Advances and Applications in Mathematical Sciences
Volume 21, Issue 12, October 2022, Pages 6859-6869

SECURITY ALGORITHMS IN BIG DATA OVERCLOUD COMPUTING ENVIRONMENT

HRIDAY KUMAR GUPTA and RAFAT PARVEEN

Department of Computer Science and Engineering
KIET Group of Institutions Ghaziabad, India
E-mail: hridayakumargupta@gmail.com

Department of Computer Science
Jamia Millia Islamia
New Delhi, India
E-mail: rparveen@jmi.ac.in

Abstract

Big data and cloud computing both are trending terms in IT industries. We might think that they both do the same thing but, both have a Big Data foundation from which to operate. Big Data refers to a large amount of data. Which, regardless of whether the data is structured, semi-organized, or unstructured, is extremely hard to handle by a single machine Cloud computing, on the other hand, is more than simply a program that systematically not only stores data and programs over the internet utilizing a network of more servers, but also delivers services such as software as a service, platform as a service, and infrastructure as a service. We are relying on third-party service provider on taking services like software, platform, and infrastructure because our data, application, and processes are executing on some third party, therefore, security becomes an issue that what is the confidentiality, integrity, and availability means where my data is stored, whether is being seen or intercepted by some other party and whether we are unable to access data resources. Authorizing security and privacy protection for big data in the cloud environment is one of the challenging and critical research in recent days. Currently, every data is accessed from a cloud environment via the internet. The data warehoused in the cloud should be provided with highly reliable security. To reduce security issues in data are stored in the cloud by encrypted form. Basically, Symmetric, Asymmetric and hash function are the three popular method for encryption and decryption. The purpose of this study is to compare the performance of various symmetric key encryption techniques in a cloud context.

2020 Mathematics Subject Classification: 94A60, 68P25, 62R07, 68M25.

Keywords: Computer science, Mathematical programming, Information and communication, General and overarching topics, Security big data, Cloud Computing, Encryption, Decryption, symmetric encryption.

Received June 8, 2021; Accepted July 12, 2022

Introduction

A. Big Data

The Big Data technology, describes the astonishingly accelerated growth of the amount of data. "According to IBM Big Data Analytics 2019, there are around 294 billion emails sent per day, over 1 billion Google searches per day with 40,000 searches per second, trillions of sensors that monitor, track, and communicate with one another, over 30 petabytes of user-generated data saved, accessed, and analysed, and more than 230 million tweets per day with 7000+ tweets per second. By 2020", at least a third of all data will be stored in the cloud [1]. We live in a time when data is being generated at an ever-increasing rate. Massive volumes of data, in the terabytes and petabytes, are generated in real time, making it challenging to access, store, and analyse all structured, unstructured, and partly structured heterogeneous and sophisticated data. For distributed and real-time processing, traditional approaches are also problematic. We require a secure system that can protect the integrity, confidentiality, and availability of processed heterogeneous data while also converting it into useful information. In order to secure big data, techniques such as logging, encryption, and honeypot detection must be necessary. [13]

B. Cloud Environment

For effective management, a cloud environment provides on-demand resource access from a common pool of resources such as hardware and software. By handing along user data to the general public [3], the data owner's control over data is reduced in a cloud environment. Maintaining control of data in transit or at rest within networks provides more benefits for data security. [7]

C. Cloud Security

To comprehend the securities offered by cryptography in the cloud environment, we consider Confidentiality, Integrity, Availability (CIA Triad). Confidentiality allows authorized user to access sensitive and protected data. Integrity is an unauthorized modification of sensitive data is noticeable. [8] This means that changes must be made by an authorized entity through an authorized mechanism. Availability ensures data must be available to the

authorized user. The problem with security and privacy in everyday life could be solved or could be minimized by the use of Big Data (BD) analysis tools and services [14]. In [21, 22, 23, 24], the researchers define the cloud as "A computing environment which provides IT-enable services to in the form of pay-as-you policy to the users"

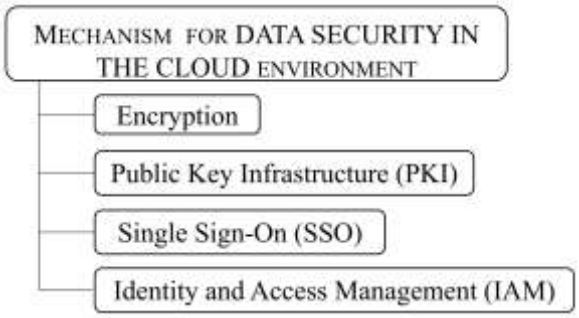

Figure 1. Data Securities mechanism in Cloud Computing.

II. Methods for Data Security in the Cloud

There are four mechanism to provide cloud securities, Encryption, Public Key Infra Structure (PKI), Single Sign-On (SSO) and Identity and Access Management (IAM).

A. Encryption

Encryption refers to the process of transforming genuine information (Plain Text) into garbage text (Cipher Text). It also refers to hiding or locking genuine information in another form.

B. Public Key Infrastructure (PKI)

PKI is a frame using two asymmetric key encryption for communication. It provides authentication using digital certificate and confidentiality using encryption of transmission.

C. Single Sign-On (SSO)

Single Sign-On may be a method of authentication that permits a user to access many apps with only one set of login credentials. Once you've got checked in, you will not need to log in again for every program connected to the system. Google may be a exemplar of single sign-on. Once we sign up to

Google, [9] we are automatically logged into all of Google's products. Sign-On may be a process of authentication that permits a user to access multiple applications with one set of login credential, once you logged in, you don not login repeatedly for each application linked to the system, a typical example of single sign-on is google. Once we logged in google then automatically log into the varied product of google.

D. Identity and Access Management (IAM)

IAM ensures secure and suitable access to resources distributed across diverse technology environment and meet increasingly savior compliance requirement in an organization. According to Wikipedia IAM is a security discipline that enable the right individual to access the right resources at right time for the right reason.

III. Classification of Cryptography Approach

The experimental investigation of Symmetric Key encryption was the emphasis of this research. The symmetric key cryptography techniques AES, DES, Triple DES, RC5, and Blow fish are based on [4, 5]. All the algorithms are based on block cipher. Table 1 lists the important properties of selected algorithms (AES, DES, Triple DES, RC5, and Blow-fish) for investigation as classified in Figure 2.

A. Data Encryption Standard (DES)

The National Institute of Standards and Technology (NIST) published the Data Encryption Standard (DES), which is a symmetric key block cypher (NIST). [2] This algorithm encrypts and decrypts data using symmetric block cypher. The two primary building blocks for such standards are encryption and a secret key. Algorithms are a complex procedural technique that governs the methods for converting plaintext to encrypted text. [11] The key is a random bit sequence that will be used in the method. Two users can communicate through encryption if they use the same algorithm and key. The receiver and sender utilize the same key in some encryption circumstances, while in others, they must use different keys for encryption and decryption.

B. Triple-Data Encryption Standard (Triple-DES)

After 1990, the speed of exhaustive key searches against DES started to irritate DES users. Users, on the other hand, did not want to replace DES

since changing encryption algorithms that are extensively used and incorporated in major security systems takes a significant amount of time and money. [10] The realistic approach was to alter the way DES.

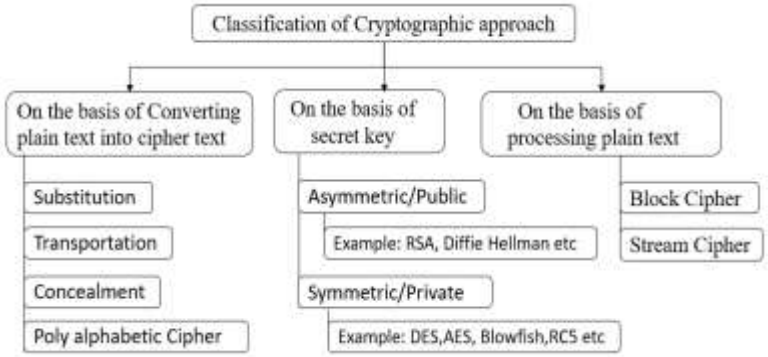

Figure 2. Data Securities mechanism in Cloud Computing.

C. Advanced Encryption Techniques (AES)

The American Encryption Standard (AES) is a symmetric cypher standard that the US government approved in 2002 to replace triple DES. In AES blocks of information are shuffled through multiple rounds of bit shifting, swapping and multiplying. AES key comes in three different length AES-128, AES-192 and AES-256. The least secure model is AES-128-bit key. AES-256 model provides highest securities but choosing a model is depends on various parameters like speed. Longer key may make the algorithm slower. The below graphical diagram describes the AES. [12] Plain text XORed with cypher key and go through various rounds depends on which model is adopted. Each round is blended with four transformation named as sub byte, shift rows, mix column, and add key round. Even Triple DES has outlived its usefulness and has been replaced with the Advanced Encryption Standard (AES) that we should all be utilizing since 2002. [2]

D. Blowfish techniques

Blowfish design in 1993 by Bruce Scheneier. It is also a symmetric key and block cipher encryption with feistal architecture. An alternative algorithm for DES and IDEA algorithms. Blowfish employs a huge number of subkeys. Before data encryption or decryption, the key must be calculated. Blowfish is a Feistel network algorithm that consists of 16 rounds (Feistel Network). [6]

Blowfish a block cipher, meaning that during encryption and decryption it divides a message into fixed length blocks. Padding concept will ne applicable if fixed length is not in 8 bytes. The HDFS storage level data are more secure by blowfish algorithm.

The processing of data using MapReduce is also applied to Blowfish algorithms. It is the fastest to encrypt and decrypt data, using the Hadoop cluster and with the help of MapReduce, Parallel processing of map and reduce function cost is reduced by the Blowfish encryption and decryption. Blowfish is also faster than DES and has a higher encryption rate. The following graphical diagram can be used to explain the entire encryption procedure. There are primarily three stages. In step one, we generate the subkey, then in step two, we initialize the substitution package, and in step three, we perform encryption, which is divided into two parts: Round and Pre-processing.

E. RC-5 Techniques

Rivest Code is referred known as RC5. The RC5 encryption technique was created by Ron Rivest in 1994 and uses a symmetric key block. It's noteworthy since it's small, fast (for basic computer operations like XOR, shift, and so on), and uses less memory. RC5 is a block cypher that processes a data block in one go and treats two-word blocks simultaneously. It requires less memory producing fast encryption and decryption due to using primitive operation like addition XOR and shift operation. It also has variable length of rounds (0 to 255) and key bits. RC5 is most suitable to secure the for modern processor like RICS architecture and devices having less memory. The method is composed of a series of iterations known as rounds r, each of which takes a different value. The initial input text or plain text, as well as the output cypher at the completion of encryption, are stored in two 32-bit registers A and B. We load plain text into registers A and B first, then apply encryption and decryption functions to it.

IV. Experimental Result and Discussion

The simulation experiment was performed on Param Shavak Desktop Supercomputer having multiple high end GPU accelerator cards, Dual server grade intel Xeon processor having 96 GB RAM and 16 TB storage. Multiple

different file types (i.e., audios, videos, text files) and size were randomly selected as dataset for experiment. The above-stated encryption algorithms were executed and were analysed for different packet sizes by taking into consideration the stimulation time for Encryption and Decryption only. The Python programming language was used to carry out the experiments. The five personalized programs are as follows: AESEnDnTime.py, DESEnDnTime.py, BlowfishEnDnTime.py TripleDESEnDnTime.py, RC5EnDnTime.py are executed to record the execution time (milliseconds) during the execution. Further observations from the experimental results: AES encryption increases the data security for cloud-based application. When using the key as 128-bit AES then it is not possible to decide the private key even if attacker has determined public key. When the end user logins to the web portal of cloud then application may be accessed but he will be unable to logout and leaves the session. The encryption time of the existing algorithms is depicted in the figure 8. The packet size is represented on the x-axis in KB, and the encryption time is represented on the y-axis in milliseconds. According to the experimental observation filled in Table 1. The blow fish algorithm has the minimum mean, median and standard deviation. In comparison to other encryption techniques, the Blow fish algorithm requires less operations to finish. Other encryption techniques, such as the Data Encryption Standard, are slower (DES) than blow fish techniques. Blow fish's key schedule is lengthy, yet this can be useful because brute force attacks are more difficult. If Throughput defined as Total Plain text in Kilo Byte /Encryption Time, then blow fish has optimal throughput. The Avalanche Effect occurs when a change in one bit of plain text or one bit of the key schedule causes a change in several bits of the cipher text, again the blow fish has minimum avalanche effect [10].

Table I. Encryption time of existing algorithms with different Packet size in KB.

Encryption Time (Milli Second)					
Packet Size(KB)	AES	DES	Blowfish	Triple DES	RC5
96	23.13	53.99	10.94	45.94	891.95
163	35.95	84.11	54.99	52.94	1333.06

380	86.06	102.94	101.94	136.95	3005.82
1163	137	405.93	210	758.93	9321.47
14373	407.93	1078.89	324.92	1451.86	118570.53
18388	501.95	4184.93	516.31	2184.39	142917.6
73292	2385.88	5301.76	1868.83	8371.5	589662.05
100694	4076.68	7795.61	2806.78	11378.86	864778.84
142359	6440.14	9592.03	3820.77	14659.85	1170950.88
355621	30821.87	27734.49	9145.16	37393.83	2777846.39
969573	646031.69	110848.64	26714.34	571424.06	6777846.39

Figure 3. Encryption time for various techniques with different packet size.

According to the experimental observation filled in Table 2 below, the blow fish algorithm has the minimum mean, median and standard deviation. In comparison to other decryption techniques, the Blow fish algorithm requires less operations to finish.

Other decryption techniques, such as the Data Encryption Standard, are slower (DES) than blow fish techniques. Blow fish's key schedule is lengthy, yet this can be Useful because brute force attacks are more difficult. If Throughput defined as Total Plaint-ext in Kilo Byte/decryption Time, then blow-fish has optimal throughput. When a change in one bit of plain text or one bit of the key schedule induces a change in numerous bits of the ciphers text, the Avalanche Effect occurs; the blow fish, once again, has the least avalanche effect [10]. The Decryption time of the existing algorithms is depicted in the diagram below. The packet size is represented on the x-axis in KB, and the decryption time is represented on the y-axis in milliseconds in figure 4.

Table II. Decryption time of existing algorithms with different Packet size in KB.

Decryption Time (milli Second)					
Packet Size(KB)	AES	DES	Blowfish	Triple DES	RC5
96	4.95	6.99	3.95	11.94	603.86
163	20.36	8	6.94	17.94	1168.84
380	24.03	18.88	17.94	144.59	2644.98
1163	37.95	69.94	35.94	530.93	9137.47
14373	152.94	652.98	363.92	1513.01	108828.15
18388	232.11	786.94	471.93	1929.83	132853.84
73292	1487.97	3522.2	1778.84	7726.51	538266.92
100694	2396.86	4614.05	2450.46	10855.04	769486.68
142359	4850.14	9177.98	3510.28	17051.83	1076702.77
355621	10283.05	19435.02	8559.57	37758.81	1383918.87
969573	265296.97	62281.45	20656.76	48792.33	1691134.96

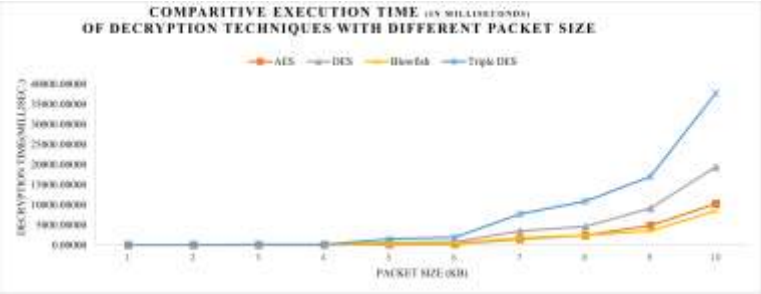

Figure 4. Decryption time for various techniques with different packet size.

Conclusion

The importance of the encryption algorithm in cloud security cannot be overstated. The performance of existing encryption approaches such as AES, DES triple DES, and Blowfish algorithms was investigated in our research. Based on the results of experiments with various file sizes and different types of data, it was determined that Blowfish consumed the least amount of

encryption time for huge files. It was also discovered that the AES algorithm's decryption is superior to those of other algorithms. It will involve image and audio data tests, with the goal of improving encryption and decryption times. Because of the variable length key, Blowfish is faster, smaller, easier to use, and more secure than DES. On huge text data files, Blowfish algorithms are almost equal (size greater than 10,000 bytes). Based on packet size during execution, the experimental results clearly show that Blowfish is a better alternative to AES and DES. As a result, using Blowfish for data constraint functions should be suitable for security implementation.

References

[1] H. K. Gupta and R. Parveen, Comparative Study of Big Data Frameworks, International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), (2019).

[2] D. Morthy, Computer Security, National Institute of Standards and Technology, Gaithersburg NIST Special Publication, (2013).

[3] M. Tajammul and R. Parveen, Key Generation Algorithm Coupled With DES for Securing Cloud Storage, International Journal of Engineering and Advanced Technology, (2019).

[4] S. Heron, Advanced Encryption Standard (AES), Network Security, (2009).

[5] G. N. Krishnamurthy, D. Ramaswamy and M. Leela, Performance Enhancement of Blowfish Algorithm by Modifying Its function, (2007).

[6] M. Yakoubov, M. Sophia and Gadepally, A survey of cryptographic approaches to securing big-data analytics in the cloud, (2015).

[7] S. Yakoubov, V. Gadepally, N. Schear, E. Shen and A. Yerukhimovich, A survey of cryptographic approaches to securing big-data analytics in the cloud, 2014 IEEE High Performance Extreme Computing Conference (HPEC), (2014).

[8] N. Islam and M. Riyas, Analysis of various encryption algorithms in cloud computing, International Journal of Computer Science and Mobile Computing, IJCSMC, (2017).

[9] M. Tajammul and R. Parveen, Key Generation Algorithm Coupled With DES for Securing Cloud Storage, International Journal of Engineering and Advanced Technology, (2019).

[10] M. Thomas and S. V. Athawale, Study of Cloud Computing Security Methods: Cryptography, SSRG International Journal of Computer Science and Engineering, (2019).

[11] B. Thakkar and B. Thankachan, A Survey for Comparative Analysis of various Cryptographic Algorithms used to Secure Data on Cloud, IJERT, (2020).

[12] G. Manogaran, C. Thota and M. Vijay Kumar, Meta Cloud Data Storage Architecture for Big Data Security in Cloud Computing, Procedia Computer Science, (2016).

[13] C. Stergiou, E. Psannis, B. Brij and Y. Ishibashi, Security privacy efficiency of sustainable Cloud Computing for Big Data IoT, Sustainable Computing: Informatics and Systems, (2018).

[14] Z. Tan et al., Enhancing Big Data Security with Collaborative Intrusion Detection in IEEE Cloud Computing, (2014).

[15] S. Rallapalli, R. R.Gondkar and U. Pavan Kumar Ketavarapu, Impact of Processing and Analyzing Healthcare Big Data on Cloud Computing Environment by Implementing Hadoop Cluster, Procedia Computer Science, (2016).

[16] U. Narayanan, V. Paul, S. Joseph, A novel system architecture for secure authentication and data sharing in cloud enabled Big Data Environment, Journal of King Saud University - Computer and Information Sciences, (2020).

[17] A. Alabdulatif, I. Khalil and Xun Yi, Towards secure big data analytic for cloud-enabled applications with fully homomorphic encryption, Journal of Parallel and Distributed Computing, (2020).

[18] H. S. Yahia, S. R. Zeebaree, M. A. Sadeeq, N. O. Salim and H. A. Hussein, Comprehensive Survey for Cloud Computing Based Nature-Inspired Algorithms Optimization Scheduling, Asian Journal of Research in Computer Science, (2021).

[19] M. U. Sana, Z. F. Javaid, H. B. Liaqat and M. U. Ali, Enhanced Security in Cloud Computing Using Neural Network and Encryption, in IEEE Access, (2021).

[20] J. Huang Z. C. Gao and K. Chen, Privacy preserving outsourced classification in cloud computing, Cluster Computing, (2018).

[21] K. Yang, X. Jia, K. Ren, B. Zhang and R. Xie, DAC-MACS: effective data access control for multiauthority cloud storage systems, IEEE Trans Inf. Forensics Secure, (2013).

[22] M. Shucheng Yu, Y. Zheng, K. Ren and W. Lou, Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption, IEEE Trans Parallel Distrib System, (2012).

[23] J. Yang and Y. Niu, A hybrid solution for privacy preserving medical data sharing in the cloud environment, Future General computer System, (2015).