Research Perspective on Security Based Algorithm in Big Data Concepts

Seema Rai, Ashok Sharma

***Abstract:** Providing a robust security for large data is one in all the first concern for most of the researchers. This paper makes an attempt to uncover all the protection solutions associated with unstructured, structured and semi structured data. also, the main aim of this paper is to cover the information related with the several encryption algorithms used to provide confidentiality, integrity, privacy and data silos. Different algorithmic program and tools play an efficient role in playacting significant analysis on huge volume, variety of big data.*

***Keywords:** Big Data, Confidentiality, Data Integrity, IoT, Privacy, Data silos*

I. INTRODUCTION

Attempting of defining a big data with its characteristics and security services is a primary concern of this paper. This paper documents how security services has provided by different authors to the big data which is coming from various domain such as geoscience, life science, scientific, municipal planner and social domain. A vital contribution in this paper is to focus on the often-ignored area of the big data. The common debate on big data is one that is both influenced and driven by the marketing strategies of massive software as well as hardware developers. The organization of this paper is depicted below: We initiated the paper by describing the big data to show that the scale is not just one dimension. Other attributes, including the rate of frequency when the data are produced, are essential in identifying big data because we know the lack of data protection can result in increased financial losses and damage to the credibility of the company. Therefore, the lines below summarized the situated difficulties posed by the big data:

- Possibility of sensitive information mining.
- Troubles of cryptographic protection
- Struggles of granular access control
- Potential presence of untrusted mappers
- Depending on the size of cloud and big data, its sources are not continuously tracked and monitored.

Some of the objectives, which are to be taken into account, are given below:

- Confidentiality:
- Security: specifies the security level of an algorithm.
- Performance Scrutinization:
- Time complexity:

Revised Manuscript Received on February 05, 2020.

* Correspondence Author

Seema Rai*, Asst. Professor, Sphoorthy Engineering College (SPHN), Hyderabad, Telangana, India, Seema61190@gmail.com

Ashok Sharma, Professor, Department of Computer Science, Lovely Professional University, Jalandhar, Punjab, India.

A. Defining big data

The Big data is a word or phrase, which actually does not mean the huge amount of data in fact it is the term which is used to say, when the input data cannot be stored and cannot be processed. for simplicity we can take the example of Gmail where the capacity of mail can be 25MB and the document is 35MB ,So in such condition 35MB cannot be stored and cannot be processed finally it becomes the big data with respect to Gmail. In this context many authors has given the description of Big Data by giving initially Big data characteristics such as 3V's,4V's and finally 5V's. According to TechAmerica foundation, Big Data describes as massive quantities of high rate, advanced and differential knowledge that need progressive methodologies and strategies to change the distribution, storage, management, analysis of data and capture (TechAmerica Foundation's Federal Big Data Commission,2012)[35].

The effect of big data volumes extends to all aspects. Therefore, there is no criterion for variety, velocity and quantity that describes big data [32]. Big data protection is one of the desired outcomes for large institution, where they can establish the utmost protection and security controlled by the data centre. (Chaowei Yang, 2017)[34] Says that some observations and study have been made to address Big Data in the digital world and related science domains. The wide accessibility of Big Data and computing capacities present social challenges of geospatial significance, and the weaving of technologies turns Big Data into geospatial science, engineering and business values. We cannot deny that there are unlimited potential opportunities available for big data in the field of health care system so (Abouelmehdi, Beni-Hessane and Khaloufi, 2018) [31] has presented some of the issue of confidentiality associated to big data in the context of health care system.

II. CONFIDENTIALITY

To protect the sensitive data against adversary act, a key attribute will be constructed and such a process we will name it as confidentiality, the confidentiality can be provided when the data is in motion or when the data is at rest i.e. when user access any application the channel must be secure by using any encryption algorithm, Accessing the data by intended user using access privileges is also one of the concern in confidentiality.

Most of the researcher use different encryption schemes to provide confidentiality to the sensitive part of big data, we will discuss them gradually to clarify number of encryption algorithm used to secure the data by different researchers.

Concept of cloud computing is widely spread as the data set is increasing need for the storage has also increased, though several protection has been maintain from user to cloud service operator ,sensitive part of the data will still be the question of fear so several frameworks and algorithm has developed and recently on of the paper has discussed about the privacy of the information in cloud i.e. means protecting the data from cloud service operator by using (SA-EDS), AD2 algorithm. According to(Li et al., 2017)[22]using the AD2 algorithm, EDcon algorithm provide the security from major threats in the cloud with less processing time, but data duplication security has not been achieved with this approach

A. Quantum Cryptography

According to (Thayananthan and Albeshri, 2015)[27], symmetric key utilization with a block cipher is more efficient because it is appropriate to supervise the big data safety as the architecture is very simple for the block cipher of the big data. Quantum cryptography offers most fortification with less difficulty which is able to increase the storage capability Complexity invariably will increase we tend to use huge blocks. Here, block cipher mistreatment GA that gives economical key search is one all told the foremost effective QC strategies in big data protection measures. The estimated result shows that there is a likelihood of change in the polarization of photon and also it suffers with lack of concept of digital signature, certified mail.

B. A security framework in G-Hadoop

(Zhao et al., 2014)[26] has used G-Hadoop framework that designs a security model has the power to forestall the foremost popular attacks, like delay, replay, and MITM attack, and guarantee a safe interaction of G-Hadoop over public net- works. In addition, it enacts entirely different methods to shield G-Hadoop's resources from maltreatment or mishandling. In its entirety, it offers the user with a consistent and happy solution to the single-sign-on approach for accessing G-Hadoop. For further enhancement, keys have been built as modifiable to raise the complexity of the attacker's cryptographic analysis. According to (Zhao et al., 2014), in attempting to share or exchange data over several administrative domains, G Hadoop security architecture uses a public key cryptographic algorithm, which is similar to the one utilized by Digital Signature. As it uses the SSL protocol, prospect of slow down connection will be more because it goes for handshaking and both the side process of encryption and decryption procedure will get establish.

C. Security with AES algorithm

Big Data is also a base for cloud computing incorporated along with Internet Of things. For enhancing the security challenges an architecture has been developed by (Stergiou and Psannis, 2017)[36] which is based on security of the network. AES algorithm has been used by the author to provide the security between the user of the network

D Attribute -Based Encryption

Attribute-based encryption (ABE) describes the identity or individuality not atomic but as a set of features. It is a public cryptographic key in which a message is authenticated for a particular recipient utilizing the public key of the receiver. Furthermore, it is inevitable attack by the side channels if ABE is used in open networks. To prevent such attacks (Wang et al., 2017)[29]has develop CP-ABE for leakage resilient scheme. But the processing time taken by ABE is comparatively more than the other encryption.

E. The General Number Field Sieve (GNFS) algorithm

The General Number Field Sieve (GNFS) algorithm was perhaps the best effective integer factoring algorithm which is currently applicable for digits greater than 110, and cloud computing was capable to offer a strong capacity to accomplish the GNFS algorithm. Concentrating on the RSA security research, (Yang et al., 2017)[37] have analyzed GNFS algorithm in the Cloud. More particularly, the latest study on addressing large and sparse linear systems over GF (2) was discussed; it was one among the time-consuming phases of the GNFS algorithm. Subsequently, Laurence T. Yang, et al have suggested a new innovative parallel block Wiedemann Cloud algorithm to lower the communication expenses for addressing large as well as sparse linear systems over GF(2). The proposed parallel block Wiedemann algorithm included improved strip, strip and cyclic partitioning that accelerated the various phases of the algorithm in a parallel manner. In their research, it is seen that the parallel block Wiedemann algorithm can significantly improve the performance of GNFS relative to many other existing conventional algorithms, both in respect of computational complexity and speeding up their conceptual and experimental tests.

F. A dynamic prime number for data stream

Existing conventional symmetric cryptographic security approch for data security have been identified both as a centralized dynamic or a static shared key (Puthal et al., 2017)[30]. A long key has to define to guard the data from potential attacker especially in static shared key. As we remember, the key's length is always directly proportionate to the time of security verification. As a result, it is evident from the desired attributes of big data streams that the security authentication should have been in real-time. For a dynamic key, distributing keys and rekeying or centralizing processor to all or any of the sources can be a little time-consuming operation. Big data sources have always been consistent in nature and large in scale. It causes it difficult to interrupt data for rekeying, source delivery, and DSM synchronization.

To overcome the difficulties, they have constructed and established a Dynamic Prime-Number Based Security Verification (DPBSV) method. The method considers a generally shared key, which is upgraded statically by generating synchronize prime numbers.

Table 1. The Above Table Shows the Category of Encryption Algorithm Used and their Soft Spot Which Makes the Chance to Improve.

Encryption algorithm	Soft spot
AD2 algorithm	Work has to be done in data duplication prevention
Quantum cryptography	Polarization of photon
Public key cryptographic algorithm	Computation overhead
AES algorithm	Algebraic structure is simple
Attribute Based encryption	Coarse-grained level
GNFS algorithm	Sieving part takes more time because it needs to generate independent relation[23]

III. INTERNET OF THINGS

Internet of Things (IoTs) can take the big data to next level by relying on to the database logic. As per the (Atzori, Iera and Morabito, 2010) [38], IoTs were defined by implying that IoT is the concept of daily life material objects being linked to the internet and also being capable of identifying themselves with other systems. The term is closely known with RFID because the methodology of communication, though it conjointly could embrace different detector methods, QR codes or wireless techniques. Furthermore (Alaba et al., 2017) [39] has explained A well-defined confidential and security policy to designed and deployed in order to ensure, and privacy for users and items confidentiality and access control. By predicting security flaws and the absence of uniformity in the IoT world, they suggest a theoretical form of design, which can help eliminate security problems to a large extent.

IV. BIG DATA ANALYSIS

Big data Analysis is generally the typically complicated method of inspecting giant and varied datasets to uncover data together with unknown correlations, client preferences, hidden patterns and market trends that may facilitate organizations build up. If we discuss about the analytics, we can observe that in present we have multiple data analytics tools such as R Programming, Tableau Public, SAS, Apache spark, Excel, Rapid Miner, KNIME, QlikView. In order to reach at current tools and trend several researches has been done I will discuss them gradually and then latest research will be taken up, So we initially start with(Sandryhaila and Moura, 2014)[16]who have developed a paradigm for data analysis based ondigital signal processingon graphs(DSPG) and has implemented with parallelization and vectorization, As we move with the result obtained by the researcher it shows that moving with DSPG is efficient for the data set which contain irregularities and noisy, the soft spot of the research is graph signal behaviour, so if we think of extension of the DSPG it is not straight forward we need to focus on behaviour of signal i.e. (localization, stationarity). The DSPG is has open the way for machine learning concepts for large data set.

If we want to understand in better way, we need to understand first, for which types analytics can be applied let us say first and foremost text analysis [17] which says to extract information from unstructured data using IE [18], text summarization [19], QA system which is based on the speech analysis [20] which chooses two approach transcript-based and phonetic based approach. Scientists are still conducting experiments about how to derive useful knowledge through video streams, the very popular use of video analytics in the protection security and monitoring systems [21].

V. INFORMATION INTEGRITY

Information integrity nothing but assuring that the data being accessed or read has neither been alleviated with, nor been altered or impaired through any of the system fault, since the time of the last authorized access mostly at a higher risk, (Patterson, 2013)[33]stated that, In a Analytics and Big Data system implementations, auditors and others in assurance roles such as(customer services, and financial reporting, disclosures, and risk management processes) can be implicated to assist implementers information integrityas well as build-in process-integrity. Carl lagoze [9] says that a big data can be differentiated from *lots of data* just by maintaining the *control zone*s, A proper control zones can only provide the data integrity and can make the science credible. As mentioned by the carl lagoze about eBird species to maintain the data integrity isarduous because it makes automated sensing process difficult.

Considering efficiency and security as concerned measurement (Liu, C., Yang, C., Zhang, X., & Chen, J. (2015)) [10] has given a study of encryption-based data integrity verification methods on cloud and IoTs data. Liu has given analysis on authenticated data structure which uses MHT [11] and RASL [12] which can authenticate indices of the data block, so depending on these analyses he has suggested Ranked Markle Hash Tree for well grained upgrades. But it is used for only authentication of variable block.

PDP [13] and PoR [14] are the two protocol mainly design for providing the data integrity in cloud storage, according to (Tan et al., 2018)PoR is more useful than the PDP though they have same functionalities because PoR can recover faulty or corrupted outsource data. Many researchers have worked on PoR with varied enhancement to it so that better integrity can be achieved, nevertheless if PoR uses with Symmetric encryption then for sure the security of data will be minimum.

Table 2. The table shows the soft Spot of, Data Integrity and Services offered by the Author

Author	Services	Soft spot	DI
Liu, C., Yang, C., Zhang, X., & Chen, J.	Authenticator-based	rMHT only suitable for authentication of variable block	□
Carl lagoze	Maintaining integrity by proper control zones	Process of automation will be difficult with eBird species dataset	□
Tan et al.	Enhancement in PoR scheme	Failing to work encryption key.	□

As per the (Zharova and Elin, 2017)[41]It is essential to create a "Big Data" rule, that recognizes and defines risks as well as threats. The author also specifies the appropriate degree of protection needed for the analysis of information by using Big Data techniques. In the perspective of the law, the algorithms of separation have identified data types, in the phase of managing them by means of Big Data as well as the mechanisms for handling this information. In that rule, the phrase national operator of' big data '" is described to operate on the sort of public-private collaboration.

Due to the presence of Russian legislation on the security and shield of private personal information, a lot of changes are necessary to safeguard data subjects in the context of Big Data technology. The functions of Roskomnadzor as a supervisory authority based on the Art. 23, of Federal Law No 152, apply only to communicate among personal information operators and their cooperation with legal powers; As noted, they are inadequate to defend the safety of private information subjects and do not fall within the limits of the powers exercised by public agencies in other nations.

Threats jeopardize are few of the fundamental safeguarding specifications in the cloud. Such risks typically represented a breach of privacy, data leakage as well as illegal or unauthorized access to data on various cloud layers. Syed AsadHussai, et al.[40] proposed an innovative multi-level classification framework for various security threats throughout different cloud services for each tier. It also recognized the different kinds of attacks and even the levels of risk correlated with the varying cloud services on these layers. Risks were graded as high, low and medium. The magnitude of these levels of risk depends on the position of the cloud layers. Threats became more serious in the lower layers in which the networks as well as the platform were concerned. The strength of these risk rates has also been correlated with the security criteria for data privacy, data encryption, authentication, multi-tenancy, and authorization of multiple cloud services. The multi-level classification framework contributed to the development of a dynamic and structured security agreement for each cloud layer, which dynamically agreed on security criteria for cloud users and providers.

To sustain the privacy, confidentiality and security in Big Data,(Wei et al., 2015) [28] have proposed an generalized signcryption scheme based on ID. The method which is used is appropriate for big data efficiency requirement because signcryption scheme can work as the encryption scheme based on the need and it doesn't have a huge burden on the difficult management of certificates, as conventional cryptographic strategies do.

VI. PRIVACY

Big data has created a number of privacy concern for the customer, using the branded social websites may also leads to privacy breach. Concerning this as an issues many authors has written the article to anonymize the data .I can put some of them which are primarly used mainly t-closeness [1],where (Li, Li and Venkata subramanian, 2007) has found the limitations of l-diversity[2] and k-anonymity[3] and succeeded in dealing with choosing t-closeness approach , which particularly figure out distribution of a sensitive feature in any equivalence class to the distribution of the feature in the overall table. But when we look on l-diversity, approach which was proposed by (Machanavajjhala et al., 2007).has mentioned that there is absence of variety in sensitive features where he has mention the weak area of k-anonymity, and worked on continuous sensitive attributes to provide more security. K-anonymity is the one which concerns examining re-identification attacks. All these researches have opened the door for future endeavors.

In order to check the amount of degree of anonymity wide spread on twitter (Peddinti, Ross and Cappos, 2017) has developed an approach of machine-based classifier system [4] which identify the account that tweet sensitive data, however this approach can be improved more because using only SVM, we will face the problem of choosing of kernel function and also algorithm will be complex.

Clustering of heterogeneous big data has led to profound thought by authors because(Zhang et al., 2017)has indicated that to maintain privacy of heterogeneous big data, Only PCM [5] is not sufficient as it works for small dataset, so they have moved to distributed HOPCM[6] which can provide clustering of huge dataset efficiently, but when it comes to privacy the process has been improved with BGV encryption by adding HOPCM and therefore named it as PPHOPCM[7],the result analysis shows that as data volume increases the execution times may increase ,it has improved only 30% over the previous HOPCM also it need more cloud server to perform better clustering and protecting the private data of user.

To improve the data quality and to provide new insight for population information in health care, governmental and social services linking the data from different sources will be fruitful, but this scenario again provides privacy issue for confidential data. Keeping privacy protection into the mind (Vatsalan, D., Sehili, Z., Christen, P., & Rahm, E. 2017) has proposed PPRL[8] which uses masked version of QID i.e. quasi identifier which will be an attribute and it will be common to all the database which are linked. When we looked into PPRL techniques ,it has found that it is getting affected by the Big data uncertainty and variety as the volume increases the technique needs an enhancement.

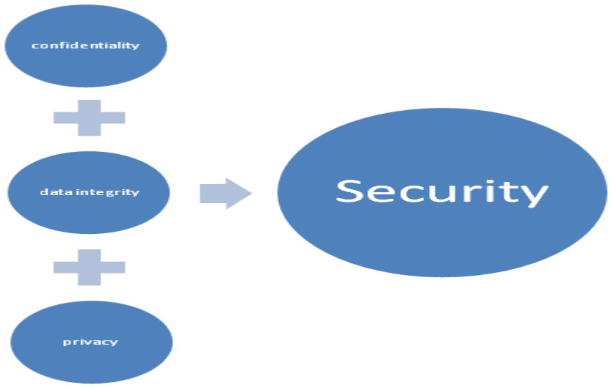

Fig 1: Groups Which Makes the Security of Issues

Table 3. Shows The 3 V's, Services Offered by Techniques, and Their Impediment

Proposed method	Services	Impediment	V

Machine based classifier	Mainly for checking degree of anonymity	SVM usage and influenced by the size of dataset	Volume
PPHOPCM	Used for protecting the privacy of large dataset	Need for more cloud server and affected by the size of data	Volume, Variety,
PPRL	Privacy for linked record from different database	The result is not as expected because of uncertainty and increased size of dataset.	Volume,

Table 4. Showing the Security Part Covered by Each Author Mentioned in This Paper

Author	Security Part
Li et al.	Data Confidentiality
Zhao et al.	Data Confidentiality
Stergiou and Psannis	Data Confidentiality
Wang et al.	Data Confidentiality
Yang et al.	Data Confidentiality
Puthal et al.	Data Confidentiality
Sandryhaila and Moura	Data analysis
Patterson	Data Integrity and privacy protection
Carl lagoze	Data integrity
Liu, C., Yang, C., Zhang, X., & Chen, J.	Data Integrity
Tan et al.	Data Integrity
Zharova and Elin	Data integrity and privacy
Syed AsadHussai	Authentication, privacy protection
Wei et al.	Data confidentiality
Zhang et al	Privacy protection
Vatsalan, D., Sehili, Z., Christen, P., & Rahm, E.	Privacy protection

VII. DATA SILOS

Data silos means putting the data or we can say the information in silos, the silos will be in different management level, individual department of organization, government sector and many more areas. Finding a collaboration among different silos will be very less, Silos is done to accomplish business achievement and for market beneficial act. For 98% of executive it is a discussion to be bothered, silos restrict the clarity and people becomes less collaborative and expressive it restricts cohesive team work. these all leads to incompatibility in working environment. (Speiser, S., &Harth, A. (2010))[24]Has developed an approach to integrate the data set with Linked Data for enabling to create LInked Data services (LIDS)from inaccessible data silos .In order to extract the terms from data silos (Lomotey& Deters, 2013)[25] has proposed a tool Known as TouchR which is based on Hidden Markov Model, furthermore there is need for development of dictionary adaptation to achieve more extraction from silos.

VII. CONCLUSION

In this paper, we collaboratively identified confidentiality as well as safety issues across every part of big data and the benefits and soft spots of previous conventional techniques in terms of big data confidentiality and safety. We principally examined the confidentiality and data integrity, privacy and Data analytics ways that are used recently in Big Data and mentioned however anonymization ways which is used for information protection additionally as given their limitations. Additionally, there are alot of varied techniques embrace concealment a needle in an exceedingly hayrick. Currently the topic Data Solis is one of the issues for business and other regional sector to carry out efficient marketing, the paper has reviewed some of the technique used for extracting the information, in this context very few researchers have reviewed the part of Solis. By the end of this paper, the reader finds the way of understanding security issues which means covering all the part of data integrity, data confidentiality, data analysis and data privacy. Any feedback from the reader side is acceptable as the author is aspirant.

REFERENCE:

1. Li, N., Li, T. and Venkatasubramanian, S. (2007). t-Closeness: Privacy Beyond k-Anonymity and l- Diversity. 2007 IEEE 23rd International Conference on Data Engineering.
2. Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M. (2007). L-diversity. ACM Transactions on Knowledge Discovery from Data, 1(1), p.3-es.
3. SWEENEY, L. (2002). k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), pp.557-570.
4. Peddinti, S., Ross, K. and Cappos, J. (2017). User Anonymity on Twitter. IEEE Security & Privacy, 15(3), pp.84-87.
5. Krishnapuram, R. and Keller, J. (1996). The possibilistic C-means algorithm: insights and recommendations. IEEE Transactions on Fuzzy Systems, 4(3), pp.385-393.
6. Q. Zhang, L. T. Yang, Z. Chen, and Feng Xia,"A High-Order Possibilistic-Means Algorithm for Clustering Incomplete Multimedia Data," IEEE Systems Journal, 2015, DOI: 10.1109/JSYST.2015.2423499.
7. Zhang, Q., Yang, L., Chen, Z. and Li, P. (2017). PPHOPCM: Privacy-preserving High-order Possibilistic c-Means Algorithm for Big Data Clustering with Cloud Computing. IEEE Transactions on Big Data, pp.1-1.
8. D., Sehili, Z., Christen, P., & Rahm, E. (2017). Privacy-Preserving Record Linkage for Big Data: Current Approaches and Research Challenges. Handbook of Big Data Technologies, 851–895.doi:10.1007/978-3-319-49340-4_25
9. Lagoze, C. (2014). Big Data, data integrity, and the fracturing of the control zone. Big Data & Society, 1(2), p.205395171455828.
10. Liu, C., Yang, C., Zhang, X., & Chen, J. (2015). External integrity verification for outsourced big data in cloud and IoT: A big picture. Future Generation Computer Systems, 49, 58–67.
11. R.C. Merkle, A digital signature based on a conventional encryption function, in: Proceedings of A Conference on the Theory and Applications of Cryptographic Techniques on Advances in Cryptology, CRYPTO'87, 1987, pp. 369–378.
12. Erway, A. Küpçü, C. Papamanthou, R. Tamassia, Dynamic provable data possession, in: Proceedings of the 16th ACM Conference on Computer and Communications Security, CCS'09, Chicago, USA, 2009, pp. 213–222.
13. Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., & Song, D. (2007). Provable data possession at untrusted stores. Proceedings of the 14th ACM Conference on Computer and Communications Security - CCS '07.
14. Shacham H., Waters B. (2008) Compact Proofs of Retrievability. In: Pieprzyk J. (eds) Advances in Cryptology - ASIACRYPT 2008. ASIACRYPT 2008.
15. Tan, C., Hijazi, M., Lim, Y. and Gani, A. (2018). A survey on Proof of Retrievability for cloud data integrity and availability: Cloud storage state-of-the-art, issues, solutions and future trends. Journal of Network and Computer Applications, 110, pp.75-86.

16. Sandryhaila, A. and Moura, J. (2014). Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure. IEEE Signal Processing Magazine, 31(5), pp.80-90.
17. Chung, W. (2014). BizPro: Extracting and categorizing business intelligence factors from textual news articles. International Journal of Information Management, 34(2), 272–284
18. KLUEGL, P., TOEPFER, M., BECK, P., FETTE, G., & PUPPE, F. (2014). UIMA Ruta: Rapid development of rule-based information extraction applications. Natural Language Engineering, 22(01), 1– 40
19. Nenkova A., McKeown K. (2012) A Survey of Text Summarization Techniques. In: Aggarwal C.,Zhai C. (eds) Mining Text Data. Springer, Boston, MA
20. Këpuska, V. and Elharati, H. (2015). Robust Speech Recognition System Using Conventional and Hybrid Features of MFCC, LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions. Journal of Computer and Communications, 03(06), pp.1-9.
21. Adams, A. and Ferryman, J. (2013). The future of video analytics for surveillance and its ethical implications. Security Journal, 28(3), pp.272-289.
22. Li, Y., Gai, K., Qiu, L., Qiu, M. and Zhao, H. (2017). Intelligent cryptography approach for secure distributed big data storage in cloud computing. Information Sciences, 387, pp.103-115.
23. Yang L.T., Xu L., Lin M., Quinn J. (2006) A Parallel GNFS Algorithm with the Biorthogonal Block Lanczos Method for Integer Factorization. In: Yang L.T., Jin H., Ma J., Ungerer T. (eds) Autonomic and Trusted Computing. ATC 2006. Lecture Notes in Computer Science, vol 4158. Springer, Berlin, Heidelberg
24. Speiser, S., &Harth, A. (2010). Taking the LIDS off data silos. Proceedings of the 6th International Conference on Semantic Systems - I-SEMANTICS '10.
25. Lomotey, R. K., & Deters, R. (2013). Terms extraction from unstructured data silos. 2013 8th International Conference on System of Systems Engineering.
26. Zhao, J., Wang, L., Tao, J., Chen, J., Sun, W., Ranjan, R., Kołodziej, J., Streit, A. andGeorgakopoulos,
27. D. (2019). A security framework in G-Hadoop for big data computing across distributed Cloud data centres.
28. Thayananthan, V. and Albeshri, A. (2015). Big Data Security Issues Based on Quantum Cryptography and Privacy with Authentication for Mobile Data Center. Procedia Computer Science, 50, pp.149-156.
29. Wei, G., Shao, J., Xiang, Y., Zhu, P. and Lu, R. (2015). Obtain confidentiality or/and authenticity in Big Data by ID-based generalized signcryption. Information Sciences, 318, pp.111-122
30. Wang, Z., Cao, C., Yang, N. and Chang, V. (2017). ABE with improved auxiliary input for big data security. Journal of Computer and System Sciences, 89, pp.41-50.
31. Puthal, D., Nepal, S., Ranjan, R. and Chen, J. (2017). A dynamic prime number based efficient security mechanism for big sensing data streams. Journal of Computer and System Sciences, 83(1), pp.22- 42.
32. Abouelmehdi, K., Beni-Hessane, A. and Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. Journal of Big Data, 5(1).
33. Gandomi, A. and Haider, M. (2019). Beyond the hype: Big data concepts, methods, and analytics.
34. Patterson, T. (2013). Information Integrity in the Age of Big Data and Complex Information Analytics Systems. EDPACS, 48(6), pp.1-10.
35. Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu & Fei Hu (2017) Big Data and cloud computing: innovation opportunities and challenges, International Journal of Digital Earth, 10:1, 13-53
36. TechAmerica Foundation's Federal Big Data Commission(2012).Demystifying Big Data: A practical guide to transforming the business of Government .Retrieved from http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf
37. Stergiou, C. and Psannis, K. (2017). Efficient and secure BIG data delivery in Cloud Computing. Multimedia Tools and Applications, 76(21), pp.22803-22822.
38. Laurence T. Yang, Gaoyuan Huang, Jun Feng, Li Xu, "Parallel GNFS algorithm integrated with parallel block Wiedemann algorithm for RSA security in cloud computing", Information Sciences, Vol. 387, pp. 254-265, May 2017
39. LuigiAtzori, Antonio Iera, Giacomo Morabito,The Internet of Things: A survey,ComputerNetworks,Volume 54, Issue 15,2010,Pages 2787-2805,ISSN 1389-1286,
40. Alaba, F., Othman, M., Hashem, I. and Alotaibi, F. (2017). Internet of Things security: A survey. Journal of Network and Computer Applications, 88, pp.10-28.
41. Syed Asad Hussain, Mehwish Fatima, Atif Saeed, Imran Raza, Raja Khurram Shahzad, " Multilevel classification of security concerns in cloud computing", Applied Computing and Informatics, Vol. 13, Issue 1, pp. 57-65, January 2017.
42. Zharova, A. and Elin, V. (2017). The use of Big Data: A Russian perspective of personal data security. Computer Law & Security Review, 33(4), pp.482-501.

AUTHOR PROFILE

Ms. Seema Rai is Assistant Professor in Computer Science and Engineering Department at **Sphoorthy engineering College** located in Nadargul Village, Balapur Mandal, Hyderabad, Telangana state , India. I have received Master of Technology Degree in Computer Science and Engineering from **Aurora's scientific technological and research academy**, India. I have 5 year experience in the education field. My research interest in the field of Security in Big Data. I have completed various courses through NPTEL on BIG DATA Computing and compiler design where I have received elite certificate, Attended many workshop in the field of teaching and learning center.

Dr. Ashok Sharma is having 18 Years of Teaching Experiences in Higher Education and He has worked in Various Reputed Institution of Higher Learning in India in different capacity.His area of Interest is Machine Learning, Cloud Computing and Data Science. He is certified in Beidou Technology from Shanghai Jio Tong University ,Shanghai, Leveraging Technology for Effective Teaching in the Classroom and Beyond from International Institute of Information Technology, Bangalore, Internet of Things and Introduction to Big Data from University of California San Diego.He has attended 30+ DST/AICTE sponsored Training in different Technologies and He is presently holding position of Associate Professor in School of Computer Science and Engineering, Lovely Professional University, Phagwara, India and 8 PhD Scholars are working under his guidance in the area of Cloud Computing, Data Science and Cognitive Behaviour Analysis.