Big Data: Privacy and Security Aspects

Shraddha Ovale[1], Vallabh Kulkarni[2], Shreya Patil[3], Priyanka Nalla[4], Pranav Patil[5]

[1]*Assistant Professor, Dept of Computer Engineering, PCCOE, Pune, Maharashtra, India*
[2,3,4,5]*Dept of Computer Engineering, PCCOE, Pune, Maharashtra, India*

---***---

Abstract - *In the ever-growing era of technology and commercialization, data has been a very important research topic for techno-savvy people. 'Data' is something that is processed by one party and provided by another party. On a commercial level, companies and consumers are the two parties in which the exchange of data occurs as per the benefits and requirements of the company and the consumer. Hence, on different stages and scales, various transactions and exchanges of data lead to concern about the privacy and security of the data. At the company level, a lot of data having complex information, high variety, increased volume, and having high velocity is exchanged among multiple parties. This data is called 'Big Data '. Such huge and complex data makes the exchange to be complicated and the preservation of privacy and security of big data requires various mechanisms. This review paper puts light on the various concerns about the protection of privacy and security of big data. The infrastructure, the mechanisms of protection, and the cycle of big data are discussed in this paper.*
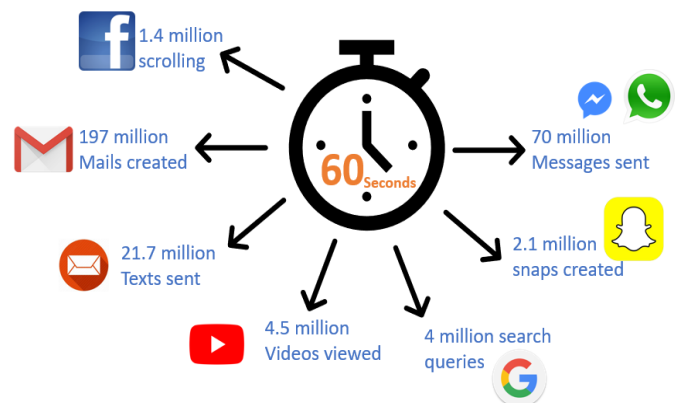
*Key Words***:   Big data; Privacy; Security; Information Security: Mechanisms; Security Protection; Infrastructure; the cycle of big data; Data Analysis; Social Applications.**
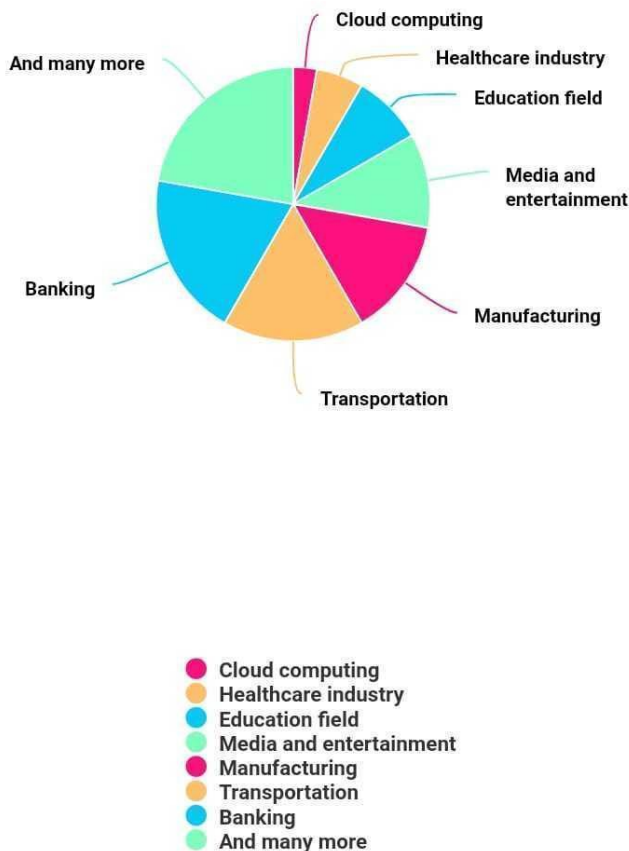
1. INTRODUCTION

In the last few years, data has become a prime significant factor for businesses, companies, governments, the medical sector, engineering sector to name a few. Data is an essential asset for these companies and organizations to carry out their business plans, marketing strategies and to make decisions. A single smartphone user generates approximately 40 Exabytes of data every month. Existence of this data is in the form of pictures, emails, music and songs, texts, videos, searches, etc. The number of users of smartphones is in the millions and thus data generated is quite a lot of traditional computing systems to process and analyze it. This collection of large and complex sets of data is termed "Big Data". As we live in the era of big data, we need to go through all the important aspects of big data. So how do we classify a set of data as Big Data? This can be explained with velocity, volume, variety, value and veracity. Big data is super beneficial once it is processed and analyzed properly.

From a privacy and security perspective, a glance is to be taken on the fact that, for marketing purposes, making   business   strategies   and   research;   many companies use data collected from consumers but they may not provide a future economic profit to consumers on a significant level. Instead this collected data may be misused to track the irrelative history of a consumer. Many businesses, companies and organizations use petabytes of data about their customers and company. These companies use technology to store and analyze data. Classification of this information (extracted from analyzed data) becomes even more critical. These issues of big data are related not only to variety and volume but also to security and privacy. A balance between data privacy and information needs is to be maintained. Hence, this review paper focuses on Big data - privacy and security solutions.

APPLICATIONS OF BIG DATA

Cloud computing
Healthcare industry
Education field
Media and entertainment
Manufacturing
Transportation
Banking
And many more

2. LITERATURE REVIEW

Paper[1] written by Mr. Apparao et al. proposed a method of data privacy and protection in transaction logs and threats in stealing the user credentials. Also, the paper had proposed and explained some newer methodologies to prevent privacy in the transaction field. Big data is not structured and it is very big in volume that it is a tough process to store such a huge amount of data. So, one technology named Robust technology comes into the picture which helps to store data and gives tools which analyze it. To increase privacy in big data, especially in transaction logs, a mechanism called the auto-tier mechanism is explained.

Paper[2] i.e. journal written by S. K. Saravanan et al. explained the concept of the use of encryption and decryption techniques to prevent the security and privacy

aspects of Big Data information. Some technologies pose a serious threat to one's information like Data mining. Encrypting and decrypting the data makes it unreadable to third parties and the data is stored safely. But, to emphasize the security in big data, 'SHA3' algorithms of encryption and decryption are used.

In paper[3], B.B Gupta et al. proposed security and privacy issues regarding the information on an online platform, log-in into apps, and cloud computing. Cyber threat is considered a major task today because the issue is being faced by all countries whether it is developed or developing. On multimedia, users share most of the memories of their life and the explosion and use of that information wrongly creates distrust among users regarding social media. So, it is a very challenging task that is taken into consideration because Big data is not only huge in volume but also it is unstructured.

In paper[4], Julio Moreno et al. explained problems and challenges related to security issues in Big Data. The author further stated that security issues were much more significant when big data was introduced. Many researchers, therefore, focused on their research on solutions for security and privacy issues. Moreover, the protection of the Hadoop system is of prime significance for researchers. If a proper framework on security in big data is carried out by the government, there will be a rapid spread of big data technology. Big Data technology will be useful for the basic development of upcoming technology.

In Paper[5], Jose Moura et al. explained the infrastructure security in big data considering the assured modeled processing of data and provided the mechanisms regarding security for databases that are not in the specific storage pattern. The author stated about the aspects of data analysis through data mining, granular access control, granular audits, and data provenance. The author also proposed 'cryptographic' solutions regarding storing data in a secure and private manner. Various case studies in secure social apps are taken into account in the research paper. A large capacity and highly distributed architecture will be required, exposed to a hostile environment.

The present state of the legal rules and regulations regarding the security and privacy of big data is proposed by Nils Gruschka et al. in their research paper[6] by analyzing various data conservation and privacy-oriented mechanisms. In addition, for complying with the 'Data regulation laws', the author presented and analyzed two authentic research projects as case studies dealing with sensitive data and actions. The research paper proposed employed techniques under the legal requirements for protecting privacy, and the influence of these techniques on the data processing phase. This paper also explained how privacy preservation can be applied and in what manner the data protection laws are implied.

Nandhini. P, in her research paper[7], depicts the various security and privacy protection methodologies proposed by the various researchers and analyses the merits and demerits of those methodologies. The security and handling of big data are also considered with algorithms and techniques to do so. Challenges with the security and privacy of transaction data, stored data, and cloud computing are discussed with the help of surveys and other research material. She concluded that there is a need to consider new technologies and modifications in the available technologies to make the data secure and to provide privacy.

In paper [8], the author Venkata Narasimha Inukollu has discussed cloud computing, the Hadoop environment, and Big data. The author has conversed about security concerns about big data in cloud computing. The author broached the issues and provocations at various levels, namely: the network, validation, data, and generic level. The possible solutions for the issues and challenges allied with cloud computing security and Hadoop are also mentioned. The role played by cloud computing in sectors of data protection, its applications as well as the related infrastructure by assistance of policies, technologies, controls, and big data tools is also mooted. Measures and proposed approaches such as encryption of files and network, logging, software format and maintenance of node, authentication of node, honeypot nodes, acrid system testing of map-reduce jobs, and framework for persuading cloud can be implemented to improvise the security issues in cloud computing.

The author 'Kudakwashe Zvarevashe' of the paper [9] has emphasized the subjects such as big data, MongoDB, Hadoop, Kerberos, NoSQL, and social networks. The paper pinpoints the list of directions of research from 2009 to 2014, which includes the topics such as security matters generated during the inventions of social networks, smart devices like smartphones, etc., and other technological advancements. The paper highlights the major security problems linked with Big Data. Using nodes like mapper nodes, reducer nodes. The use of encryption with Kerberos protocol makes the data more secure. However, security and privacy issues are faced in different forms such that Kerberos might not be enough to fully secure the data.

The paper written by Dongpo Zhang [10] tells us about the impact, dependence, and importance of big data in the daily lives of human beings. It elaborates the need for strengthening the privacy and security of huge data transfers, to protect and secure the privacy. Big data is essential for ordinary people and very necessary in industrial, economical, and commercial sectors. This article states how to ensure the security of big data and the protection of privacy of lives. One of the most important aspect mentioned in this research paper is that to ensure the authenticity and reliability of data.

According to author, the citizens of a country should also be made aware so that the privacy and security is protected.

Paper Id	Related Work	Mechanism/ Algorithm	Results	Advantages	Disadvantages
[1]	Cloud security alliance membership and surveyed security practical oriented interviewed by them to get a basic idea of high-security issues	Auto-tiering mechanism proposed	Proposed more ways to secure data from hacking	Proposed new mechanisms and techniques	Auto-tiering mechanism is not explained in detail
[2]	Explained use of encryption and decryption to prevent security	SHA3 algorithm	Proposed encryption and decryption technique	Data mining is explained in brief	PPDM is not explained in brief
[3]	Paper explains mainly on multimedia explosion and difficulty in handling it	---	Cyber threat is considered a major issue faced by countries.	Explains well about cyber threat	Explanation is very basic
[4]	Explains how Hadoop technique is big concern for researchers	---	Displays all possible problems in Big data.	Explained well about threats and problems related to security issues related to Big data .	Does not provide solutions to threat
[5]	The infrastructure security in big	Case Study	A high capacity and highly distribut	Cryptographic solutions are propose	Implementation of solutions is not

Ref					
	data considering the assured modeled processing of data.		ed architecture will be necessary when exposed to a hostile environment.	d.	discussed.
[6]	The present state of the legal regulations regarding big data security is proposed.	Data regulation	Data Regulation Laws should be considered when dealing with security and privacy of big data.	Proposed employed techniques under the legal requirements for protecting privacy.	Legal Actions on violation of laws is not explained.
[7]	Study of various security and privacy methodologies and their merits and demerits	Survey Study	Modifications required in the available technology	Challenges with the security and privacy of transaction data, stored data, and cloud computing are discussed.	The holistic Framework is not explained in detail.
[8]	Research about data storage on clouds, the Hadoop environment and the security issues in cloud computing.	---	Cloud environments can be secured for complex business operations by applications of proposed approaches.	Data Security and Privacy issues at various levels are discussed.	Third-Party involving is not discussed thoroughly.
[9]	Security issues generated during the inventions of social network	---	Use of encryption with Kerberos protocol makes the data more secure.	Concept of encryption is introduced to protect data hacking.	Kerberos may not be sufficient for fully securing the data.
	s, smart devices like smartphones, and other technological advancements.				
[10]	The need for strengthening the privacy and security of huge data transfers, to protect and secure the privacy.	Awareness of common people	The citizens of a country should be made aware about data hacking along with other measures.	Focus is on authenticity and reliability of data and common people.	Full supervision is proposed which may be violation of privacy itself.

3. CONCLUSION

This review paper presented a view about various papers which include various security and privacy aspects. Not only big data in mobiles and multimedia but also in cloud storage of data. Also, the security aspects of mobiles and cloud-based data transfers are taken into account. It is helpful in knowing about many open questions related to security. The huge data transactions and mining as well as the significance of big data in the networks encourages researchers to look into various security and privacy problems that are essential to be resolved crucially by developing efficient and effective solutions. Massive work is essential prior to this solution getting quixotically taken on. We hope the papers dealt with in this review paper will be a good provision to the trends and issues coming to the picture regarding the security and privacy of big data in transfers and computing. Modifications and research are still required to set the seal on the security and privacy of big data. The current ways and mechanisms to protect data are insufficient to handle big data and will fail to protect it from misuse. Thus finding out the solutions to current problems and making big data transactions more secure and confidential is a need of technology.

REFERENCES

[1] Mr. Yannam Apparao, Mrs. Kadiyala Laxminarayanamma. "Security Issue on "Secure Data Storage and Transaction Logs" In Big Data ".

[2] SHASHANK -PG Student II year MCA S.K.Saravanan, Assistant Professor (Sel.Gr) G.Rekha, Assistant Professor (Sen.Gr). "Information Security in Big Data using Encryption and Decryption".

[3] B. B. Gupta1 & Shingo Yamaguchi 2 & Dharma P. Agrawal 3."Advances in Security and Privacy of Multimedia Big Data in Mobile and Cloud Computing ".

[4] Julio Moreno *, Manuel A. Serrano and Eduardo Fernández-Medina. "Main Issues in Big Data Security".

[5] Jose Maura, Carlos Serrao. "Security and Privacy Issues of Big Data".

[6] Nils Gruschka† , Vasileios Mavroeidis† , Kamer Vishi† , Meiko Jensen∗ †Research Group of Information and Cyber Security, University of Oslo, Norway. "Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR".

[7] Nandhini. P. "A Research on Big Data Analytics Security and Privacy in Cloud, Data Mining, Hadoop and MapReduce".

[8] Venkata Narasimha Inukollu1, Sailaja Arsi1, and Srinivasa Rao Ravuri3, "SECURITY ISSUES ASSOCIATED WITH BIG DATA IN CLOUD COMPUTING".

[9] Kudakwashe Zvarevashe1, Mainford Mutandavari2, Trust Gotora3, "A Survey of the Security Use Cases in Big Data" from International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 5, May 2014.

[10] Dongpo Zhang, "Big Data Security and Privacy Protection" from Advances in Computer Science Research, volume 77.