International Journal of Distributed and Cloud Computing
8 (1 & 2) 2020, 09-16
http://www.publishingindia.com/ijdcc

Hybrid Searchable Encryption Scheme for Data Security in Cloud

Prachi Gupta
M.Tech Student, Department of Computer Science & Application, Kurukshetra University, Kurukshetra, Haryana, India. Email: prachigpt.gupta7@gmail.com

Abstract: Cloud Computing is an emerging technology which provides on-demand services based on the pay-as-you-go model which reduces IT costs significantly. Cloud comes with many benefits such as Business continuity, Collaboration efficiency, scalability, resource sharing etc. But still, it constitutes several data security issues such as data confidentiality, integrity, authentication, privacy etc. Thus, to protect sensitive data, the document is encrypted at the data owner side before transmitting it to a cloud environment. Cryptography is necessary to provide data security. Although this overpowers the advantages of cloud computing, whenever data user wants to access a part of data it needs to decrypt the data before using. In this paper, a Hybrid Searchable Encryption Scheme is proposed which provide data security while supporting multi-keyword search over the encrypted data by using a Vector Space Model (VSM). Where sense embedding technique is used to create VSM for the document. To provide data confidentiality combination of AES and Blowfish encryption is used. Sha-3 and RSA are used for data integrity. Steganography is applied for secure transfer of keys and message digest. The experimental result shows that the proposed scheme achieves better data security by providing data confidentiality, integrity and authentication along with multi-keyword search at the client end.

Keywords: Cloud computing, Hybrid encryption, Searchable encryption, Sense embedding based search, Vector space model.

I. Introduction

Cloud Computing is a fast-growing technology that provides hardware as well as software services through a shared pool of resources. Even though the cloud is considered as a very good environment but still it constitutes many security issues such as data confidentiality, integrity and authentication [1]. Confidentiality is the process in which data is hidden from unauthorized access and should be interpreted only by the intended user. Integrity refers to the process in which data is secure from any tampering while transmission. Authentication is defined as the process in which only the intended user has access to data. There are three types of cloud deployment model (Fig. 1): Public Cloud, Private Cloud and Hybrid Cloud:

- *Public Cloud:* Public Cloud deals to provide cloud services to the general public. It is considered unsafe because it is accessible to the general public. A cloud service provider is an owner with all ownership of the public cloud with its own policies. Advantages of public cloud include low cost, no maintenance, scalability and high reliability.
- *Private Cloud:* Private Clouds are those in which sharing of data is done within a particular organisation only and no data shared with other organisation. Private cloud is managed by the organisation itself or by any third party. It is considered to be the safest as the number of users are limited under this category.

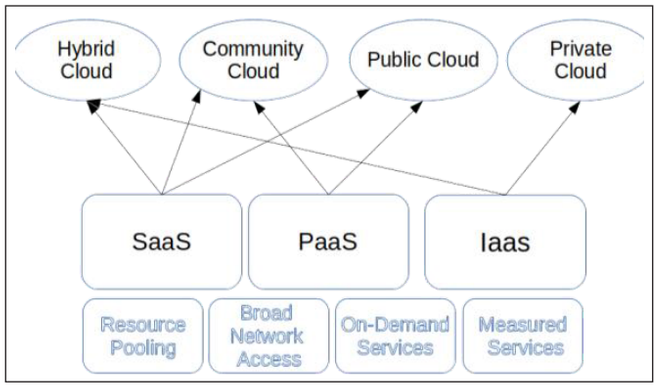

Fig. 1: Cloud Deployment Models

- *Hybrid Cloud:* Hybrid Cloud is a combination of both public and private cloud. The hybrid cloud contains the features and advantages of each cloud model.

Service models in the cloud can also be categorised into 3 categories:

- *Software as a Service (SaaS):* Using SaaS, cloud consumers can release their application on an environment hosted by cloud. The environment can be accessed through networks via different clients. Eg. SalesForce.com, Google Mail, Google Docs etc.

- *Platform as a Service (PaaS):* Paas provides a development platform which supports full software lifecycle allowing cloud users to develop cloud service (SaaS) and application. Google AppEngine is an example of PaaS.
- *Infrastructure as a Service (IaaS):* IT infrastructures such as processing, storage, networks etc are provided to cloud consumers via IaaS. Extensive use of virtualisation is done here to integrate physical resources in an ad-hoc manner. Amazon EC2 is an example of IaaS.

Data security is one of the major issues that come with the usage of the cloud environment thus, to provide security to data using cryptography is necessary. Cryptography is a technique in which data in readable form is converted to unreadable data. There are many cryptographic techniques used for providing security to data. Cryptography is divided into two main categories:

- *Symmetric Technique:* For both the process of encryption and decryption of data a single is used. It is faster because only a single key is used. AES, DES, Blowfish, Twofish are examples of symmetric encryption.
- *Asymmetric Technique:* In this, two separate keys are used one for encrypting the data and other for decrypting the data. It is slower as compared to the previous technique. RSA is one of the examples for this technique.

For maximum security of data, advantages of both symmetric and asymmetric technique need to take into account which leads to the formation of a system known as hybrid encryption. The main goal of data security is to achieve data confidentiality by using symmetric technique, integrity through message digest and authentication via asymmetric technique along with secure transfer of keys from user to client-side. For key transfer mechanism. a technique known as steganography is used. Steganography is a technique which hides the existence of data by concealing the data into a source object. An object may be an image file, audio file, video file or text file.

Whereas encrypting of data before storing to the cloud comes hinders advantages provided by the cloud environment. Each time encrypted data needs to be decrypted at the data user end to perform manipulation. Efforts have been made to conquer this issue by providing a way in which multi-keyword search could be run over encrypted data.

Searchable encryption is a method in which encrypted data is allowed to be searched in a private manner without decrypting the data. For the same, Embedding is defined which is a technique in natural language processing where documents can be represented in form of vectors.

Vector Space Model is a way in which documents are represented in form of vectors. VSM assumes that each word in a document is independent. For each word *w* present in document D. VSM of document cane represented as below:

$$D_{ij} = (w_{11}, w_{12}, w_{13}, ..., w_{nn})$$

For representing the document's features appropriately in vector space many attempts have been made. TF-IDF model is one of the most popular approaches for the same.

Word embedding is another way to represents a document in form of a real-valued vector by capturing a relationship between words present in the document. However, word embedding is unable to consider polysemous word (word having multiple meaning). Sense embedding on the other hand can be used to represent a polysemous word as well.

In this paper, a scheme proposed which provide data with three main security factor i.e. data confidentiality, integrity and authentication along with a feature to search over encrypted data which is achieved through sense embedding technique.

II. Related Work

In this section, we will describe some previously developed hybrid encryption scheme and searchable encryption scheme. Hybrid Encryption combines two or more cryptographic technique to achieve security goals. Searchable encryption is described as a procedure in which encrypted data is stored over cloud and data user is allowed to search that encrypted data.

A. Hybrid Encryption in Cloud

For increasing security of document stored in a cloud environment, a combination of cryptographic encryption scheme are used. Nagasai and Supriya [2] implemented three hybrid models on the text and their time complexity is compared. Combination of AES and RSA algorithms are used along with OTP technique to develop three different models. Olumide *et al.* [3] proposed a system in which a combination of AES and fully homomorphic encryption scheme is used to create a hybrid encryption scheme. Divya *et al.* [4] in their paper proposed a scheme which provides better security on the cloud while transmitting data. Blowfish, RSA and SHA-2 algorithms are used in combination to create hybrid encryption. Citra *et al.* [5] developed an efficient algorithm which receives data security issues data confidentiality. integrity and authentication. A combination of hybrid encryption and steganography is used to achieve security.

B. Searchable Encryption

Further in this section, the previously developed encryption scheme is introduced. In view of the single keyword search over encrypted data, the first idea was developed by Song *et al.* [6]. Symmetric searchable encryption was developed where the search time of scheme was proportional to the size of the collection of the document because searching requires a scan of the whole document. In 2006, Curtmola *et al.* [7] introduced searchable symmetric encryption (SSE). Two SSE were introduced by the author first one was Non-adaptive SSE (SSE-1) and the second one was Adaptive SSE (SSE-2). Addressing issues in previous work in which adversary is not

allowed to search over data, the author achieved a search time which is linear to the document that was matching to search query.

Public key encryption with keyword search (PEKS) implies identity-based encryption in which a user's identity is used to encrypt as well as decrypt the data. Boneh *et al.* [8] in 2004 developed PEKS, in which a user's public key was used to encrypt the data containing certain keywords for performing a search. A keyword is also encrypted which unable server to get keyword location in the document. Z. Deng *et al.* [9] developed a scheme in which multiple users were allowed to search the keyword (MSESKA). 6 polynomial-time algorithms were used to develop multi-user attribute.

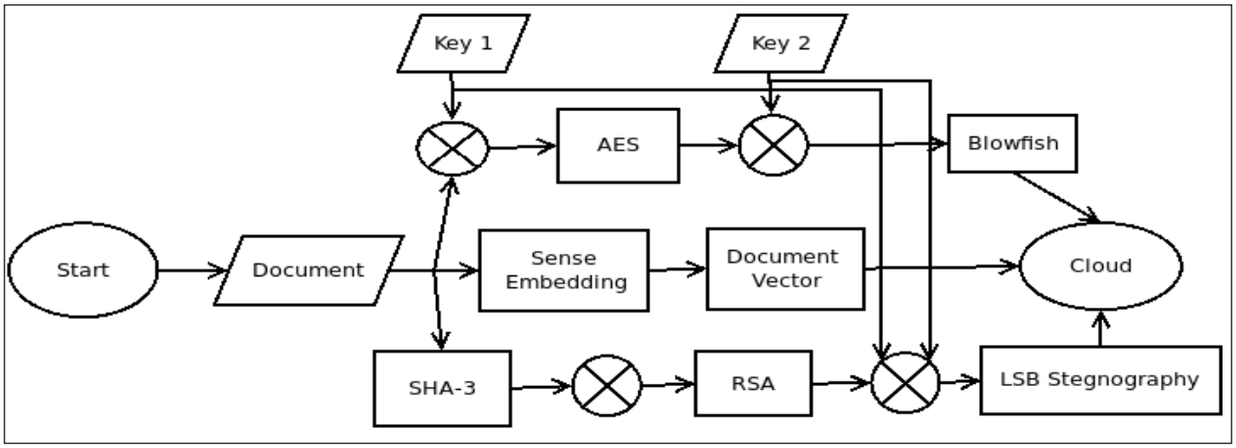

Fig. 2: Encryption Phase

In Predicate encryption, the user does not need any private key related to the public key to search on encrypted data. X. A. Wang *et al.* [10] in their paper claimed predicate encryption to be superior to traditional public-key encryption as the former scheme achieves more flexible and sophisticate functionality. Hidden vector encryption is also a type of predicate encryption. J. Karz *et al.* [11] introduced hidden vector encryption that supports a conjunctive combination of equality along with comparison and subset queries.

Schemes described above support only a single keyword search. Further studies have been conducted to extend search to multi-keyword search over data with a ranking of documents. N. Cao *et al.* [12] achieved a multi-keyword search result by using inner product similarity. Sun *et al.* [13] developed a searchable encryption scheme with document ranking using the vector space model. Aritomo and Watanbe [14] achieved a searchable ranked based scheme using a word embedding to create a vector for documents.

III. Word Embedding V/S Sense Embedding

Word embedding can be considered as a set of language modelling vocabulary words and phrases are mapped to real-valued vectors where vectors are represented in low-dimensional space. Sigmoid function of vector cosine similarity is to define the prediction probability. Two major limitations we can consider while using the word embedding 1) They don't consider the different meaning of a word and conflate the different meaning for a single word into a single vector, and 2) They ignore the wealth of information that is given by the semantic resources thus make their representations only on the basis of distributional statistics that is obtained from corpora. Several types of research have been made to eliminate these two limitations but none was successful in eliminating these two limitations at a time.

Word similarity measurement is used evaluate methods in lexical semantics and semantic similarity. For a given set of words, similarity judgement is created by the system for each pair. These judgements should be closest to those that are provided by humans.

Table I: Spearman Correlation Performance

Approaches	WS-Rel	YP-130	MEN	RG-65	WS-Sim	Average
GloVe	0.559	0.577	0.763	0.769	0.666	0.737
Word2Vec	0.476	0.343	0.665	0.732	0.707	0.644
ESA	-	-	-	0.749	-	-
PMI-SVD	0.523	0.337	0.726	0.738	0.659	0.695
Pilehvar *et al.* [21]	0.457	0.710	0.690	0.868	0.677	0.677
Our Approach	0.703	0.639	0.805	0.871	0.812	0.794

Different word similarity datsets are used to evaluate the performance of different algorithms. YP-130, RG-65, WS-353, WS-Rel and MEN. Our approach of creating sense vectors are compared with several other approaches namely GloVe, Word2Vec, ESA, PMI-SVD and Pilehvar *et al.* [21]. Table I shows the comparison between different approaches on 5 different datasets. Spearman correlation performance is done for all the considered approaches. Our approach for creating sense vector shows high reliability on both similarity and relatedness measurement task. Right most column in the table shows the average performance of all the approaches. This comparison shows that our approach is beneficial. Moving from word to sense embedding can significantly improve the effectiveness and accuracy of the representation.

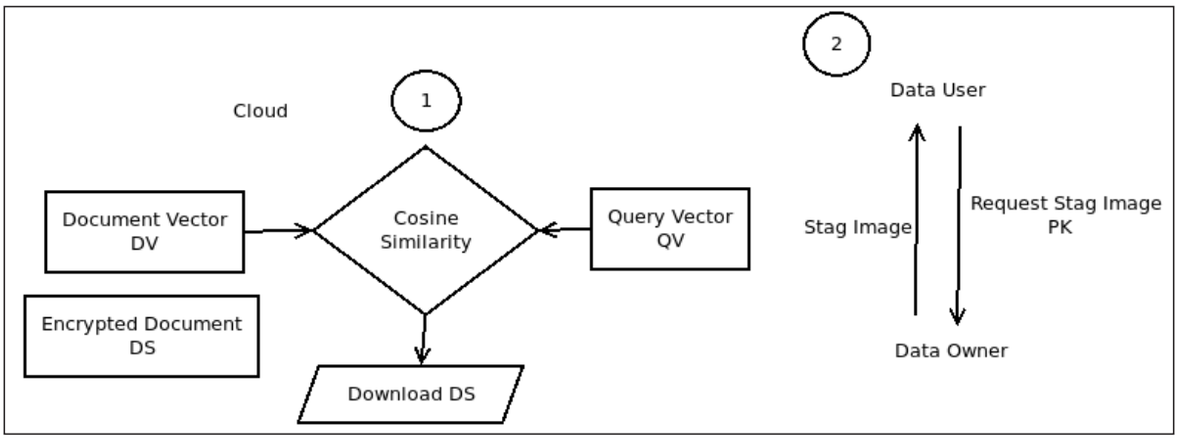

Fig. 3: Search Phase

IV. Problem Definition

A. Scheme Model

The scheme proposed contains three conscious things:

First the data owner which has the original document DC and it generates the corresponding sense embedding vector DV generated through sense embedding technique, second the data user and third one an untrusted cloud server. Document vector represents the features of the document. Sensitive data is protected from an intruder by first encrypting the document termed as secure document DS by data user before uploading it to cloud.

As shown in Fig. 2, firstly AES encryption is used and after then Blowfish encryption is applied to the output of AES encryption. To provide a document with data integrity and authentication message-digest MD of the document is created which is further encrypted using asymmetric encryption termed as secure message digest MDS. The keys (Key 1 and Key 2) and the message digest are transported using steganography technique. While the document vector DV and encrypted document DS is stored on the cloud.

In the searching session, initially, a trapdoor is created by the data used to get authorized by the data owner. The trapdoor is made which contains the query vector QV and is sent to the cloud server. The similarity between the query QV and the document vector DV is calculated by the cloud server and the calculated score is returned to the data user. The data owner according to the calculated score request for the encrypted document from the cloud server.

B. Threat Model

In the proposed scheme, the cloud server is considered, to be honest, but curious which is responsible for the introduction of searchable encryption in our scheme. This means cloud server is responsible to handle the request and pass the data to the data owner in an honest manner, but is curious at the same time. Hence, the cloud tries to recover maximum information from the document. In our scheme, we wish to let cloud infer minimum information while analyzing the data it receives.

C. Model Preludes

- *Word Embedding:* It is a technique where documents are converted into their corresponding vector. Word embedding creates a relation between the words present in the document and tries to represent those words in low dimensional real-valued vector. Doc2Vec and Word2Vec [15] are the two most popular algorithms for representing word embedding of the document. While the most common method to measure the similarity is the cosine measure in which the inner product of the vector is calculated.
- *Sense Embedding:* Word Embedding has a problem that it doesn't consider the polysemous word (word having multiple meaning). Sense Embedding in contrast to word embedding creates embedding for a document where polysemous words are also considered. There are many techniques [16], [17], [18] to do so but in our model, we have considered SensEmbed [19] to implement sense embedding to the document to create VSM.
- *Symmetric Encryption:* Symmetric Encryption is a technique which converts the plain text to cipher text

using only a single key that is a private key. It is also known as private key encryption. There are many symmteric encryptions available but AES and Blowfish are considered to be two of the best encryption algorithm. For first encryption, AES is considered and later Blowfish is applied to the output of AES encryption.

- *Asymmetric Encryption:* In contrast to symmetric encryption which only needs a single key for encryption as well as decryption, asymmetric encryption uses two separate keys each for encryption and decryption process. Asymmetric encryption is considered to be more secure as is allows user to share only the public key and not the private key. But asymmetric encryption is slower than symmetric encryptions when talking about time efficiency. RSA is one of the most popular algorithms used in this category.

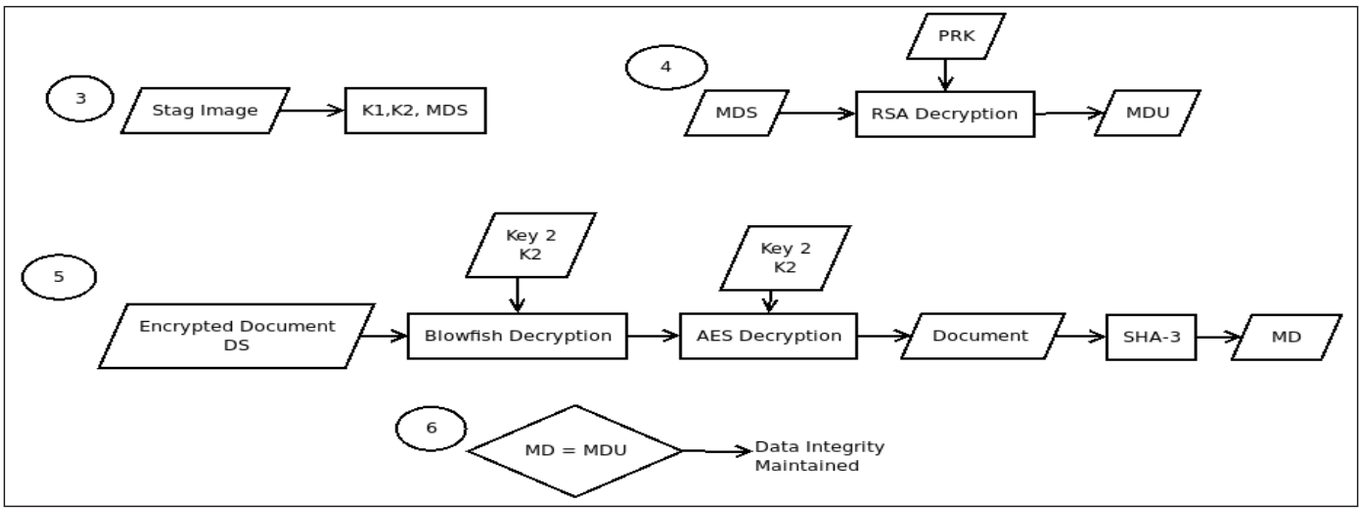

Fig. 4: Decryption Phase

- *Message Digest:* Message digest function creates a fixed-length one-way message of the document which is used to check the integrity of the document. If the document is modified in any way while the transfer from the data owner to data user then its message digest will definitely vary from the message digest of the original data. Many algorithms are there to create message-digest but one of the which SHA-3 is considered for our scheme.
- *Steganography:* Steganography is a method in which data is hidden under some other source. The source may be an image, audio or video. One of the most commonly used steganography technique is the LSB image steganography in which the transmitted data is stored in the Least Significant Bit of the source.

V. Scheme Methodology

In this section, description of the scheme methodology of plaintext is given which describes the data owner side scenario. The overview of the proposed scheme is simple and shown in Fig. 2.

A. Encryption Phase

The data user first creates a query vector QV by using same sense embedding technique as that of the data owner and send that query vector QV to the cloud server. As a measure to calculate the similarity between the query document QV and the document vector DV by the cloud server, cosine similarity between the two vectors that are query vector QV and document vector DV is generated.

Setup (K1, K2, PK): Firstly the system is initialized, two secret key K1 & secret key K2 is created by the data owner. While public key PK is obtained from the data user.

These keys are used for AES encryption, Blowfish Encryption and RSA encryption respectively. These AES and Blowfish encryption are used to encrypt the document and RSA encryption technique is used for encrypting message digest.

BuildScheme (DV, DS, MDS): The data owner runs the sense embedding technique to create sense vector DV for the document. Further, the message digest of the document is created using SHA-3 and then encrypted using RSA encryption using the key PK. The document is considered for encryption now using AES and Blowfish with the key K1 and K2 respectively.

Transport (K1, K2, MDS, IMG): The keys used for encryption that are K1 and K2 and the encrypted message digest termed as message digest secure MDS is transport using image IMG in which these three things are hidden using LSB steganography.

Table II: Notations

Q	Query
QV	Query Vector
DV	Document Vector
DS	Encrypted Document (Secure)
MD	Message Digest of Document
MDU	Message Digest Created by Data User
MDS	Encrypted Message Digest (Secure)
PK	Public Key of Data User
PRK	Private Key of Data User
IMG	Original Image
K1	Key 1 for AES
K2	Key 2 for Blowfish

B. Search Phase

Cloud server contains the encrypted document DS and the document vector DV for the document. Step 1 gets initiated when the data owner starts searching the cloud for the document of use.

Step 1: Search (QV, DV): The query vector QV which represents the vector of an interested word in query Q is received from data user and is used to calculate similarity with the document vector DV which contains the vector of polysemous words present in the document. This procedure is done without decrypting the encrypted document DS.

C. Decryption Phase

This section describes the data user's end. Where the decryption process for the required document takes place. Data user generates a query vector using the interested words as shown below:

GenQuery (Q, QV): The data user generates a query, for the keywords of interest W a query vector is generated QV which contains the interested words in query Q. QV is further used to search the plaintext vector DV. QV is sent to a cloud server for further working.

As described in part B of section IV Search (QV, DV) is generated using the DV present with the cloud server and QV provided by the Data user. If the search is successful the secure document for which Search (QV, DV) was generated is downloaded.

Step 2: The data user asks for the stag image from Data Owner as shown in Fig. 3 where the image consists of K1, K2 and secure message digest. Afterwards, Transport (K1, K2, MDS, IMG) is initiated and is transferred from Data owner to data user.

Step 3: Using the stag image, secure message digest (MDS), K1 and K2 are recovered as shown in Fig. 4.

Step 4: Secure message-digest MDS is decrypted to recover the message digest MDU using the data users own private key PRK.

Step 5: Using the keys K1 and K2 the document is recovered from secure document DS. First blowfish decryption is initiated using K2 and AES decryption process is applied to the output of blowfish decryption. Using the decrypted document, a message digest is generated using the same message digest function as that used at data owners side.

Step 6: MDU and MD are compared and considered for checking data integrity. If both MD and MDU are equal then data integrity is maintained.

VI. Performance Analysis

In this section, performance of the proposed scheme is analyzed by implementing the scheme and measuring multiple metrics. The proposed scheme was implemented in python 3.6.

- *Sense Embedding Generation:* Sense embedding for the document is created using babelfy library in python while considering the babelnet dataset. Further for creating the vector, word2vec in used from gensim python library. The time is not considered during creating the embedding for document. This scheme is open to any sense embedding technique so considering the time for one technique is not of much meaning. Desired secure document is downloaded only when similarity between the query vector and document vector is matched. If the similarity is matched then the desired secure document is download and decrypted using the AES and Blowfish Encryption.

(a) Original Image

(b) Stag Image

Fig. 5: Steganography Result

- *Data Confidentiality:* To make sure that the document which is to be stored on cloud server can be read only by data user the document is encrypted before storing to cloud. For purpose of encryption of document to make it secure, AES and Blowfish algorithms are used. Both AES and blowfish are part of crypto library in python.
- *Data Integrity:* To ensure integrity of data which means that the data is not tampered during the whole procedure while transfer from the data owner to data user. Message digest was created using SHA-3 which is also a part of Crypto python library.
- *Data Authentication:* For authenticating data user, asymmetric encryption is used which make use of data using a public key to encrypt the message digest and data users private key is used to decrypt the same. RSA is also the part of Crypto python library.
- *Key Transfer:* For transfer of keys and secure message digest between the data owner and data user, LSB steganography is used. Original image Fig. 5a and Stag image Fig. 5b is shown in Fig. 5.

TABLE III: COMPARISON BETWEEN PROPOSED SCHEME AND PREVIOUS SCHEMES

Scheme → / Parameters ↓	Proposed Scheme	Daisuke Aritomo and Chiemi Watanabe [14]	Divya Prathana and Ajit Kumar Santra [20]
Confidentiality	YES	NO	YES
Integrity	YES	NO	YES
Authentication	YES	NO	YES
Key Transfer	YES	NO	NO
Search Attribute	YES	YES	NO

The proposed scheme is compared with two of the existing scheme [14]. [20] Based on the parameters described above that are Data confidentiality, Data integrity, Data Authentication, Key transfer and Searchable encryption, the comparison between schemes is shown in Table III.

The proposed scheme protects the document from malicious users at the time of transmission and also in cloud server storage. The proposed scheme increases the difficulty level for intruders or hacker to decrypt the document through double encryption and RSA. In the scheme, the encrypted document can be searched with multiple keywords which make it convenient and time-efficient.

VII. CONCLUSION

In this paper, Hybrid Searchable Encryption Scheme in Cloud was proposed which is based on using a combination of symmetric and asymmetric encryption while providing a way to generate multi-keyword based search over encrypted data. The proposed scheme makes the following contribution: Data confidentiality of a document is maintained using a combination of two symmetric encryption algorithm that is AES and Blowfish. Data integrity is maintained by creating a message digest of the document which is encrypted using RSA encryption to ensure data authentication. Concept of steganography is used for secure transfer of keys from data owner end to data user end. Existing sense embedding are easily adopted to create a document vector. A scheme which meets requirements in such a way that the cloud server is honest but curious is proposed. A secure search scheme is proposed by creating a vector of the document using sense embedding technique which makes data user search over encrypted data.

REFERENCES

[1] A. Hendre, and K. P. Joshi, "A semantic approach to cloud security and compliance," 2015 *IEEE 8th Int. Conf. Cloud Comput.,* New York, NY, 2015, pp. 1081-1084, doi: 10.1109/CLOUD.2015.157.

[2] N. L. Kodumru, and M. Supriya, "Secure data storage in cloud using cryptographic algorithms," *2018 4th Int. Conf. Comput. Commun. Control and Automat.,* Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697550.

[3] A. Olumide, A. Alsadoon, P. W. C. Prasad, and L. Pham, "A hybrid encryption model for secure cloud computing," *2015 13th Int. Conf. ICT and Knowl. Eng. (ICT and Knowl. Eng., 2015),* Bangkok, 2015, pp. 24-32, doi: 10.1109/ICTKE.2015.7368466.

[4] D. P. Timothy, and A. K. Santra, "A hybrid cryptography algorithm for cloud computing security," *2017 Int. Conf. Microelectronic Devices, Circuits and Syst. (ICMDCS),* Vellore, 2017, pp. 1-5, doi: 10.1109/ICMDCS.2017.8211728.

[5] C. Biswas, U. D. Gupta, and M. M. Haque, "An efficient algorithm for confidentiality, integrity and authentication using hybrid cryptography and steganography," *2019 Int. Conf. Elect., Comput. and Commun. Eng. (ECCE),* Cox'sBazar, Bangladesh, 2019, pp. 1-5, doi: 10.1109/ECACE.2019.8679136.

[6] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," *Proc. 2000 IEEE Symp. Secur. and Privacy (S&P'2000),* Berkeley, CA, USA, 2000, pp. 44-55, doi: 10.1109/SECPRI.2000.848445.

[7] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," In *Proc. 13th ACM Conf. Comput. and Commun. Secur. (CCS'06),* Association for Computing Machinery, New York, NY, USA, 7988, 2006, doi: https://doi.org/10.1145/1180405.1180417.

[8] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," In C. Cachin, and J. L. Camenisch, Eds., Advances in

Cryptology EUROCRYPT 2004. EUROCRYPT 2004. *Lecture Notes in Computer Science,* vol. 3027, 2004. Springer, Berlin, Heidelberg, 2004. [Online]. Available: https://doi.org/10.1007/978-3-540-24676-3 30

[9] Z. Deng, K. Li, K. Li, and J. Zhou, "A multi-user searchable encryption scheme with keyword authorization in a cloud storage," *Future Generation Computer Systems*, vol. 72, pp. 208-218, Jul. 2017, doi: https://doi.org/10.1016/j.future.2016.05.017.

[10] X. A. Wang, F. Xhafa, W. Cai, J. Ma, and F. Wei, "Efficient privacy preserving predicate encryption with fine-grained searchable capability for cloud storage," *Computers and Electrical Engineering*, vol. 56, pp. 871-883, Nov. 2016.

[11] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," In *Annu. Int. Conf. Theory and Appl. Cryptographic Techn. (EUROCRYPT 2008),* 2008, pp. 146-162.

[12] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," In *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 1, pp. 222-233, Jan. 2014, doi: 10.1109/TPDS.2013.45.

[13] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," In *Proc. 8th ACM SIGSAC Symp. Inf., Comput. and Commun. Secur.,* Association for Computing Machinery, New York, NY, USA, 2013, pp. 71-82, doi: https://doi.org/10.1145/2484313.2484322.

[14] D. Aritomo, and C. Watanabe, "Achieving efficient similar document search over encrypted data on the cloud," *2019 IEEE Int. Conf. Smart Comput. (SMARTCOMP),* Washington, DC, USA, 2019, pp. 1-6, doi: 10.1109/SMARTCOMP.2019.00020.

[15] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013. arXiv:1301.3781.

[16] S. Rothe, and H. Schütze, "AutoExtend: Extending word embeddings to embeddings for synsets and lexemes," In *Proc. 53rd Annu. Meeting Association Comput. Linguistics and 7th Int. Joint Conf. Natural Lang. Process.,* Beijing, China, Jul. 26-31, 2015, pp. 1793-1803.

[17] A. Neelakantan, J. Shankar, A. Passos, and A. Mccallum, "Efficient non-parametric estimation of multiple embeddings per word in vector space," 2015. 10.3115/v1/D14-1113.

[18] A. Trask, P. Michalak, and J. Liu, "sense2vec - A fast and accurate method for word sense disambiguation in neural word embeddings," 2015. arXiv abs/1511.06388.

[19] I. Iacobacci, Mohd. T. Pilehvar, and R. Navigli, "SensEmbed: Learning sense embeddings for word and relational similarity," In *Proc. 53rd Annu. Meeting Association Comput. Linguistics and 7th Int. Joint Conf. Natural Lang. Process.,* Beijing, China, Jul. 26-31, 2015, pp. 95-105.

[20] D. P. Timothy, and A. K. Santra, "A hybrid cryptography algorithm for cloud computing security," *2017 Int. Conf. Microelectronic Devices, Circuits and Syst. (ICMDCS),* Vellore, 2017, pp. 1-5, doi: 10.1109/ICMDCS.2017.8211728.

[21] Mohd. T. Pilehvar, D. Jurgens, and R. Navigli, "Align, disambiguate and walk: A unified approach for measuring semantic similarity," *ACL 2013 - Proc. 51st Annu. Meeting Association Comput. Linguistics,* vol. 1, 2013, pp. 1341-1351.