security worries in different areas of Big Data as they develop, since time to time audit of security and protection issues help comprehend the more extensive part of Big Data in innovative progression that nearly everybody will be a piece of.

References:

[1]. Norshidah Mohamed, lli Hawa Ahmad "Information Privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia

[2]. Workshop Report: Big Data Security And Privacy Sponsored by the National Science Foundation, , The University of Texas at Dallas, September 16-17, 2014

[3]. P. Institute, "Third Annual Benchmark Study on Patient Privacy and Data Security," Ponemon Institute LLC, 2012.

[4]. P. Groves, B. Kayyali, D. Knott and S. V. Kuiken, "The 'big data' revolution in healthcare," McKinsey & Company, 2013.

[5]. Boyd and K. Crawford. Six Provocations for Big Data. SSRN eLibrary, 2011.

[6]. Andrea Zanella, Nicola Bui, Angelo Castellani,Lorenzo Vangelista, Michele Zorzi, "Internet of Things for Smart Cities" IEEE Internet Of Things Journal, Vol. 1, No. 1, February, 2014

[7]. C. Perera et al., "Context Aware Computing for the Internet of Things: A Survey," IEEE Comm. Surveys & Tutorials, vol. 16, no. 1, 2013, pp. 414–454

[8]. C.Perera R. Ranjan, Lizhe Wang; S.U. Khan, A.Y. Zomaya, " Big Data Privacy in the Internet of Things Era", IT Pro May/June 2015

[9]. P. Middleton, P. Kjeldsen and J. Tully, "Forecast: The Internet of Things, Worldwide," Gartner, 2013.

[10]. M. Hanson, H. Powell, A. Barth, K. Ringgenberg, B. Calhoun, J. Aylor and J. Lach, "Body Area Sensor Networks: Challenges and Opportunities," Computer, pp. 58-65, 2009.

[11]. H. Sundmaeker et al., "Vision and Challenges for Realizing the Internet of Things," Cluster of European Research Projects on the Internet of Things, 2010, www.internet-of-things -research.eu/ pdf/IOT_Cluster book_March_ 2010.pdf.

[12]. Delphine Christin, Pablo Sanchez Lopez, Andreas Reinhardt, Matthias Hollick and Michael Kauer" Share with Strangers: Privacy Bubble as user centered privacy control for mobile content sharing applications "Elsevier 2012.

Srikant Singh"Big Data and its privacy and security concerns " "International Journal of Engineering Science Invention (IJESI), vol. 07, no. 8, 2018, pp 53-56