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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

An Implementation of Privacy Preserving Search on Cryptographic Cloud

KratiMehto, Rahul Moriwal

M.Tech Scholar, Dept. of C.S.E., AITR, RGPV University, Indore, M.P., India

Assistant Professor, Dept. of C.S.E., AITR, RGPV University, Indore, M.P., India

ABSTRACT: Cloud computing is an innovative computational manner for satisfying the need of new generation computing and the storage solutions. That offers scalable computing performance as well as storage solution therefore more than one cloud service providers are collaborating together for offering the scalable solutions. Additionally the data outsourcing techniques are developed to reduce the overhead of maintenance and reducing the computational cost. But data hosting on third party servers is always untrusted. Therefore keep preserve the data on third party servers in secure manner need a cryptographic solution for data storage.

In the similar direction the proposed work is provide the solution for enhancing cryptographic storage solution, authentication process and the data negotiation or retrieval techniques. In order to achieve the efficient solution for cryptographic cloud storage on third party server the MD5 and AES algorithms are used to develop a hybrid cryptographic technique. For improving the authentication mechanism an authentication process is involved through the authentication server. Finally for providing ease in data retrieval technique the feature selection and keyword based efficient search technique is proposed. The proposed keyword based search technique implements the KNN (k-nearest neighbour) algorithm for retrieving the documents from the storage.

The implementation of the proposed technique is performed using JSP (java server pages). Additionally for deployment of the given implementation a public cloud namely OpenShift services are used. After implementation and testing the performance of the implemented system is evaluated in terms of precision, recall and f-measures for finding the query relevance of the data retrieval. Additionally for finding the cryptographic performance the time and space complexity is evaluated. The obtained experimental results demonstrate the effective and efficient computing technique, which is adoptable for third party data storage and retrieval processes.

KEYWORDS: cryptography, security, privacy preserving, data retrieval, cryptographic cloud**.**

I. INTRODUCTION

Rapidly increasing need of computational efficiency and the scalable storage technique which can handle a tons of data leads to work with the cloud computing. That offers the individuals and organizations to write the applications for the cloud platforms and storage the huge amount of data. But the storage on the local disk increases the overhead of maintenance and complexity of cloud servers. In order to manage the data the storage of data can be performed on the third party servers. These servers are specifically designed for storage services and frequent data access.

But data storage on the third party server leads to harm in privacy and security aspects of data therefore the cryptographic techniques are utilized for enhancing the security and privacy of data and data owner.

The increasing demand of computation and the expectation of scalable computing and storage lead to develop new technology cloud computing offers both of them. Additionally need of secure and privacy preserving techniques also fulfil by using cloud. The cloud environment provides support for efficient computing and enables to provide the storage solutions at the remote end. In this presented work the cloud storage and security both are the main aim to achieve. Therefore the following issues are considered for investigation and new approach development:

- Data security
- Data owner and client privacy management
- Searchable data space

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

II. RELATED WORK

Attribute-based encryption was firstly introduced by Sahai and Waters [1], in which they suggested that one's identity can be viewed as a combination of several attributes expressing the characteristics of the user in the form of access policy by using Boolean expressions such as AND, OR, or NOT. Later studies are broadly categorized into key-policy attribute-based encryption (KP-ABE) [1, 2, 3] and ciphertext-policy attribute-based encryption (CP-ABE) [4–6] studies. In KP-ABE, the access policy is associated with keys corresponding to attributes implying that an encryptor is not authorized to grant access to the encrypted content except of descriptive attributes for the data by the encryptor's choice. On the other hand, CP-ABE is complementary to KP-ABE by enabling encryptor to specify access policy combined with the ciphertext. Both schemes allow secure one-to-many communications such as targeted broadcasts for a specific group and individual user according to their attributes, Some studies [7-10] suggested modification of ABE schemes by hiding the access policy. These schemes operate on the assumption that the data owner directly delivers the ciphertext to the receiver without an intermediate third party. In other words, when adopting those approaches directly in cloud storage, decryption keys can be exposed to an unauthorize third party. Hence, they are not feasible for data retrieval services in the cloud storage systems because the test procedure allows the CSP to learn which attributes the user has.

III. PROPOSED ALGORITHM

The proposed cryptographic data storage technique is implemented with the help of AES and MD5 algorithm.

MD5

MD5 algorithm is an essential contribution in data security, which generates a 128 bit length hash for complete data. The MD5 algorithm includes 5 step processes for generating hash [22].

Steps 1 –append padded bits:

- The message is padded so that its length is congruent to 448, modulo 512.
    - Means extended to just 64 bits shy of being of 512 bits long.
- A single "1" bit is appended to the message, and then "0" bits are appended so that the length in bits equals 448 modulo 512.

Step 2 –append length:

- A 64 bit representation of b is appended to the result of the previous step.
- The resulting message has a length that is an exact multiple of 512 bits.

Step 3 – Initialize MD Buffer

- A four-word buffer (A, B, C, and D) is used to compute the message digest.
    - Here each of A, B, C, D, is a 32 bit register.
- These registers are initialized to the following values in hexadecimal:
    - word A: 01 23 45 67
    - word B: 89 ab cd ef
    - word C: fe dc ba 98
    - word D: 76 54 32 10

Step 4 –Process message in 16-word blocks.

- Four auxiliary functions that take as input three 32-bit words and produce as output one 32-bit word.
    - F(X,Y,Z) = XY v not(X) Z

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

- G(X,Y,Z) = XZ v Y not(Z)
- H(X,Y,Z) = X xor Y xor Z
- I(X,Y,Z) = Y xor (X v not(Z))

- Process message in 16-word blocks cont.
  - if the bits of X, Y, and Z are independent and unbiased, the each bit of F(X,Y,Z), G(X,Y,Z), H(X,Y,Z), and I(X,Y,Z) will be independent and unbiased.

Step 5 –output

- The message digest produced as output is A, B, C, D
- That is, output begins with the low-order byte of A, and end with the high-order byte of D

AES

AES is based on a design principle known as a substitution-permutation network, combination of both substitution and combination, and is fast in both software and hardware. Unlike its predecessor DES, AES does not use a Feistel network. AES is a variant of Rijndael which has a fixed block size of 128 bits, and a key size of 128, 192, or 256 bits. By contrast, the Rijndael specification per se is specified with block and key sizes that may be any multiple of 32 bits, both with a minimum of 128 and a maximum of 256 bits [13].

Proposed Hybrid cryptography

The proposed hybrid cryptographic overview is given in the figure 1, in this diagram the data is accepted during data outsourcing processing. And using the MD5 algorithm a 128 bit hash key is generated. This key and original data is provided in the AES encryption algorithm and the cipher text is prepared. Cipher text is used for transmission over the network. At the end of third party server the cryptographic data is stored in server and during the user request the data is decrypted. The finalized cipher text is combination of cipher text and the 128 bit key which is divided in two major parts cipher text and key which are used with the AES to recover the original data.

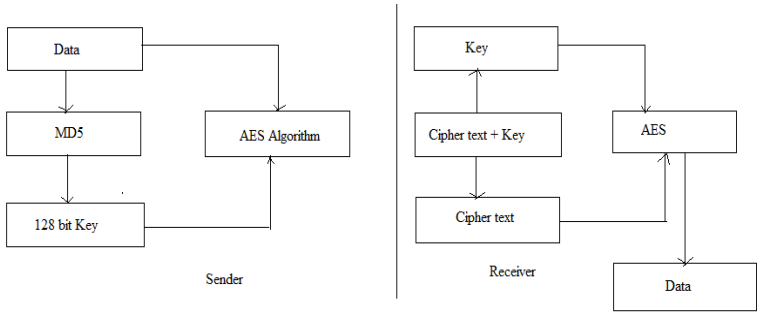

Figure 1 proposed hybrid cryptography

IV. PSEUDO CODE

The given figure 2 shows the proposed security technique that involve the following steps for authentication and data preserving technique.

1. Client node is an end client system who wants to store or retrieve the data from the secure cloud server. In this step the end client initiate the authentication by making the request from the server.
2. In this phase the secure server system trigger the authentication server. After initializing the authentication process user provide input credentials. If the user id and password is matched with the database then data attributes and security questions are ask by the authentication server. If user passes through this and got correct input by the user, OTP (one time password) is applied to make secure the communication between client and server.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

3. After authenticating the user access the system ask for user id, password and OTP again here the OTP works as the salt for the encryption and validation.
4. In this step user initiate the communication and data request from the server, during this the MD5 and AES algorithm is organized for encrypting the data additionally the following information is preserved into the attribute MAP.

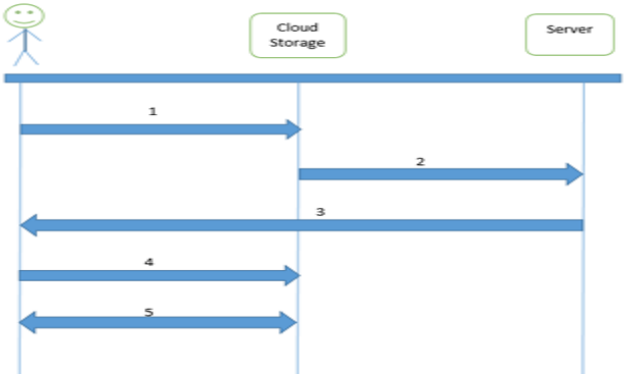

Figure 2 authentication process

The below given working steps can summarized as:

1. User ID-Id is used to identify user uniquely. It is provided by server to user when they interact with server first time. And after that when user wants to access the system.
2. Password
3. Session key
4. Text file features as frequent token
5. Original file name
6. Mapped file name

5. In final step user can access the information and data using the server in this step the KNN algorithm is applied over MAP data for finding the user targeted information from search space.

V. SIMULATION RESULTS

This section provides the understanding about the evaluated performance. The performance evaluation of the proposed technique is performed on the various key parameters that reflect the effectiveness and search relevancy.

A. **Encryption time**

The total amount of time consumed during the encryption is termed here as the encryption time. The evaluated performance of the algorithm for storing the data using cryptographic manner is listed using figure 3.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

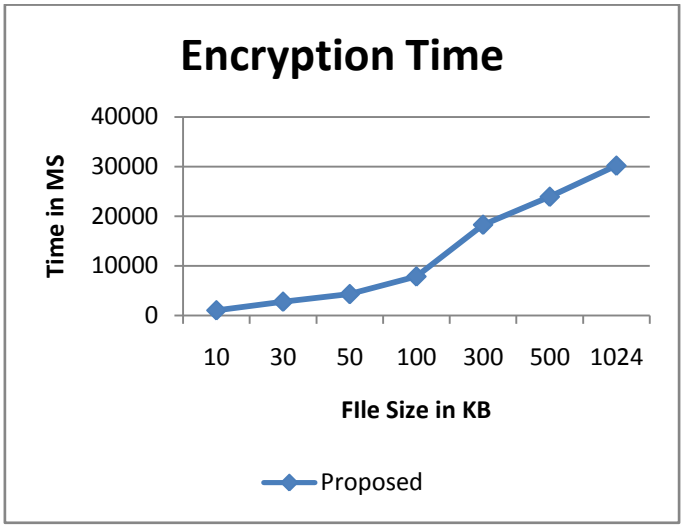

Figure 3 encryption time

Here during experiments the uploading time is not included with the encryption time for finding the absolute encryption time of the proposed system. According to the given figure 3 the encryption time is reported using Y axis and the file size is given using X axis. Encryption time of the proposed technique is depends on the amount of data produced for encryption.

B. Response time

The response time shows the efficiency of a system, and also shows how the system responds during the query input to make search with the system. The amount of time required to return results after submitting the user query is termed here as the response time or the search time. The response time of the proposed cryptographic search methodology is given using figure 4.

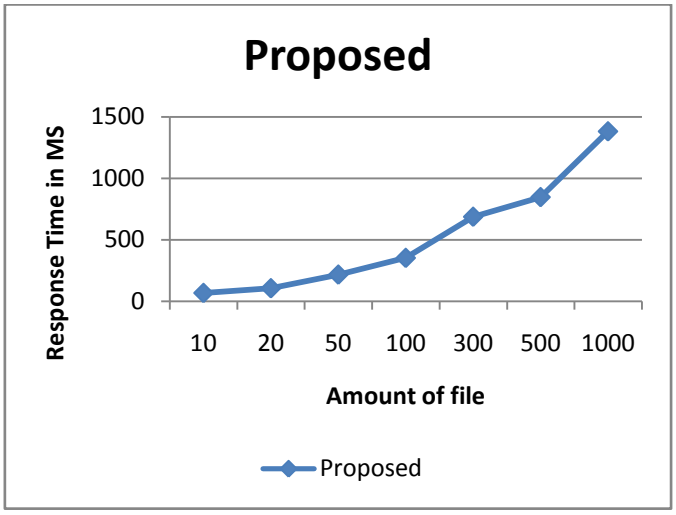

Figure 4 response time

The response time of the search system is given using the figure 4, in this diagram the amount of file in host is given using the X axis and the Y axis shows the amount of time required to retrieve the results from the input query.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

C. Precision

In any search system the precision is a fraction of search results which is most relevant to the input query. The provided precision of the proposed cryptographic search system are given using figure 5. This can be evaluated by the user feedback and can be evaluated using the following formula.

$$precision = \frac{releventdocument \cap retrieveddocumets}{retrieveddocuments}$$

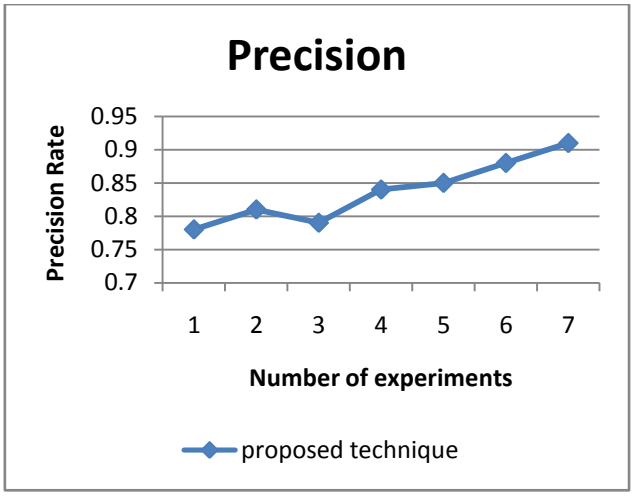

Figure 5 precision rate

Figure 5 shows the precision of the proposed system, according to the precision rate the performance of the proposed cryptographic search technique is increases as the amount of files are increases in the database. In this diagram the different numbers of experiments are demonstrated by increasing the amount of files in server. According to the obtained results the precision of the proposed multi-keyword search for encrypted data is acceptable due to their precision rate.

D. Recall

The search recall values are measured in this section, that is an accuracy measurement in terms of relevant document retrieved according to the input search query. This can be evaluated using the following formula.

$$recall = \frac{relevantdoucment \cap retrieveddocuments}{relevantdocuments}$$

The recall rate of the system is given using figure 6 in this diagram the X axis shows the number of experiments performed and the Y axis shows the recall of the proposed system.

According to the obtained search results the proposed system not much fluctuating and provides the consistent performance during different kinds of key word search through the cryptographic data storage. Therefore the proposed

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

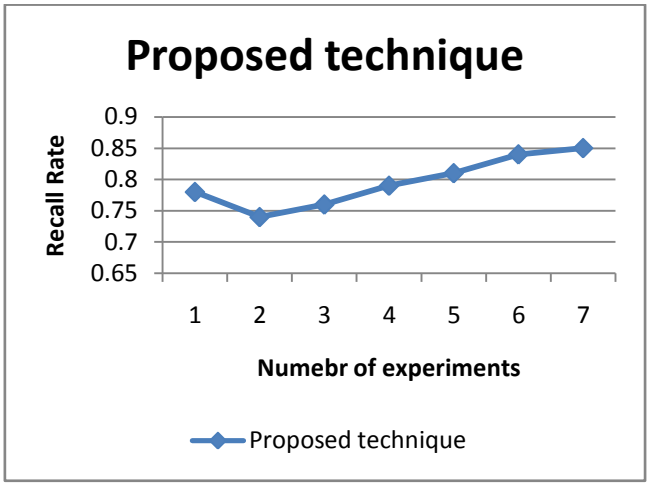

Figure 6 recall rate

model is adoptable for improving the performance of the multi-keyword search in cryptographic storage.

E. F-measures

That is estimated using the precision and recall values estimated using the search or document retrieval technique. That represents the harmonic mean of the system, and can be evaluated using the following formula.

$$f - measure = 2.\frac{precision.recall}{precision + recall}$$

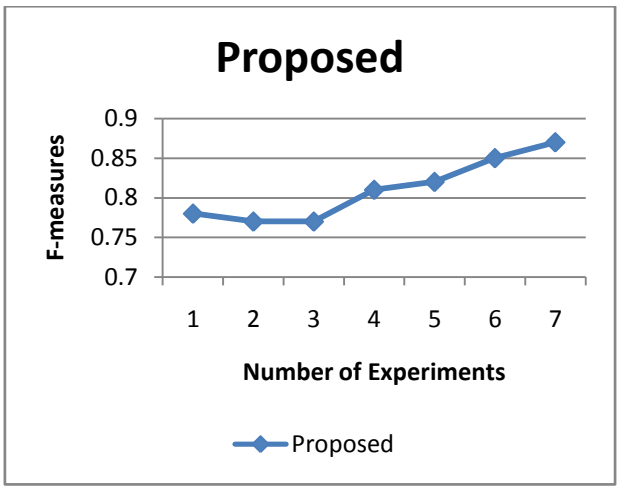

Figure 7 f-measures

The figure 7 shows the f-measures of the proposed cryptographic search system the f-measures is a combination of the precision and recall that shows, how consistent the results are obtained through the proposed systemAccording to the obtained performance the proposed search system found consistent and efficient, therefore the proposed technique is adoptable for the different data retrieval techniques over the cryptographic clouds.

F. Memory consumption

The amount of main memory required to execute the proposed algorithm is termed as the memory consumption. The

memory consumption of the proposed system is defined in the figure 8. In this diagram the X axis shows the amount of files stored on the cryptographic host and the Y axis shows the amount of main memory required in terms of KB.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 12, December 2015

According to the obtained results the performance of the system is depends on the amount of files stored on host additionally that is not much fluctuating during the evaluation of search outcomes.

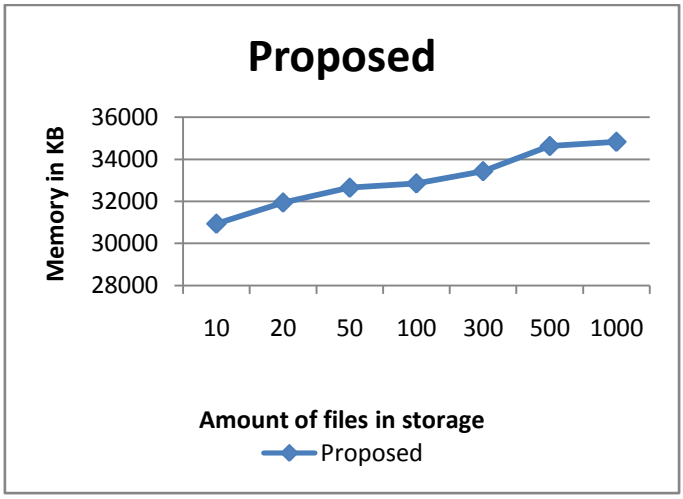

Figure 8 memory consumption

VI. CONCLUSION AND FUTURE WORK

The proposed cloud based cryptographic data storage and retrieval technique is implemented and provides the efficient mechanism for data retrieval process. The proposed technique is in near future extended for the data outsourcing form more than one cloud platform. In addition of that can be extended for improving the data categorization for huge digital library.

REFERENCES

1. Sahai A, Waters B. Fuzzy identity-based encryption. In: Cramer R, editor. Advances in cryptology, EUROCRYPT 2005, vol. 3494. Berlin/Heidelberg: Springer; 2005. p. 557–57. Hong-ryeol Gil1, Joon Yoo1 and Jong-won Lee2,'An On-demand Energy-efficient Routing Algorithm for Wireless Ad hoc Networks', Proceedings of the 2nd International Conference on Human. Society and Internet HSI'03, pp. 302-311, 2003.
2. Katz J, Sahai A, Waters B. Predicate encryption supporting disjunctions, polynomial equations, and inner products. 27th annual international conference on Advances in cryptology, EUROCRYPT'08. Berlin, Heidelberg: Springer-Verlag; 2008. p. 146–62.DilipKumar S. M. and Vijaya Kumar B. P. ,'Energy-Aware Multicast Routing in MANETs: A Genetic Algorithm Approach', *International Journal of* Computer *Science and Information Security* (IJCSIS), Vol. 2,2009.
3. Goyal V, Pandey O, Sahai A, Waters B. Attribute-based encryption for fine-grained access control of encrypted data. In: Proceedings of the 13th ACM conference on Computer and communications security, CCS '06. New York, NY, USA: ACM; 2006. p. 89–98.
4. Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption. In: IEEE symposium on security and privacy; 2007. p. 321–34.
5. Waters B. Ciphertext-policy attribute-based encryption: An expressive, efficient, and provably secure realization. In: Catalano D, Fazio N, Gennaro R, Nicolosi A, editors. Public Key Cryptography, PKC 2011, vol. 6571. Berlin/Heidelberg: Springer; 2011. p. 53–70.
6. Cheung L, Newport C. Provably secure ciphertext policy abe. In: Proceedings of the 14th ACM conference on Computer and communications security, CCS '07. New York, NY, USA: ACM; 2007. p. 456–65.
7. Lewko A, Okamoto T, Sahai A, Takashima K, Waters B. Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption. Advances in Cryptology, EUROCRYPT 2010, vol. 6110. Springer; 2010. p. 62–91.
8. Kapadia A, Tsang PP, Smith SW. Attribute-based publishing with hidden credentials and hidden policies. In: The 14th annual Network and Distributed System Security Symposium (NDSS 07) to appear; 2007.p.179–92.
9. Yu S, Ren K, Lou W. Attribute-based content distribution with hidden policy. In: 4th Workshop on secure network protocols, 2008. NPSection 2008; 2008. p. 39–44.
10. Nishide T, Yoneyama K, Ohta K. Attribute-based encryption with partially hidden encryptor-specified access structures. Applied cryptography and network security, vol. 5037. Berlin/Heidelberg: Springer; 2008. p. 111–29.
11. HASHING ALGORITHMS, binarywarriors.googlecode.com/ files/ HASH_ ALGORITHMS.pdf
12. Hamdan.O.Alanazi, B.B.Zaidan, A.A.Zaidan, Hamid A.Jalab, M.Shabbir and Y. Al-Nabhani, "New Comparative Study Between DES, 3DES and AES within Nine Factors", JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010, ISSN 2151-9617
Hamdan.O.Alanazi, B.B.Zaidan, A.A.Zaidan, Hamid A.Jalab, M.Shabbir and Y. Al-Nabhani, "New Comparative Study Between DES, 3DES and AES within Nine Factors", JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010, ISSN 2151-9617.