JOURNAL
COMPUTER SCIENCES INSTITUTE

JCSI 23 (2022) 140–144
Received: 31 March 2022
Accepted: 8 May 2022

A Novel Inconsequential Encryption Algorithm for Big Data in Cloud Computing

Ravi Kanth Motupalli [a,*], Krishna Prasad K.[b]

[a]*Research Scholar, Institute of Computer Science & Information Sciences, Srinivas University, Mangalore, Karnataka, India and Assistant Professor, Department of CSE, VNR VJIET, Hyderabad, Telangana, India.*
[b]*Associate Professor, Institute of Computer Science & Information Sciences, Srinivas University, Mangalore, Karnataka, India.*

Abstract
In the digitalized era of the information technology the expansion of the data usage is very high accounting for about enormous data transaction in day to day life. Data from different sources like sensors, mobile phones, satellite, social media and networks, logical transaction and ventures, etc add an gigantic pile to the existing stack of data. One of the best way to handle this exponential data production is the Hadoop network. Thus in the current scenario big industries and organizations rely on the Hadoop network for the production of their essential data. Focusing on the data generation and organization, data security one of the most primary important consideration was left unnoticed making data vulnerable to cyber attacks and hacking. Hence this article proposes an effective mixed algorithm concept with the Salsa20 and AES algorithm to enhance the security of the transaction against unauthorised access and validates the quick data transaction with minimal encryption and decryption time. High throughput obtained in this hybrid framework demonstrates the effectiveness of the proposed algorithmic structure over the existing systems.

Key words: Hadoop network; data security; cyber-attacks; Salsa20

*Corresponding author
Email address: ravikanth_m@vnrvjiet.in (Ravi Kanth. M)

1. Introduction

In this digital era the data generated every day for multiple purpose accounts for a very large pile. Data may be of any time from either sensors or devices, Public media or concerns, Cognitive data or transaction data, etc are getting added to the cascading data forming a big torrent called big data. This avalanche of data are the data expansion in the last decade from little to huge volume. As there are enormous amount of data in every information Big data receives its limelight and demonstrated outstanding performance in the field of Digital information technology. These large volume of data can be either in an organized or unorganized format. Uncertainty of the big data is estimated by the organized format of the digital data that is generated in current scenario with the advancement of the novel and modern communication techniques. This is just an initial phase there is a bright future of expansion for the big data in the upcoming decades. The estimated revenue forecast of the big data in US has been illustrated in the Figure 1. This shows the relevance of the big data investement and its progressive expansion.

Big Data is delivered in correspondence with the transmission of data in a broader spectrum accumulated from different sources.Thus Big Data is obligated to set up the data mining computations. The primary requirement of this framework is to assure the security of the data along with its structural analysis. The Information engineering has developed its analysis in an advanced way migrating from the centralized framework to the distributed framework. In this corresponding study mobile data is taken for the consideration. With the developed gadget usage and the production the amount of storage space in the mobile phones are increased to bigger levels. With high storage device there comes issues with the performance and heat dissipation factors. To avoid such issues and to centralize the storage Big data in cloud storage was introduced to the mobile users. Hence we consider the analysis of the big data stored in the Hadoop Distributed File System (HDFS) [1]. The users of the big data are concerned in gaining more knowledge about the data structure to earn a good profit but fail to consider the security of the data which is the most important attribute of the cognitive communication.

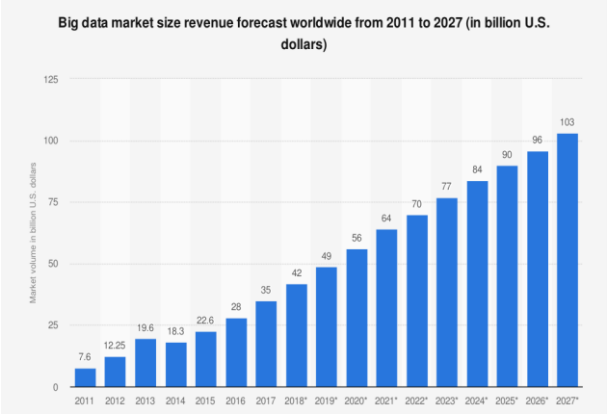

Figure 1: Revenue forecast of Big Data (Satsita, 2021).

Security is an primary factor to be assured in every storage system to save the data from the unfortunate attacks and known hackers. Upon investing it is known that HDFS has no organized security framework to secure the data. Hence the current study intends to pro-

pose a inconsequential encryption algorithm to assure the security of the data. HDFS is an distributed network which is built at less expenses with high tolerant against the faults [2]. When enormous data is being used in the HDFS, high throughput is achieved making it qualified to handle the Big Data. The size of the HDFS varies from Giga bytes to Tera Bytes which is facilitated to support the larger files [3]. Thus in a single instance larger data files.

Hadoop is anticipated to handle large amount of data , irrespective of its structure. Hadoop works on the MapReduce structure to handle the index of the web search[4,5]. Hadoop has an inbuilt security framework that makes use of cryptographic technique to mix the data and save it in the cloud. The sensitive data is encrypted to enhance the security where the raw data is transferred in to the cipher text and transformed in to arbitrary data.

The essential intension of this article is to design a technique through which the secure transformation of the data can be done with easy recovery of lost data without increasing the overhead in the computation, communication and storage. Hence in this study an hybrid algorithm mixing AES and Salsa20 algorithm was proposed. In this hybrid algorithm the size of the key is developed up to 256 bits. Apache ranger is used to achieve the cross control over the security of the data. In the proposed work Salsa20 encryption algorithm is executed before moving the HDFS and the data mining computation is done using the mapper class.

2. State of Art

Big Data is a messy, Jumbled and large which has an uncontrollable approach. The accumulated information is segregated and classified to identify the data type in an organized form to be used by a business or an organization[6,7]. Several studies and researches were done to enhance the security of the data stored in the cloud. There are broadly two types of encryption system which was commonly in practise known as symmetric and asymmetric key encryptions[8,9]. Lin.H.Y et al attempted to propose an hybrid encryption system mixing AES and pairing of the HDFS system to improve the security[10]. An novel encryption technique based on the new instruction encryption was proposed by Acharya et al [11] to upright the use of a trusted environment.

Privacy is the term which is used to denote the protected handling and storage of data in the cloud network [12]. Cloud computing has occupied the digital market due to the specialized features like consistency, less expensiveness, rigid against faults and scalability [13,14]. With the suggested usage of symmetric encryption , blowfish encryption algorithm was used in the IOT clod controlling FPGA based devices [15]. Data masking technique was identified as the simple and efficient security enhancing technique [16]. Hamid et al in his paper highlighted three important key points like security in Big Data, retrieval of information using context aware concepts and ontology integrated big data [17].

Sachin et al had made an conceptual study on the big data security and supported the arguments uplifting the use of high volume data and RDBMS [18]. Bhargavi et al proposed that the original DS is used for the effective use of Big data [19]. The biggest challenge in the Big Data security is the information security , where the data is vulnerable to data leaks, Unauthorized access, DoS attacks, etc. The impression of handling the unstructured enormous data was changed by the Hadoop.

Aditham et al [20] developed an algorithm to detect the attack on the network in two steps. Initially the control of the algorithm was framed in the first step and the process to merge the replicating nodes was developed in the second step. Reddy et al [21] specifically developed detection algorithm for the cloud environment where the access control was the main objective. But this method failed to assure security to the data when providing restricted control over the access. In the proposed framework four stages of the security was ensured. In the initial stage the security for the access control was ensured with the existing technology. In the second or registration process both the receiver and the transmitter were registered to the trust center and service provider to obtain the secret key. In the third stage authentication of the user is validated against the entry of the user with the service provider so as to ensure the access control over the data. In the last phase the secret key of the user is updated or revoked upon the detection of the leakage or misusage.

3. Challenges in the Cloud Storage Security

The data in the cloud is stored in a remote location where the attack or hacking of the cloud environment without proper security features happen, there may be collateral loss of the data assets and information. The physical damage to the cloud server machines may result in unrecoverable data loss. In order to assure the data security the information is replicated and stored in different location where it should be managed effectively to avoid the confusion and data replication issues to the end users. Though the cloud servers are handled by the third entity the service provider themselves should have restricted access to the data to ensure the privacy of the data stored. Hence in order to assure the data privacy through limiting the data access to the service providers, all the data hiding and recovery process of encryption, decryption and scrambling are given to the user side. The security and privacy of the data in big data are considered to be the two giant attributes of its performance analysis

4. Research Methods

The architecture for storing data in cloud has several framework. The proposed framework has four parameters named as trust center, data transmitter, receiver and server in cloud environment. Trust center is an sturdy, vigorous, and enormous entity that is framed for the assurance of security of data in the cloud environment. The relation between the receiver, transmitter and the server was effectively and efficiently managed by this

entity. This entity is also responsible for tracking the secret key in all the transaction and has powers to revoke the key when misused. Each transmitter should register themselves in this trust center to obtain the benefits of secure communication. Similarly each receiver should register in to the trust center to obtain authenticated access to the data.

Processes like slicing, transformation of data, encryption, decryption, scrambling and unscrambling, consolidation, and recovery are the basic processes involved in any data transfer. Using an effective secret key the transmitter uploads or stores data in the cloud.

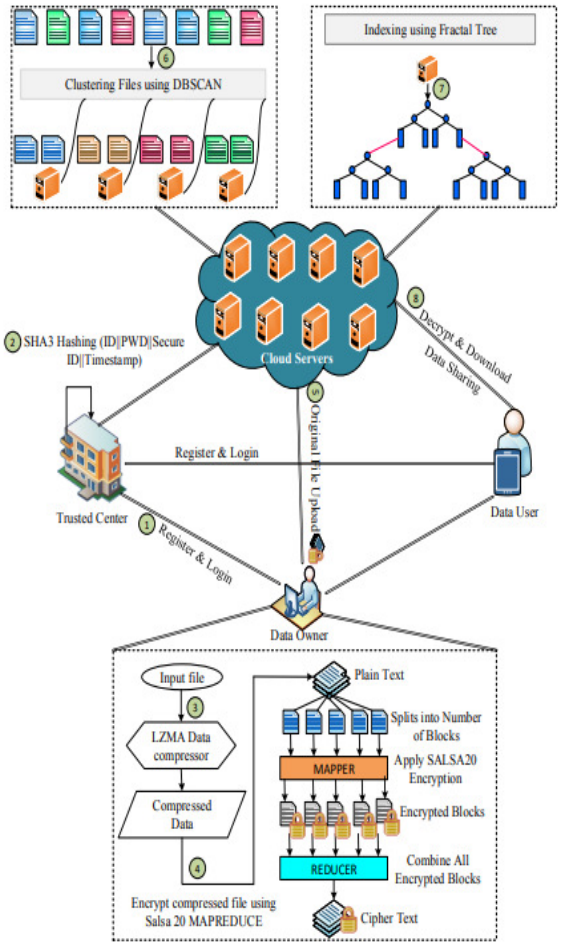

Figure 2: Proposed framework for the inconsequential encryption algorithm in Big Data.

When transferring the data raw information is scrambled and converted in to cipher text. Then the cipher text is decrypted and unscrambled at the other end using the secret key provided by the owner to recover the original data at the receiver end.

The process flowchart is illustrated in the Figure 3. Salsa20 algorithm is an efficient stream cipher which can effective transform key bits from 256 to 264 streams with occasionally susceptible 64 byte blocks per stream. This 64 byte block was represented in 16 word format which is obtained as a result of 320 reversible alterations, one word for each modification [22,23]. The corresponding output was amended to the producing 64 byte output. In each change, two words, one original word and the rotated version of the other was xor-ed twice. Thus each round is performed as 4 parallel quarter rounds.

There is separate function for each quarter round. There is row round function for each quarter which modifies the rows. Then column round function is processed changing the column. Then a double round function is performed for the specified number [24,25]. In this article 20 rounds of Salsa 20 algorithm was proposed. Then the resulting 64byte is transformed in to 16 word format using little endian method.

Thus a 64byte output is obtained. Encryption tools along with the authentication and platform manager was used to improvise the Hadoop security. In this proposed module 10 rounds of Hadoop clustering with specified configuration was done. Deployment validation is highly prioritized based on the running concerns and the Apache ranger is given steady administration platform to configure and strategize the secured data in each cluster [26,27]. The ranger key management service is used for the implementation of HDFS encryption. Apache officer is used for the authenticated data access through the centralized fine grain approval system [28].

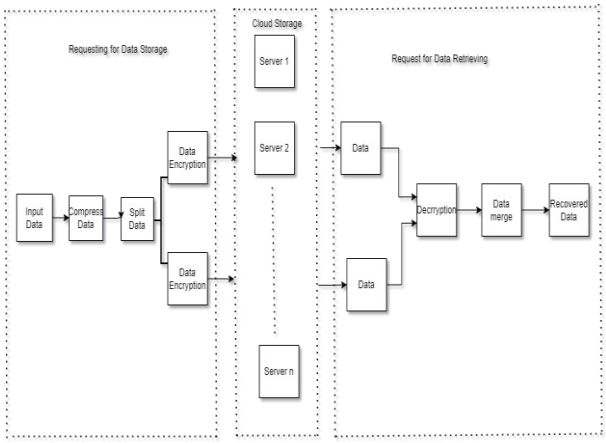

Figure 3: Proposed framework flowchart.

5. Results and Discussion

For the experimental procedure 10 cluster was created where the first 8 nodes serve as MapReduce clients and as Data Node servers. The last two nodes function as MapReduce scheduler and Name Node manager. The first eight nodes are a 8V processor operating with 8GB random access memory with Ethernet NIC. The last two nodes are 16 V processor with 32GB memory. Hadoop replica was not allowed in this system. The sequenced process of encryption and decryption of enormous data pushes and stacks them into a HDFS.

The encryption time was calculated in milli seconds. To make the system more responsive and more dynamic the execution period should be small enough. Similarly, decryption time should be less enough to accelerate the system. The third important parameter for consideration is Avalanche effect where a small change in the input has big and drastic effect at the output. The encryption time analysis of the proposed system with other encryption and decryption techniques was shown in the Fig-

ure 4. This illustrates that the proposed methodology is more suitable for the foresaid applications.

From the Figure 4, it is evident that the encryption and decryption time of the proposed model is short enough to support the processing of big data files quickly and actively. The throughput of the decryption process is observed to higher than that of the encryption process. The throughput is calculated as the ratio of plain text to the encryption time in milli seconds.

Performance analysis was done with the other encryption algorithms like AES and blowfish algorithm. From the result it is evident that the proposed technique has higher encryption than that of the other prevailing techniques.

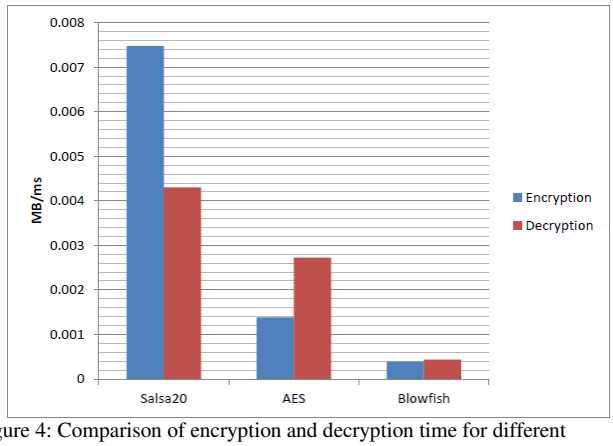

Figure 4: Comparison of encryption and decryption time for different algorithms.

Due to the smaller key size the throughput of the blowfish algorithm was very less. The security of the data in the proposed model is measured with the parameter called avalanche effect. A small change in the input stream causing a big impact on the output stream is called avalanche effect. This method allows the data scrambling in a reversible format to enhance the security of the data. Thus the overall performance of the proposed Salsa20 hybrid algorithm is found to have minimal running time, increased throughput, with the augmented avalanche effect guarantying the security of the stacked data in HDFS.

6. Conclusion

The overall concept of the proposed framework is to secure the data transaction in the cloud system through an effective encryption algorithm in the Hadoop network. In this proposed model Salsa20 hybrid algorithm was used for effective encryption and efficient key management. The overall performance efficiency of the proposed hybrid algorithm is accessed based on the speed of transmission and the less computational time for multiple nodes in a cluster. Thus the proposed framework is guaranteed for the authenticated data access, confidential data storage and controlled access of the Hadoop network. The low computational time and the improvised throughput facilitated the immediate transaction of the data with utmost security. This study is majorly focused on the mobile data transaction to the cloud storage. The future work may include designing framework for the most complicated IoT transaction like defence data storage in the cognitive cloud, Big Data accessing in the educational domain and the large database secured in the healthcare sector.

References

[1] Big Data market revenue forecast worldwide 2011-2027, by Wikibon; Silicon ANGLE Available: https://www.statista.com/statistics/254266/global-big-data-market-forecast/ [04.02.2019].

[2] Apache Hadoop 3.2.0-Hdfs Architecture, Available: https://hadoop.apache.org/docs/r3.2.0/hadoop-project-dist/hadoop-hdfs/HdfsDesign [04.01.2019].

[3] D. Borthakur, The Hadoop distributed file system: Architecture and design, Hadoop Project Website 11 (2007) 21.

[4] M. Dhatrak, H. Panadiwal, Privacy-Preserving Mining using Data Encryption scheme for Hadoop Ecosystem, In International Journal of Advanced Research in Science, Engineering and Technology 5(4) (2018) 5578-5585.

[5] S. Singh, M. Sharma, The Prototype for Implementation of Security Issue in Big Data Application using Hadoop Server, International Journal of Computer Applications 145(13) (2016) 9-13.

[6] S. Ghemawat, J. Dean, MapReduce: Simplified data processing on large clusters, ACM Commun. Mag. 51(1) (2008) 107-113.

[7] Y. Kumar, R. Munjal, H. Sharma, Comparison of symmetric and asymmetric cryptography with existing vulnerabilities and countermeasures, (IJAFRC) 1:6 (2014) 2348-4853.

[8] B.H. Lee, E.K. Dewi, M.F. Wajdi, Data security in cloud computing using AES under HEROKU cloud, In The 27th Wireless and Optical Communications Conference (WOCC2018) (2018) 1-5.

[9] P. Mahajan, A. Sachdeva, A Study of Encryption Algorithms AES, DES, and RSA for Security, Global Journal of Computer Science and Technology Network, Web & Security 13(15) (2013) 15-22.

[10] H.Y. Lin, S.T. Shen, W.G. Tzeng, B.S.P. Lin, Toward data confidentiality via integrating hybrid encryption schemes and Hadoop Distributed File System, in the Proceedings of the 2012 IEEE 26th International Conference on Advanced Information Networking and Applications (AINA) (2012) 740-747.

[11] J. Cohen, S. Acharya, Towards a Trusted Hadoop Storage Platform: Design Considerations of an AES Based Encryption Scheme with TPM Rooted Key Protections," IEEE 10th International Conference on and Autonomic and Trusted Computing (UIC/ATC), Ubiquitous Intelligence and Computing (2013) 444-451.

[12] J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, R. Buyya, Ensuring security and privacy preservation for cloud data services, ACM Computing Surveys 49(1) (2016) 1-39.

[13] W. Song, B. Wang, Q. Wang, Z. Peng, W. Lou, Y. Cui, A privacy preserved full-text retrieval algorithm over

encrypted data for cloud storage applications, Journal of Parallel and Distributed Computing 99 (2017) 14-27.

[14] P.V. Bharati, T. Sita Mahalakshmi, Data storage security in cloud using a functional encryption algorithm, In Emerging Research in Computing, Information, Communication and Applications, Springer Singapore (2016) 201-212.

[15] K.N. Prasetyo, Y. Purwanto, D. Darlis, An Implementation of data encryption for internet of things using blowfish algorithm on FPGA. International Conference on Information and Communication Technology (2014) 75-79.

[16] A. Abo-alian, N. L. Badr, M. F. Tolba, Data Storage Security Service in Cloud Computing: Challenges and Solutions, In Multimedia Forensics and Security, Springer International Publishing (2017) 25-57.

[17] H. Bagheri, A.A. Shaltooki, Big Data: Challenges, Opportunities and Cloud Based Solutions, International Journal of Electrical and Computer Engineering (IJECE) 5(2) (2015) 340-343.

[18] S.A. Thanekar, K. Subrahmanyam, A.B. Bagwan, A Study on MapReduce: Challenges and Trends, Indonesian Journal of Electrical Engineering and Computer Science (IJEECS) 4(1) (2016) 176-183. DOI: 10.11591/ijeecs.v4.i1.pp176-183.

[19] I. Bhargavi, D. Veeraiah, T.M. Padmaja, Securing BIG DATA: A Comparative Study Across RSA, AES, DES, EC and ECDH, In Computer Communication, Networking and Internet Security. Lecture Notes in Networks and Systems 5 (2017) 355-362.

[20] S. Aditham, N. Ranganathan, A System Architecture for the Detection of Insider Attacks in Big Data Systems. IEEE Transactions on Dependable and Secure Computing (2017) 1–1. doi:10.1109/tdsc.2017.2768533.

[21] Y. Reddy, Big Data Processing and Access Controls in Cloud Environment. 2018 IEEE 4th International Conference on Big Data Security on Cloud (Big Data Security), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS). doi:10.1109/bds/hpsc/ ids18.2018.00019.

[22] D. J. Bernstein, The Salsa20 family of stream ciphers, In New Stream Cipher Designs. Berlin, Germany: Springer, (2008) 84–97.

[23] S. Potteti, N. Parati, Secured Data Transfer For Cloud Using Blowfish, International Journal Of Advances In Computer Science And Cloud Computing 3(2) (2015) 17-22.

[24] P. Ghosh, V. Thakor, P. Bhathawala, Data Security and Privacy in Cloud Computing Using Different Encryption Algorithms, International Journal of Advanced Research in Computer Science and Software Engineering 7(5) (2017) 469-471.

[25] K.Sekar, M Padmavathamma, Comparative Study of Encryptio Algorithm over Big Data in Cloud Systems, International conference on Computing for Sustainable Global Development (INDIACom) (2016) 1571-1574.

[26] B.T. Reddy, K.B. Chowdappa, S.R. Reddy, Cloud Security using Blowfish and Key Management Encryption Algorithm, International Journal of Engineering and Applied Sciences (IJEAS) 2(6) (2015) 59-62.

[27] G. Saini, N. Sharma, Triple security of data in cloud computing, International Journal of Computer Science and Information Technologies 5(4) (2014) 5825-5827.

[28] D.S. Elminaam, H.M. Abdual-Kader, M.M. Hadhoud, Evaluating the performance of symmetric encryption algorithms, Int. J. Netw. Secur. 10(3) (2010) 216-222.