Ensure Security of Healthcare Big Data by Evaluation of Performance of AES and DES by Hadoop Streaming

Deepti Goyal,[1] Dr. Ankit Kumar[2]
[1]Research Scholar, Department of Computer Science, Baba Mastnath Univesity, Asthal Bohar, Rohtak,, Haryana
[2]Assistant Professor,department of Computer Science, Baba Mastnath University, Asthal Bohar, Rohtak, Haryana

Abstract
Immense utilization of Internet and web applications leads to developing enormous data in our daily routine life. Big data becomes an imperative mechanism for managing and collecting data from multiple sources and analyzing it in a reliable and high performance platform. Healthcare industry demands a framework which gives a higher evolution environment for big data as a security of sensitive data form large amounts of data. These Cryptography techniques provides the high level security of sensitive healthcare data. In this paper we compare encryption techniques to ensure the security of large data form leakage, unauthorized and unauthenticated access. Evaluation of healthcare big data compare on the basis of time taken for encryption and decryption of multiple sizes data files by AES and DES encryption algorithms.

Keywords: Apache Hadoop**,** Encryption techniques, Big Data, Security, AES, DES

I. INTRODUCTION

This Voluminous data captured and produced in a very large scale in the present day to day life by various sources like social media (Facebook, Twitter, Instagram), economic data, Governmental confidential data, Private details of humans on web, Scientific research, Banking Industry, Stock Exchange, Defense academy data and data about patients history in healthcare industry That voluminous data is considered as big data. Security becomes first priority due to its privacy and confidentiality. A wide range of data managed by healthcare industry every day from various operations systems like clinical and operational systems for example: Laboratory Information Management System (LIMS), Electronic HealthRecord (EHR) etc. (Ali, Shrestha, Soar, & Wamba, 2018) Medical experts and practitioners day to day developing new applications to facilitate the healthcare Industry or stakeholders to increase the opportunity value of healthcare data.

Big Data must be preserved from leakage, unauthorized and unauthenticated access plus manipulation. Security and analysis of HealthCare data have become an essential or drastic problem in all over the world. Hadoop Framework developed for storing and analyzing big data, HDFS Store a large amount of data in an unstructured form. MapReduce is a technology of Hadoop used for manipulations and processing of big data. (Ali, Shrestha, Soar, & Wamba, 2018) Hadoop basically apply encryption techniques to forge an adequate security mechanism to the big Healthcare data Which gives exact information about the time taken for the encryption and decryption of data with file and block size at a particular time. (Ali, Shrestha, Soar, & Wamba, 2018) (Naveen, Sharma, & Nair, 2019)

II. OBJECTIVE OF STUDY

Volume, Velocity and Variety are three basic concerns of Big Data. Security of Big Data on these three aspects become very much essential. For that, various symmetric encryption algorithms applied for continuously viewing the Security status of big data. Hadoop and MapReduce become essential tools for structuring and ensuring the security of big data from various symmetric encryption techniques.

The objective of the study is to enhance the security of big Healthcare data of patients and their medical history by Encryption and decryption time. We also find out how much the data is secured when a patient diagnoses online by various internet devices for remote location analysis. This study gives frequent persistence to the medical data as a free analysis of patients online and the data becomes safe. (Subramaniyaswamy, Vijayakumar, Logesh, & Indragandhi, 2015)

Apache Hadoop and Hadoop Distributed File System

(HDFS)

Hadoop is an open source framework for Big Data analytics. It works on big data files for process it and gives and analytical review on Big Bata. Apache Hadoop becomes an efficient and effective solution for analytical review of big healthcare data. Hadoop consists of many components like Hadoop Distributed File system (HDFS) which enhance security of data. HDFS apply various encryption techniques for the secured transmission of data. MapReduce, a technology of Hadoop designed for managing huge data clusters. MapReduce contains one main node called Job Tracker and the rest of all slave nodes as a Task Tracker. (Subramaniyaswamy, Vijayakumar, Logesh, & Indragandhi, 2015) (Meena & Sujatha, 19 june, 2019)

Hadoop runs big data on its own file system, Hadoop Distributed File System (HDFS), which includes two distinct functions: the Name node, which stores metadata, and the Data node, which stores data from various applications. MapReduce is a component of the Hadoop process that executes enormous data sets and summarizes the results.

III. REVIEW OF LITERATURE

The authors offer a monitoring system for the healthcare industry that uses the AES encryption methodology to deliver safe communication based on IoT devices. The proposed monitoring system obtains health data from biosensors every 10 seconds and stores it in a cloud platform after encryption with AES. After that, only authorized users with an id and a password can use and access the data. (Naveen, Sharma, & Nair, 2019) The authors' proposed technique, dubbed attribute based honey encryption (ABHE), provides a superior security solution for huge amounts of data. The ABHE method is detailed in detail in this article, with a step-by-step procedure in a clear and simple format. Big data files are password and login protected in HDFS, and then specified attributes from the file are matched, and the data is encrypted. The huge data can only be encrypted or decrypted by only those people who have authentication credentials. The security layer of the ABHE approach is impenetrable by intruders. (Kapil, et al., February, 2020)

In this project, unstructured data is structured and processed using the Map Reduce technique, and user taste is automatically anticipated via collaborative filtering. The application of collaborative filtering and sentiment analysis offers recommendation creation for any number of data providers, and map reduction is the most efficient technique for processing big volumes of data. The developed approach can be improved and the recommendation generating process can be re-create even more efficient and streamlined by employing Emoticon Based Clustering and Tagging Techniques. (Subramaniyaswamy, Vijayakumar, Logesh, & Indragandhi, 2015)

The researcher present a brief summary of these algorithms. A comparison study of DES, AES, and EB64 evaluated at key size, block size, scalability, algorithm, encryption, decryption, power consumption, security, key used, rounds, hardware, and software implementation. These algorithms were evaluated on SMS data of varied sizes. Considering and analyzing how long it would take the three ways to encrypt and decrypt SMS of various sizes. When comparing the EB64 method to DES and AES, it was apparent that the EB64 algorithm takes the least amount of time to encrypt and decode data, while DES takes the greatest time. (Logunleko, Adeniji, & Logunleko, February, 200)

Cryptography Algorithms:

Cryptography refers to using encryption and decryption algorithms by executing their codes Data protection and security with the help of unprotected channels. Encryption Algorithm transforms plaintext to cipher text. Cryptography consists of two key schemes for encryption. Symmetric and asymmetric encryption. Symmetric Encryption refers to using the same key to both encrypt and decrypt a message. Asymmetric encryption refers to using different keys for encrypting or decrypting a message by private and public keys. Senders of the message use recipients' public-key for the encryption or code of the message while transmitting the message (Naisuty, et al., 2020). On the contrary, sender's private key used for decrypt message by recipient of message

Advanced Encryption Standard

AES algorithm expressed as a symmetrical block cipher text algorithm. Originally AES developed by Dr. Joan Diemen and Dr. Vincent Rijndael in 2000. AES works as a symmetrical key encryption key, It means that both sender or receiver use the same key for encryption and decryption of messages. It takes 128 bits of plain text at a time and converts it into 128,192 and 256 block cipher text. Also it contains 10 Rounds for encryption and decryption so comparison becomes easy. Due to the small number of rounds it is so fast and flexible. (Logunleko, Adeniji, & Logunleko, February, 200) (Naisuty, et al., 2020) Each round of AES algorithm requires four types of operations to make changes in the array and its state: Mix Columns, XOR Key, Shift Rows and sab byte.

Data Encryption Standard (DES)

DES worked as symmetric key encryption from both sides. DES refers to the same key for encryption and decryption of big data. It takes a fixed length block size for plaintext and transforms it into cipher text by applying various algorithms, which also a fixed length size 64 bits. DES contains 16 rounds for encryption of plaintext for security of data. It also contains a 64 bit key size which is also fixed for the plaintext and cipher text. (Logunleko, Adeniji, & Logunleko, February, 200)

IV. RESEARCH METHODOLOGY

MapReduce

MapReduce methodology process data by dividing big data into small chunks. In clusters. It is a component of Apache Hadoop, and analyzes big data by two basic functions: mapper and reducer. Big data checked for interdependencies for removing complex problems when the results and datasets are being merged. (Kousalya & Parvez, 2018) Basically MapReduce technology used for aligned process of big data sets on different clusters for sorting, filtering perform by mapper function and calculate results by merging the data is called reducer function.

Hadoop Streaming

The amount of digital data produced every day is expanding exponentially as a result of the rise of digital media, the Internet of Things, and other technologies. As a result of this predicament, the data analyst now faces a significant task in building next-generation tools and technologies to store and manage these data. Here's where Hadoop streaming enters the IT world as a major revolving door.

Hadoop Framework is considered as a totally Java based framework as it works only in Java. Hadoop streaming facilitates programs written in other than java such as Python, Ruby, and Perl Etc. to execute and process big data files. (Kousalya & Parvez, 2018) Python gives an immense power of security for Big healthcare data by writing encryption and decryption symmetric and asymmetric algorithms and processing the code on Hadoop framework by MapReduce Framework. A brief description of these two algorithms AES and DES has been performed on their characteristics based like their Block size, Round for encryption and decryption, speed of processing, Block Cipher text and Key Size. (Kapil, et al., February, 2020) (Shastri & Deshpande, 2020) Table 1. Show the description of each of the encryption techniques.

Experimental Cluster Setup

AES and DES encryption techniques coded in python 3.6. Hadoop 3.10.0 installed in machine consist Windows 7 with 8GB RAM 64 bit-processor (Intel i5) and Hadoop commands run on Window PowerShell. A single-node Hadoop cluster configured for better utilization of HDFS, MapReduce and various other components of Hadoop. In this study, we compare AES and DES Symmetric Key encryption techniques by time taken for encryption and decryption and size of big data files before and after encryption and decryption (64MB, 128MB, and 256 MB) in different block sizes. Through Web UI of Hadoop clusters we can optimize the time of encryption and decryption of Big Data files and also observe size of files before and after encryption and decryption of files. (Kapil, et al., February, 2020)

- Encryption Time: The time taken for converting plain text into cipher text.
- Decryption Time: The time taken for converting cipher text in to plain text.
- File Size: Contains the words in documentation of file countable in Binary Units.
- Encryption Techniques: It implies the algorithm or Executable statement for encode or decode the plain text into cipher text and cipher text in to plain text

Comparison of Encryption Algorithms

A brief description of these two algorithms AES and DES has been performed on their characteristics based on five attributes like their Block size, Round for encryption and decryption, speed of processing, Block Cipher text and Key Size. Table no. 1. Show the description of each of the encryption techniques.

Table 1. Comparison of AES and DES.

Basis of Comparison	AES Algorithm	DES Algorithm
Block Cipher	Symmetric Block Cipher	Symmetric Block Cipher
Key Size	128, 192 and 256 bits	64 bits
Block Size	128 bits	64 bits
Round	10,12 or 14	16

Speed	Very Fast	Slow

V. RESULT ANALYSIS AND COMPARISON

The Execution of work carried out an experiment on above two symmetric encryption algorithms AES and DES. The Study conduct two attempts of each algorithms on different file sizes and then evaluate the time for encryption and decryption and calculate the results on the basis of performance.

Table 1. Shows AES and DES symmetric encryption on different files sizes. Average Encryption Time calculated on the basis of two attempts of encryption techniques.

Table 2. Similarly, Shows Decryptions time of two attempts of AES and DES and then calculate average time taken for accurate comparison of results.

Size of file before and after encryption of data files shown in Table 3. Files stored on HDFS first and then identify the storage space occupy by the stored big data file. Little bit difference in storage space while put the Dataset File in to HDFS environment from local system.

Table 1. Encryption time take by AES and DES

Size of	AES Algorithm (in Seconds)			DES Algorithm (in Seconds)		
	1st Attempt	2nd Attempt	AVG. Time	1st Attempt	2nd Attempt	AVG Time
32 MB	34	33	33.5	38	39	38.5
64 MB	37	39	38	44	43	43.5
128 MB	47	44	45.5	50	51	50.5
256 MB	78	80	79	101	78	89.5

Table 2. Decryption time take by AES and DES

File's Size	AES Algorithm (in Seconds)			DES Algorithm (in Seconds)		
	1st Attempt	2nd Attempt	AVG.Time	1st Attempt	2nd Attempt	AVG Time
32 MB	36	35	35.5	41	40	40.5
64 MB	40	36	38	43	40	41.5
128 MB	49	41	45	51	42	46.5
256 MB	70	71	70.5	74	69	71.5

Table 3. File Size before and after Encryption

Size of Files	File Size in HDFS Storage	File Size After AES Encryption	File Size After DES Encryption
32 MB	32.38	46.02	43.89
64 MB	64.08	88.42	86.21
128 MB	128.07	174.09	171.82
256 MB	256.91	377.05	350.57

By below computation from Fig. 1. And Fig. 2. , It is observed that AES takes less time for encrypt and decrypt (Naveen, Sharma, & Nair, 2019)the big data files as compared to DES. It is also observed that AES occupies much more space as compared to DES after encryption of data. Initial time for all sizes of files being same, so by above Tables and Figures it identifies that decryption time become fell down after particular size of file.

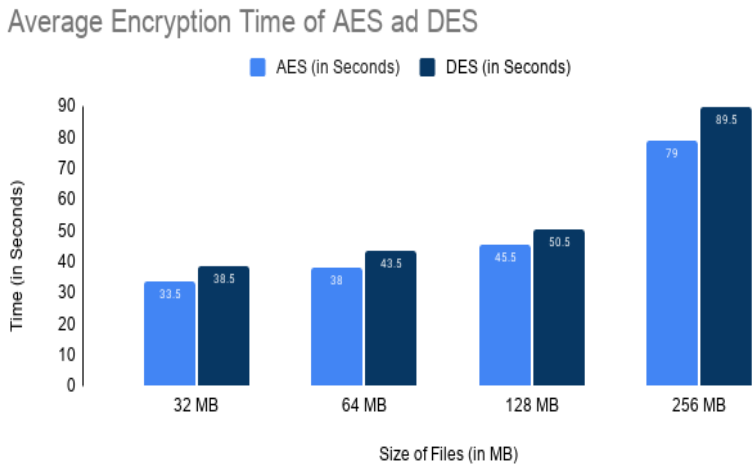

Figure 1. Average Encryption Time of AES and DES

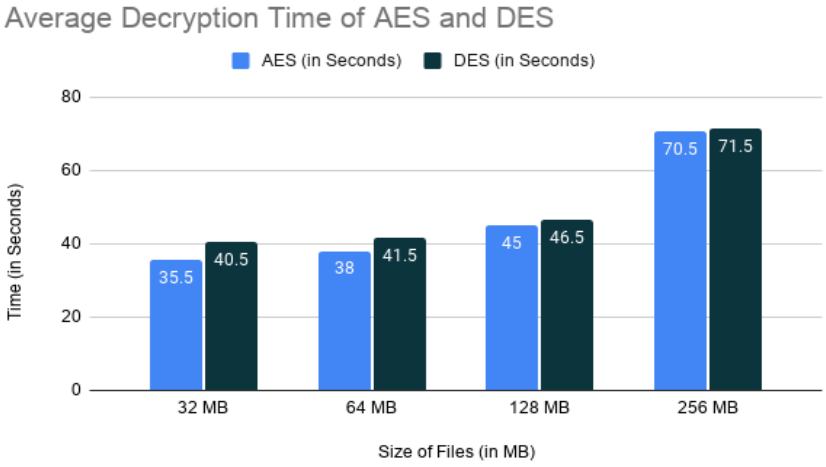

Figure 2. Average Decryption Time of AES and DES

As a result of above comparison and analysis of big healthcare data, AES plays an important role for the security purpose of medical data of patients. Patients freely use internet services and its web app for the better and advanced treatment for diagnose themselves. With the Internet devices, remote location diagnose and consultancy of high level diseases become possible for those who not travel so far without the fear of theft of their data online.

VI. CONCLUSION

Encryption algorithms plays a vital role in information security. However, in this study healthcare big data processed by MapReduce technology, which considered as a very efficient technology for large and huge data processing on various clusters paralleled. Encryption and Decryption of Big Data analyzed due to leakage or unauthenticated access of Healthcare data of patients. By comparing and analysis the processing results of AES and DES from the above computation and evaluations, time taken for encryption and decryption on different sizes of big data file , it is clearly defined that AES take less timing for encryption and decryption of data files of different sizes while DES takes comparatively more time. Tables and Figures says that time processing of files increases with the size of Big Data files.

VII. REFERENCES

[1]. Ali, O, Shrestha, A, Soar, J., & Wamba, S. F. (2018). Cloud computing enabled Health care Opportunities, Issues, and applications: A systematic Review,. *International Journal of Information Management,*, 146-158.

[2]. Arputhamary, D. B., & Benita, A. (October 2020). Hybrid Encryption of Big Data Security using Improved Ellipitic Curve Cryptography. *International Journal of Computer Science and Mobile Computing*, *9*(10), 83-94.

[3]. Begum, G., & Huq, S. (2020). Sandbox security model for Hadoop system,. *Journal of Big Data, 7*, 10.

[4]. Kapil, G., Aggarwal, A., Attaallah, A., Algarni, A., Kumar , R., & Khan, R. (February, 2020). Attribute based honey-encryption algorithm for securing Big Data: Hadoop Distributed file system perspective. *PeerJ Computer Science*.

[5]. Kareem, S. W., Yousif, R. Z., & Mohammed Jihad, s. A. (2020). An Approach for Enhancing data Cofidentiality in Hadoop. *Indonesion journal of electrical engineering and computer science, 20*(3), 1547-1555.

[6]. Kousalya, K., & Parvez, S. (2018). Effective processing of unstructured data using python in Hadoop map reduce. *International Journal of Engineering & Technology, 7*, 417-419.

[7]. Logunleko, K., Adeniji, O., & Logunleko, A. (February, 200). A Comparative study of symmetric cryptography mechanism on DES, AES and EB64 for information Security. International Journal of Scientific Research in Computer Science and Engineering, *8*(1), 45-51.

[8]. Meena, K., & Sujatha, J. (19 june, 2019). Reduced Time Compression in big data using Map Reduce Approach and Hadoop. *Journal of Medical System, 43*, 239.

[9]. Naisuty, M., Hidayanto, A. N., Harahap, N. C., Rosyiq, A., Suhanto, A., & Hartono, G. M. (2020). Data protection on hadoop distributed file system by using encryption algorithms: a systematic literature review. *Journal of Physics*, *1444*, 012012.

[10]. Naveen, Sharma, R., & Nair, A. R. (2019). IoT-based Secure Healthcare Monitoring System. *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT),*, 1-6.

[11]. Shastri, A., & Deshpande, M. (2020). *Big Data Analytics in Healthcare.* Switzerland: Springer Nature.

[12]. Subramaniyaswamy, V., Vijayakumar, V., Logesh, R., & Indragandhi, V. (2015). Unstructured data analysis on big data using MapReduce. *Procedia Computer Science,, 50*, 456-465.