

Literature Analysis and Survey about State of the Art Privacy and Security Issues of BigData

P.Nagaprasannalakshmi, M.Tech, Student of Computer Science and Engineering ,Vasireddy Venkatadri Institute of Technology,Nambur.

R.Ashok, M.Tech, Assistant Professor, Computer Science and Engineering,Vasireddy Venkatadri Institute of Technology,Nambur.

Abstract: Since a decade, the concept big data has emerged as a warm research subject matter. The growing amount of massive data additionally increases the risk of breaching the privacy of people. Since massive information calls for excessive computational power and large storage, distributed structures are used. Due to the involvement of different users in these systems, the risk of privacy violation is extended. There have been a number of privacy-preserving mechanisms advanced for privacy protection at special ranges of a massive facts existence cycle. The intention of this paper is to offer a complete assessment of the privacy renovation mechanisms in bigdata and problems identified in existing mechanisms. Specifically, in this paper we illustrate the infrastructure of big data and the state-of-the-art privacy-maintaining mechanisms in every stage of the large information existence cycle. Furthermore, we present challenges and future research directions associated with privacy protection in huge facts.

Keywords: Privacy and security, Bigdata, Threat Prevention, Secured Access

I. Introduction

Throughout the most recent couple of years, data has turned out to be a prominent amongst the most critical resources for organizations of all emerging fields. In addition to the fact that, data is critical for organizations identified with the software engineering industry, yet additionally for associations, for example, governments, healthcare, education, or the engineering sector. Data is the fundamental relating to doing their everyday operations, and furthermore helping the organizations' administration to accomplish their objectives and settle on the best choices based on the data extricated from them [1]. It is evaluated that of the amount of data which is recorded world's history, 90 percent has been made over the most recent couple of years. In 2003, total 5 hexabytes

of data generated by people, whereas this measure of data is now a days, generated in 2 days.

This tendency toward growing the volume and detail of the statistics that is amassed through organizations will not exchange within the close to destiny, because the dramatic improvement of social networks, multimedia, and the internet of things (IoT) is producing an amazing flow of information [3]. We're residing inside the era of BigData as shown in figure-1. Furthermore, this information is frequently unstructured, signifying that conventional systems are not able to analyzing it. Enterprise systems are willing to extract extra beneficial records from this excessive quantity and form of data [4]. There is a need of new analysis paradigm with which is used to analyze and understand the statistics is, consequently emerged in an effort to reap not only private, however also for public sectors data, called as BigData.



Fig.1. Bigdata generation from different real life sources

Since we will focus on the security and privacy issues [2, 4], we are willing to know what they mean. Security means existing in a safe position without any attacks or threats on data, whereas Privacy is nothing but protecting the data from external unauthorized access or attention. In this article 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.

II. Related Work

Today famous Internet organizations like Google, Facebook, Amazon, eBay, Microsoft, and Yahoo!— are occupied with Big Data[3, 4] in some frame and regard data as a important resource and wellspring of value generator. Google is a particularly decent case as it depends on the accessibility of the data it gathers from its own administrations to subsidize its operations as well as to prepare its own data processing algorithms and to feed their process mechanisms, for example, voice recognition, translation, and location-based services. However, big data has a lot of bigger and wider pool of organizations than these big corporations solely. it's been extended to any company and government agencies that rely upon datasets of huge knowledge for statistical algorithms and totally different data mining techniques to research these massive datasets and ultimately up decision making and enhancing potency to require higher selections.

BigData Trends: Today Bigdata is available from different environments like Media/Entertainment, Healthcare, video surveillance, transportation, logistics, telecommunications, retail environment, sensors etc.

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.

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.

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].

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

III. Security and Privacy issues

Bigdata is being tremendously used with its large amount of datasets through different data sources. It is usually giving rise to the security and privacy considerations in special domains. As significant information involves knowledge and knowledge so that it will be used for special purposes, the protection and privacy of a man or woman is at threat. Here in this part we in short talk about in regards to the security and privacy disorders in various domains of massive data.

*HealthCare:*As healthcare industry is growing, so are the privacy and security considerations with it. BigData is a collection of huge and tricky datasets and getting adopted in the healthcare tremendously, security and privacy problems in healthcare turns into essential to care for.

Most healthcare data facilities are HIPPA certified, although this certification does not assure sufferer's report protection as HIPPA is extra excited about guaranteeing security insurance policies and procedures than enforcing them. The foremost purpose of the legislation is to make it simpler for individuals to maintain health insurance, preserve the confidentiality and safety of healthcare information and support the healthcare enterprise control administrative costs.

Social Media: Social media is the media (content) that we add or download, whether or not that could be a web publication, video, audio, slideshows, podcast, newsletter or an e-book and so on. Public social media are for the general public and individuals share their know-how with each different. This expertise might be in the form of textual content, pictures, audios, and movies.

When observing privacy problems in social media in big data [4], there's a necessity to differentiate that of the numerous Big data applications domains area unit being mentioned. because the ancient big data applications like physics and alternative e-sciences typically operate non-personal info and intrinsically don't have privacy problems because this sort of knowledge isn't of private connection. The privacy important big data applications exist within the new domains of social net among that datasets of social media is vital from the safety and privacy purpose of read.

IOT : The IOT [6] could be a recent communication paradigm that envisions a close to future, during which the objects of daily life are equipped with microcontrollers, transceivers for data communication, and appropriate protocol stacks that may build them ready to communicate with each other and with the users, changing into associate degree integral a part of the web.



Fig.2. Bigdata security and privacy concepts

As per NIST [1, 5] the below are the top threats for security and privacy of big data processing applications are specified in figure 2. This section lists and explains about the same below:

Architecture: In order to assure the security and privacy the design of system architecture is an important phase of big data application development. Identifying the vulnerabilities occurring locations and mitigating them is possible at that time of architecture design. Apart from this many other security issues we can tackle at the time of architecture design

Identity and Access: User identity and authorization plays a vital role in protecting privacy and security of bigdata applications. Several proved encryption mechanism helps at this time to design the application secure and robust. Vulnerabilities at this stage of development leads to unauthorized access and data tampering.

Availability: Availability of an application and its resources for processing the user requests is called as the uptime value. To maintain the SLA agreed uptime, system should follow through security principles and to be implemented with application. Any violations of security

principles lead to reduce the uptime value which affects the availability.

Incident response : Incident response is a composed way to deal with tending to and dealing with the fallout of a security break or cyberattack [6], otherwise called an IT incident, PC incident, or security incident. The objective is to deal with the circumstance in a way that breaking points harm and lessens recuperation time and expenses.

Data Protection: Data protection is that the method of safeguarding vital data from corruption, compromise or loss. The importance of information protection will increase because the quantity of information created and hold on continues to grow at unprecedented rates. There's conjointly very little tolerance for period of time which will create it not possible to access vital data.

Data governance: Data governance (DG)[7, 8] is that the overall management of the provision, usability, integrity AND security of knowledge employed in an enterprise. A sound information governance program includes an establishment or council, an outlined set of procedures and a concept to execute those procedures.

IV. Conclusion

In this paper we have explored security and protection issues in various locations of Big Data. It has additionally been said that “what are the diverse wellsprings of datasets that constitutes the Big Data”. Security and protection issues identified with human services, web-based social networking, IOT time and informal organization has been taken into thought for consideration. In future we will audit protection and security worries in different spaces of Big Information as they develop, since time to time survey of security and protection issues help comprehend the more extensive part of Big Data

V. References

- [1]. Stephen Northcutt, Karen Frederick, Scott Winters, and Lenny Zeltser. Inside Network Perimeter Security: The Definitive Guide to Firewalls, VPNs, Routers, and Intrusion Detection Systems. Sams; 2 edition, 2005.
- [2]. M. V. Ramana Murthy, P. Ram Kumar, E. Devender Rao, A. C. Sharma, S. Rajender, and S. Rambabu. "Performance of the Network Intrusion Detection Systems." IJCSNS International Journal of Computer Science and Network Security, VOL.9, No.10, 2009: 198.
- [3]. R. Bird, I. Gopal, A. Herzberg, P. A. Janson, S. Kuttan, R. Molva, and M. Yung,. "Systematic design of a family of attack-resistant authentication protocols." IEEE Journal on Selected Areas in Communications, Vol. 11, Issue 5, 1993 : 679.
- [4]. O. Dain and R. Cunningham," Fusing a heterogeneous alert stream into scenarios", In Proc of the 2001 ACM Workshop on Data Mining for Security Applications, (Nov. 2001), pp. 1-13.
- [5]. Mayer-Schönberger, V.; Cukier, K. Big Data: A Revolution that Will Transform How We Live, Work, and Think; Houghton Mifflin Harcourt: Boston, MA, USA, 2013.
- [6]. Sagioglu, S.; Sinanc, D. Big data: A review. In Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 20–24 May 2013; pp. 42–47.
- [7]. Hashem, I.A.T.; Yaqoob, I.; Anuar, N.B.; Mokhtar, S.; Gani, A.; Ullah Khan, S. The rise of “big data” on cloud computing: Review and

open research issues. *Inf. Syst.* 2015, 47, 98–115.

- [8]. Sharma, S. Rise of Big Data and related issues. In *Proceedings of the 2015 Annual IEEE India Conference (INDICON)*, New Delhi, India, 17–20 December 2015; pp. 1–6.