Asian Journal of Computer Science and Technology
ISSN: 2249-0701 (P) Vol.8 No.S1, 2019, pp.33-35
© The Research Publication, www.trp.org.in
DOI: https://doi.org/10.51983/ajcst-2019.8.S1.1972

Survey of Security and Privacy Issues in Big Data Analytics

G. A. Mylavathi[1], N. M. Mallika[2] and K. Mohanraj[3]
[1]Assistant Professor, Department of Computer Science, Gobi Arts & Science College, Tamil Nadu, India
[2]Dean of Science & Assistant Professor in Computer Science, Sri Vasavi College, Tamil Nadu, India
[3]Research Scholar, Sri Vasavi College, Tamil Nadu, India
E-Mail: nm.mallika@gmail.com, kmohamca@gmail.com

(Received 2 December 2018; Accepted 22 January 2019; Available online 1 February 2019)

***Abstract* - Due to the reasons such as the rapid growth and spread of network services, mobile devices, and online users on the Internet leading to a remarkable increase in the amount of data. Almost each trade is making an attempt to address this large information. Big data phenomenon has begun to gain importance. However, it's not solely terribly tough to store massive information and analyses them with ancient applications, however conjointly it's difficult privacy and security issues. For this reason, this paper discusses the massive information, its scheme, considerations on massive information and presents comparative read of massive information privacy and security approaches in literature in terms of infrastructure, application, and data. By grouping these applications associate overall perspective of security and privacy problems in massive information is usually recommended.**

***Keywords:* Big data, Hadoop Security, Cloud Security Monitoring Auditing, Key Management Anonymization**

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 S. Marchal, J. Xiuyan, R. State, T. Engel, [1]. It is predicted that the data will increase 300 times, from 130 Exabyte's in 2005 to 40,000 Exabyte's in 2020 L. Liu, J. Lin, [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 A. Gupta, A. Verma, P. Kalra, L. Kumar, [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 A. Gupta, A. Verma, P. Kalra, L. Kumar, [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 L. Chang Liu, R. Ranjan, Y. Chi, Z. Xuyun, W. Lizhe, C. Jinjun, [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 H. Chingfang, Z. Bing, Z. Maoyuan, [5]. Identifying the sources of problems will result in more efficient use of big data. For this reason, this paper examines and classifies studies on security and privacy breaches and solutions in big data. This perspective would lead to an understanding of important research areas and the development of new methods. In addition, the use of big data in analysis would make the systems become safer. Section II presents a brief summary of big data. Section III contains categorization of big data concerning security and privacy studies in literature. The results obtained with security and privacy issues in big data are discussed in Section IV, and section V explains how to use big data to maintain security. Finally the conclusion highlights the importance and requirements to secure big data communication.

II. DEFINITION AND CHARACTERISTICS OF BIG DATA

Big knowledge refers to massive and complicated datasets that typical software system is insufficient for managing. There area unit numerous explanations of massive knowledge via Vs.5Vs are typically used to characterize of Big Data as volume, velocity, variety, veracity and value. Volume is that the size of knowledge; speed is that the high speed of data; selection indicates heterogeneous data sorts and sources; truthfulness describes consistency and trustworthy of data; and worth provides outputs for gains from large data sets. Identifying characteristics of the information is useful in extracting its hidden patterns. Big data is classified into ten categories in terms of data type, data format, data source, data consumer, data usage, data analysis, data store, data frequency, data processing propose, and data processing method .

A. Big Data Security and Privacy Approaches in Literature

Traditional solutions square measure deficient once coping with huge knowledge to confirm security and privacy. Encryption schemes, access permissions, firewalls, transport layer security is broken; cradle of knowledge of knowledge of information is unknown; even anonymised data is re-identified. For these reasons, advanced techniques and technologies square measure developed to guard, monitor and audit huge knowledge processes in terms of infrastructure, application and knowledge. Considering the connected literature, this paper has classified security and privacy problems for giant knowledge underneath five titles as Hadoop security, cloud security, watching and auditing, key management andanonymization.

B. Hadoop Security

Hadoop may be a distributed method framework and it had been not originally developed for security. It was meant to operate in trusted environments. As Hadoop has become a preferred platform, security precautions have began to be developed. In addition, it has started to receive academic interest.

When developing a Hadoop system that guarantees security and privacy of information on the cloud, two techniques were proposed to prevent a hacker who wants to get all data in cloud. A trust mechanism has been implemented between user and name node which is component of HDFS and manages data nodes. According to this mechanism user should evidence himself to access name node. Firstly, user sends hash function then name node produces hash function too and it compares these two generated functions.

If compare results correct, accessing system is provided. In this step, SHA-256 which is one of the hashing techniques isused for authentication. Random secret writing techniques like RSA, Rijndael, AES and RC6 has been also used on data in order that a hacker does not gain an access whole data. Map Reduce is executed encryption/decryption process in this approach. Finally, these 2 techniques area unit tested mistreatment twitter stream for indicating the way to maintain of security problems.

Another unit that cause the security weakness is Hadoop Distributed File System (HDFS).Three strategies to extend HDFS security has been developed. In order to attain authentication issue, Kerberos mechanism based on Ticket Granting Ticket or Service Ticket have been used as first method. To handle name node problems as final method, two name node shave been proposed: one of them is master and the other is slave. If something happened to master node, administrator gives data from slave name node on condition that Name Node Security Enhance (NNSE) permission. Therefore, latency and Knowledge availableness issues succeeded in secure method.

C. Cloud Security

The widespread use of cloud computing for such reasons as broad network access, on-demand service, resource pooling and being elastic have made a proper environment for big data. However, cloud hosts traditional threats and new attacks.

Data storage on clouds is one of the main problems nowadays. Therefore, some precautions must be taken by the service provider. Because of this, a secure way to handle and share big data on cloud platform has been presented. Data has been encrypted and compressed to prevent security issues. It also takes precautions in case of a natural disaster and uses three backup servers for this purpose.

D. Big Data Analytics for Security

Big data analytics aims to obtain beneficial information from large scale and complicated data. The increase of stored or streamed data and development of analysis systems has led to using these activities in information security. The anomaly detection, intrusion detection, fraud detection, advanced persistent threats (APT) detection, and forensics from big data has been accomplished by examining the logs, system events, network traffic, website traffic, security information and event management (SIEM) alerts, cyber-attack patterns, business processes and other information sources . To detect these attacks, large volume and variety of data is accumulating and associate with network history. The advantageous uses of big data, such as performing without deletion of logs after a certain period, running complex queries on large and unstructured datasets, and facilitating human-computer interactions via visual interfaces, for security is becoming quicker and cheaper than traditional methods.

A method to detect malware using big data has been proposed. For this purpose, Large Iterative Multitier Ensemble (LIME) classifiers have been used to handle big data. The performance of LIME classifier with other base classifier was evaluated. The results showed that the presented method performed better than base classifiers. In another study, a framework with big data environment has been suggested such as big data analytic tool and NoSQL database for android application security assessment.

The white-box, black-box and mobile environment forensic approaches have been used to determine security assessment level. Then authors inserted assessment results into CouchDB, which is one of the NoSQL database. Using this database, they tried to discover security issues or visualization with big data analytic tools like Sckit, Matplotlib. Finally, they used SOA controller to publish their results via web service.

III. CONCLUSION

Big data needs extra requirements for security and privacy in data gathering, storing, analyzing, 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] S. Marchal, J. Xiuyan, R. State, T. Engel, “A Big Data Architecture for Large Scale Security Monitoring”, *Big Data (BigData Congress)*, pp. 56 – 63, Anchorage, AK, 2014.

[2] L. Liu, J. Lin, “Some Special Issues of Network Security Monitoring on Big Data Environments”, *Dependable, Autonomic and Secure Computing (DASC)*, pp. 10 – 15, Chengdu, 2013.

[3] A. Gupta, A. Verma, P. Kalra, L. Kumar, “Big Data: A security compliance model”, *IT in Business, Industry and Government (CSIBIG)*, pp. 1 - 5, Indore, 2014.

[4] L. Chang Liu, R. Ranjan, Y. Chi, Z. Xuyun, W. Lizhe, C. Jinjun, “MuR-DPA: Top-Down Levelled Multi-Replica Merkle Hash Tree Based Secure Public Auditing for Dynamic Big Data Storage on Cloud”, *Computers*, Vol. 64, No. 9, pp. 2609 – 2622, 2015.

[5] H. Chingfang, Z. Bing, Z. Maoyuan, “A novel group key transfer for big data security”, *Applied Mathematics and Computation*, Vol. 249, pp. 436–443, 2014.

[6] S. Junggab, K. DongHyun, R. Hussain, O. Heekuck, “Conditional proxy re-encryption for secure big data group sharing in cloud environment”, *Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 541 - 546, Toronto, ON, 2014.

[7] M.R. Islam, M.E. Islam, “An approach to provide security to unstructured Big Data”, *Software, Knowledge, Information Management and Applications (SKIMA)*, Dhaka, pp. 1-5, 2014.

[8] T. Omer, P. Jules, “Big Data for All: Privacy and User Control in the Age of Analytics”, *Northwestern Journal of Technology and Intellectual Property*, Vol. 11, No. 5, 2013.

[9] J. Sedayao, R. Bhardwaj, N. Gorade, “Making Big Data, Privacy, and Anonymization Work Together in the Enterprise: Experiences and Issues”, *Big Data (BigData Congress)*, pp. 601 – 607, Anchorage, AK, 2014.

[10] T. Vijey, A. Aiiad, “Big Data Security Issues Based on Quantum Cryptography and Privacy with Authentication for Mobile Data Center”, *Procedia Computer Science*, Vol. 50, pp. 149–156, 2015.

[11] B. Matturdi, X. Zhou, S. Li, F. Lin, “Big Data security and privacy: A review”, *Big Data, Cloud & Mobile Computing, China Communications* vol.11, No. 14, pp. 135 – 145, 2014.