Securing BIG DATA: A Comparative Study Across RSA, AES, DES, EC and ECDH

I. Bhargavi, D. Veeraiah and T. Maruthi Padmaja

Abstract Now a days cloud computing has emerged as a cost-effective platform for providing IT or business services over the Internet. However, the services provided by the cloud are of third parties, ensuring the security and privacy of the customers. BIG Data is critical at cloud storage. Several security frameworks using cryptographic methodologies have been proposed to address this issue. However, there is no wide comparative study across the cryptographic methods that ensure the security of BIG DATA at the cloud data center. This paper presents a comparative study across fundamental cryptographic methodologies used in securing BIG Data at cloud storage. This work assumes that the cloud storage is erected with HADOOP based data center. In order to carry out comparative study, we have considered ECC, RSA, AES, DES and Elliptic curve Diffie-Hellman to verify the BIG Data security at cloud based data center.

Keywords Big Data · HADOOP · RSA · AES · DES · EC · ECDH

1 Introduction

Big data is a term for datasets that are so large, complex and that cannot be analysed with traditional computing technologies. The quantity of computed data being generated is increasing exponentially from different application sources like retail, logistics and financial databases, social networks, sensors, internet of things etc.

In order to explicate the data and know its characteristics, it is very important to securely store, manage and share the huge amount of complex data. Now this sort

I. Bhargavi (✉) · D. Veeraiah · T. Maruthi Padmaja
Department of Computer Science and Engineering, VFSTR University, Guntur, India
e-mail: bhargaviinduri10@gmail.com

D. Veeraiah
e-mail: d.veeraiah@gmail.com

T. Maruthi Padmaja
e-mail: padmaja.tu2002@gmail.com

S.C. Satapathy et al. (eds.), *Computer Communication, Networking and Internet Security*, Lecture Notes in Networks and Systems 5,
DOI 10.1007/978-981-10-3226-4_36

of facility made available via the distributed platform which is popularly known as cloud computing.

The main feature of cloud computing is on-demand network access to computing resources, on pay per use basis, which are provided by cloud service providers. Common deployment models for cloud computing include Platform as a Service (PaaS), Software as a Service (SaaS), Infrastructure as a Service (IaaS). The PaaS provides platform to customers to develop, run and manage applications without owning the respective infra. The SaaS provides businesses with applications that are stored and run on virtual servers in the cloud.

In the IaaS model, client will pay on a per-use basis for the use of equipment to support computing operations that include storage, hardware, servers and networking equipment.

Security over cloud services is however in its maturing phase, the data in the cloud would be at risk for a large number of security vulnerabilities. The cloud administrators do not have any clue over the data where it is being stored and in what format. Therefore, in this scenario the users must be ensured that proper security measures have been adapted to protect their information mainly from data leakage and data tampering. Further, processing/analysing the huge data at cloud data center is a critical issue. Recently, several distributed frameworks like HADOOP [1, 2], Google File System [3] have been developed for storing and processing the BIG DATA. However, HADOOP distributed framework is quite popular among industry and research communities. HADOOP includes two sets of functionalities, (i) The HADOOP Distributed File System (HDFS) to store large and unstructured datasets, (ii) The Map Reduce framework for processing huge data. Usually, HADOOP works with applications having thousands of nodes and petabytes of data.

Security mechanisms are not incorporated at HADOOP. Several works have been reported the usage of cryptographic algorithms to encrypt the data and store the data at HDFC. Encryption is used to provide security for sensitive information. Encryption algorithm performs various substitutions and transformations on the original message or data and transforms it into cipher text which ultimately becomes a random message. Various cryptographic algorithms are available and used in information security. There are different types of algorithms: (i) Symmetric-key algorithms [4, 5] such as Data Encryption Standard (DES) [6], Triple DES [7] and Advanced Encryption Standard (AES) [7] (ii) Asymmetric-key algorithms [8] such as RSA [7] and Elliptic Curve Diffe-Hellman (ECDH).

This paper is organized as follows. Section 2 depicts the framework for HADOOP based cloud data center. Section 3 presents the work related to securing BIG DATA at HADOOP based cloud data center. Section 4 discusses the performances of encryption and decryption operations performed using different cryptographic algorithms for securing BIG DATA at HADOOP based cloud data center. Finally, Sect. 5 concludes the work.

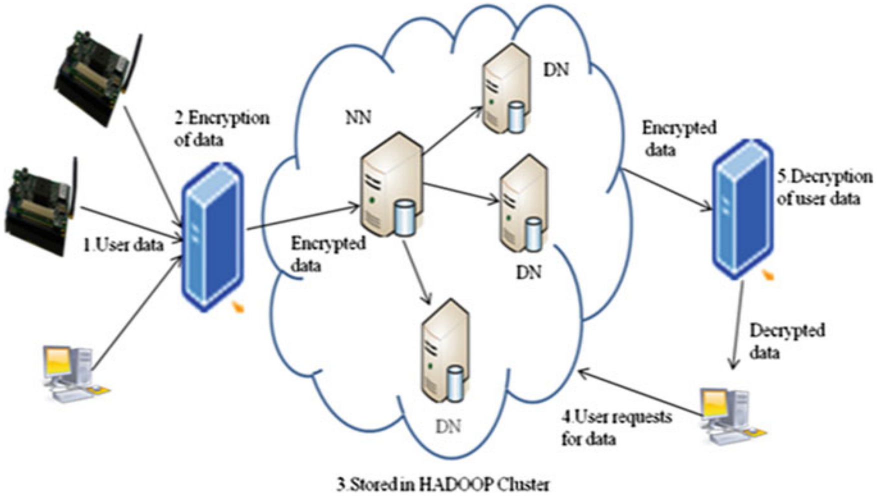

Fig. 1 A HADOOP based cloud data center architecture for big data analytics

2 HADOOP Based Cloud Data Center

As shown in Fig. 1, first, the user data are taken from different sources and encrypted on the server using either symmetric or asymmetric cryptographic mechanisms. Soon after encryption, the data is stored on cloud i.e. it will be stored in a cluster via HADOOP File System (HDFS). In HADOOP, the NameNode (NN) is responsible for the data distribution to DataNodes (DN). Whenever the user requests for data, the encrypted data is given by the server for decryption. Then user takes the encrypted data and decrypts using corresponding keys.

3 Related Work

The HADOOP architecture assumes secure network and hence no security framework is incorporated at base level. As a first step Park and Lee [2] introduced a secure HADOOP framework with AES based encryption/decryption. Research work cited in [9, 10] demonstrated the adoption of Kerberos based authentication mechanism to secure the data in HDFS storage. Zhou and Wen [11] applied 'Cipher Text Policy' and Attribute Based Encryption (CP_ABE) to provide access control credentials for valid cloud users. Here, CP_ABE uses an encrypted data access control structure rather using user's personal identity. The user can perform the decryption provided, the user identification attribute matches with access control structure. In this mechanism the cipher text and corresponding cipher key generated via CP_ABE method are transmitted to the Namenode. The Namenode further re-encrypts the cipher text and distributes the file blocks to Datanodes. Here, the key distribution seems to be simple with less user intervention due to the centralized key distribution which is based on CP_ABE at Namenode. However, the original file is also sent to Namenode for re-encryption. Therefore, the security to the client file is not guaranteed. Cohen and Acharya [12] proposed an AES based New Instruction (AES-NI) encryption framework for data encryption and integrity validation by making use of Trusted Platform Module (TPM). Further, an advanced cryptographic mechanism like homomorphic encryption is also widely used to secure BIG DATA at cloud storage. Using fully homomorphism Jin et al. [13] devised a security mechanism for cloud storage. In this method, agent technology is used for encryption and user authentication. However, fully homomorphic encryption may not be fully applicable to address the real world requirements in Big data Scenario. The hybrid encryption schemes were also devised to secure data at HDFS. Lin et al. [14] proposed a hybrid encryption method where the users' data file is symmetrically encrypted by a unique key *k* and this *k* is then asymmetrically encrypted with the owner's public key.

To encrypt users' files this mechanism uses the DES algorithm initially in order to generate the 'data key'. Later on, RSA is used to encrypt the already generated 'data key' and the user keeps the private key to decrypt the 'data key'. Here Yang et al. [15] assumed that the generated private key using RSA is still vulnerable. Consequently, they have used IDEA (International Data Encryption Algorithm) to further encrypt the secret key. Although, this hybrid encryption method seems to secure the data, it increases the computational complexity to the extent. Saini and Naveen [6] presented a steganography based hybrid scheme to make the encrypted data completely not visible to the outside users.

4 Comparative Study of Cryptographic Algorithms Over HADOOP Based Cloud Data Center

Table 1, depicts the comparison of encryption and decryption of differing file sizes using various cryptographic algorithms. Here the time unit is shown in seconds (s). From our experiments we have observed that RSA could not perform over the files

Table 1 Comparison across RSA, AES, DES, EC and ECDH with respect to encryption and decryption times

Algorithms	Key length (bits)	File size (MB)	Encryption time (SECs)	Decryption time (SECs)
AES	128	50, 100, 150	4, 6, 9	14, 14, 14
DES	56	50, 100, 150	4, 6, 9	14, 14, 14
EC	256	50, 100, 150	7, 15, 18	14, 14, 14
ECDH	256	50, 100, 150	58, 115, 286	180, 360, 555

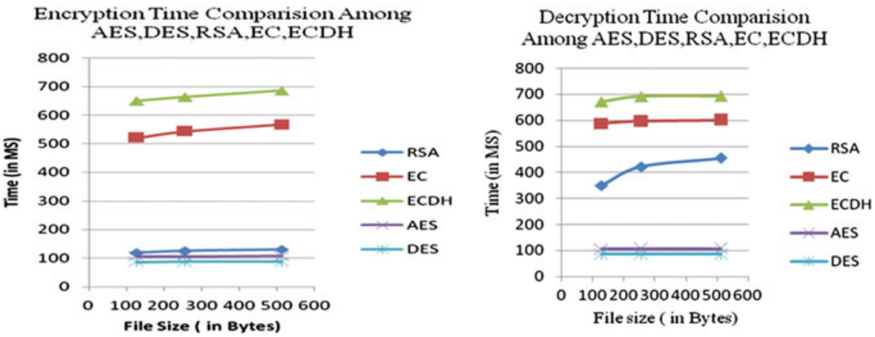

Fig. 2 Encryption and decryption times for cryptographic

with larger sizes (considered file sizes 50, 100 and 150 MB). Hence, the results reported here are in the file sizes 512, 1024, 2048, 4096 Bytes. However, the encryption and decryption using AES and DES algorithms even performed well for the files in smaller sizes (512, 1024, 2048, 4096 Bytes) which is shown in Fig. 2. Thus, AES, DES, EC and ECDH are considered for HADOOP based experiments.

In order to study the behavior of considered algorithms on HADOOP we have considered three different scenarios of HADOOP setting in securing the BIG DATA. The performance of these algorithms with respect to HDFS storage is measured in terms of writing and read speeds of the files with varying sizes in to and from HADOOP's HDFS. Usually HDFS reading takes more time than writing as data is stored in different blocks. Further, HADOOP supports replication factor that reflects how many times the original data can be replicated across HDFS nodes.

4.1 Scenario 1

Same node act as a master and slave of considering replication factor as one. From Fig. 3, it can be observed that as the plain text size increases the time required to write and read to and from HDFS is also increasing. In this scenario encrypted file writing and reading using EC and ECDH yields more time than AES and DES. This is because of larger encrypted files generation by EC and ECDH.

4.2 Scenario 2

By considering a three node cluster, among the three, one node act as Namenode and two others are data nodes. i.e., one master node and two slaves and here the replication factor are considered as one. The writing and reading speeds in this scenario are shown in Fig. 4. Here writing time is increased compared to the first

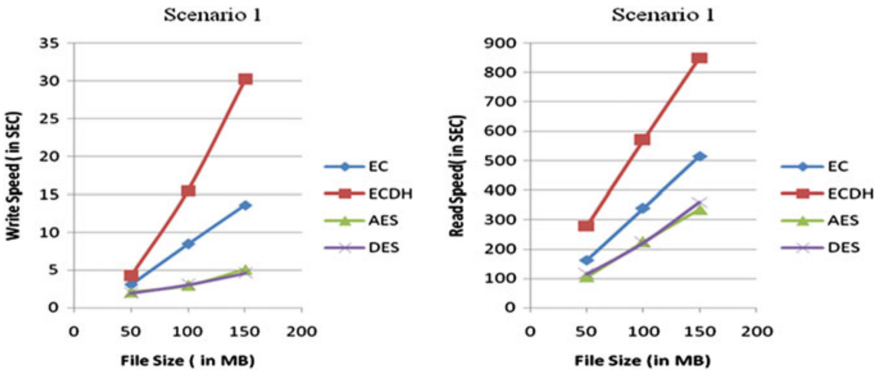

Fig. 3 Write and read speeds in Scenario 1

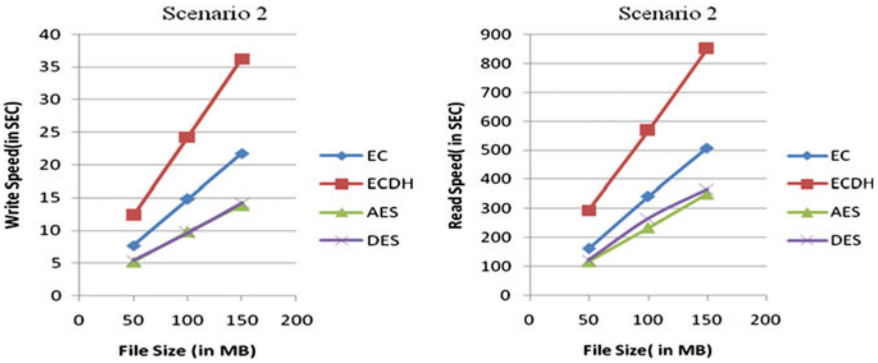

Fig. 4 The write and read speeds in Scenario 2

scenario because in previous scenario all blocks are stored in one node, but in this scenario, data blocks are stored in three different nodes. In this scenario, read speed is nearly same as the previous scenario.

4.3 Scenario 3

This scenario is similar to Scenario 2 except replication factor is set to three. The writing and reading speeds in this scenario are depicted in Fig. 5. In this scenario, the writing time is increased when compared with previous scenarios because of replication factor. In previous scenario only one copy of the data is stored, but here three copies of data are stored due to three replication blocks. Here also the read speed is similar to previous one and two scenarios.

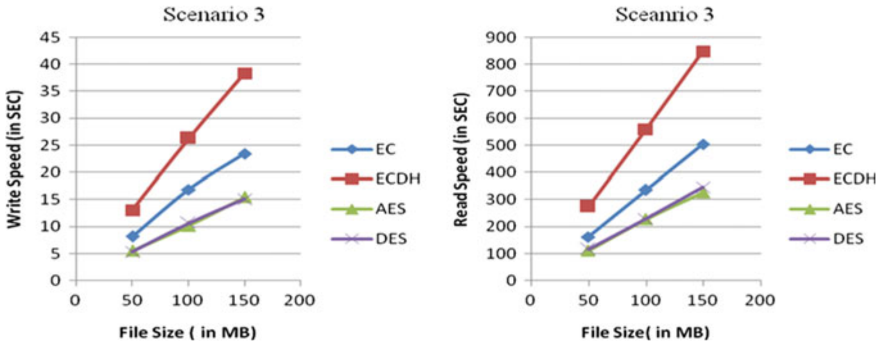

Fig. 5 Write and read speeds in scenario 3

5 Conclusion

To observe the behavior of encryption and decryption operations on BIG DATA in the HADOOP environment, we have used symmetric and asymmetric crypto algorithms over varied file sizes. We could conclude from the experiment that ECDH yielded highest reading and writing speed compared to EC, AES and DES. EC yielded a second highest speed and AES, DES yielded least speed for the same size of file blocks. Further, it is identified that the read speed is consistent even with the increase in replication factor.

References

1. Tom White, “Hadoop: The Definitive Guide”, Third Edition, O’Reilly, (2012).
2. S. Park, Y. Lee, “Secure Hadoop with Encrypted HDFS,” Chapter Grid and Pervasive Computing, Vol. 7861 of the series Lecture Notes in Computer Science, pp 134–141, (2013).
3. Bo Li, Mengdi Wang, Yongxin Zhao, Geguang Pu, Huibiao Zhu, Fu Song “Modeling and Verifying Google File System Modeling and Verifying Google File System”, 16th International Symposium on High Assurance Systems Engineering, Pages: 207–214, IEEE, (2015).
4. Sourabh Chandra, Siddhartha B, Smita Paira. “A Study and Analysis on Symmetric Cryptography” ICSEMR, pp 1–8, IEEE (2014).
5. Rejani. R, Deepu.V. Krishnan, Study of Symmetric key Cryptography Algorithms, Volume 2 Issue 2, pp 45–50, IJCT (2015).
6. Garima Saini and Naveen Sharma, “Triple security of data in cloud computing,” International Journal of Computer Science and Information Technologies, Vol. 5, No 4, pp. 5825–5827, (2014).
7. Behrouz A. Forouzan “Cryptography and Network Security”, Tata McGraw-Hill Companies, (2007).
8. Sourabh Chandra, Sk Safikul Alam, Smita Paira and Goutam Sanyal. “A comparative survey of symmetric and asymmetric key cryptography”, International Conference on Electronics, Communication and Computational Engineering (ICECCE), pp 83–93, IEEE (2014).

9. D. Das, O. O' Malley, Sanjay Radia, Kan Zhang, "Adding Security to Hadoop," Hortonworks Technical Report I, (2010).
10. O. O'Malley, Kan Zhang, Sanjay Radia, Ram Marti, and Christopher Harrell. Hadoop security design. https://issues.apache.org/jira/secure/attachment/12428537/securitydesign.pdf, October, (2009).
11. H. Zhou and Q. Wen, "Data Security Accessing for HDFS Based on Attribute-Group in Cloud Computing," In Proc. of International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2014), pp. 525–528, (2014).
12. 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, pp. 444–451, (2013).
13. S. Jin, S. Yang, X. Zhu, and H. Yin, "Design of a Trusted File System Based on Hadoop," In Proc. of Trustworthy Computing and Services, ed: Yuyu Yuan, Xu Wu, Yueming Lu, pp. 673–680, (2013).
14. 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 Proceedings of 26th International Conference on Advanced Information Networking and Applications, IEEE Computer Society Washington, DC, USA, pp. 740–747, (2012).
15. C. Yang, W. Lin, and M. Liu, "A Novel Triple Encryption Scheme for Hadoop-Based Cloud Data Security," in Proc. of 4th Emerging Intelligent Data and Web Technologies (EIDWT), pp. 437–442, (2013).