ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 6, June 2018

A Comparative Study on Various Encryption and Hashing Algorithms for Ranked Keyword Search over Encrypted Cloud Data

Saba, Prof. S. C. Karande

Department of Computer Engineering, Maharashtra Institute of Technology, Pune, India

ABSTRACT: With the fast progression of digital data exchange in electronic way, Information Security is becoming much more important in data storage and transmission. Information Confidentiality has a prominent significance in the study of ethics, law and most recently in Information Systems. With the evolution of human intelligence, the art of cryptography has become more complex in order to make information more secure. In recent years, a lot of applications based on internet are emerged such as on-line shopping, stock trading, internet banking and electronic bill payment etc. Such transactions, over wire or wireless public networks demand end-to-end secure connections, should be confidential, to ensure data authentication, accountability and confidentiality, integrity and availability.

KEYWORDS: AES, RSA, MD5, SHA

I. INTRODUCTION

Encryption is the process of scrambling a message so that only the intended recipient can read it. Encryption can provide a means of securing information. As more and more information is stored on computers or communicated via computers, the need to insure that this information is invulnerable to snooping and/or tampering becomes more relevant. Encryption is one of the principal means to guarantee security of sensitive information. Encryption algorithm performs various substitutions and transformations on the plaintext and transforms it into ciphertext. Many encryption algorithms are widely available and used in information security. Encryption algorithms are classified into two groups: *Symmetric-key* and *Asymmetric-key* encryption.

Symmetric key encryption is a form of cryptosystem in which encryption and decryption are performed using the same key. It is also known as conventional encryption whereas asymmetric encryption is a form of cryptosystem in which encryption and decryption are performed using the different keys – one a public key and one a private key. It is also known as public-key encryption

Message Digest is one way where a master fingerprint has been generated for the purpose of providing a message authentication code (hash code) **.** The Data integrity is measured by MD5 by the help of 128 bit message, that message is given by user to create a fingerprint message is of variable length; the main thing is that it is irreversible. MD5 is the extension of MD4 algorithm which is quite faster because of its three rounds and MD5 contains four rounds which makes its slower. It's a one way hash function that deals with security features.

A. Multi-keyword

The multi-keyword defines the method of searching more than one keyword at a particular given time. The file that is retrieved after the keyword search must contain at least one of the keywords that were being searched. The maximum numbers of the keywords that can be matched are that all the keywords that were searched by the user.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 6, June 2018

B. Ranked Search

When the user searches any number of keywords, the files that are retrieved are in a bulk amount. This can be further refined by ranking the documents using some parameter. The parameters can be set by the users that will be using. This can be done by considering the following parameters; recently used documents, recently uploaded document, most popular documents, etc. The users can also define the number of entries of documents that they would like to see.

II. RELATED WORK

The author Wei Zhang et al [1] describes a system which helps in data retrieval from the cloud. As the data can be sensitive and hence when it gets into the hands of wrong people, it can turn out to be harmful for both the owners and the receivers. Hence to overcome such a situation, encryption and decryption of data can be done for the safe exchange of any amount and type of data.

Michael Armbrust et al [2] explain about the various advantages and uses of the cloud storage. It helps in a large amount of storage space without the use of the resources in the real time. The resources used to carry out storage operations are never owned by the users and hence the users have to pay per use. This strategy helps in eco friendly and green computing as well.

Dawn Xiaodong Song et al [3] discusses about the four of the most important concepts of provable security, query isolation, controlled searching and hidden query. The provable security helps in keeping the data secure. This is done as the untrusted server cannot understand about the plaintext from the encrypted text that is uploaded. The query isolation also helps in maintaining the secrecy.

Cong Wang et al [4] discusses over the drawbacks of the traditional cloud file retrieval system. There are majorly two drawbacks of this system. The user when tries to retrieve a file from the cloud, the user has to download all the files that are related to the keyword or the query that the user has entered.

Jin Li et al [5] discusses about the fuzzy keywords that might occur during a search. Fuzzy keywords are the words which are spelled wrong and are entered into the query box. They can be done using the simple spell checking. These kinds if words should be given suggestion and have to show results which might be possibly near the word. For example, if a query is made for Lonfon, the fuzzy keyword correction should show any data that might contain Lonfon or London as it is the nearest word which might be correct. It may also consider every possible keyword. This might include *Lonfon, L*onfon, etc.

Qin Liu et al [6] discusses about the ADL. The aggregation and distribution layer (ADL) is a middleware layer between the users and the cloud. It was envisioned such that an ADL will be deployed in an organization that has outsourced the data operations to a cloud. The ADL will aggregate queries from multiple users and send a combined query to the cloud. Due to this combined query, the cloud will need to execute the query only once and return all matched files to the ADL.

The comparative analysis of RSA, 3DES, AES and DES algorithm was discussed in [7]. This has highlighted a few of the features which make AES algorithm favourable in our case. AES algorithm is faster than the RSA and hence this would help in faster recovery of the data that is stored over the cloud server. The next paper compares the MD5 algorithm and the SHA-1 [8]. Though SHA-1 is more secure for hashing, MD5 is efficient and faster than the SHA-1.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 6, June 2018

III. ENCRYPTION ALGORITHMS

A. AES Algorithm

It is a symmetric algorithm which is used to convert plain text into cipher text. The DES is a weaker algorithm, hence AES is used in place of it as the 56 bit key of DES is no longer safe against attacks based on exhaustive key searches. The 64-bit block AES is used with 128-bit block with128-bit keys. It has four 4 column-major order matrix of bytes. The number of rounds depends upon the key length.

- 10 cycles of repetition for 128-bit keys.
- 12 cycles of repetition for 192-bit keys.
- 14 cycles of repetition for 256-bit keys.

Each round consists: Sub byte, Shift byte, Mix columns, Add Round key.

- Sub byte: In the Sub bytes step, each byte a[i,j] a[i,j] in the state matrix is replaced with a SubByte S(a[i,j]) S(a[i,j]) using an 8-bit substitution box, the Rijndael S-box. This operation provides the non-linearity in the cipher. The S-box used is derived from the multiplicative inverse.

- Shift Byte: The Shift rows step operates on the rows of the state; it cyclically shifts the bytes in each row by a certain offset. For AES, the first row is left unchanged Each byte of the second row is shifted one to the left. Similarly, the third and fourth rows are shifted by offsets of two and three respectively. For blocks of sizes 128 bits and 192 bits, the shifting pattern is the same. Row n is shifted left circular by n-1 bytes.

- Mix columns: In the Mix columns step, the four bytes of each column of the state are combined using an invertible linear transformation. The mix columns function takes four bytes as input and outputs four bytes, where each input byte affects all four output bytes. Together with ShiftRows, MixColumns provides diffusion in the cipher.

- Add Round key: In the Add Round key step, the subkey is combined with the state. For each round, a subkey is derived from the main key using Rijndaels key schedule; each subkey is the same size as the state. The subkey is added by combining each byte of the state with the corresponding byte of the subkey using bitwise XOR.[14]

The above steps are done until the very final round. When the last round is executed, every step is executed except the mix columns. When decrypting, the first round exempts the mix columns. Then the rest of the process is carried normally.

B. DES Algorithm

DES is one of the most widely accepted, publicly available cryptographic systems. The Data Encryption Standard (DES) is a block Cipher which is designed to encrypt and decrypt blocks of data consisting of 64 bits by using a 64-bit key.

Although the input key for DES is 64 bits long, the actual key used by DES is only 56 bits in length. The least significant bit in each byte is a parity bit, and should be set so that there are always an odd number of 1s in every byte. These parity bits are ignored, so only the seven most significant bits of each byte are used, resulting in a key length of 56 bits. The algorithm goes through 16 iterations that interlace blocks of plaintext with values obtained from the key. The algorithm transforms 64-bit input in a series of steps into a 64-bit output. The same steps, with the same key are used for decryption. There are many attacks and methods recorded till now those exploit the weaknesses of DES, which

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 6, June 2018

made it an insecure block cipher. The algorithm processes with an initial permutation, sixteen rounds block cipher and a final permutation.

C. RSA Algorithm

It is one of the best known public key cryptosystems for key exchange or digital signatures or encryption of blocks of data. RSA uses a variable size encryption block and a variable size key. It is an asymmetric cryptosystem based on number theory, which is a block cipher system.

It uses two prime numbers to generate the public and private keys. These two different keys are used for encryption and decryption purpose. Sender encrypts the message using Receiver public key and when the message gets transmit to receiver, then receiver can decrypt it using his own private key. RSA operations can be decomposed in three broad steps; key generation, encryption and decryption.

RSA have many flaws in its design therefore not preferred for the commercial use. When the small values of p & q are selected for the designing of key then the encryption process becomes too weak and one can be able to decrypt the data by using random probability theory and side channel attacks. On the other hand if large p & q lengths are selected then it consumes more time and the performance gets degraded in comparison with DES.

Key Generation Procedure

1. Choose two distinct large random prime numbers p & q such that $p \neq q$.
2. Compute $n = p \times q$.
3. Calculate: phi (n) = (p-1) (q-1).
4. Choose an integer e such that $1<e<\text{phi}(n)$
5. Compute d to satisfy the congruence relation $d \times e = 1 \bmod \text{phi}(n)$; d is kept as private key exponent.
6. The public key is (n, e) and the private key is (n, d). Keep all the values d, p, q and phi secret.

IV. HASHING ALGORITHMS

A. MD5 Algorithm

MD5 processes a variable-length message into a fixed length output of 128 bits. The input message is broken up into chunks of 512-bit blocks; the message is padded so that its length is divisible by 512. The padding is as follows: first a single bit, 1, is appended to the end of the message, which is followed by a trial of 0s. The message digest algorithm is a hash function that takes a bit sequence of any length and produces a bit sequence of a fixed small length. The output of a message digest is considered as a digital signature of the input data. MD5 algorithm operates on a 128-bit state, divided into four 32-bit words, denoted A, B, C, and D. The processing of a message block consists of four rounds; each round is composed of 16 similar operations based on a nonlinear function F, modular addition, and left rotation.

- F(B,C,D) = (B AND C) OR ((NOT) B AND D)
- G(B,C,D) = (B AND D) OR (C AND (NOT) D)
- H(B,C,D) = B XOR C XOR D
- I(B,C,D) = C XOR (B OR (NOT) D)

B. SHA Algorithm

In cryptography, **SHA** is a cryptographic hash function which takes an input and produces a 160-bit (20-byte) hash value known as a message digest - typically rendered as a hexadecimal number, 40 digits long. The output of SHA is a message digest of 160 bits in length. It works for any input message that is less than 264 bits. This is designed to be computationally infeasible to: a) Obtain the original message , given its message digest. b) Find two messages

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 6, June 2018

producing the same message digest. The output of SHA is a message digest of 160 bits in length. It works for any input message that is less than 264 bits.

V. COMPARISON OF METHODS

Factors	RSA	DES	3DES	AES
Created By	Ron Rivest, Adi Shamir, and Leonard Adleman In 1978	IBM in 1975	IBM IN 1978	Vincent Rijmen, Joan Daemen in 2001
Key Length	Depends on number of bits in the modulus n where n=p*q	56 bits	168 bits (k1, k2 and k3) 112 bits (k1 and k2)	128, 192, or 256 bits
Round(s)	1	16	48	10 - 128 bit key,12 - 192 bit key,14 - 256 bit key
Block Size	Variable	64 bits	64 bits	128 bits
Cipher Type	Asymmetric Block Cipher	Symmetric Block Cipher	Symmetric Block Cipher	Symmetric Block Cipher
Speed	Slowest	Slow	Very Slow	Fast
Security	Least Secure	Not Secure Enough	Adequate Security	Excellent Security

Fig. 1: Difference between RSA, DES, 3DES, AES

Keys For Comparison	MD5	SHA
Security	Less Secure than SHA	High Secure than MD5
Message Digest Length	128 Bits	160 Bits
Attacks required to find out original Message	2^{128} bit operations required to break	2^{160} bit operations required to break
Attacks to try and find two messages producing the same MD	2^{64} bit operations required to break	2^{80} bit operations required to break
Speed	Faster, only 64 iterations	Slower than MD5, Required 80 iterations
Successful attacks so far	Attacks reported to some extents	No such attach report yet

Fig. 2: Difference between MD5 and SHA

Keys For Similarities	MD5	SHA
Padding	✓	✓
Message bit	✓	✓
Members (Hash Family)	✓	✓
Resource Utilization (same)	✓	✓
Fingerprint	✓	✓

Fig. 3: Similarities between MD5 and SHA

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 6, June 2018

VI. CONCLUSION

After the comparison of various encryption and hashing algorithms, the combination of AES algorithm and MD5 was found to be the best out of the above. The AES algorithm is safer and faster than the others. It also has the advantage of single public key which is shared with both the parties. Though the MD5 algorithm is less safer than the SHA algorithm, it is much faster than the other and hence with the combination of the two, a much more safer system can be built.

REFERENCES

[1] W. Zhang, Y. Lin, S. Xiao, J. Wu, and S. Zhou, Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing, IEEE Transactions on computers, vol. 65, no. 5, May 2016.
[2] D. Song, D. Wagner, and A. Perrig, Practical techniques for searches on encrypted data, in Proc. IEEE Int. Symp. Security Privacy, Nagoya, Japan, Jan. 2000, pp. 4455.
[3] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, Secure ranked keyword search over encrypted cloud data, in Proc. IEEE Distrib. Comput. Syst., Genoa, Italy, Jun. 2010, pp. 253262.
[4] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, Fuzzy keyword search over encrypted data in cloud computing, in Proc. IEEE INFOCOM San Diego, CA, USA, Mar. 2010, pp. 15.
[5] Q. Liu, C. C. Tan, J. Wu, and G. Wang, Efficient information retrieval for ranked queries in cost-effective cloud environments., in Proc. IEEE INFOCOM, 2012, pp. 25812585.
[6] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, Dav id Patterson, Ariel Rabkin, Ion Stoica and Matei Zaharia, A view of cloud computing, Communications of the ACM, April 2010.
[7] G.Singh and Supriya, A study of encryption algorithms (RSA, DES, 3DES and AES) for information security., International Journal of Computer Applications,Volume 67 No.19, April 2013.
[8] P. Gupta and S. Kumar, A comparative analysis of SHA and MD5 algorithm., International Journal of Computer Science and Information Technologies, Volume 5- Nov 2014.
[9] Mahajan, P. and Sachdeva, A., 2013. A study of encryption algorithms AES, DES and RSA for security. *Global Journal of Computer Science and Technology*.
[10] Patil, P., Narayankar, P., Narayan, D.G. and Meena, S.M., 2016. A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA and Blowfish. *Procedia Computer Science*, *78*, pp.617-624.