ISSN: 2393-8528

Contents lists available at www.ijicse.in

International Journal of Innovative Computer Science & Engineering

Volume 2 Issue 4; September-October-2015; Page No.12-15

# Big Data Security and Privacy Issues: A Review

**Nitin Kr. Agrawal**[1], **Dr. Aprna Tripathi**[1]

[1]Department of Computer Applications, Institute Of Engineering & Technology, Mangalayatan University

E-mail: nitin.agrawal@mangalayatan.edu.in, aprna.tripathi@mangalayatan.edu.in

**ARTICLE INFO**

Received: 01 Oct. 2015
Accepted 26 Oct. 2015

**Corresponding Author:**

**Nitin Kr. Agrawal**

Department of Computer Applications, Institute Of Engineering & Technology, Mangalayatan University

**E-mail:**

nitin.agrawal@mangalayatan.edu.in

**ABSTRACT**

Big Data has emerged as a necessity in the present world. Most of the human beings are connected to one another through different modes of communications. People share information in different forms. The information that connects people is growing tremendously in large volume which is creating security and privacy concerns. As Big Data technologies are emerging at very fast pace, it is also creating space for security and privacy issues. Until these issues are not addressed properly, it may create obstacles to the fulfillment of expected growth and opportunities and long term success of Big Data. In this paper we review the various domains of Big Data such as healthcare, social media, Internet of things (IOT) and social networking for security and privacy related issues.

**Keywords:** Big Data, security, privacy, healthcare, social media, social networking, IOT

## INTRODUCTION

The term "Big Data" refers to the collection of very large and complex data sets, which in the current scenario exceeds the existing computational, storage and communication capabilities of conventional methods or systems.

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. Database systems can also now be accessed via the web and data management services have been implemented as web services. 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 [1]. '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 [2]. 'Big Data' is best understood as a more powerful version of knowledge discovery in databases or data mining, which has been defined as 'the nontrivial extraction of implicit, previously unknown, and potentially useful information from data' [3]. The McKinsey Global Institute (MGI) recently defined Big Data as 'datasets whose size is beyond the ability of typical database software to capture, store, manage, and analyze' [4]. Since we will focus on the security and privacy issues, we need to understand what they mean. 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 we will review the security and privacy issues in various domains as follows: Security and privacy issues in Healthcare, Social Media, IOT, and Social Networks. Before that we will also see that what are the various data sources on which datasets of Big Data depends.

## 2. Big Data: Dependence on Data

All the biggest Internet companies like— Google, Facebook, Amazon, eBay, Microsoft, and Yahoo!—are engaged in Big Data in one form or another and treat data as a major asset and source of value creation. Google is an especially good example 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 [5].

However, Big Data has much bigger and wider pool of organizations than these big companies only. It has been extended to any company and government agencies that depend on datasets of Big Data for statistical algorithms and different data mining techniques to analyze these large datasets and ultimately improving decision making and enhancing efficiency to take better decisions. Data includes in:

**a. Media/entertainment:** The media/entertainment industry has moved to digital recording, production, and delivery in the recent times and is now collecting large amounts of rich content.

**b. Healthcare:** The healthcare industry is recording the data in electronic medical records and images, which is used for short-term health monitoring and long-term epidemiological research programs (Epidemiology is the study of the patterns, causes, and effects of health and disease conditions in defined populations).

**c. Video surveillance:** Video surveillance is still transitioning from CCTV (Closed Circuit Television) to IPTV (Internet Protocol Television) cameras and recording systems that organizations will analyze for behavioral patterns (security and service enhancement)[6].

**d. Transportation, logistics, retail, utilities, and telecommunications:** Sensor data is being generated at fast rate from fleet GPS (Global Positioning system) transceivers, RFID (Radio Frequency Identification) tag readers, smart meters, and cell phones (call data records [CDRs]); which is used to optimize operations and drive operational BI(Business Intelligence) to realize immediate business opportunities [7].

**e. Data through Social Networks:** Social Networks such as Facebook, Twitter, Whatsapp, Viber, etc., are used by hundreds of millions of Internet users all over the world. All these social networks are free services. We are using free services of such sites in order to stay connected online or offline with our friends by sharing photos, videos, bookmarks, blogs etc [8].

## 3. Big Data: Security and Privacy issues

Big Data is being tremendously used with its vast amount of datasets through different data sources. It is also giving rise to the security and privacy concerns in different domains. As Big Data includes data and information that will be used for different purposes, the security and privacy of an individual is at risk. Here in this section we briefly discuss about the security and privacy issues in various domains of Big Data which are as follow:

**3.1. Healthcare:** The healthcare industry harnesses the power of big data, security and privacy Issues are at the focal point as emerging threats and vulnerabilities continue to grow. In healthcare, several factors provide the necessary force to harness the power of big data. Harnessing the power of big data analysis and genomic research with real-time access to patient records could allow doctors to make informed decisions on treatments. In recent times, technological breakthroughs have played a significant role in empowering proactive healthcare. For instance, real-time remote monitoring of vital signs through embedded sensors (attached to patients) allows health care providers to be alerted in case of any problem or difficult situation [9].

**3.1.1. Security and Privacy issues in HealthCare:** As healthcare industry is growing, so are the security and privacy concerns with it. Following information need to be consider when we talk about security and privacy issues related to it:

**a.** Big Data is a collection of large and complex datasets and getting adopted in the healthcare significantly, security and privacy issues in healthcare becomes necessary to deal with. Most healthcare data centers are HIPPA certified, though this certification does not guarantee patient's record safety as HIPPA is more focused on ensuring security policies and procedures than implementing them.

[HIPPA (Health Insurance Portability and Accountability Act): It is the federal Health Insurance Portability and Accountability Act of 1996. The primary goal of the law is to make it easier for people to keep health insurance, protect the confidentiality and security of healthcare information and help the healthcare industry control administrative costs.]

**b.** A study on patient privacy and data security showed that 94% of hospitals had at least one security breach in the past two years. In most cases, the attacks were from an insider rather than external [10].

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

The wider acceptance and use of Big Data across the world has given many dimensions to the real-time monitoring and it has given opportunities to get connected anywhere, anytime, with almost anything in near future. Below mention information must be considered for the future of Big Data in healthcare:

**a.** Big Data has become an emerging force for the growth of IOT. Gartner estimates 26 billion IOT devices will be functional by 2020 and the amount of traffic generated by such devices will be large enough to place it in the category of big data [11].

**b.** In addition, with the introduction of BSN (Body Sensor Networks) and their direct application to healthcare industry, care providers will be able to monitor vital parameters, medication effectiveness, and predict an epidemic. Body sensors generate massive amount of data, and linking such healthcare data from disparate resource-constrained networks will be crucial for driving healthcare analytics. Hence, healthcare providers have enormous opportunities to revolutionize healthcare by harnessing the power of Big Data [12].

**c.** The McKinsey Global Institute estimates a $100 billion increase in profits annually, if
Big Data strategies are leveraged to the fullest potential [13].

**3.2. Social Media:** Social media is the media (content) that we upload or download – whether that is a blog, video, audio, slideshows, podcast, newsletter or an eBook etc. Public social media are for the public and people share their information with each other. This information could be in the form of text, images, audios, and videos.

**3.2.1 Security and Privacy issues in Social Media:** Public social media has given rise to the vast amount of data. This data is shareable and is in various forms. This has increased the security and privacy concerns with such vast amount of data. Following information gives an overview of security and privacy issues involved in Big Data environment:

**a.** The amount of social media in terms of text, images, audios and videos, etc, is growing too fast and the way it is increasing, it seems that there is no end to this ongoing trend due to which it will be difficult to secure the personal data and privacy of the personal data will be too difficult to handle.

**b.** When looking at privacy issues in social media in Big Data, there is a need to differentiate which of the many Big Data applications domains are being discussed. As the traditional Big Data applications such as astronomy and other e-sciences usually operate on non-personal information and as such do not have privacy issues because this kind of information is not of personal relevance [14]. The privacy critical Big Data applications exist in the new domains of social web among which datasets of social media is important from the security and privacy point of view.

**3.3. Internet of Things (IOT) Era:** The IOT is a recent communication paradigm that envisions a near future, in which the objects of everyday life will be equipped with microcontrollers, transceivers for digital communication, and suitable protocol stacks that will make them able to communicate with one another and with the users, becoming an integral part of the Internet [15].

**3.3.1 Security and Privacy issues in IOT:** IOT is a network of networks in which a massive number of objects, sensor devices are connected through the ICT (Information Communications Technology) infrastructure to provide value added services. The Internet of Things connects people and things anytime, anyplace with anything and anyone; ideally using any path or network and service [16]. The security and privacy issues in IOT can be defined and understood as follows:

**a.** Online service consumers are aware that when they use free online services (such as emails, social networking websites, newsfeeds, etc.), they automatically become data sources for the businesses, which can analyze their data to improve customer satisfaction. And the worse thing is that this data of the online consumers can be sold to the any third party for further analysis without the concern of the consumers of these online services [17].

**b.** In the IOT era, the amount of user data that can be collected will be significantly higher than in the past. For example, recent wearable technologies such as Google Glass, Apple iWatch, Google Fit, Apple Health Kit, and Apple Home Kit can collect sensitive information about users ranging from their health conditions to financial status by observing or recording their daily activities [18].

**3.4 Social Networks:** Social networking has emerged as a new era of communication, whether near or far, anyone can be connected through social networks to anyone they want to. People share images videos, audios, text, etc. almost on daily basis and this causes the need for Big Data in social networks as the data generated by the social networking sites and other social networks can reach up to many gigabytes in volume daily and this data has to be handled with extreme care. Facebook, Twitter, Google, Mobile Phone Companies, Retail Chains, and Government agencies are the best examples of social media where people share their information.

**3.4.1. Security and Privacy issues in Social Networks:**
Social networking sites provide the authority to its users to use the privacy settings so that a user can set privacy as per his requirements from the privacy settings that are provided by the particular social networking site the user is using. For example, Facebook provides so many options for privacy settings like ‘Who can see my stuff?’, ‘Who can contact me’, ‘How do I stop someone from bothering me?’ etc., that a user can use to make his profile private and secure as per his needs. Following issues are of major concern:

**a.** Privacy concern here is that these privacy settings are at the user’s end, what about the other end where the social networking sites are handled and developed? Does a user get privacy at that other end as well?

**b.** A social network user uploads images, videos, audios, text, etc with those whom he wants to get connected. Is it actually safe from the perspective of privacy in such a large datasets of Big Data? As social networks are getting more popular day by day, it has given rise to new classes of security and privacy concerns in the Big Data era.

To face and treat the new challenges in Big Data that has been developed and getting developed in near future, new approaches that must involve the current and future aspects of social and technological solutions will be required. Multiple theories in research have been taken into consideration for observing privacy related information that is available online. Below mentioned are the work that has been adopted for security and privacy concerns:

To represent private sphere of the users and automatically confine the access, privacy bubbles were used as the original boundary between users to share the online pictures [19]. To protect the privacy in social networking sites privacy measure is used as protective behavior and antecedents were used to enable privacy such as self' efficacy, gender, perceived vulnerability [20].

**Conclusion and future directions**

In this paper we have reviewed security and privacy issues in different domain of Big Data. It has also been mentioned that what are the different sources of datasets that constitutes the Big Data. Security and privacy issues related to healthcare, social media, IOT era and social network has been taken into consideration for review. In future we will be reviewing privacy and security concerns in various domains of Big Data as they emerge, since time to time review of security and privacy issues help understand the broader aspect of Big Data in technological advancement that almost everyone will be a part of.

**References:**

1. Workshop Report: Big Data Security And Privacy Sponsored by the National Science Foundation, , The University of Texas at Dallas, September 16-17, 2014
2. Big Data: The End of Privacy or a New Beginning? By Ira S. Rubinstein, January 25, 2013
3. U M Fayyad, 'From Data Mining to Knowledge Discovery, An Overview,' in U M Fayyad, Advances in Knowledge Discovery and Data Mining 6 (Menlo Park: AAAI, 1996), cited in Tal Z Zarsky, 'Mine Your Own Business! Making the Case for the Implications of the Data Mining of Personal Information in the Forum of Public Opinion' (2003) 5 Yale Journal of Law and Technology.
4. McKinsey Global Institute, 'Big Data: The Next Frontier for Innovation, Competition, and Productivity' May 1, 2011.
5. John Markoff ('How Many Computers to Identify a Cat? 16,000' NY Times B1, June 26 2012
6. ShashiRekha .H., Dr. Chetana Prakash, Kavitha G. "Understanding Trust and Privacy of Big Data in Social Networks : A Brief Review ", 2014
7. Richard L Villars, Mathew East Wood and Carl W. Olofson "Big Data what it is and why you should care. June 2011.
8. Won Kim, Ok-ran jeong, Sang –won lee. "On social websites". Sept 2009.
9. Harsh Kupwade Patil and Ravi Seshadri, 'Big data security and privacy issues in healthcare', 2014
10. P. Institute, "Third Annual Benchmark Study on Patient Privacy and Data Security," Ponemon Institute LLC, 2012.
11. P. Middleton, P. Kjeldsen and J. Tully, "Forecast: The Internet of Things, Worldwide," Gartner, 2013.
12. 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.
13. P. Groves, B. Kayyali, D. Knott and S. V. Kuiken, "The 'big data' revolution in healthcare," McKinsey & Company, 2013.
14. Boyd and K. Crawford. Six Provocations for Big Data. SSRN eLibrary, 2011.
15. 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
16. 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
17. 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
18. 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.
19. 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.
20. Norshidah Mohamed, Ili Hawa Ahmad "Information Privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia