How to Cite:
Sumithra, R., & Parameswari, R. (2022). Data privacy and data protection security algorithms for big data in cloud. *International Journal of Health Sciences*, *6*(S2), 7613–7621. https://doi.org/10.53730/ijhs.v6nS2.6834

Data privacy and data protection security algorithms for big data in cloud

Ms. R. Sumithra
Research scholar, Department of Computer Science, School of Computing Sciences, VISTAS, Chennai, India
Corresponding author email: sumithraca@gmail.com

Dr. R. Parameswari
Professor, Department of Computer Science, School of Computing Sciences, VISTAS, Chennai, India
Email: dr.r.parameswari16@gmail.com

***Abstract*---**Data security is the primary concern in the world, while storing large and confidential data in the cloud. Cloud Computing is a powerful internet-oriented technology which gives demand-based services to the users. But security is the major problem here when we use cloud computing networks. So, using secured cryptography algorithms only will protect data from attackers and gives security to the big data in cloud. It is mainly used for transmission of data in secured format. This paper proposed new enhanced hybrid security algorithm to give better security for Big Data in cloud. They are Hybrid Matrix hash key generation algorithm (HMHA) and Anonymous Multi-Fusion Shuffling Algorithm (AMSA).

***Keywords*---**encryption, decryption, cryptographic, big data, cloud.

Introduction

Cloud Computing is a very rapid growing technology in the world nowadays. Because many large companies showing interest as well as getting benefits when using cloud computing to store and process their large confidential data in the cloud. Cloud Computing not only giving storage services, but it also provides different services based on demand to the user like IAAS, PAAS, SAAS. It is mainly opted to big data when the user wants to store large complex data like structure, unstructured, semi structure data, Cloud Computing provides best services to the user, rather than typical data base services for big data in cloud computing, Big data and cloud computing both are very rapid emerging technology in software industries.

Literature Review

Marin Matsumoto, Masato Oguchi et al. (2020), introduced Speeding Up Sensor Data Encryption with a Common Key Cryptosystem combined with Fully Homomorphic Encryption on Smartphones. In this paper proposed combination of common key cryptosystems and fully homomorphic encryption, which is a public key cryptosystem, to perform high-speed encryption and statistical analysis on cloud services even on devices with limited computational resources. Dilip Venkata Kumar Vengala et al. (2020), describes uploaded user data can well be amassed in disparate CS to ameliorate the user data's security. It can be done by, first, extracting the features of the user data along with the CS. After that, the Hybrid Meerkatclan algorithm (HMCA) chose the CS, which is optimally centered on the features. Next, the SHA512 algorithm performs the data deduplication of the input user data. Subsequent to data deduplication, the input file is compressed and encrypted using a two-stage lempel–ziv algorithm and optimized CP-ABE-ECC algorithm. E. K. Subramanian ,Latha Tamilselvan et al. (2020), uses an Elliptic curve with Diffie–Hellman (ECDH) algorithm for encryption and decryption of data to improve the data security in the cloud. This algorithm reduced the computational complexity and encrypted data efficiently. In experimental analysis, the performance of proposed ECDH is calculated using evaluation parameters such as encryption time, decryption time, computation overhead and key generation time. Yoshita Sharma, Himanshu Gupta, Sunil Kumar Khati et al. (2019), Introduced an enhanced A Security Model for the Enhancement of Data Privacy in Cloud Computing and in this research paper the use of multiple encryption technique outlines the importance of data security and privacy protection.

Related Works

In this related work analyzed and implemented numerous algorithms and various investigation parameters among the several cryptographic algorithms used for Big Data in cloud computing security such RSA, AES, DES, Blowfish, Homeomorphic and DLPK algorithm. Proposed system implemented based on the existing algorithms pros and cons and new methodologies overcomes all the draw backs of the existing system and gives best result in all the aspect of the metrics and parameter for the performance evaluation result.

Proposed System

Data storage in the cloud is becoming the major security issues in nowadays. For this, we trust third party data storage in the cloud is becoming the major security issues in cloud nowadays. because here for data persist in the cloud trusting third party is not advisable for data private hand security. In the proposed system provides the security to their own data in the cloud data may be very large and complex but using this new methodology giving best security data in the world cloud. In Proposed Systems Increased the Block Size to provide higher level security for data in the cloud. It provides high security in three proposed new hybrid security algorithms. These new methodologies algorithms process the data quickly rather than other algorithms Proposed methodologies will be giving unbreakable security to the user data in the cloud by using newly implemented

hybrid algorithms. So that intruders cannot break the data in the cloud on account of high security.

Hybrid Matrix Hash Key Generation Algorithm (HMHK) Algorithm

HMHK is the new methodology algorithm for providing best security to the data in the cloud. This Proposed new algorithm is Hybrid Matrix Hash Key Generation Algorithm (HMHK).This algorithm is used hybrid model concept. It is combination of AES and newly implemented Dynamic key cast grid counterchange algorithm. It will be using to protect the data with the help of dynamic secret key and on top of that secret key using matrix manipulation and Transpose technique, generating new key for encryption and decryption. It increases the Block size as 2048 bits of the data to give better security.

HMHK algorithm pseudo code has been given below.

sk=generatedsecretkey();
ANS_value= shuffledDynamicGeneratedAlphaNumericSpecialCharacter();
|A|=formDynamicMatrix(ANS_value);
$| A^T |=|A|$
k=encryptkey(sk,$| A^T |$)
CT=encrypt(PT,k);
sk=encryptkey(sk,$| A^T |$)
PT=decrypt(CT,sk);

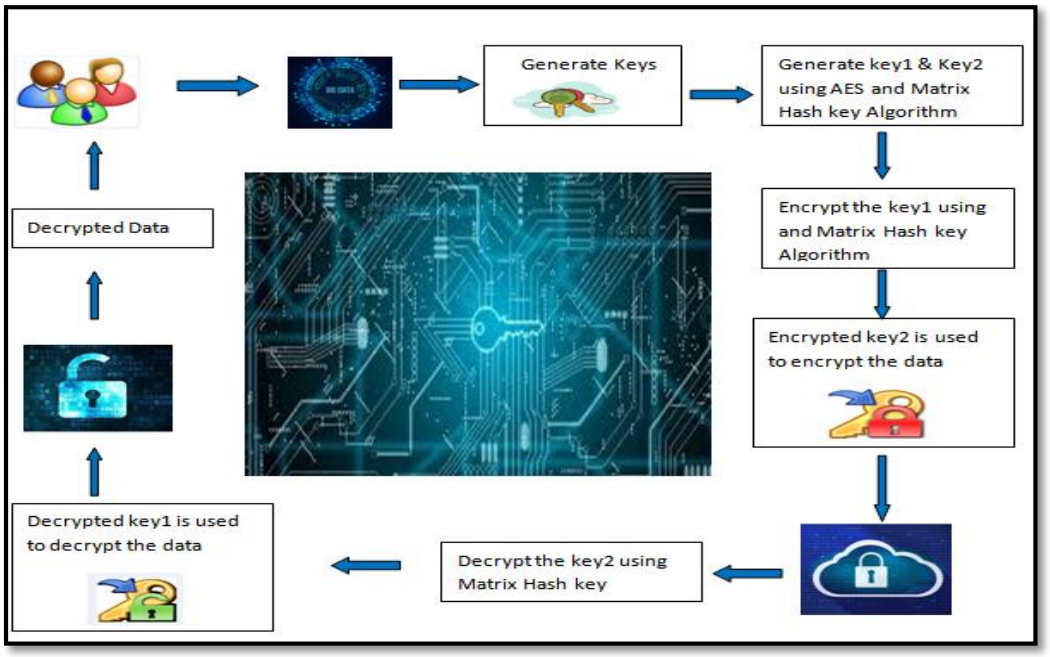

Fig 1: HMHK Algorithm

Anonymous Multi - Fusion shuffling (AMSA) Algorithm

AMSA is the new methodology algorithm for providing best security to the data in the cloud. This Proposed new algorithm is Anonymous Multi - Fusion shuffling Algorithm. It is the fastest as well as secured algorithm rather than other algorithm. Irrespective of specific algorithm, it is employing various encryption algorithm in one new proposed system. It supports various encryption algorithm such as DES, Blowfish, AES, RSA, Homomorphic, DLPK, HMHA, AMSA. When encrypt the data, it will select one algorithm randomly using the Anonymous Multi - Fusion shuffling Algorithm with Random Selector List then it will do the encryption and decryption over the data and store that data in cloud.

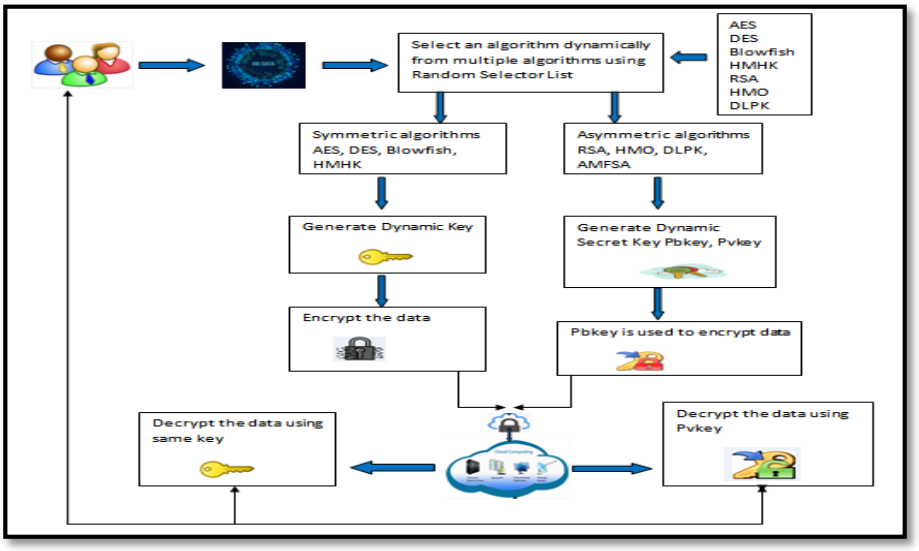

Fig 2: Anonymous Multi - Fusion shuffling (AMSA)

Result and Discussions

The efficiency of the existing algorithms and new methodologies of the proposed algorithms derived based on the processing time of the original, cipher text data and compared with different metrics and parameter.

HMHK - Encryption Time

Table1: Encryption time compared with file size and Milliseconds

Algorithms	DES (ms)	Blowfish (ms)	HMO (ms)	RSA (ms)	AES (ms)	HMHK (ms)
500 B	10000	5000	10000	5000	1800	1500
1 KB	20000	6000	70000	980000	7000	3300
5 KB	30000	10000	900000	980000	10000	6100
10 KB	100000	20000	950000	980000	15000	11000
100 MB	500000	40000	990000	990000	45000	35000
125 MB	500000	450000	990000	990000	50000	36000

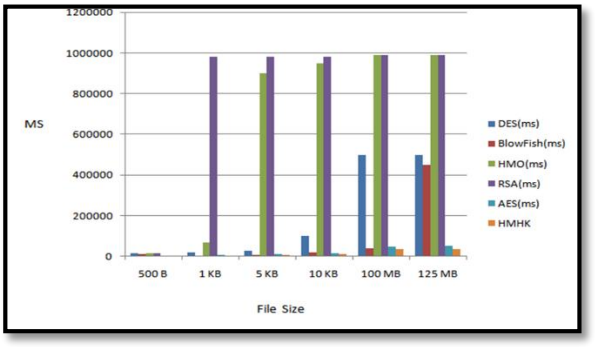

Fig 3: Encryption time compared with file size and Milliseconds

HMHK - Decryption Time

Table 2: Decryption Time compared with file size and Milliseconds

Algorithms	DES(ms)	lowfish(ms)	HMO(ms)	RSA (ms)	AES (ms)	HMHK(ms)
500 B	10000	5000	10000	3000	1500	1000
1 KB	19000	6000	70000	4000	7000	2200
5 KB	25000	12000	77000	8000	9000	4500
10 KB	30000	20000	89000	16000	15000	11000
100 MB	50000	40000	91000	80000	35000	28000
125 MB	46000	40000	99000	89000	37000	29000

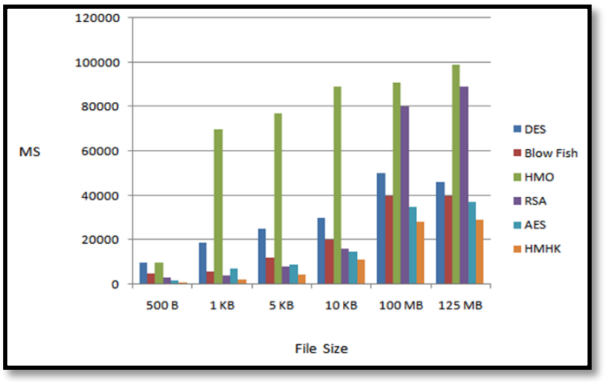

Fig4: Decryption time compared with file size and Milliseconds

HMHK - Throughput Time

Table 3: Throughput encryption time compared with file size and Milliseconds

Algorithms	DES(ms)	Blowfish(ms)	HMO(ms)	RSA(ms)	AES(ms)	HMHK(ms)
500 B	0.05	0.01	0.05	0.01	0.31	1.20
1 KB	0.512	0.17	0.15	0.01	0.146	1.21
5 KB	0.17	0.51	0.005	0.005	0.512	1.49
10 KB	0.15	0.51	0.01	0.01	0.68	1.65

100 MB	0.21	0.25	0.103	0.103	2.27	2.12
125 MB	0.256	0.285	0.103	0.103	2.56	3.15

Table 4: Throughput decryption time compared with file size and Milliseconds

Algorithms	DES	Blow Fish	HMO	RSA	AES	HMHK
500 B	0.05	0.1	0.05	0.16	0.33	0.42
1 KB	0.05	0.17	0.01	0.25	0.14	0.40
5 KB	0.2	0.42	0.06	0.64	0.56	0.86
10 KB	0.34	0.51	0.11	0.64	0.68	0.74
100 MB	1.04	2.56	1.12	1.28	2.92	3.26
125 MB	1.78	3.2	1.29	1.43	3.45	3.49

Comparison result of the existing algorithms such as DES, AES, Blowfish, HMO, RSA and the new HMHK Algorithm, HMHK algorithm gives the best result in all the metrics aspects like encryption time, decryption time, Throughput and memory consumption. Based on the above result and discussion, HMHK algorithm provides the elevated security and reduced the processing time.

AMSA - Encryption Time

Table 5: Encryption time compared with file size and Milliseconds

Algorithms	DES(ms)	Blowfish(ms)	HMO(ms)	RSA(ms)	AES(ms)	AMSA(ms)
500 B	10000	5000	10000	5000	1800	1200
1 KB	20000	6000	70000	980000	7000	3100
5 KB	30000	10000	900000	980000	10000	5400
10 KB	100000	20000	950000	980000	15000	10000
100 MB	500000	40000	990000	990000	45000	33000
125 MB	500000	450000	990000	990000	50000	34000

AMSA - Encryption Time

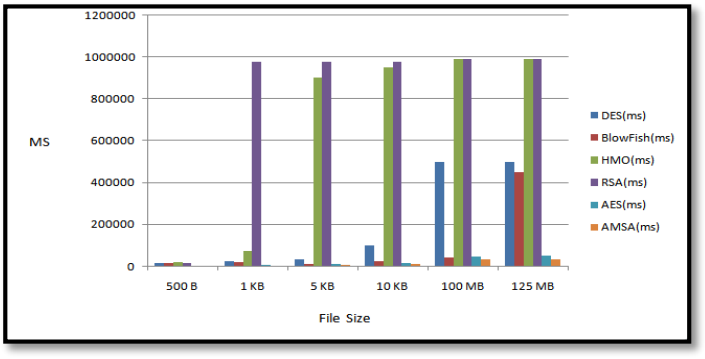

Fig5: Encryption time compared with file size and Milliseconds

Table 6: AMSA Decryption time compared with file size and Milliseconds

Algorithms	DES(ms)	Blowfish(ms)	HMO(ms)	RSA(ms)	AES(ms)	AMSA(ms)
500 B	10000	5000	10000	3000	1500	900
1 KB	19000	6000	70000	4000	7000	2000
5 KB	25000	12000	77000	8000	9000	4000
10 KB	30000	20000	89000	16000	15000	10000
100 MB	50000	40000	91000	80000	35000	25000
125 MB	46000	40000	99000	89000	37000	26000

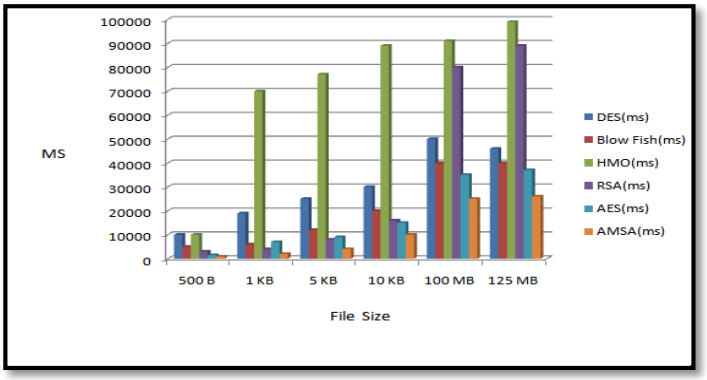

Fig6: AMSA Decryption time compared with file size and Milliseconds

AMSA – Throughput Encryption Time

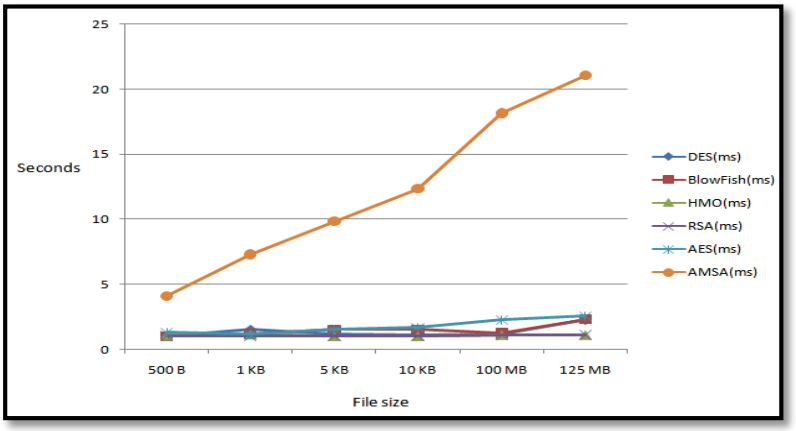

Fig7: Throughput encryption time compared with file size and Milliseconds

AMSA – Throughput Decryption Time

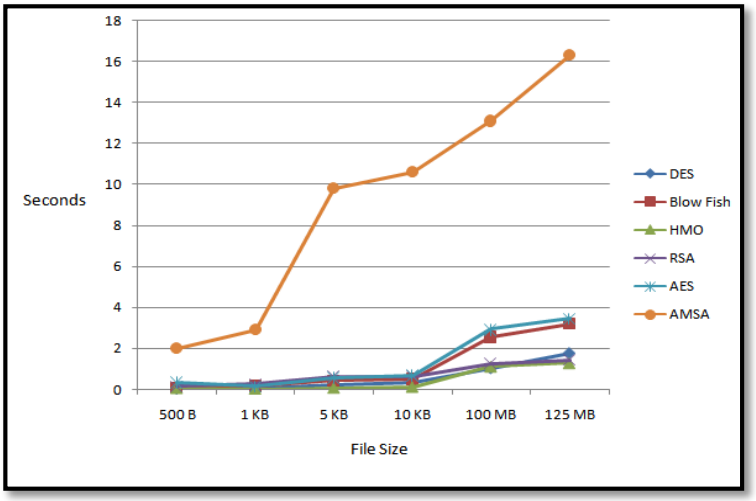

Fig8: Throughput Decryption time compared with file size and Milliseconds

Comparison result of the existing algorithms such as DES, AES, Blowfish, HMO, RSA ,the new HMHK Algorithm and the new AMSA Algorithm , AMSA algorithm gives the best result in all the metrics aspects like encryption time, decryption time, Throughput and memory consumption. Based on the above result and discussion, AMSA algorithm provide the elevated protection over the data in cloud and make available defense mechanism against the intruder's and reduced the processing time.

Conclusion

A traditional security gives many security policies and controls over the data in cloud, even though but is still inadequate for holistically tackling the security challenges established by Big Data. In order to address these problems modeled by statistics accumulation, establishment's requirement discovery new ways to safeguard their critical tools, methods, and measures used to obtain, preserve, and evaluate data of certain apprehension to expand the strength of its safety organization, as most commercially available security package have access to all material and consequential data available on the system. The additional protection consists of multiple features and offering by the third party are not valuable in the Big Data environment. Additionally, experts and other operators of information technology need more successful protection training that will help them identify with the specific hazards they deal with, that how their defense ranges will affect the consequences for those newly introduced methodologies gives best security to their own data by giving own algorithms that resolves our securities issues in the cloud.

References

1. N. Miloslavskaya, A. Tolstoy, and S. Zapechnikov, “Taxonomy for unsecure big data processing in security operations centers,” in Future Internet of Things and Cloud Workshops (FiCloudW), IEEE International Conference on, pp. 154–159, 2016.
2. N. Chaudhari and S. Srivastava, “Big data security issues and challenges,” in Computing, Communication and Automation (ICCCA), 2016 International Conference on, pp. 60–64, 2016.
3. M. Paryasto, A. Alamsyah, B. Rahardjo, et al., “Big-data security management issues,” in Information and Communication Technology (ICoICT), 2014 2nd International Conference on, pp. 59–63, 2014.
4. Cho C, Baek M, Won Y (2018) Guaranteeing the integrity and reliability of distributed personal information access records. J Ambient Intell Hum Comput. https://doi.org/10.1007/s12652-018-0871-7
5. Parmar RR, et al. Large-Scale Encryption in the Hadoop Environment: Challenges and Solutions. IEEE Access 2017;5:7156–63. https://doi.org/10.1109/ ACCESS.2017.2700228.
6. Yoder M. Cloudera's process for handling security vulnerabilities. Available: https://blog.cloudera.com/blog/2016/05/clouderas-process-for-handling-securi ty-vulnerabilities/. [Accessed 8 January 2019].
7. Raj Samani CB. McAfee threats report 2018. Available: https://www.mcafee.com/e nterprise/en-us/assets/reports/rp-quarterly-threats-dec-2018.pdf. [Accessed 2 February 2019].
8. CyberPedia. An overview of DoS attacks. Available: https://www.paloaltonetworks.com/cyberpedia/what-is-a-denial-of-service-attack-dos. [Accessed 20 January 2019].
9. Verizon. Data breach digest: perspective is reality. Available: https://enterprise.ve rizon.com/resources/reports/data-breach-digest/. [Accessed 2 February 2019].
10. Millman R. Thousands of Hadoop clusters still not being secured against attacks. Available: https://www.scmagazineuk.com/thousands-of-hadoop-clusters-still -not-being-secured-against- attacks/article/637389/. [Accessed 5 February 2019].