Indonesian Journal of Electrical Engineering and Computer Science
Vol. 20, No. 3, December 2020, pp. 1547~1555
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v20.i3.pp1547-1555

An approach for enhancing data confidentiality in hadoop

Shahab Wahhab Kareem[1], Raghad Zuhair Yousif[2], Shadan Mohammed Jihad Abdalwahid[3]
[1,3]Department of Information System Engineering, Erbil Polytechnic University, Iraq
[2]Department of Information Technology, Catholic University in Erbil, Iraq

Article Info

Article history:

Received April 24, 2020
Revised May 31, 2020
Accepted Jun 7, 2020

Keywords:

BigData
Cryptography
ElGamal
Hadoop
RSA

ABSTRACT

The quantity of prepared and stored data in the cloud is rising dramatically. The common storage devices at both hardware and software levels cannot satisfy the necessity of the cloud. This fact motivates the requirement for principles that can manage this difficulty. Hadoop is an expanded principle proposed to succeed in the big data challenge that usually utilizes MapReduce structure to prepare huge quantities of data of the cloud system. Hadoop has no policy to guarantee the protection and secrecy of the data collected in the hadoop distributed file system (HDFS). In the cloud, the security of sensitive data is a significant issue in which data encryption schemes play an avital rule. This research proposes a hybrid system between two popular asymmetric key cryptosystems (RSA, and ElGamal) to encrypt the data collected in HDFS. Thus before storing data in HDFS, the proposed cryptosystem is utilized to encrypt the data. The user of the cloud might upload data in two ways, non-safe or secure. The hybrid method presents more powerful computational complexity and less latency in comparison to the RSA cryptosystem alone.

Corresponding Author:

Shahab Wahhab Kareem,
Department of Information System Engineering,
Erbil technical Engineering College, Erbil Polytechnic University,
Kurdistan Regin, 120 Meter St, Erbil 44001, Iraq.
Email: shahab.kareem@epu.edu.iq

1. INTRODUCTION

Cloud computing has drawn growing concentration for the latest several years. Cloud computing presents clients including a broad range of sources such as computing platforms, computing power, storage, and internet applications. Google, Amazon, Microsoft, IBM, etc. are the most important cloud accessible into the businesses now. Among an increasing number of organizations using resources inside the cloud, data from several users require to stay preserved. Cloud computing is directly being utilized in a huge amount in different areas. In daily life, enormous amounts of data are generated. Users use cloud computing services to save this very large amount of data. Some of the important challenges cloud computing faces are to secure, preserve and prepare the data that is the user's property [1, 2]. Big data leads to the processing and retrieval of huge data collection. Big data must also be involved by the group of highly valuable also sensitive data from social sites and issues of government and security. This collected data has to be encrypted by applying suitable algorithms to secure them. The features of Big Data can be classified in term of four V's [3]: Volume, Velocity, Variety, and Veracity. Every issue has its job of surviving in Big data. Thus, volume: the volume of data generated and might be collected it, could be in the level of many terabytes or Petabytes in size. Variety: This is the set of data and its applications. structure, semi-structured and unstructured. Velocity: This means the input and the output rates of data streams produced also collected into the system.

In these circumstances, a concept is produced in a direction that the big data systems can finally, store data separately of the incoming or outgoing rate. Veracity: It's the term quality of data, this context is

also Refers to data confidentiality, integrity, data privacy, and availability. Companies need a grantee that the data and the analyses transferred on the data are accurate. Big data processing has become most pivotal for several governments and business applications with an impossible rate of data produced, collected and analyzed by computer systems. Thus, several factors have played in data huge increase like the emerge of IoT, object localization and tracking, besides the growing adoption of healthcare tools that gather private statistics. This prevalence of big data has some limitations. The data collected normally includes some personal information about persons or it is including secrets that would be problematic if they are discovered by the opponent. Illegal groups organize covered markets for the ownership and investment of stolen personal information [3]. Government intelligence services rely on personal, corporate and adverse government eavesdropping and competitive advantage systems. Most recent, extremely advertised cyber-attacks against economic attacks demonstrate this potential for damage, and government targets, it spends millions of dollars to these organizations and effects critical damage to the concerned individuals and organizations. Moreover, protection across cloud services is in its growing phase, a large number of protection vulnerabilities would risk data in the cloud. The cloud managers have no evidence as to where and in what format the data is being stored. Thus, enough security measures must be adapted to preserve information principally from data leakage and manipulation. In addition, processing/analyzing huge data at the cloud data center is a critical issue. Several distributed structures like hadoop have recently been prepared [4]. Government intelligence services rely on personal, corporate and adverse government eavesdropping and competitive advantage systems. Most recent, highly publicized cyber-attacks against commercial attacks demonstrate this potential for damage, and government targets, it pays millions of dollars to these organizations and causes severe damage to the affected individuals and organizations [5].

Furthermore, protection across cloud services is under its developing stage; a huge quantity about safety vulnerabilities would risk data in the cloud. The cloud administrators have no clue as to where and in what format the data is stored. Thus, adequate security measures must be modified to preserve the data, essentially of information leakage plus manipulation. Also, processing/analyzing enormous data in the data center is a dangerous problem in the cloud. Different spread structures like hadoop have recently been available [6, 7], like Google File System [8], which is developed to store and process big data. However, the distributed hadoop framework is popular among industry and research communities. Hadoop includes two sets of functionalities, (i) For storage of large and unstructured data sets (HDFS), has been employed, and (ii) Map Reduce frame-work for hug data manipulation. Hadoop usually works with applications that have thousands of data nodes and petabytes. As literature survey Yang *et al.* [8]. Suggest triple encryption scheme for enhancing the security of hadoop. Thus the encryption of HDFS files is achieved by using DEA (data encryption algorithm), whereas RSA has been used in the encryption of data key. Eventually the RSA private key is secured using IDEA (international data encryption algorithm). Zhou et al. [9] They present CP-ABE (Cipher text policy attribute based encryption) scheme for access control instead of the traditional schemes like PKI, which requires all relevant customer data to be sent to the resource provider, thus destroying the privacy of the user, and takes more bandwidth and overhead processing. Jam *et al*. [10] point out that currently, the core technology of cloud computing are services security and data privacy. A security mechanism based on Kerberos protocol for authentication firewalls of perimeter level security was presented [11]. Security leak was handled by implementing the Apache sentry for access control, triple encryption of data using RSA, DES, IDEA algorithms, was proposed in protecting file system based on fully homomorphic encryption. Parmari *et al*. [12], proposed a novel method which can be used to secure hadoop, a cost-effective technique works in hadoop cluster to give it 3-D security. Usama, *et al*. [13], proposed data compression and encryption for hadoop. Hence a combined compression and encryption scheme was presented based on tent map and piece-wise linear chaotic map (PWLM), the proposed approach implements a masking pseudorandom keystream that strengthens the encryption process. The proposed algorithm, providing robust encryption and compression schemes.

Hadoop does not incorporate security mechanisms. The Application of ciphering algorithms in hadoop data encryption, then storing them at HDFS has reported in several works. Ciphering schemes perform different replacements and do some manipulation on the clear message to transforms it into ciphertext, which must be random and incomprehensible. Different ciphered schemes were developed and employed for the sake of information security. Hence the two main categories are: (i) Symmetric-key (secret key) cryptosystems [14] like advanced encryption standard (AES), data encryption standard (DES), and triple DES (ii) asymmetric-key (public key) algorithms [15] like elliptic curve diffie-hellman (ECDH) and RSA. The proposed approach is considered as an attempt to improve what was presented by the paper [13] at both of encipherment/decipherment procedures for securing files of big data-based hadoop-integrated AES and OTP algorithms [16]. An architecture to secure Hadoop was examined in paper [15]. Thus for data encryption and decryption, AES encryption/decryption classes are added. Implement two HDFS pairing integrations and HDFS-RSA [17] applied since various amazing kinds of

extensions from HDFS. Analyses demonstrated adequate expenses for understanding processes also significant overhead for recording actions [18]. The following is the organization of this paper: Section 2 outlines the security framework. Section 3, based on HDFS and mapreduce, presents the big data at hadoop. Section 4 discusses the proposed optimized hybrid encipherment algorithm and compare it with the classical public-key cryptosystems before applying it to secure big data at hadoop. Section 5 presents the discussion of the simulation results. Finally, section 6 list the conclusions.

2. SECURITY ISSUES

Information protection is of more interest and importance to all within the world of technology now, whether you are a smartphone or a computer user, this is why information security remains of the common essential in our daily experience, and in the IT technology areas [3-19]. Information security research contains many ideas including issues that every IT specialist must master or possess amazing basics of experience and knowledge of information security, they are just a few essentials for all those affected in the IT sector. Such as cybersecurity analysts, forensic analysts, network administrators, system administrators, and application developers. The lack of knowledge in this important field of information security is likely to improve unsafe applications or create insecure networks and attackers easily penetrate them, which is why knowledge of information security is so important in our daily lives. Regardless of the job selection, you will find yourself in the IT sector [2, 4].

Information security is described as preserving systems and information from illegal access, use, disruption, disclosure, alteration, or damage to implement integrity, confidentiality, and availability. Careful implementation of information security controls is vital to protecting enterprise information assets as well as its reliability, right location, employees, and other tangible and intangible assets [20]. Big data is about data storage, data processing, data recovery. Many technologies, such as memory management, transaction management, visualization and networking, are used for these purposes. These technologies security issues are also applicable to big data. Big data's four major security issues are authentication, data level, network level and generic matters [1, 21].

2.1. Authentication level issues

A lot of clusters and nodes are present. Each node has priorities or rights that are different. Administrative nodes can access any data. But sometimes it will steal or manipulate the critical user data if any malicious node has organizational priority. Many nodes are joining clusters for faster execution with parallel processing. Any malicious node can disturb the group in the event of no authentication. Logging in big data plays an important role. If logging not provided, no activity that modifies or deletes data will record. If the new node joins the cluster, the absence of logging will not recognize it. Users may also sometimes use malicious data unless the log provided.

2.2. Data level issues

In Bigdata the Data is an essential part and operates an important role. The data is just any of the important social media sites and personal information about us or government. The principal concerns that can be labeled within the data level are the availability and integrity of data such as data security and distribution. Big data ecosystems like hadoop collect data as they are without encryption to improve efficiency. If the hacker reaches the machines, they cannot be ended. Information collected in the distributed data store for immediate access in multiple nodes with duplicates. But if the hacker removes or prepares any cloning or information from a different node, it will be hard to retrieve that data.

2.3. Network-level issues

In Cluster, there are several nodes, and they used the nodes to process or compute data. The data processing can be prepared anyplace within the nodes in cluster. It is challenging to conclude which node data is prepared. It will be complicated because of this challenge on which node protection should be implemented. Two or extra nodes can communicate or share their data/resources via the network. RPC (remote procedure call) often used for network communication. But until and unless it is encrypted, RPC will not be secure.

2.4. General level issues

Utilized different technologies inside the big data ecosystem to prepare the data for amazing popular protection mechanisms for protection objectives. Common mechanisms have been produced. Thus with the new shared form of big data, these machines may not be completed well. As big data utilizes various data

processing, data storage, and data recovery technologies, there may be some complexities due to these various technologies.

3. HADOOP

The architecture of hadoop contains essentially two principal elements which are: hadoop distributed file system (HDFS) to collect Big Data and MapReduce for analyzing Big Data [22]. HDFS is a data management system utilized to the distributed storage of huge datasets on the hadoop cluster in with a default block size of 64 MB [23]. Following collecting the input files in HDFS, then it managed with MapReduce software. Ultimately, the events moved to the output folder of HDFS [24]. MapReduce in hadoop is an application software designed for processing huge volumes of data sets over machine sets [9]. In the hadoop, MapReduce is a core scheme used for developing a collection of work. Every input data, which occupies during the cluster on a distributed file system, is separated into groups of similar size to promote and analyze in a proper, and almost error-free manner the enormous volumes from processing the data under parallel at enormous companies concerning tools. As specified by the name, MapReduce includes a couple –stages like data consideration within hadoop, the first stage is the map, and the second stage implies reducing, i.e. a large amount from data sets is transformed inside structured key-value pairs and provided since inputs [24].

Figure 1 presents the MapReduce computation data flow. The mapper doesn't write immediately on disk but utilizes the benefit from buffering some writings. Each mapper becomes a round buffer of memory with default size is 100 MB which can do changed by developing each part of (io. sort. mb). That makes a rapid flush. If the buffer is loaded up before specific inception, it initiates the transfer to the disk the content of the barrier. Before every spill seems on the drive, every thread separations these data based on the reducers that demand ongoing background thread performs any sort of in-memory within the key-based partition before the spill takes place to the disk. If a mixer is started, it applies the output of the in-memory kind [24].

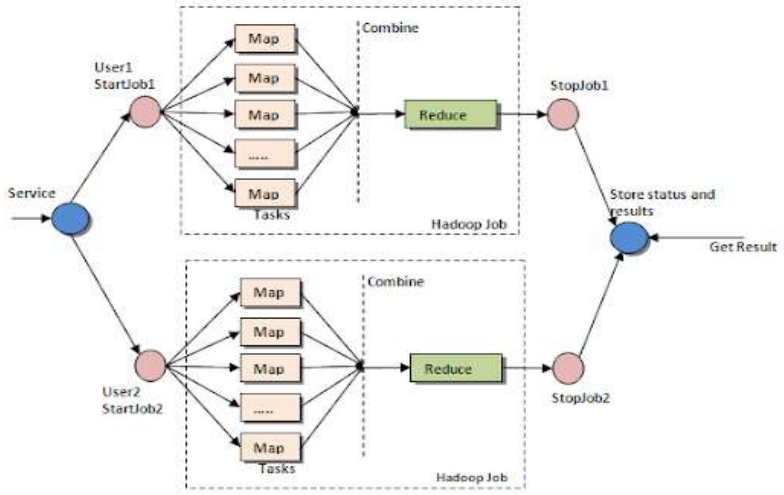

Figure 1. MapReduce data stream [1]

4. PROPOSED ALGORITHMS

The principal supplier of large-scale cloud data processing and storage is hadoop, therefore, practices any methods of encryption to guarantee protection. This paper presents modern technology—this procedure depended on combining two public-key cryptosystems (RSA and Elgamal) [2]. Hybridization's a way to overcome the weaknesses of using every cryptosystem individually also to develop assurance is responsible for keys generations (public and private keys). Then the proposed hybrid system is used while the file caching in HDFS encrypted it using the unorganized data for the file. The HDFS begins transmitting an encrypted file on the data nodes. These stages are shown in Figure 2 ElGamal cryptosystem [25] elies on the computational difficulty of solving the discrete logarithm problem, while RSA relies on the computational difficulty of factoring large primes. As such:

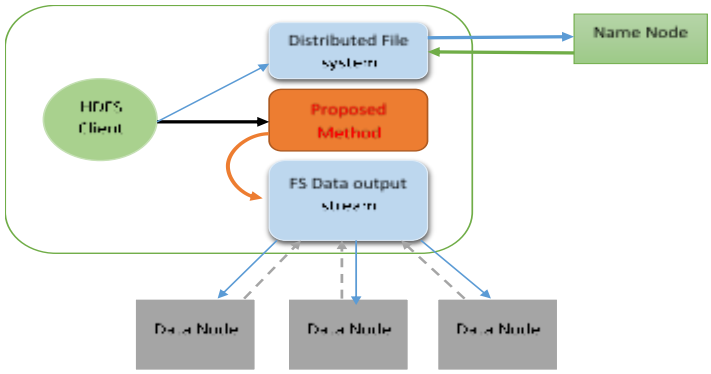

Figure 2. Encryption procedure in HDFS

RSA encryption is faster than ElGamal but RSA decryption is slower than ElGamal. HDFS consists of a Name Node that stores Metadata which manages the namespace of the file system and controls clients' access to the encrypted file. The encrypted file is made up of one or more blocks stored in a set of data nodes. The proposed hybrid encryption system is described in the fig 3. below: It is examined that all the files written to HDFS must be encrypted before collecting it. The HDFS client

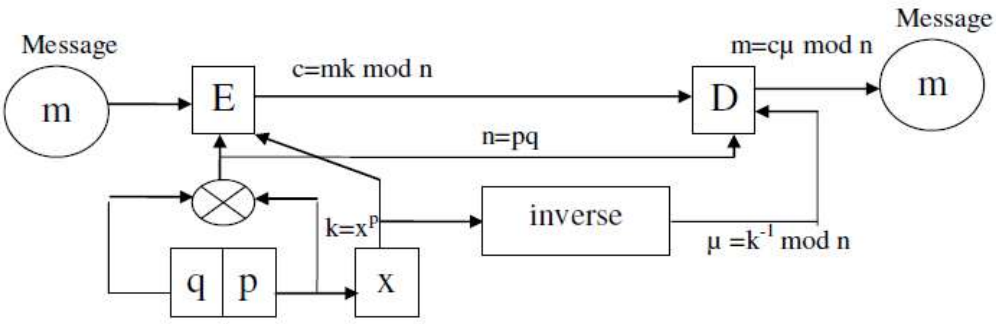

Figure 3. Process of hybrid public key algorithm

This method is hybrid between RSA and ElGamal cryptosystem named RSA-ElGamal Hybrid cryptosystem. The key generation process involves both scenario (RSA, ElGamal) .Thus a two large primes (p, q) are generated with the same size to compute $n = p*q$ (public key), then chose random integer(x) with in the range :1<x<p to compute (k) such that:

$$k = x^p * \mathrm{mod}(n). \quad (1)$$

At that level the set of private keys is established which is the vector=[p, q, k], then compute:

$$\mu = k^{-1} \mathrm{mod}(n) \quad (2)$$

Eventually to encrypt message (m) the cipher text c is computed as:

$$c = m * k \,\mathrm{mod}(n). \quad (3)$$

To recover Plain text (message) c:

$$m = c * \mu \bmod(n). \quad (4)$$

The keys (public and private) generation procedure is described the Algorithm 1 below: (Algorithm 2) below shows in detail the encryption procedures:

Algorithm 1: - Key Generation of RSA-ElGamal Hybrid algorithm

INPUT: Select large random prime numbers p and q

OUTPUT: A public key, (n), and a private key(p,q,k)

User A sends the message to user B.

1. Generate two large random (and distinct) primes p and q, each roughly the same size.

2. Compute n=p*q .

3. Select a random integer x, 1<x<p, and compute $k=x^p$ mod n.

4. User's public key are (n); user's private key are (k, p, q).

The encryption process is depicted by the algorithm below:

Algorithm2:-Encryption process of RSA-ElGamal Hybrid algorithm

INPUT: Plaintext to encrypt, and receiving user's public key (n, e).

OUTPUT: Encrypted cipher-text.

User A sends the message to user B.

To encrypt B should do the following:

(a) Obtain A's authentic public key (n).

(b) Represent the message as an integer m in the interval [0, n-1].

(c) Compute c=(m*k) mod n.

(d) Send the cipher-text c to A.

The decryption procedure is described by the algorithm below:

Algorithm3:-Decryption process of RSA-ElGamal Hybrid algorithm

INPUT: Received encrypted cipher-text and the receiver's private key (p, q, k).

OUTPUT: Original plaintext.

To recover plaintext m from c, B should do the following:

(a) Compute $\mu=k^{-1}$mod n.

(b) Compute m=c*μ mod n=m mod n

After applying proposed encryption scheme, data is stored in the cloud. Thus via hadoop file system (HDFS) data will be stored in a clusters. Whenever the user requests data, the server will introduce the encrypted data to the decryption procedure. The user then uses the private key to retrieve the decrypted data using a hybrid system which is the proposal of this paper.

5. EXPERIMENTAL RESULTS AND ANALYSIS

HDFS and MapReduce functions have been used for performance evaluation of encrypted HDFS. Each node has i5 core, 8 GB of memory, and 2TB of hard disk. Encryption Time: The time taken by the RAS, ElGamal alone or the hybrid algorithm to encrypt the hadoop divided dataset files into ciphertext using a key. It is calculated in seconds. Decryption time: The time taken by the RAS, ElGamal alone or the hybrid system to decrypt the hadoop splitted dataset files back into the plaintext using the private key. It is calculated in seconds. Thus the Encryption Time is equivalent to system current time before encryption subtracted from it the system current time after encryption. Whereas the Decryption Time is equivalent to the system current time before decryption subtracted from it the system current time after decryption. Figure 4, depicts the results of the comparison between encryption schemes, the RSA alone, ElGamal alone and the Hybrid (RSA-ElGamal) cryptosystem with different file sizes. It's clear that the proposed method showed efficient time consumption compared to the RSA for files size stars from 100 Mb and ends with 1 GB with step size of 100 Mb. But it gives comparable results with ElGamal method for file size less than 400 Mb but after that size the gab is increased between the proposed hybrid system and ElGamal method. And hence, the proposed method (hybrid system) in encryption stage is faster than the default RSA. Figure 5 shows the running time for RSA and for proposed method in decryption stage. The encrypted files applied to this stage are with different sizes. By comparing RSA with RSA-ElGamal method (the proposed

method) it's obvious that decryption time needed by proposed method is shorter than that needed by RSA or ElGamal methods alone. The computational complexity for the proposed RSA-ElGamal method in encryption stage for input message unit is n.

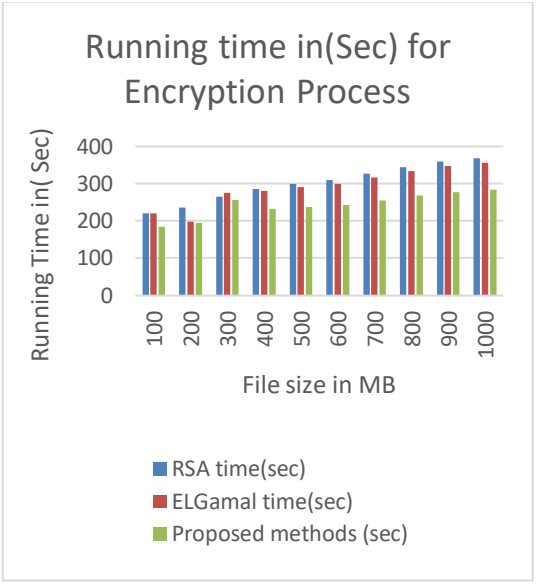

Figure 4. Time of encryption process

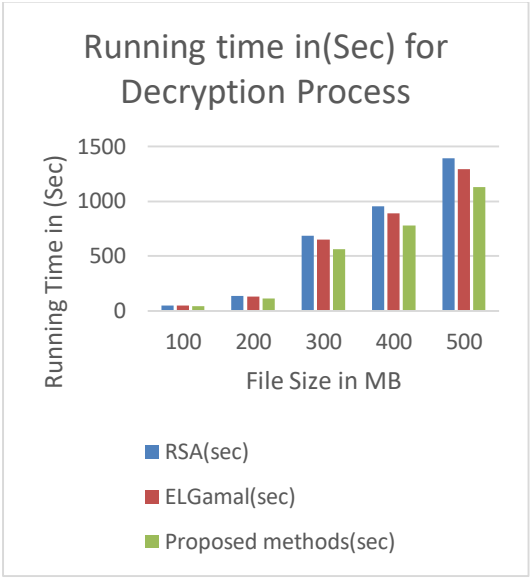

Figure 5. Time of decryption process

$$T(C)=O(\log n)^3+O(\log n) \text{ bit operation} \quad (5)$$

In the decryption scheme the computational complexity is calculated as depicted in (6) below:

$$T(M)=2\ O(\log n)^3+O(\log n) \text{ bit operation} \quad (6)$$

While the RSA computational complexity for encrypyion and decryption respectiviltiy are:

$$T(C)=O(\log n)3 \quad (7)$$

$$T(M)=O(\log n)3 \quad (8)$$

As well as the ElGamal computational complexity for encrypyion and decryption respectiviltiy are:

$$T(C)=2\ O(\log n)3+O(\log n) \quad (9)$$

$$T(M)=2\ O(\log n)3+O(\log n) \quad (10)$$

6. CONCLUSION

The hadoop permits to succeed in the difficulties faced with big data in companies and industries, it is not mentioned to security mechanism. An intruder or snoop may compromise the information collected within hadoop. The authenticity of information is continuously at stake since hadoop does not implement a protection mechanism. Before collecting the data in HDFS, the proposed RSA-ElGamal cipher asymmetric key algorithm encode the content of the files through securing it of the different network intrusions. The files or data can be saved in hadoop without worrying about security issues by implementing the encryption algorithm to the files before it is collected in hadoop. The proposed RSA-ElGamal cipher supports most cloud computing system service models such as service infrastructure (IaaS), service software (SaaS), and service platform (PaaS). It also helps data administration and security issues (authentication, integrity, availability, and confidentiality) in security and key management for data transfer. The proposed method showed excellent time consumption with different file sizes in the encryption and decryption stages with higher complexity(double the computational complexity in decryption stages.

REFERENCES

[1] M. Bhandarkar, "MapReduce programming with apache Hadoop," *in International Symposium on Parallel & Distributed Processing (IPDPS)*, Atlanta, 2010.

[2] S. W. Kareem, "Hybrid Public Key Enthcryption Algorithms For E-Commerce," Erbil: University of Salahaddin–Hawler, *thesis*, 2009.

[3] Goodubaigari Amrulla, et al., "A Survey of : Securing Cloud Data under Key Exposure," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 7, no. 3, pp. 30-33, 2018.

[4] Marti Motoyama, et al., "An analysis of underground forums," in *ACM SIG- COMM Conference on Internet Measurement Conference IMC '11*, New York, pp. 71-80, 2011.

[5] Shahab Wahhab Kareem, Yahya Tareq Hussein, "Survey and New Security methodology of Routing Protocol in AD-Hoc Network," in *Qalaai Zanist Journal*, Erbil, 2017.

[6] Bo Li, et al., "Modeling and Verifying Google File System Modeling and Verifying Google File System," in *16th International Symposium on High Assurance Systems Engineering*, 2015.

[7] Danish Ahamad, MD Mobin Akhtar, Shabi Alam Hameed, "A Review and Analysis of Big Data and MapReduce," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 1, pp. 1-3, 2019.

[8] Chao Yang, Weiwei Lin, and Mingqi Liu, "A Novel Triple Encryption Scheme for Hadoop-based Cloud Data Security," in *Fourth International Conference on Emerging Intelligent Data and Web Technologies*, pp. 437-442, 2013.

[9] Huixiang Zhou, and Qiaoyan Wen, "A new solution of data security accessing for Hadoop based on CP-ABE," in *5th International Conference on Software Engineering and Service Science*, 2014.

[10] Masoumeh RezaeiJam, et al., "A Survey on Security of Hadoop," in *4th International Conference on Computer and Knowledge Engineering (ICCKE)*, pp. 716-721, 2014.

[11] Roojwan S. Ismael, Rami S. Youail, Shahab Wahhab Kareem, "Image Encryption by Using RC4 Algorithm," *European Academic Research*, vol. II, no. 4, pp. 5833-5839, 2014.

[12] R. R. Parmar, et al., "Large-Scale Encryption in the Hadoop Environment: Challenges and Solutions," in *IEEE Access*, vol. 5, pp. 7156-7163, 2017. doi: 10.1109/ACCESS.2017.2700228.

[13] Usama M, and Zakaria N, "Chaos-Based Simultaneous Compression and Encryption for Hadoop," *PLoS One*, vol. 12, no. 1, 2017.

[14] Sourabh Chandra, Siddhartha B, and Smita Paira, "A Study and Analysis on Symmetric Cryptography," in *ICSEMR*, 2014.

[15] H. Mahmoud, A. Hegazy and M. H. Khafagy, "An approach for big data security based on Hadoop distributed file system," *2018 International Conference on Innovative Trends in Computer Engineering (ITCE)*, Aswan, pp. 109-114, 2018. doi: 10.1109/ITCE.2018.8316608.

[16] S. Park and Y. Lee, "Secure Hadoop with Encrypted HDFS," in *International Conference on Grid and Pervasive Computing*, pp. 134-141, 2013.

[17] C. Yang, W. Lin, and M. Liu, "A novel triple encryption scheme for Hadoop-based cloud data security," in *4th Int. Conf. Emerg. Intell. Data Web Technol. EIDWT 2013*, pp. 437-442, 2013.

[18] Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri, "Security Issues Associted With big Data In Cloud Computing," *International Journal of Network Security & Its Applications (IJNSA)*, vol. 6, no. 3, pp. 45-56, 2014.

[19] P. Merla and Y. Liang, "Data analysis using Hadoop MapReduce environment," in *IEEE International Conference on Big Data (Big Data)*, Boston, pp. 4783-4785, 2017.

[20] Shadan Mohammed Jihad Abdalwahid, Raghad Zuhair Yousif, Shahab Wahhab Kareem, "Enhancing Approach Using Hybrid Pailler And Rsa For Information Security In Bigdata," *Applied Computer Science*, vol. 15, no. 4, pp. 63-74, 2019.

[21] Raghad Z.Yousif, Shahab.Wahhab Kareem, Ammar. O.Hasan, "Design Security System Based on AES and MD5 for Smart Card," in charmo university, Sulaimanyia, 2016.

[22] A. Bhardwaj, et al., "Analyzing BigData with Hadoop cluster in HDInsight azure Cloud," *2015 Annual IEEE India Conference (INDICON)*, New Delhi, pp. 1-5, 2015. doi: 10.1109/INDICON.2015.7443472.

[23] A. K. Dubey, V. Jain and A. P. Mittal, "Stock market prediction using Hadoop Map-Reduce ecosystem," *2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom)*, New Delhi, pp. 616-621, 2015.

[24] Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," in : *Proceedings of the 6th Symposium on Operating Systems Design and Implementation*, vol. 6, 2008.

[25] T. Elgamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," in *IEEE Transactions on Information Theory*, vol. 31, no. 4, pp. 469-472, July 1985. doi: 10.1109/TIT.1985.1057074.

BIOGRAPHIES OF AUTHORS

Shahab Wahhab Kareem I received my BSc in Control and Computer Engineering from University of Technology Baghdad in 2001, and MSc in Software Engineering from Salahadeen University in 2009 and Achieving Ph. D in Computer Engineering from Yasar University Izmir, Turkey. My research interests include Machine learning and Big data. I'm a lecturer at the Information System Eng. (ISE) Department (2011-till now)

Dr. Raghad Zuhair Yousif Born in Baghdad 1975. Obtained a bachelor's degree in electronic Engineering and communications from the University Of Baghdad College Of Engineering in 1998. He earned a master's degree in Electronic Engineering and Communications from Mustansriyha College of Engineering in 2001. Achieving Ph.D. in Communications and Information Technology form university of Technology Department of Electronica and Electrical Communication in 2005. Awarding Assistant Professor since April 2011 currently working as a university professor at Salahaddin University College of Science.

Shadan Mohammed Jihad Abdalwahid I received my BSc in Software Engineering from University of Baghdad in 2001, and MSc in Computer Engineering from Yasar University Izmir, Turkey. My research interests include Cyber security and Cyber warefar. I'm a lecturer at the Information System Eng. (ISE) Department (2011-till now)