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. 5, Issue 11, November 2017

Privacy Preserving on multi-keyword search with Lucene Indexer over Encrypted Data in Cloud

Nilam Jamdare, Prof. K. Vishal Reddy

M.E Student, Dept. of CSE., DIEMS, Aurangabad, India

Professor, Dept. of CSE., DIEMS, Aurangabad, India

ABSTRACT: It is a desirable technique for cloud users to take full advantage of the data encrypted in the cloud by searching for what they need through input keywords. The exact keyword search schemes on the encrypted data have been well addressed for better efficiency and accuracy of recovery. However, existing research on keyword search is based primarily on single-entry keywords, where the search for multiple words has not yet been resolved and the expansion of the word-based search application has not yet been proposed. Key that is, range search based on search. In this document, for the first time, we propose a novel diffuse multi-word search system that supports the rank query by exploiting the encryption of order preservation and the hash sensitive to the localities. Our scheme manages the matching of keywords by algorithmic design to support the recovery classification of returned encrypted files. You can also perform searches with constraints indexed by predefined keywords and eliminate the increase in calculation and search overhead for searches of multiple words compared to traditional keyword search schemes. As an expansion of the application of our scheme, the rank query can be achieved by creating a safe Lucene Index (LI). Therefore, the scheme can achieve a search of several classified keywords as well as a range query in data encrypted in the cloud through BF of two layers per document. The extensive security analysis and experimental results in the real-world data set show that our proposed scheme can safely achieve the design goals for keyword search in encrypted data. To the best of our knowledge, this is the first attempt to achieve a classification of recovery results and range query based on search on data encrypted in the cloud.

KEYWORDS: Cloud computing, blowfish, searchable encryption, privacy preserving, keyword search, ranked search.

I. INTRODUCTION

Cloud computing is the long-dreamed vision of computing as a utility, where cloud customers remotely store their data in the cloud to enjoy high-quality applications and services on demand from a shared set of configurable computing resources. Its great flexibility and economic savings are motivating both people and companies to outsource their complex local data management system in the cloud. To protect the privacy of data and to oppose unsolicited access in the cloud and beyond, data owners must encrypt confidential data, for example, emails, personal health records, photo albums, tax documents, etc., before subcontracting them to the commercial public cloud.

This however, obsoletes the traditional data utilization service based on the search for plain text keywords. The negligible solution of downloading all data and deciphering locally is clearly impractical, due to the large amount of bandwidth cost in cloud-scale systems. In addition, apart from eliminating local storage management, storing data in the cloud is useless unless it can be easily searched and used. Therefore, it is very important to explore the preservation of privacy and the effective search service on data in the encrypted cloud. Considering the potentially large number of on-demand data users and a large number of data documents outsourced to the cloud, this problem is particularly challenging as it is extremely difficult to also meet the requirements for performance, usability of the system and scalability. The classification of the document is provided for quick search, but the priorities of all data documents.

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. 5, Issue 11, November 2017

Remain the same so that the provider of the cloud service and the third party remain on the sidelines of important documents, thus maintaining the privacy of the data.

Classified search can also elegantly eliminate unnecessary network traffic by sending only the most relevant data, which is very desirable in the cloud paradigm "pay as you use it". For the protection of privacy, such classification operation, however, should not filter any information related to keywords. In addition, to improve the accuracy of the search results and to improve the user's search experience, it is also necessary for the classification system to support the search for multiple keywords, since the keyword search often produces results that are too coarse. As a common practice indicated by today's search engines (for example, Google search), data users tend to provide a set of keywords instead of just one as an indicator of their search interest to recover the most relevant data. Along with data privacy and efficient search schemes, real privacy is obtained only if the user's identity remains hidden from the cloud service provider (CSP), as well as from the external user on the cloud server.

II. RELATED WORK

Ensures that multiple words are found in encrypted data in the cloud: In computing cloud computing, business owners tend to instigate complex information management systems from local sites. Still, the public cloud is more flexible and economical. To ensure the security of stored data, it is necessary to encrypt data before storing data. It is necessary to retrieve encrypted data. The specialization of cloud storage should allow multiple keywords in the solitary search query, and generate data in order of relevance in [1]. The primary purpose is to search for multiple search solutions. Based on cloud encryption (MRSE) data, while keeping system privacy in the cloud computing paradigm. There are a multitude of keyword meanings, measuring the similarity of an effective. "Matching" (as much as possible) is used to capture the relevance of the data for search terms. Especially That is, the number of search terms that appear in the document to determine the measure of quantitative similarity of the document to search terms is used in the MRSE algorithm. User ID (ID) will not be hidden. For this reason, those who enter the cloud service provider are known. There may be risks in certain situations where confidentiality is required. So this barrier will overcome in the proposed system.

Confidentiality in finding remote encrypted data: Consider the problem: User U wants to store the file in an encrypted format on the remote file server S. After that, user U wants to get some files efficiently. Encrypted with specific words kept. Secret keywords and not compromise the safety of remote archive files. For example, users may want to store encrypted email messages on servers that Yahoo and other major carriers manage and retrieve certain messages while traveling on mobile devices. [2] The solution to this problem is provided under Clearly defined security requirements. This scheme is effective because it is not a public key cryptosystem. In fact, this method does not depend on the encoding method used for remote files. U users can also send new secure files from previous queries, but they can also search for future searches. The main problem is to keep the data remotely on another server and retrieve it from anywhere via a mobile device, laptop, etc.

Cryptographic Cloud Storage: When the advantages of using a public cloud infrastructure are clear, it presents significant security and privacy risks. In fact, it seems that the biggest obstacle to the adoption of cloud computing is the concern for confidentiality and integrity of information. [3] A general description of the benefits of hosted services. Encrypted data is offered, for example, by reducing the legal exposure of customers and cloud providers and compliance. In addition, cloud services that can be built from encrypted storage services such as secure backups, file system health registries, secure data exchange, and electronic discovery have been described in brief[10].

Efficient and secure multifunctional search in encrypted data in the cloud: on the one hand, users who do not necessarily have prior knowledge of thedata encrypted in the cloud, have to process each recovered file to find the ones that most closely match their interests; On the other hand, the invariable recovery of all the files containing the keyword queried generates unnecessary network traffic, which is absolutely undesirable in the current cloud of pay-per-use paradigm. This document has defined and solved the problem of searching for keywords classified as effective but secure over data encrypted in the cloud [4]. The classified search greatly improves the usability of the system by returning the matching files in a sort order with respect to certain criteria of relevance (for example, keyword frequency), approaching one more step to the practical deployment of hosting services. data that preserve privacy in

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. 5, Issue 11, November 2017

Cloud Computing. For the first time, the document defined and resolved the challenging multi-keyword classified search problem that preserves privacy over encrypted data in the cloud (MRSE) and establishes a set of stringent privacy requirements for a data utilization system in the reality cloud The proposed classification method proves to be efficient to return highly relevant documents corresponding to the search terms presented. The idea of the proposed classification method is used in our proposed system in order to improve the security of the data in Cloud Service Provider.

Provide preservation of privacy in cloud computing: privacy is an important issue for cloud computing, both in terms of legal compliance and user confidence and should be considered at each stage of the design. The [5] document states the importance of protecting the privacy of people in cloud computing and provides some privacy preservation technologies that are used in cloud computing services. Paper says that it is very important to consider privacy when designing cloud services, if this involves the collection, processing or sharing of personal data. From this document, the main theme is to preserve the privacy of the data. This document only describes the privacy of the data, but does not allow the indexed search or hide the identity of the user. Therefore, these two drawbacks are overcome in our proposed system.

Privacy preserving the exchange of data with the anonymous identification assignment: in this document, an algorithm for the anonymous exchange of private data between N parts is developed. This technique is used iteratively to assign these ID numbers of nodes ranging from 1 to N. This assignment is anonymous because the identities received are unknown to the other members of the group. In [6], existing and new algorithms for assigning anonymous identifications are examined with respect to the tradeoffs between communication and computational requirements. These new algorithms are built on a safe sum data extraction operation using Newton's identities and Sturm's theorem. The main idea extracted from this document is to assign an anonymous identification to the user in the cloud. Enable efficient search for keywords on encrypted data in cloud computing: in this document, the main idea is to formalize and solve the problem of the effective search for keywords on encrypted data while maintaining the privacy of the keywords [7]. This basic idea is taken, but it is for searches with multiple keywords (MRSE scheme) in our proposed system. In [8], we propose the design of a secure cloud storage service that solves the reliability problem with an almost optimal overall performance.

Achieving a secure, scalable and accurate data access control in the cloud computing: Achieving a fine granularity, scalability and confidentiality of access control data simultaneously is a problem that has not yet been solved. The document [9] addresses this challenging open issue, on the one hand, defines and applies access policies based on data attributes and, on the other hand, allows the owner of the data to delegate most of the computing tasks involved in data Detailed access control to servers in the cloud that are not trusted without revealing the underlying data contents. In [10], the authors have proposed a public audit system that preserves privacy for the security of data storage in the Cloud Computing scheme. It uses the Homomorphic linear authenticator and randommasking to ensure that the TPA will not learn any knowledge about the data content stored on the cloud server during the efficient audit process, which eliminates the user's burden of the tedious cloud and possibly expensive audit task. It also relieves the user's fear of their outsourced data leakage.

III. PROPOSED SYSTEM

Taking into account a data hosting service in the cloud that involves three different entities, the owner of the data, the user of the data and the server of the cloud. The owner of the data first registers in the cloud using cloud computing services. The owner of the data has a collection of F data documents to be outsourced to the server in the encrypted C form. To enable search capability on C for effective data utilization, the data owner will first build a search index I using F's Lucene Indexer before outsourcing, and then outsource both the index I and the collection of encrypted documents C to the cloud server.

The work deals with efficient algorithms to assign identifiers (ID) to users in the cloud in such a way that the FILE identifiers are anonymous using a distributed calculation without central authority as the data is encrypted.

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. 5, Issue 11, November 2017

Since there are N nodes, this assignment is essentially a permutation of the integers {1 ... .N} with each FILE that is known only by the node to which it is assigned. Our main algorithm is based on a method of anonymously sharing simple data and results in methods for the efficient exchange of complex data.

To search the collection of documents for certain keywords, an authorized user who has an identification and a specific designation acquires a corresponding K through our search control mechanisms.Upon receiving T from a data user, the server in the cloud is responsible for searching the index I and then returns the corresponding set of encrypted documents. To improve the accuracy of document retrieval, the cloud server must classify the search result according to some classification criteria (for example, coordinate match) and assign anonymous FILE ID [6] to the user in the cloud to Make the data cloud more secure. In addition, to reduce the cost of communication, the user of the data can send an optional k number together with the trap door T, so that the server in the cloud only sends the top-k documents that are most relevant to the query of search.

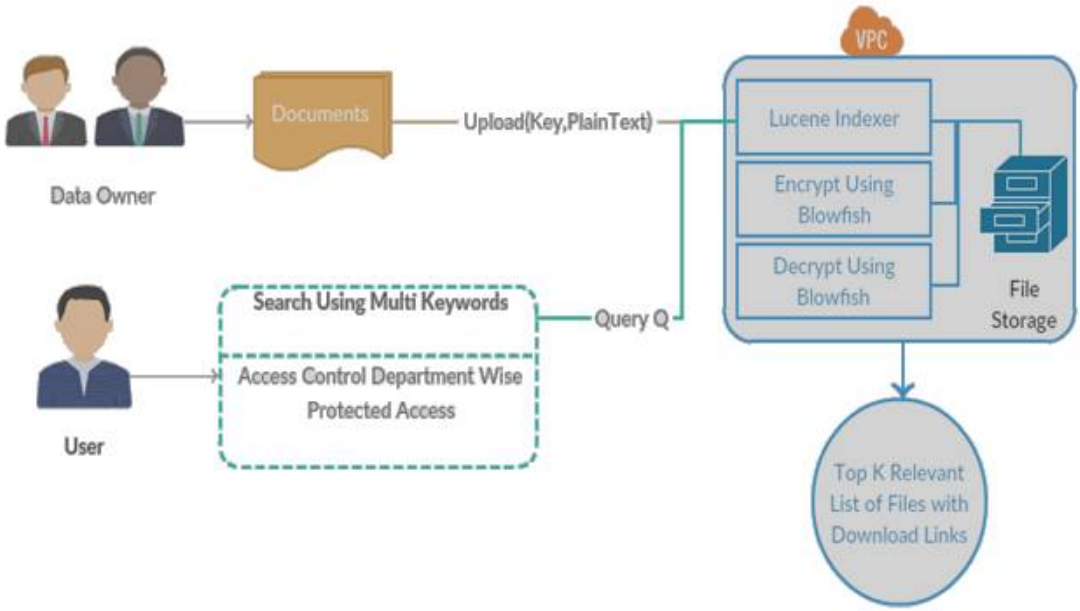


Figure 1 Proposed Architecture

Finally, the access control mechanism is used to manage the decryption capabilities provided to users and the data collection can be updated in terms of inserting new documents, updating existing ones and deleting existing documents.

Encryption Algorithm

Blowfish is a popular security algorithm that was developed by Bruce Schneier in the advent of the year 1994. The algorithm works on the same line as DES and consumes block blocks with blocks of a size of 64 bits. Blowfish became quite popular after its arrival, just because Bruce Schneier [1] himself is one of the most famous experts in cryptology and, above all, the algorithm is not patented, open source is free and available for its use and modifications. Blowfish is a 64-bit block cipher with a variable length key. Define 2 different boxes: S boxes, one box P and four boxes S [3].

Taking into account that P box P is a one-dimensional field with 18 values of 32 bits. The tables contain variable values; those can be implemented in the code or generated during each initialization. The frames S S1, S2, S3 and S4 each contain 256 32-bit values. Blowfish is a symmetric encryption algorithm, which means that it uses the same secret key to encode and decrypt messages. Blowfish is also a block cipher [5], which means that it divides the message into

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. 5, Issue 11, November 2017

blocks of fixed length during encryption and decryption. The block length for Blowfish is 64 bits; Messages that do not have a size of multiples of eight bytes must be filled.

Blowfish consists of two parts: key expansion and data encryption. During the expansion stage of the key, the key entered becomes several matrices of sub-keys in a total of 4168 bytes. There is the matrix P, which is eighteen boxes of 32 bits, and the boxes S, which are four matrices of 32 bits with 256 entries each. After initialization of the string, the first 32 bits of the key are XORed with P1 (the first 32-bit box in the matrix P). The second 32 bits of the key are XORed with P2, and so on, until all 448 or fewer key bits have been XORed. Cycle through the key bits returning to the beginning of the key, until the entire set P has been processed.XORed with the key.

Encrypt the zero string with the Blowfish algorithm, using the modified P matrix above, to get a block 64 bits. Replace P1 with the first 32 output bits, and P2 with the second 32 output bits (from the 64-bit block). Use the 64-bit output as input again in the Blowfish encryption, to get a new block of 64 bits. Replace the following values in the matrix P with the block. Repeat for all the values in the matrix P and all the squares S in order.Encrypt the whole zero chain using the Blowfish algorithm [12], using the modified P matrix above, to obtain a block of 64 bits. Replace P1 with the first 32 output bits and P2 with the second 32 output bits (from the 64-bit block).

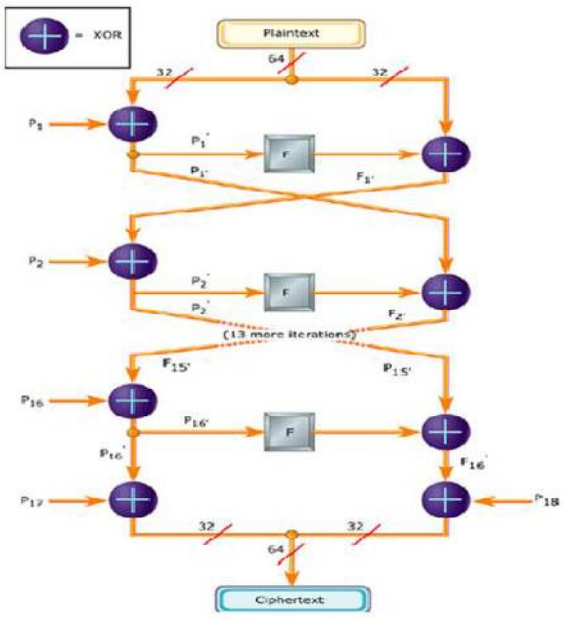


Figure 2 Fiestal Network

Use the 64-bit output as input again in the Blowfish encryption, to get a new block of 64 bits. Replace the following values in the matrix P with the block. Repeat for all the values in the matrix P and all the squares S in order.

Modified Algorithm:

This system basically uses the Blowfish encryption algorithm [12] to encrypt the data file. This algorithm is a 64-bit block cipher with a variable length key. This algorithm has been used because it requires less memory. It uses only simple operations, therefore, it is easy to implement. It is a 64-bit block cipher and is a fast algorithm for encrypting data. It requires a 32-bit microprocessor at a rate of one byte for every 26 clock cycles. It is a variable length key block encryption of up to 448 bits. Blowfish contains 16 rounds. Each round consists of XOR operation and a function. Each round consists of key expansion and data encryption. The key expansion generally used to generate initial contents of a matrix and the data encryption uses a network of 16 round of Feistal [14]. Simple text and key are the entries of this

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. 5, Issue 11, November 2017

algorithm. 64 bit Normal text is taken and divided into two 32-bit data and in each round the given key is expanded and stored in 18 p-array and gives 32bit key as input and XORed with previous round data. The functionality consists in dividing a 32-bit input into four bytes and using them as indexes in an S matrix. Search results are aggregated and XOR together to produce the result. In round 16 there is no function. The output of this algorithm must be 64-bit encrypted text. It is having a function to iterate 16 times of network. Eachround consists of a permutation dependent on the key and a key and a substitution dependent on the data. All operations are XOR and additions in 32-bit words. The only additional operations are four index data search tables indexed for each round.

Function F

Divide xL into four eight-bit quarters: a, b, c and d
F (a, b, c, d) = ((S1,a + S2,b) XOR S3,c) + S4,d
Thus, each round includes the complex use of addition modulo 232 and XOR, plus Substitution using S-Boxes. The function divides a 32 bit input into four bytes and uses those as indices into an S-array. The lookup results are then added and XORed together to produce the output.

Encryption Algorithm

Divide x into two 32-bit halves: xL, xR
For i = 1to 32:
xL = XL XOR Pi
xR = F(XL) XOR xR
Swap XL and xR Swap XL and xR (Undo the last swap.)
xR = xR XOR P17 xL = xL XOR P18
Recombine xL and xR

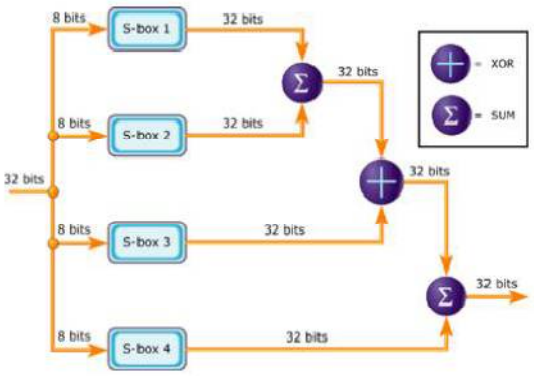


Figure 3) Modified Function F(x)

Decryption
For decryption, the same process is applied, except that the Pi subclasses must be supplied in reverse order. The nature of the Feiestel network [12] ensures that each half is exchanged for the next round (except, here, for the last two sub-words P17 and P18). The proposed algorithm of Blowfish can achieve an efficient data encryption of up to 4 bits per clock. In this design, we avoid limited I / O restrictions by modifying the 64-bit I / O to 16 bits. The proposed architecture should satisfy the need for high-speed data encryption and can be applied to several devices, respectively.
Decryption Algorithm
Divide x into two 32-bit halves: xL, xR
For i = 1 to 16:
xL = xL XOR P19-i
xR = F (xL) XOR xR
Swap xL and xR

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. 5, Issue 11, November 2017

Next i
Swap xL and xR (Undo the last swap.)
xR = xR XOR P2
xL = xL XOR P1
Recombine xL and xR

Performance Analysis

Java Profiling Results

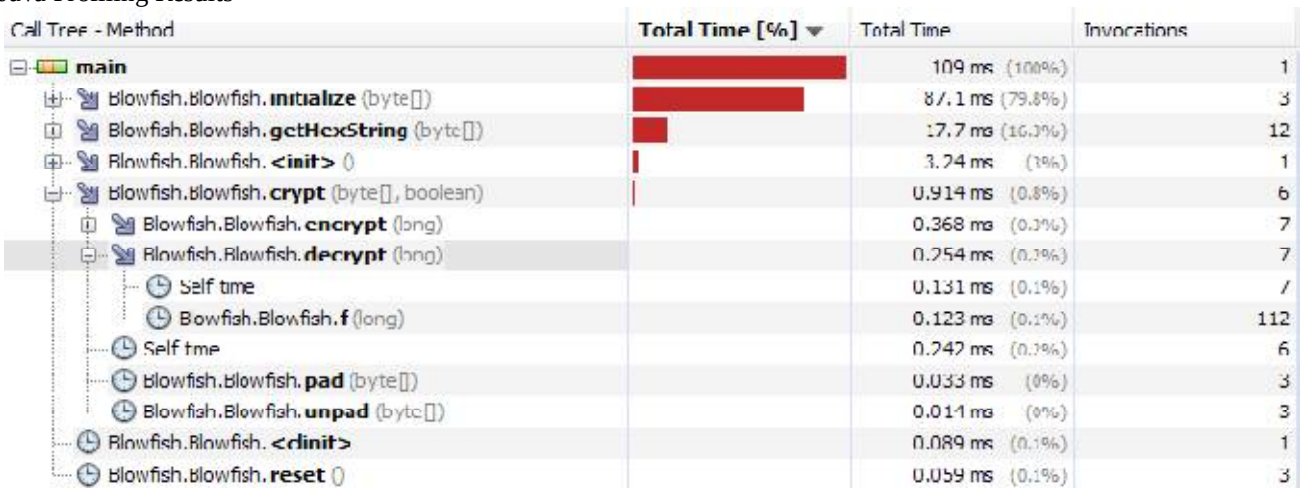


Figure 4a) Hex String Initialization

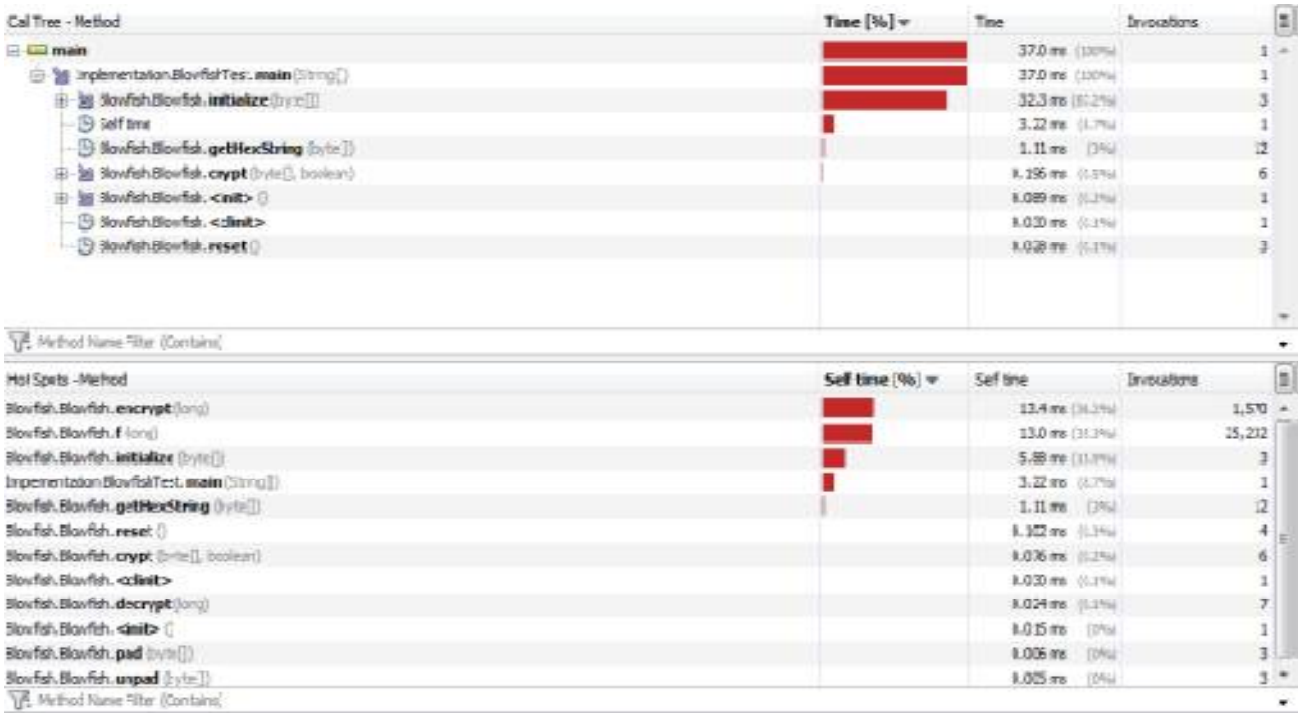


Figure 4.2 Encryption time of standard algortihm& modified algorithm

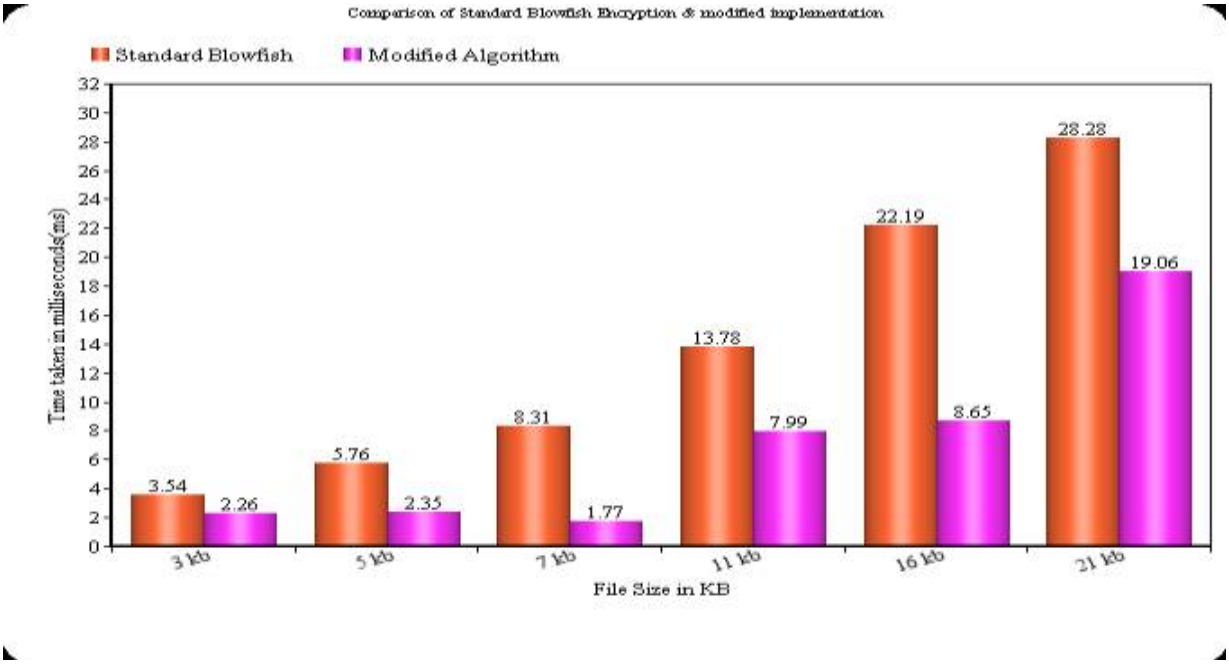
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

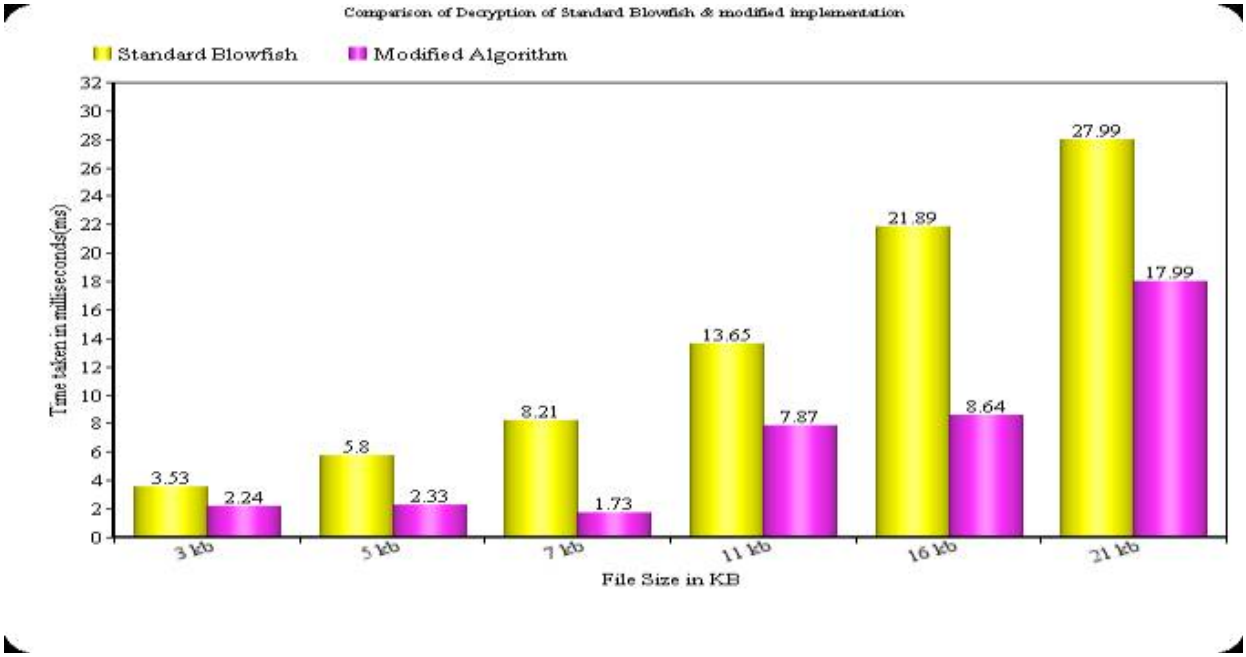
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 11, November 2017



Graph 4.1 Comparison of encryption of standardblowfish & modified implementation



Graph 4.2 Comparison of decryption of standard blowfish & modified implementation

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. 5, Issue 11, November 2017

Table 1 Through put efficiency

File	Size	Throughputin millisecond (Standard Blowfish Algorithm)	Through put in millisecond (Modified Blowfish Algorithm)
Block Specimen 1	3 KB	3.53	2.24
Block Specimen 2	5 KB	5.80	2.33
Block Specimen 3	7 KB	8.21	1.73
Block Specimen 4	11 KB	13.65	7.87
Block Specimen 5	16 KB	21.89	8.64
Block Specimen 6	21 KB	27.99	17.99

IV. CONCLUSION AND FUTURE SCOPE

The previous work [1] focused mainly on providing privacy to the data in the cloud in which multiple-word classified search was used on the data of the coded cloud using an efficient similarity measure of the coincidence of coordinates. The previous work [4] also proposed a basic idea of the use of a safe calculation of the internal product. There was a need to provide more real privacy than this document presents.

Blowfish has a better performance than other common encryption algorithms used. Since Blowfish has not any known security weak points so far, this makes it an excellent candidate to be considered as a standard encryption algorithm. In this system, strict privacy is provided by assigning access restriction and the files and user details are hidden from the cloud service provider and from the external user to protect the user's data in the CSP cloud and from the external user. Therefore, by hiding the attributes of the user and the data of the organization, the confidentiality of the data is maintained.

This work may be useful in several contexts designing ‘secure’ symmetric block cipher algorithms. As a future work we can consider a ranked scheme that can support latent semantic searching that can use only True positive values as vectors for indexing and thereby try to increase the efficiency of the system.

REFERENCES

[1] AnkathaSamuyelu Raja Vasanthi ,” Secured Multi keyword Ranked Search over Encrypted Cloud Data”, 2012
[2] Y.-C. Chang and M. Mitzenmacher, “Privacy Preserving Keyword Searches on Remote Encrypted Data,” Proc. Third Int’l Conf. Applied Cryptography and Network Security, 2005.
[3] S. Kamara and K. Lauter, “Cryptographic Cloud Storage,” Proc. 14th Int’l Conf. Financial Cryptograpy and Data Security, Jan. 2010.
[4] Y. Prasanna, Ramesh .”Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data”, 2012.
[5] Jain Wang, Yan Zhao ,ShuoJaing, and Jaijin Le, ”Providing Privacy Preserving in Cloud Computing”,2010.
[6] Larry A. Dunning, Ray Kresman ,“ Privacy Preserving Data Sharing With Anonymous ID Assignment”,2013.

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. 5, Issue 11, November 2017

[7] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, " Keyword Search Over Encrypted Data in Cloud Computing," Proc. IEEE INFOCOM, Mar. 2010.
[8] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou, "LT Codes-Based Secure and Reliable Cloud Storage Service," Proc. IEEE INFOCOM, pp. 693-701, 2012.
[9] S. Yu, C. Wang, K. Ren, and W. Lou, "Achieving Secure, Scalable, and Fine-Grained Data Access Control in Cloud Computing," Proc. IEEE INFOCOM, 2010.
[10] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing," Proc. IEEE INFOCOM, 2010.
[11] N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou, "Privacy preserving Query over Encrypted Graph-Structured Data in Cloud Computing," Proc. Distributed Computing Systems (ICDCS), pp. 393-402, June, 2011.
[12] Bruce Schneier, "Applied Cryptography", John Wiley & Sons, Inc. 1996
[13] The homepage of description of a new variable-length key, 64-bit block cipher http://www.counterpane.com/bfsverlag.html
[14] Patterson and Hennessy, "Computer Organization & Design: The Hardware/ Software Interface", Morgan Kaufmann, Inc. 1994
[15] B. Schneier, "Description of a New Variable-Length Key, 64-bit Block Cipher (Blowfish)," Fast Software Encryption: Second International Workshop, Leuven, Belgium, December 1994, Proceedings, Springer-Verlag,1994, pp.191-204.
[16] S. Vaudenay, "On the Weak Keys in Blowfish," Fast Software Encryption, Third International Workshop Proceedings, SpringerVerlag, 1996, pp. 27-32.
[17] P. Karthigai Kumar and K. Baskaran. 2010. An ASIC implementation of low power and high throughput blowfish crypto algorithm. Microelectron. J. 41, 6 (June 2010), 347-355.
[18] TingyuanNie; Chuanwang Song; XulongZhi; , "Performance Evaluation of DES and Blowfish Algorithms," Biomedical Engineering and Computer Science (ICBECS), 2010 International Conference on , vol., no., pp.1-4, 23- 25 April 2010. [8] TingyuanNie; Teng Zhang; , "A study of DES

BIOGRAPHY

Nilam Panduarang Jamdare is a Research Assistant in the Computer science and Engineering Department, College of Deogiri Institute of Engineering and Management Studies Aurangabad ,Babasaheb Ambedkar Marathwada University. She received Bachelor of Engineering (BE) degree in 2014 from BAMU, Aurangabad, MS, India. Herresearch interests are Cloud Computing , Top K query, Algorithms etc.