International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2016, pp. 323-329
e-ISSN:2278-621X

A STUDY ON ENCRYPTION DECRYPTION ALGORITHM FOR BIG-DATA ANALYTICS IN CLOUD

SreenivasaB.L[1], Manish Kumar[2], Mohammed Nueed Shaikh[3] and Dr. S Sathyanarayana[4]

Abstract- Big data and cloud computing is the backbone to the modern Data Storage and access mechanism. The Big Data in cloud is the most important research problem, researchers are trying to find the solution for it. And one of the issue is to give a perfect security for big data in cloud computing, so that big data could be handled in the recent systems and managed with the cloud computing. In this paper, we study and compare the different encryption and decryption algorithm. We have summarized in the table the confidentiality, integrity and availability of encryption and decryption security algorithm for big data in cloud computing.

Keyword: encryption, decryption, confidentiality, integrity, security.

I. INTRODUCTION

Big data is one area where we can store, extract and process large amount of data .In the near future large amount of data will be stored on the cloud environment but there will be security issue. In big organization big data act like a virtual machine, were one single machine will be holding all the data another machine will have access to it. Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public networks, to provide dynamically scalable infrastructure for application, data and file storage.

Data generation is growing rapidly and many organizations demand efficient solutions to store and scrutinizingthese big amounts of data that are largely generated from various resources such as internet, social media, personal storage and sensor data [1]. For this purpose, big data technologies are utilized in cloud computing to provide significant benefits, such as the availability of automated tools to assemble, connect, configure and reconfigure virtualized resources on demand. These makes it much easier to meet organizational goals as organizations can easily deploy on cloud services [1].

Hackers try to access the confidential data using different decoding techniques .As the organizations data which is to be saved in the cloud storage server gets in hands with different attackers who are trying to access your data using organizations cloud servers. However, with the increased adoption of web-based, mobile and cloud-based applications, sensitive data has become accessible from different platforms. These platforms are highly vulnerable to hack, especially if they are low-cost or free.

[1] *Aloysius Institute of Management and Information Technology Mangalore, Karnataka, India*
[2] *Aloysius Institute of Management and Information Technology Mangalore, Karnataka, India*
[3] *Aloysius Institute of Management and Information TechnologyMangalore, Karnataka, India*
[4] *First Grade Women's CollegeMysore, Karnataka.*

In this paper, we study and compare the different encryption and decryption algorithm. We have summarized it in the table the confidentiality, integrity, availability, Adversary type, interaction, Security, Performance analysis and Time Complexity of encryption and decryption security algorithm for big data in cloud computing. Section 1 describes about big data in cloud and security, section 2 explains about challenges in big data and cloud security, Section 3 explains about the security algorithm, in section 5 we have summarised the security algorithms in the table and in section 6 we have concluded that how the encryption and decryption is carried on big data in cloud.

II. NINE MAINCHALLENGES IN BIG DATA AND CLOUD SECURITY

Nowadays, organizations are collecting and processing massive amounts of information. The more data is stored, the higher security has to be ensure. A lack of data security can lead to great financial losses and reputational damage for the company. Here are some of the challenges that the cloud and big data face [2]:

- Most distributed systems' computations have only a single level of protection, which is not recommended.
- Non-relational databases (NOSQL) are actively evolving, making it difficult for security solutions to keep up with demand.
- Automated data transfer requires additional security measures, which are often not available.
- When a system receives a large amount of information, it should be validated to remain trustworthy and accurate; this practice doesn't always occur, however.
- Unethical IT specialists practicing information mining can gather personal data without asking users for permission or notifying them.
- Access control encryption and connections security can become dated and inaccessible to the IT specialists who rely on it.
- Some organizations cannot – or do not – institute access controls to divide the level of confidentiality within the company.
- Recommended detailed audits are not routinely performed on Big Data due to the huge amount of information involved.
- Due to the size of Big Data and cloud, its origins are not consistently monitored and tracked.

There are various criteria which we have taken into consideration and have listed them below [2].

Adversary types: Determines whether an algorithm is malicious or non-malicious.

Confidentiality: The data sent should be read by the receiving user only.

Integrity: The message which is sent to the receiver by the sender should be received at the same time without any delay.

Availability: All the 3 goals are the requirement of secure digital communication but the process of achieving these goals should not hinder the performance of the applications. Thus, it means these processes should have overhead (in terms of speed and memory) as low as possible.

Security: specifies the security level of an algorithm.

Performance Scrutinizing: determines the performance of an algorithm when tested.

Time Complexity: it quantifies the amount of time taken by an algorithm to run as a function of the length of the string representing the input.

III. DIFFERENT CLOUD SECURITY ENCRYPTION TECHNIQUE

A. Homorphic Encryption algorithm:

Homorphic encryption is an encryption algorithm which allows specific types of computations to be carried out on plaintexts and generate an encrypted result which, when decrypted, matches the result of operations performed on the plaintexts. RSA is the first encryption algorithm with the homorphic property. In the algebraic concept we can define homorphic as a structure preserving map between two algebraic structures such as groups [3].
Agroupisaset,G,togetherwithanoperation(calledthegrouplawofG)
thatcombinesanytwoelementsaandbtoform another element,denotedaob. Toqualify asagroup,thesetandoperation,(G, o),mustsatisfyfourrequirements knownasthegroupaxioms [3]
• Closure: Forallaand binG,theresultoftheoperation,aob, isalsopresent in G.
• Associativity:Foralla ,b andcinG,(aob)oc=ao(boc)
• Identity element: There existsan element inG,suchthatforeveryelementa in G,the equalityeo a=a
o e=aholds.Suchan elementisunique,andthus onespeaksoftheidentityelement.
• Inverseelement:ForeachainG, thereexistsanelementbinGsuchthat (ao b)=(boa)=e, where isthe
identityelement. Identity of g is written as 1.
Theresultofanoperation maydependontheorderoftheoperands. Inother words,theresultofcombining elementawithelementbneednotyieldthesame resultas combiningelementbwithelementa;theequation aob=boamaynot always be true. This equationalwaysholds inthegroupofintegersunderaddition, becausea+b=b+aforanytwointegers(commutativityofaddition).Groups forwhichthecommutativityequationa+b=b+aalwaysholdsarecalledaBoolean groups.

B. Verifiable computation algorithm (outsource computing):

Verifiable computation (VC) algorithm is the one which permits a frail (week) customer to send his information on the cloud without much stressing over the security issues. This is the most secure algorithm which helps the user to send his data to the cloud storage device [4].

Algorithm:

- *VC* = (KeyGen, ProbGen, Compute, Verify) consists of four algorithms as follows:
- KeyGen(F, λ) → (PK, SK):it generates two keys the public and private key based on the security parameter λ. the public key encodes the target function f and is sent to the server computer for the secret key is kept private by the client.
- ProbGenSK(x) → (σx, τx): The problem generation algorithm encodes the function input x into two values, public and private, using the secret key SK. The public value σx is given to the worker to compute F(x) with, while the secret value τx is kept private by the client.
- ComputePK(σx) →σy: The worker computes an encoded value σy of the function's output y = F(x) using the client's public key PK and the encoded input σx.
- $Verify_{SK}(\tau x,\sigma y) \rightarrow y \cup \perp$: The verification algorithm converts the worker's encoded output σy into the actual output of the function F using both the secret key SK and the secret "decoding"τx. It outputs y = F(x) if the σy represents a valid output of F on x, or outputs ⊥ otherwise.

C. Message digest algorithm:

The MD5 function is a cryptographic algorithm that takes an input of arbitrary length and produces a *message digest* that is 128 bits long. The digest is sometimes also called the "hash" or "fingerprint" of the input. MD5 is used in many situations where a potentially long

message needs to be processed and/or compared quickly. The most common application is the creation and verification of digital signatures. How MD5 works [5].

The MD5 algorithm first divides the input in **blocks** of 512 bits each. 64 Bits are inserted at the end of the last block. These 64 bits are used to record the length of the original input. If the last block is less than 512bits, some extra bits are 'padded' to the end.Next, each **block** is divided into 16 **words** of 32 bits each. These are denoted as M_0 ... M_{15}.MD5 uses a buffer that is made up of four **words** that are each 32 bits long [5].

The table MD5 further uses a table K that has 64 elements. Element number i is indicated as Ki. The table is computed beforehand to speed up the computations. The elements are computed using the mathematical sin function [5]:

Ki = abs(sin(i + 1)) * 232

Four auxiliary functions

In addition MD5 uses four auxiliary functions that each take as input three 32-bit words and produce as output one 32-bit word. They apply the logical operators and, or, not and xor to the input bits.

F(X,Y,Z) = (X and Y) or (not(X) and Z)

G(X,Y,Z) = (X and Z) or (Y and not(Z))

H(X,Y,Z) = X xor Y xor Z

I(X,Y,Z) = Y xor (X or not(Z))

The contents of the four buffers (A, B, C and D) are now mixed with the words of the input, using the four auxiliary functions (F, G, H and I). There are four rounds, each involves 16 basic operations. After all rounds have been performed, the buffers A, B, C and D contain the MD5 digest of the original input.

D. Key rotation algorithm:

The key rotation algorithm is the best algorithm in which we can store our data or send our data on the cloud environment using a shared symmetric key.Here the data owner provides the permission to encrypt the data using the shared symmetric key method.in this method a cloud user can store his data using the encryption method .the cloud service provider with the support of the data owner who gives support to convert plain data to cipher data and store in cloud .the same way he gives the permission to retrieve data from the cloud data centre in an encrypted manner and gives it back to the user in a decrypted manner.

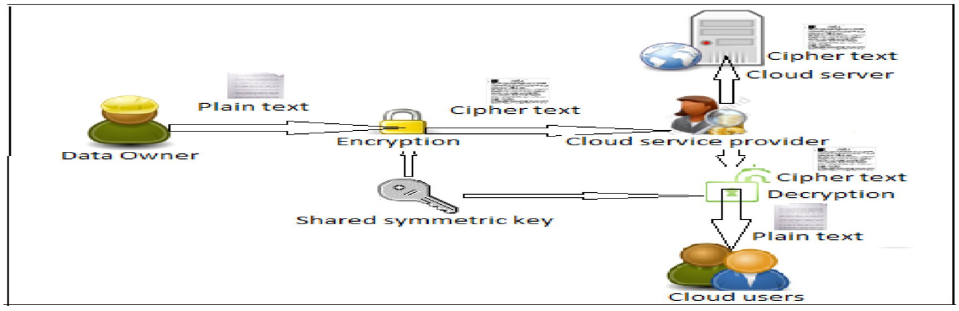

Figure 1. Algorithm

E. DES Algorithm-

The Data Encryption Standard (DES) is the name of the Federal Information Processing Standard (FIPS), Which Describes the data encryption algorithm (DEA).it has a 64-bit block size key during execution. DES is a symmetric cryptosystem, specifically a 16-round FeistelCipher.While communicating, both sender and receiver must know the same secret

key, which can be used to encrypt and decrypt the message, or to generate and verify a Message Authentication Code (MAC). The DES can also be used for Single – user encryption, such as to store files on a hard disk in encrypted form .The DES has a 64-bit block size and uses a 56 bit key during execution. In Cipher Block Chaining mode of operation of DES, each block of ECB encrypted cipher text is XOR ed with the next plain text block to be encrypted, thus making all the blocks dependent on all the previous blocks .this means that in order to find the plaintext of a particular block, you need to know the cipher text, the key and the cipher text for the previous block.

The first block to be encrypted has no previous cipher text, so the plaintext is XORed with a 64bit number called the initialization vector.So if data is transmitted over network or phone line and there is a transmission error, the error will be carried forward to all the subsequent blocks since each block is dependent upon the last .this mode of operation is more secure than ECB (electronic code book) because the extra XOR step adds one more layer to the encryption process.

F. Rijndael Encryption Algorithm:

Rijndal is the block cipher algorithm recently chosen by the National Institute of Science and Technology (NIST) as the Advanced Encryption Standard (AES). It supersedes the Data Encryption Standard (DES). Rijndael is a standard symmetric key encryption algorithm used to encrypt sensitive information. it is an iterated block cipher. Therefore, the encryption or decryption of a block of data is accomplished by the iteration (a round) of a specific transformation (a round function). Rijndael also defines a method to generate a series of sub keys from the original key. The generated sub keys are used as input with the round function.

Rijndael is designed based on the following 3 benchmark:

1. Resistance against all known attacks;
2. Speed and code compactness on a wide range of platforms;
3. Design simplicity

Rijndael is the best combination of security, performance, efficiency, ease of implementation and

Flexibility. The Rijndael algorithm supports key sizes of 128, 192 and 256 bits, with data handled in 128-bit blocks. Rijndael uses a variable number of rounds, depending on key/block sizes, as follows: 9 rounds if the key/block size is 128 bits, 11 rounds if the key/block size is 192 bits, and 13 rounds if the key/block size is 256 bits. Rijndael is a substitution linear transformation cipher. It uses triple discreet invertible uniform transformations (layers). Specifically, these are: Linear Mix Transform; Non-linear Transform and Key Addition Transform. Even before the first round, a simple key Addition layer is performed, which adds to security. Thereafter, there are Nr-1 rounds and then the final round. The transformations form a State when started but before completion of the entire process.

Algorithm

- Key Expansion

Round keys are derived from the cipher key using Rijndael's Key schedule.

- Initial Round

Add Round Key each byte of the state is
combined with the round key using bitwise xor.

- Rounds
  Sub Bytes
  Shift Rows

Mix Columns
Add Round Key

- Final Round (no Mix Columns)
  Sub Bytes
  Shift Rows
  Add Round Key
- The Sub Bytes Step

IV. Summary of Encryption/Decryption Algorithm

Cryptographic technique	Adversary type	Confidentiality	Integrity	Requires interaction	Security	Performance analysis	Time Complexity
Homorphic encryption	Malicious	True	False	False	Not a good choice	Encryption takes time	( ) O~ (Y)[3]
Verifiable computation(vc)	Malicious	False	True	False	Well secured. One private key which is kept secret	encryption takes more time	[4] O(nlog2 n).
MD5-(Message-Digest algorithm 5)	Non Malicious	True	True	False	Well secured as that of all message digest	Performance is up to the mark	2^{104} time complexity [5]
Key rotation	Non malicious	True	True	False	Highly efficient	Highly efficient	Big o[6]
Double Ds Algorithm	Non malicious	True	True	True	Depend-s on having large key space[7]	Best to secure data with huge	Big o[7]
Rijndael Encryption Algorithm	Non malicious	True	True	True	Highly efficient	Efficient	Big o[8]

V.CONCLUSION

The Big Data in cloud is the most important research problem, researchers are trying to find the solution for it. And one of the issue is to give a perfect security for big data in cloud computing, so that big data could be handled in the recent systems and managed with the cloud computing. In this paper, we study and compare the different encryption and decryption algorithm. We have summarized in the table the confidentiality, integrity and availability of encryption and decryption security algorithm for big data in cloud computing. This helps us to understand the algorithm which provides maximum security for big data in cloud computing. In future we will analyse and implement the compared algorithms. And check the level of security they provide for big data analytics in cloud.

REFERENCES

[1] http://aircconline.com/ijnsa/V8N1/8116ijnsa04.pdf

[2] http://www.datacenterknowledge.com/archives/2016/01/19/nine-main-challenges-big-data-cloud-security/

[3] http://ijns.jalaxy.com.tw/contents/ijns-v19-n3/ijns-2017-v19-n3-p449-457.pdf

[4] http://www.cse.wustl.edu/~jain/cse571-09/ftp/l_07hsh.pdf

[5] https://eprint.iacr.org/2013/279.pdf

[6] http://research.ijcaonline.org/volume91/number8/pxc3895081.pdf

[7] https://www.isaca.org/Groups/gProfessional-English/big-data/GroupDocuments/Big_Data_Top_Ten_v1.pdf

[8] http://www.inform ationweek.com/big -data/big-dataanalytics/big-databrings-big-security-problems/d/d-id/1252747.

[9] http://wireilla.com/papers/ijcis/V2N2/2212ijcis03.pdf

[10] https://eprint.iacr.org/2013/279.pdf

[11] http://www.sciencedirect.com/science/article/pii/S0166218X05002428

[12] http://www.nku.edu/~christensen/3DES.pdf

[13] http://www.datacenterknowledge.com/archives/2014/08/14/data-security-encryption-in-the-cloud/

[14] https://en.wikipedia.org/wiki/Analysis_of_algorithms

[13] https://crypto.stanford.edu/~dwu4/papers/XRDS2015.pdf

[14] http://ijecs.in/issue/v3-i4/2%20ijecs.pdf