I. J. Computer Network and Information Security, **2024, 4, 52-64**
Published Online on August 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2024.04.04

Secure Data Storage and Retrieval over the Encrypted Cloud Computing

Jaydip Kumar
Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India
E-mail: jaydipkumar2001@gmail.com
ORCID iD: https://orcid.org/0000-0001-9350-8775

Hemant Kumar*
Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India
E-mail: hemant20192@gmail.com
ORCID iD: https://orcid.org/0009-0005-4040-4071
*Corresponding author

Karam Veer Singh
Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India
E-mail: kvsingh.bhu@gmail.com
ORCID iD: https://orcid.org/0000-0003-1935-0011

Vipin Saxena
Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India
E-mail: profvipinsaxena@gmail.com
ORCID iD: https://orcid.org/0000-0003-1035-1704

Received: 21 December 2022; Revised: 27 February 2023; Accepted: 28 April 2023; Published: 08 August 2024

Abstract: Information security in cloud computing refers to the protection of data items such as text, images, audios and video files. In the modern era, data size is increasing rapidly from gigabytes to terabytes or even petabytes, due to development of a significant amount of real-time data. The majority of data is stored in cloud computing environments and is sent or received over the internet. Due to the fact that cloud computing offers internet-based services, there are various attackers and illegal users over the internet who are consistently trying to gain access to user's private data without the appropriate permission. Hackers frequently replace any fake data with actual data. As a result, data security has recently generated a lot of attention. To provide access rights of files, the cloud computing is only option for authorized user. To overcome from security threats, a security model is proposed for cloud computing to enhance the security of cloud data through the fingerprint authentication for access control and genetic algorithm is also used for encryption/decryption of cloud data. To search desired data from cloud, fuzzy encrypted keyword search technique is used. The encrypted keyword is stored in cloud storage using SHA256 hashing techniques. The proposed model minimizes the computation time and maximizes the security threats over the cloud. The computed results are presented in the form of figures and tables.

Index Terms: Genetic Algorithm, Biometric Authentication, Secure Cloud Storage, Encrypted Keyword Search.

1. Introduction

Cloud computing is a rapidly growing technology with a wide range of applications in today's digital world and it has been increasingly in demand in the recent years for the real-time processing and storing personal and professional data over the cloud through high-speed internet connectivity. Despite the fact that the cloud is only few years old, it has deeply impact on the computing industry. It is based upon the public, private, or hybrid-based technology used over cloud, hence called as public, private and hybrid cloud. The firm does not have to worry about cloud hosting management or maintenance in the public cloud because it is handled by the service provider while on the other hand private cloud is owned by a single firm. Hybrid cloud is a combination of public and private clouds. Although the user is free from data center administration but concerned about data security because the data is in the hands of service

provider. Infrastructure as a service, software as a service, web-based cloud computing, platform as a service, managed service, utility services, and other services are all examples of cloud computing. The generated data can be stored on the storage devices connected through switches by means of topological structure which are managed and control by the service provider. The fundamental benefit of the cloud is that whenever a user, or an organization, or a company employee needs to access the data around 24x7, authentication is performed first. Passwords and PIN's technology are used for authentication, but the main disadvantage is that users must remember the passwords and PIN's and users often use the same passwords everywhere so that users may remember every time. Across the internet, most of hackers are breaking such types of Passwords and PIN's daily and it is not difficult work for a hacker. Therefore, multi-factor authentication is now employed by cloud services and mobile phones.

Cloud computing is a computing model in which resources like processing power, storage, net-working and software are abstracted and made available as services over the internet in a remotely accessible manner. National Institute of Standards and Technology (NIST) defines the five key characteristics of cloud computing which are briefly described below:

- Rapid Elasticity: The ability to scale resources up and down as needed is defined as elasticity. The cloud looks to be endless to the consumer, and the consumer can purchase as much or as little as per choice.
- Measured Services: The cloud provider manages the cloud computing and monitors certain parts of the cloud service. Billing, access control, resource optimization, capacity planning, and other duties all require measured services.
- On-demand Self Service: Without interacting with the cloud provider, a consumer can access cloud services across 24x7.
- Ubiquitous Network Access: The capabilities of the cloud provider are available across the network and can be accessed by a variety of clients using standard procedures.
- Resource Pooling: It permits a cloud provider to use a multi-purpose architecture to serve the customers. According to consumer demand, physical and virtual resources are assigned and reassigned. There is a notion of location independence in that the client has little control or awareness of the specific location of the resources given, but may be able to choose one.

With rising security concerns and rapid improvements in networking, communication, and mobility, the demand for dependable user authentication techniques has grown. The security requirements for securing personal and professional contents in cloud computing communications are determined by how manages access control, including user authentication and other measures.

Biometrics, which is defined as the science of recognizing an individual based on physical or behavioral characteristics which is slowly gained acceptance as a legitimate technique of determining an individual's identification. Biometric system has been used to establish identity in a variety of commercial, civilian, and forensic applications. Because of its uniqueness and stability, biometrics technique is commonly utilized in identity identification systems. Biometric templates, unlike traditional keys, are notoriously difficult to cancel and republish. In fact, in recent years, fingerprint-based identity identification systems have gotten a lot of attention.

Sensitive data is frequently encrypted before being outsourced to preserve data privacy, which makes effective data usage a difficult undertaking. Traditional searchable encryption approaches allow a user to search over encrypted data using keywords and selectively retrieve files of interest, but these techniques only support specific keyword searches. That is, tiny errors and format inconsistencies are not tolerated. When user's searching inputs exactly match the predefined keywords, fuzzy keyword search considerably improves system usability by returning matching files or the closest feasible matching files based on keyword similarity semantics when the exact match fails.

Many different Internet-based services are provided by cloud computing which is more popular in this current era. As a result, many attackers and malicious individuals constantly attempt to gain access to a user's confidential data without having the necessary permissions. As a result, cloud data security has recently become a major problem. DNA computing, which is based on the biological concept of DNA, is an advanced emerging discipline for strengthening data security.

The main key challenges related to this research work are the security key which is not sufficient to prevent data leakage from the cloud computing. The password or PIN is also not enough to provide higher level of security and encryption/decryption of data using classical techniques which can be easily breakable through the Quantum system. One of the most important issues related to cloud computing in current days are the finding or searching desired data from huge amount of cloud storage which takes time.

The main objective of this research work is the cloud computing needs as a secure system to manage unwanted users or intruders who get access to cloud environment. The cloud computing environment needs a secure framework which provide overall security to the cloud user such as authentication, encryption/decryption and secure search over the cloud computing.

The contribution of this research work is to propose a secure access control system using a bio-metric authentication system which improves the data accessibility as well as design a data hiding technique using genetic algorithm to avoid data leakage over communication channel. Thereafter a design is proposed for a data search technique using fuzzy system to get data in lessor time from the huge cloud storage.

2. Related Works

In the current scenario, the security of images on cloud is a very challenging task. The cloud user who uploaded personal or professional images on cloud has no control of security. Let us describe some of the important literature related to present work. Kakkad et at.[1] have presented an idea to secure images on cloud using biometric authentication system and implemented a proposed model in two basic steps such as image compression and encryption. The compression of images was done using standard discrete wavelet transformation method and encryption of images was done by hybrid of SHA and Below Fish. To access the original as well as biometric data have faced several challenges related to security. It needed to prevent unauthorized access of cloud data. The biometric authentication technique is more secure than other traditional security techniques. Bhattasali et al. [2] have provided a brief survey of security and privacy based on biometric remote authentication system related to cloud and also suggested that there must be a balance between security and privacy during design of authentication system. The traditional searchable encryption systems allowed a user to safely search over encrypted data through the keywords and retrieve the only file. These retrieval methods only offer exact keyword search and zero tolerance on typos error. Li et al. [3] have implemented a cloud data retrieval model based on fuzzy keyword search. The model returned exactly matched keyword or closest matched keywords file. The traditional identification methods have drawbacks which include easily losing valuable information and unsatisfactory performance because of complex inputs including image rotation, incomplete input images and poor image quality. Due to this Zhu et al. [4] have proposed a new fingerprint authentication scheme which was based on a set of assembled invariant moment features to ensure the secure communication. The concept of cloud computing is still in its development stage because of problems with security and privacy issues and other factors. Ali and Sridevi [5] have proposed an API for the cloud computing which covered authentication problem using Bio-Signals and one round Zero Knowledge problem (ZKP) and AES algorithm for encryption and decryption of user id.

Network security, which supports the computer network, but it has many challenges in the real-world problem such as database hijacking and spoof matching process of web-based network. Manzoor et al. [6] have proposed an authentication security system using fingerprint and pin-number. Applied fingerprint features and user mobile number are stored to the database in encrypted form. If any intruder masks the fingerprint data and use it, the system generates matching ratio 80% to 90% and at the same time system generate alert and send to registered mobile number. For cloud computing, biometric recognition technologies are interesting build strong cryptography and personal authentication method. However, if the biometric is compromised the process cannot be canceled easily. Shi et al. [7] has implemented a fuzzy commitment protocol for fingerprint recognition using minutiae-based sector encoding scheme for cloud-assisted IoT. The minutiae of a fingerprint are classified into many designed sectors are encoded it and also the key was encrypted using BCH and Hash mapping.

A promising technology that enabled the user to search over encrypted data is called as Searchable Encryption (SE). However, the majority of SE strategies are unable to handle shared records with hierarchical structure. Miao et al. [8] have proposed an attribute-based keyword search over hierarchical data (ABKS-HD) scheme using ciphertext-policy attribute-based encryption (CP-ABE) technique and also proposed an improved scheme (ABKS-HD-I, ABKS-HD-II) for multi keyword search and user revocation. The majority of the searchable encryption schemes currently use the exact keyword search. These techniques fail to produce the desired result when the data user makes spelling mistakes. In searchable encryption, the cloud server may provide the user with an incorrect result in order to reduce computing costs. Ge et al. [9,10] have proposed first exact keyword search scheme and then extend it to fuzzy keyword search and implemented linked list as a secure index for efficient storage.

To manages the secure searching stockpiling recovery of client information Tariq and Agarwal [11] have used wild card security technique in fuzzy keyword searching in encrypted cloud data and also used double encryption technique to authenticate the server. Multi-keyword fuzzy search produces the correct results even if the search keyword contains minor spelling or typos errors. But due to high processing expense and lack of support for dynamic file changes, the fuzzy keyword search is inappropriate. Zhong et al. [12] have implemented a multi-keyword fuzzy searches scheme to handle the dynamic file update for cloud computing. The Locality sensitive Hashing (LSH) and Bloom filters techniques are used to generate index and query vectors. Li, Wang and Yu [13] have proposed security model using locality sensitive hashing (LSH) and Bloom filter for the closest keyword matching encrypted document search.

The biggest challenge is the field of cloud computing is security and sharing without the losing of cloud data confidentiality. To overcome from this problem Ge et al. [14] have proposed ciphertext-policy attribute-based system with keyword search and data sharing (CPAB-KSDS). The authors have proved that the proposed scheme is secure against the ciphertext as well as keyword attacks. Due to all services provided by the cloud computing are web-based, therefore many attackers and malicious user try to get data from the cloud and sometime attacker replace the data with fake data. Namasudra et al. [15] have implemented a DNA based model by the 1024-bit secret key based on DNA computing and also used user's attribute, Media Access Control (MAC) and decimal encoding using ASCII value.

The use of biometric identification that protects privacy and effectively offers secure cloud servers and storage of sensitive data has gained more attention. Due to more attention frauds cases are increasing day by day. Jasmin and Jasper [16] have designed a Privacy-Preserving Multi-Biometric Identification (PPMBI) technique which avoid higher level of security issues. The authors have also integrated fingerprint and finger vein feature for enhancing data security. To protect keyword privacy most of the current fraud's detection schemes uses keyword primary leakage as well as

access pattern and search pattern simultaneously. Due to data leakage from above discussed issues Li, Cao and Ren [17] have proposed secure keyword search index called as global index pair using k-nearest neighbor search and also have used a scheme to generate trapdoor with data user's computing of inverse matrix.

The Fuzzy Searchable Encryption (FSE) schemes are not for the practical use because of the given reason such as Inflexibility, inefficiency and limited robustness and also it is very difficult for them to avoid linear analysis attacks in the known background model. To remove the above problem Liu et al. [18] have implemented matrix-based multi-keyword fuzzy search (M2FS) scheme for approximate keyword search. Fu et al. [19] have proposed two semantic information retrieval schemes such as Bidirectional Encoder Representations from Transformers (BERT) based SSRB-1 and SSRB-2. The documents are trained by BERT and keyword vectors are generated to hold information of documents which improve the accuracy of information retrieval. The digital data which stored in cloud storage like text, images, video, audio etc. have still faced lots of security breaches such as data loss, quality issues etc. Due to this Joseph and Mohan [20] have proposed DNA-based cryptography to improve data security and also have used Grey Wolf Optimization (GWO) technique to generate optimal encrypted data. The strong authentication system avoids illegal user access to the cloud services which is the main requirements of cloud computing. In this regard, Kaur et al. [21] have presented a one-way Hash and nonce-based two-factor secure authentication scheme using the password and OTP verification system. This system has avoided brute force attack, session and account hijacking attacks, MITM and replay attacks.

3. Background

In the proposed work, there are following three entities as described below in brief:

- Cloud Service Provider (CSP): supports infrastructure and offers cloud services by deploying several servers with enough power and memory space and has overall administrator or central authority;
- Data Owner's (DO's): are the entities that depend on the CSP to manage the data and keep sensitive or general data in the cloud environment;
- Data User's (DU's): are the authorized entities who wish to access data or any other service from a cloud server.

In addition to above, the proposed work has the following goals:

- to achieve an efficient and scalable data storage scheme for the cloud environment that must provide a strong security to the user's confidential data;
- to provide the data access rights for cloud users by using fingerprint authentication system;
- to design Genetic Algorithm for encryption and decryption of cloud data;
- to search the desired data from the huge amount of cloud storage with search of fuzzy keyword;
- to achieve the security goal with SHA256 hashing functions to encrypt keyword.

3.1. SHA256 Algorithm

As the next generation of SHA functions, SHA-256 is a cryptographic hash function which was introduced in 2000 and approved as a FIPS standard in 2002 [22]. With a digest length of 256 bits, SHA-256 is an example of a secure hash operation coming under the SHA-2 category. This algorithm's main function is to take an arbitrary-length input and produce outputs that are a fixed length. In order to accomplish this, a number of procedures should be taken, including padding the message for the preprocessing phase, block decomposition, and hash computation. The SHA-256 algorithm takes a message that is less than 2^{64} bits long and generates a 256-bit message digest. The digest is a short summary of the message, and any change to the message is very likely to cause of change in the digest. SHA-256 algorithm provides 128-bit security. The six logical operations which work with message are used in the SHA-256 algorithm [23]:

$$Ch(x, y, z) = (x \wedge y) \oplus (\sim x \wedge z) \tag{1}$$

$$Mj(x, y, z) = (x \wedge y) \oplus (x \wedge z) \oplus (y \wedge z) \tag{2}$$

$$\sum_{0}(x) = ROTR^{2}(x) \oplus ROTR^{13}(x) \oplus ROTR^{22}(x) \tag{3}$$

$$\sum_{6}(x) = ROTR^{6}(x) \oplus ROTR^{11}(x) \oplus ROTR^{25}(x) \tag{4}$$

$$\sigma_{0}(x) = ROTR^{7}(x) \oplus ROTR^{18}(x) \oplus SHR^{3}(x) \tag{5}$$

$$\sigma_1(x) = ROTR^{17}(x) \oplus ROTR^{19}(x) \oplus SHR^{10}(x) \quad (6)$$

where ROTR and SHR stand for the rotate right and shift right functions, respectively, where ∧, ∼ and ⊕ are the bitwise AND, NOT and XOR operations. The 8 registers are updated through a number of rounds. There are 64 rounds in the SHA-256 compression algorithm. Arithmetic addition, a round dependent constant named K_i, two linear functions called 0 and 1, and two nonlinear functions called Ch and Mj are all parts of every round function [24].

3.2. Fuzzy String Matching

Finding strings which approximately match a pattern is a method called fuzzy string matching, which also refers to fairly efficient string matching. The number of basic operations which properly translate the string into a match is known as the Levenshtein (LS) distance, and it is used to calculate the proximity of matches. Various practical uses for fuzzy string matching exist. Spell-checking, text re-use detection, spam filtering, as well as a number of applications in the bioinformatics field, such as matching DNA sequences, are good examples [25]. The number of insertions, deletions, and substitutions needed to change the response string into the target string is measured by the LS metric. For example, a participant is transcribing a record of the word "water" but writes out "wayer" by mistake. Target then equals "water," Response equals "wayer," and there is only one substitution needed to change Response into Target [26].

The Fuzzy-Wuzzy package supports the following four popular methods of fuzzy matching logic:

- Ratio: uses only distance-based Levenshtein matching;
- Partial Ratio: Based on the best substring matches;
- Token Sort Ratio: Before matching, the strings are tokenized and sorted alphabetically;
- Token Set Ratio: Compares the intersection and remainder after tokenizing the strings.

3.3. Fingerprint Authentication

In the past few years, the integration of identity and authentication systems into daily life has grown exponentially, including access control, online payments, bank account access, and device unlocking. It has become biometric recognition which is an essential field in the cyber-security. A new kind of small area sensor called as a "touch sensor" has been widely used and integrated into portable equipment to enable fingerprint authentication. The present minutia approach for fingerprint authentication with touch sensors could not be significant amount of minutia and extracted [27]. Graph-based and minutiae-based fingerprint-based authentication methods are both available. User enrollment and user authentication are the two key processes in a fingerprint-based authentication system. An acquisition system initially takes a picture of the user's fingerprint. The image is characterized by a number of image processing techniques in order to identify and extract the information. The user is already considered as enrolled when the extracted details are saved in a data base. The user gives a fingerprint image for user authentication, which is processed once again to find and extract the details. The standard details kept in the database are then compared to these minutiae. Based on how many minutiae match, a reference score is determined. If the score rises above a particular level, the user is recognized as authenticated [28, 29]. The steps for the fingerprint matching algorithm are given below.

Step 1: Input two fingerprint images one from fingerprint scanner and other from database;
Step 2: Perform preprocessing operations such as enhancement, fingerprint edge detection;
Step 3: Detect geometric feature points in both fingerprint images using various detectors;
Step 4: Describe the key points using flannbasedmatcher() and KnnMatch ().
Step 5: Perform matching and geometrically verify the two images;
Step 6: Provide the matching score.

3.4. Genetic Algorithm

Genetic Algorithm (GA) has been extensively used to resolve optimization issues with or without applied restrictions. GA is widely used in computer sciences, mathematics, and the natural sciences. GA is used in computer science to solve optimization and security issues which are both limited and unrestrained. Because GA is capable of addressing NP-hard problems, it can significantly reduce the enormous computational complexity by resolving optimization problems quickly.GA is a bio-inspired computing technique that continuously improves each individual solution for the chosen population. The fundamental GA processes are population generation, crossover, and mutations [30]. The initialization of a population with appropriate sizes and carefully selected fitness function, which is necessary to provide an appropriate result that requires the use of a genetic algorithm. Three operators employed the chromosomes to create a new population with a higher fitness value [31]. The GA operators are elaborated in given below:

- Selection selects the most promising elements using fitness function;
- Crossover combines the two parents to form the children for next generation;
- Mutation mutates the small random changes in chromosomes to get a new solution.

The four nucleotide bases of DNA are Adenine (A), Cytosine (C), Thymine (T), and Guanine. The nucleic acid in DNA contains Genetic information (G). The gene is used in the DNA sequence that contains the genetic information of all living things [32].

4. Proposed Method

Due to the enormous enhancement in the cloud computing, it needs more security for data. The security of data depends on access authentication system, encryption or decryption of data as well as secure searching. To make cloud computing more secure, in the present work the authentication approach as well as cloud data encryption and secure keyword search based on the fuzzy logic are applied. The fingerprint-based authentication system is integrated with cloud based proposed model for the access control to avoid the illegal activities over the cloud computing. The fingerprint authentication system consists of Fast Library in Approximate Nearest Neighbor (FLANN) and KNN based image matching algorithm to authenticate the user is genuine or not. The secure data storage in cloud computing uses the DNA Genetic algorithm for encryption and decryption of data. Due to the huge amount of data in the cloud storage it is very serious problem to get desired data from the cloud. Fuzzy keywords are used to get desired searched data based on keyword and the following notations are used in the algorithm.

Table 1. Notations for used in the algorithm

Notation	Description	Notation	Description
C_{id}	Customer ID	K_{w2}	Second Keyword
F_{in}	Fingerprint	K_1	Hash value of K_{w1}
D_m	Device Model No	K_2	Hash value of K_{w2}
C_{ip}	Current IP	K_s	Plaintext of Search Keyword
G_{eo}	Geo location	$F_{ks\_ratio}()$	Fuzzy Keyword Search Algorithm
D_t	Date & Time	U_n	User Name
DB	Database	E_m	User's Email
M	Message	F_{in}	Fingerprint
M_c	Ciphertext of Message	B_{data}	Binary data
K_{w1}	First Keyword	M_c	Ciphertext of message

The proposed framework is analyzed and implemented to give better result as compared to other existing methodologies. The pseudo code of the proposed model and algorithms are given below.

Pseudo code for proposed algorithm

IF registered_user→signIn
  IF (C_{id} & finger = = TRUE)
    C_{id}, D_m, C_{ip}, G_{eo}, D_t→ DB
    IF (Upload == TRUE)
      M ← Message
      M_c←DNA_{inc}(M, Key)
      K_1← SHA256 (K_{w1}, Key)
      K_2← SHA256 (K_{w2}, Key)
      M_c, K_1, K_2→ DB
    IF (Download == TRUE)
      K_s← Search Keyword
      K_3 ← SHA256 (K_s, Key)
      IF (F_{ks_ratio}(K_3) = = 100%)
      M ←DNA_{dec} (M_c)
      ELSE reject
  ELSE reject
ELSE →SignUp
 U_n ←User_name
 E_m← Email
 F_{in} ← Input Fingerprint
 U_n, E_m, F_{in} → DB

In the proposed algorithm, the already registered user can only sign-in, otherwise need to register before upload or download from the server. During the sign-in process the user can sign-in using user id and fingerprint, if the server authenticated successfully, the system captures the user's customer id, device model number, current IP address,

geographical location and current date and time for fraud detection if any occur in future. The administrator can track the intruders through the captured details. After successfully authenticated, if user wants to upload data to the cloud server, the message M is encrypted using genetic algorithm with the help of symmetric key and stored in M_c. The two different keywords K_{w1} and K_{w2} are also encrypted through the SHA256 with the same symmetric key. The encrypted ciphertext of message and keywords are stored in the cloud data storage.

Algorithm 1: Fingerprint Authentication

Step 1: Compute keypoints and descriptor such as K_{p1}, D_{esc1}, K_{p2}, D_{esc2} for both original and Database fingerprint image using Scale Invariant Feature Transform (SIFT) algorithm.
Step 2: Calculate the matches using flannbasedmatcher() and KnnMatch() algorithm.
Step 3: Now store each matches in p and q, and store p to M_{points} if p.distance< 0.1 * q.distance
Step 4: Store length of K_{p1} in K_{points} if len(K_{p1}) <= len(K_{p2}) otherwise store K_{p2} in $K_{pionts.}$
Step 5: Now calculate the match percentage
IF len(M_{points})/K_{points}> 0.95
$M_{Percent}$= len (M_{Points})/ k_{points}* 100
ELSE
$M_{percent}$ = len (M_{points}) / K_{points}

In the fingerprint authentication, the keypoints and descriptor of the both database and current captured images are calculated using the SIFT algorithm and by using the flannbasedmatcher() and KnnMatch() algorithm, the authenticator find the match points between both images. After getting matches if p.distance< 0.1 * q.distance then store match point in M_{points} and store the length of K_{points} in K_{p1} if len(K_{p1}) <= len (K_{p2}). Now calculate the match percentage through the match points and key points.

Algorithm 2: DNA Encryption/Decryption

A- DNA-Genetic encoding pseudo code

B_{data1} ← convert secret data (M) into binary
Reshape B_{data1}
Group each two adjacent bits
DNA-Bases ← B_{data1}
While (Round != 0) do
Encrypt (B_{data2}, key)
Reshape B_{data2}
Crossover operation
Mutation
M_c= Reshape B_{data2}

The encryption algorithm is done by the genetic algorithm, the secret data (M) is converted into binary format and reshape it. Each binary digit is converted into DNA bases and compute optimized encrypted information using crossover, mutation operators.

B-DNA-Genetic decoding pseudo code

B_{data3} ← convert M_c into binary
Reshape B_{data3}
While (round != 0)
Mutation
Crossover operation
Reshape B_{data3}
Decrypt (B_{data3}, key)
Secret Data M = Reshape B_{data3}

In the DNA-Genetic decoding algorithm, the encrypted ciphertext M_c is reshaped and compute the original information using mutation, crossover with the secret key and get the final information.

Algorithm 3: Fuzzy Keyword Search Algorithm

Step 1: Let's take an input file F_{input} with keywords list (K_1, K_2, K_3,…., K_n)
Step 2: Compute apriori math file $F_{apriori}$ with the database_list {D_{w1}, D_{w2}, D_{w3}, ….D_{wn}}
Step 3: Load $F_{apriori}$ file in Append mode (a+)
Step 4: Compute F_{input} in Read mode (r)
Step 5: Read and process F_{input}
Split {K_1, K_2, K_3,...,K_n} into a list of mathematical functions {F_1, F_2, F_3… F_n}

Return $F_{filtered}$= {F_{w1}, F_{w2}, F_{w3} …. F_{wn}}

Step 6: Get ratio by applying *fuzz_partial_ratio*($F_{apriori}$, $F_{filtered}$)

Now match {D_{w1}, D_{w2}, D_{w3}, ….D_{wn}} with {F_{w1}, F_{w2}, F_{w3} …. F_{wn}}

For F_{wi} in $F_{filtered}$:

For D_{wi}in$F_{apriori}$:

IF (*fuzz_partial_ratio*(F_{wi}, D_{wi})> 75 &&F_{wi}.*isalpha*() &&*len*(F_{wi})>=3):

IF (F_{wi}not in $F_{apriori}$) :

$F_{apriori}$.append (Database_list (i, F_{wi}))

i++

ELSE :

Skip F_{wi}=F_{wi+1}

Step 7: Repeat step 6 unlit not found matched keyword

In this algorithm the encrypted keyword is fetched from the large number of databases. The collection of searched keywords F_{input} and the database keywords list $F_{apriori}$ are loaded in the different lists and each keyword is splitted into single words and computed a mathematical function {F_1, F_2, F_3… F_n}. Now apply the fuzzy_partial_ratio function for getting matching ratio percentage.

5. Results and Discussion

The above hybrid algorithm is proposed to enhance the private information of user's which is stored in cloud data storage. This hybrid model is the combination of different highly secure techniques to reduce the data security breaches. Biometric techniques are used to avoid illegal activities and DNA Genetic algorithm is used to encrypt or decrypt secret information of users which need to store in cloud storage. Fuzzy keyword search techniques are also used to get searched information from the huge amount of cloud data.

The proposed model is implemented in the Python 2.7.17 platform. Hardware specifications can be used to analyze the encryption and decryption times of various cryptographic processes. It uses an Intel(R) Core (TM) i5-10210U CPU at 1.60 GHz and 8GB of RAM. The operating system is Ubuntu 20.04 LTS. Consider an example with input text "Welcome to BBAU" and take any two keywords first is "BBAU Lucknow" and the second is "BBAU". Any user wants to upload the above information over cloud storage. The Genetic algorithm encrypts the information and produces the encrypted ciphertext as "GTCTGCTAAGCCACTATTTATAGATGAACTTTTGAACATTTTAGAACCAACTGTGACCCTCGGC" and both keywords also encrypted using SHA256 hash function using any symmetric key. The hash value of both keyword is "043F2C3C37BA26DE339FDE1D43F2FED0B7F981EA23BE63D87380496B7CDA70E2" and "87F00546256E3FE6C6969EDC9E2911E200AF51CA806FCC148099B8540B9F1BED". The process of data encryption takes 5 milliseconds which is shown in the fig. 1 and all the information is and stored over the cloud storage.

If any user wants to retrieve the data related to "BBAU Lucknow" keyword. The system gets the SHA256 hash value of the searched keyword and uses the fuzzy keyword search algorithm to retrieve data related to search keyword and gets the ciphertext of original information and takes 10 milliseconds as decryption time and also takes 8 milliseconds for keyword search which is shown in fig. 2. Now decrypts the ciphertext and gets the original information which was stored over the cloud storage.

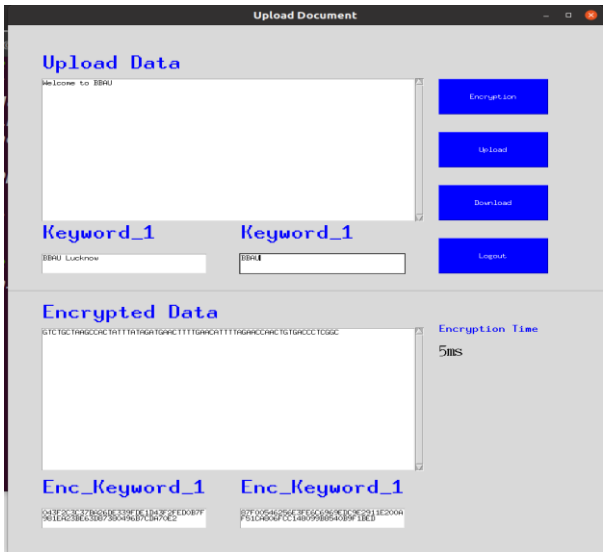

Fig.1. Representation of information to be uploaded over cloud

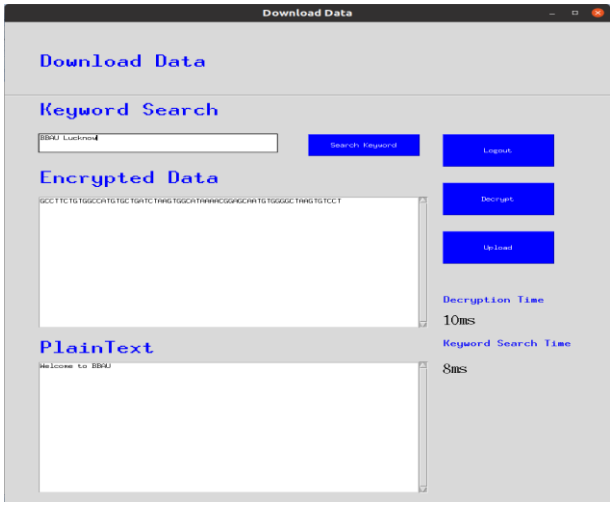

Fig.2. Representation of information to be downloaded from cloud

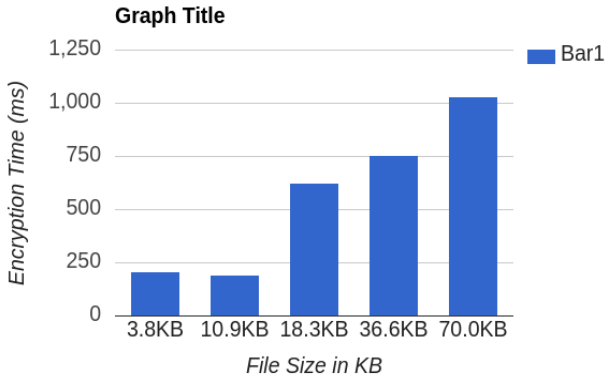

Fig.3. Performance evaluation of DNA-Genetic algorithm

The encryption time of the proposed algorithm is given in fig. 3 and the number of keywords searched time is also given in fig. 4. The table 2 shows the encryption time compare with another existing algorithm [33]. The comparison table shows that, the proposed algorithm takes less time in compared to other existing algorithm. The table 3 shows that the decryption time as compared with other exiting [33] algorithms. The proposed algorithm takes less time as compared to the existing algorithm.

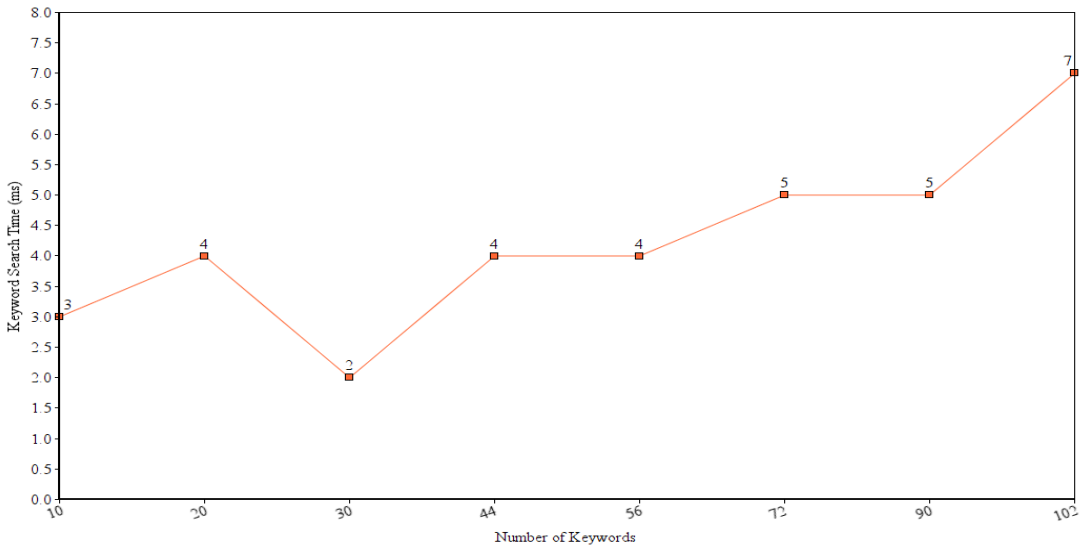

Fig.4. Performance evaluation of fuzzy keyword search

Table 2. Encryption time (millisecond)

Data Size in KB	8 KB	16 KB	32 KB	64 KB
ECC [33]	70	77	79	81
Elgamal [33]	12	14	23	45
DNA-Genetic (Proposed)	6	11	16	22

Table 3. Decryption time (millisecond)

Data Size in KB	8 KB	16 KB	32 KB	64 KB
ECC [33]	24	20	24	25
Elgamal [33]	6	9	11	22
DNA-Genetic (Proposed)	3	6	10	19

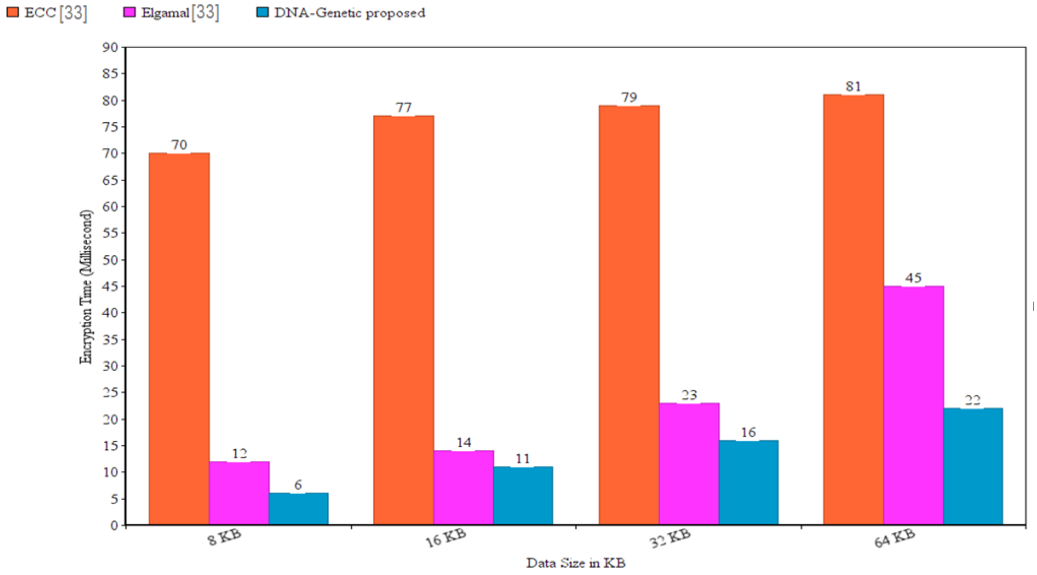

Fig.5. Comparison of computation time of encryption between proposed V/S existing

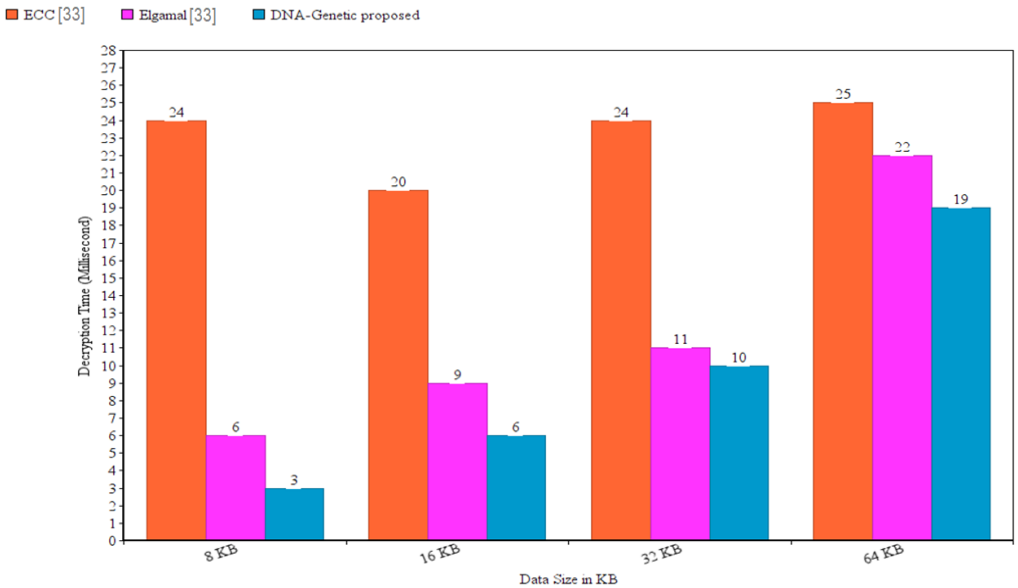

Fig.6. Comparison of decryption computation time of between proposed V/S existing

The estimation of various operations allows for the analysis of computing efficiency. In this study, we compare the durations of encryption and decryption. Encryption and decryption times are counted in milliseconds (ms). Kilobytes (KB) are used to measure data sizes. For the comparison of our proposed system with other cryptographic techniques, we took into account various data sizes. To determine the encryption and decryption times, we take into account several

examples of data sizes of 8 KB, 16 KB, 32 KB, and 64 KB. The fig. 5 shows the comparison graph of existing algorithm [33] with proposed. Fig. 5 shows that the proposed scheme takes less encryption time as compared to Elliptic Curve Cryptography (ECC) and Elgamal cryptosystems. Also, the suggested plan offers increased security and fortifies the system against different threats. Comparing the hybrid method to the ECC and Elgamal cryptosystems, Fig. 6 shows that the hybrid technique decrypts data faster. Comparatively speaking, the suggested scheme's decryption time is less than that of other existing techniques. Consequently, compared to existing cryptographic algorithms, our proposed scheme offers superior security with a hybrid approach. As a result, the proposed hybrid scheme performs better than others in terms of encryption and decryption times.

6. Conclusions

In cloud computing, storing data on the cloud server makes it vulnerable to various security attacks. To ensure security, a hybrid approach is proposed with the help of genetic, SHA-256, and fuzzy keyword search techniques. To provide security to information over the cloud, every provider needs to secure three components of cloud computing, such as authentication, encryption and decryption of data, and secure search, while getting access to a huge amount of data. To make cloud computing more secure, the proposed hybrid approach is a successful and secure model that enables the client to store data in a secure manner, and all three components are covered to make cloud computing trustworthy. To authenticate any user, the fingerprint authentication system is used. After successful authentication of the user, access is granted to the user. For hiding information from unauthorized users as well as data administrators, a genetic algorithm for encryption and decryption of data and a fuzzy keyword search technique are used. The proposed model minimizes computation time, maximizes the security of cloud computing, and makes users trustworthy. This research may encourage researchers to develop new security schemes with effective performance in the context of cloud computing by applying emerging cryptographic techniques.

References

[1] V. Kakkad, M. Patel, and M. Shah, “Biometric authentication and image encryption for image security in cloud framework,” Multiscale and Multidisciplinary Modeling, Experiments and Design, vol. 2, no. 4, pp. 233–248, May 2019, doi: https://doi.org/10.1007/s41939-019-00049-y.

[2] T. Bhattasali, K. Saeed, N. Chaki, and R. Chaki, “A survey of security and privacy issues for biometrics based remote authentication in cloud,” in Computer Information Systems and Industrial Management, Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 112–121.

[3] X. Li et al., “VRFMS: Verifiable Ranked Fuzzy Multi-keyword Search over Encrypted Data,” IEEE Transactions on Services Computing, pp. 1–1, 2022, doi: https://doi.org/10.1109/tsc.2021.3140092.

[4] Y. Zhu, X. Yin, and J. Hu, “FingerGAN: A Constrained Fingerprint Generation Scheme for Latent Fingerprint Enhancement,” IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–14, 2023, doi: https://doi.org/10.1109/tpami.2023.3236876.

[5] H. S. Ali and R. Sridevi, “Credential-Based Authentication Mechanism for IoT Devices in Fog-Cloud Computing,” ICT Analysis and Applications, pp. 307–318, 2022, doi: https://doi.org/10.1007/978-981-16-5655-2_30.

[6] A. Manzoor, M. A. Shah, H. A. Khattak, I. U. Din, and M. K. Khan, “Multi-tier authentication schemes for fog computing: Architecture, security perspective, and challenges,” International Journal of Communication Systems, p. e4033, Jun. 2019, doi: https://doi.org/10.1002/dac.4033.

[7] S. Shi, J. Cui, X.-L. Zhang, Y. Liu, J.-L. Gao, and Y.-J. Wang, “Fingerprint recognition strategies based on a fuzzy commitment for cloud-assisted IoT: A minutiae-based sector coding approach,” IEEE Access, vol. 7, pp. 44803–44812, 2019.

[8] Y. Miao, J. Ma, X. Liu, X. Li, Q. Jiang, and J. Zhang, “Attribute-based keyword search over hierarchical data in cloud computing,” IEEE Trans. Serv. Comput., pp. 1–1, 2017.

[9] X. Ge, J. Yu, C. Hu, H. Zhang, and R. Hao, “Enabling efficient verifiable fuzzy keyword search over encrypted data in cloud computing,” IEEE Access, vol. 6, pp. 45725–45739, 2018.

[10] D. Kumar, M. Kumar, and G. Gupta, “An outsourced decryption ABE model using ECC in Internet of things,” Internat. J. Uncertain. Fuzziness Knowledge-Based Systems, vol. 29, no. 06, pp. 949–964, 2021.

[11] H. Tariq and P. Agarwal, “Secure keyword search using dual encryption in cloud computing,” Int. J. Inf. Technol., vol. 12, no. 4, pp. 1063–1072, 2020.

[12] H. Zhong, Z. Li, J. Cui, Y. Sun, and L. Liu, “Efficient dynamic multi-keyword fuzzy search over encrypted cloud data,” J. Netw. Comput. Appl., vol. 149, no. 102469, p. 102469, 2020.

[13] M. Li, G. Wang, S. Liu, and J. Yu, “Multi-keyword fuzzy search over encrypted cloud storage data,” Procedia Comput. Sci., vol. 187, pp. 365–370, 2021.

[14] C. Ge, W. Susilo, Z. Liu, J. Xia, P. Szalachowski, and F. Liming, “Secure keyword search and data sharing mechanism for cloud computing,” IEEE Trans. Dependable Secure Comput., pp. 1–1, 2020.

[15] S. Namasudra, D. Devi, S. Kadry, R. Sundarasekar, and A. Shanthini, “Towards DNA based data security in the cloud computing environment,” Comput. Commun., vol. 151, pp. 539–547, 2020.

[16] R. M. Jasmine and J. Jasper, “A privacy preserving based multi-biometric system for secure identification in cloud environment,” Neural Process. Lett., vol. 54, no. 1, pp. 303–325, 2022.

[17] Y. Li, Q. Cao, K. Zhang, and F. Ren, “A secure index resisting keyword privacy leakage from access and search patterns in searchable encryption,” J. Syst. Arch., vol. 115, no. 102006, p. 102006, 2021.

[18] Q. Liu, Y. Peng, J. Wu, T. Wang, and G. Wang, “Secure multi-keyword fuzzy searches with enhanced service quality in cloud

computing," IEEE Trans. Netw. Serv. Manag., vol. 18, no. 2, pp. 2046–2062, 2021.
[19] Z. Fu, Y. Wang, X. Sun, and X. Zhang, "Semantic and secure search over encrypted outsourcing cloud based on BERT," Front. Comput. Sci., vol. 16, no. 2, 2022.
[20] M. Joseph and G. Mohan, "A novel algorithm for secured data sharing in cloud using GWOA-DNA cryptography," Int. J. Comput. Netw. Appl., vol. 9, no. 1, p. 114, 2022.
[21] S. Kaur, G. Kaur, and M. Shabaz, "A secure two-factor authentication framework in cloud computing," Secur. Commun. Netw., vol. 2022, pp. 1–9, 2022.
[22] Z. A. Al-Odat, S. U. Khan, and E. Al-Qtiemat, "A modified secure hash design to circumvent collision and length extension attacks," Journal of Information Security and Applications, vol. 71, p. 103376, Dec. 2022, doi: https://doi.org/10.1016/j.jisa.2022.103376.
[23] M. Kammoun, M. Elleuchi, M. Abid, and M. S. BenSaleh, "FPGA-based implementation of the SHA-256 hash algorithm," IEEE Xplore, Jun. 01, 2020. https://ieeexplore.ieee.org/abstract/document/9196134 (accessed Nov. 28, 2022).
[24] S. Banik et al., "WARP: Revisiting GFN Lightweight 128-Bit Block Cipher," Selected Areas in Cryptography, pp. 535–564, 2021, doi: https://doi.org/10.1007/978-303-0-81652for -0_21.
[25] T. Xiao, D. Han, J. He, K.-C. Li, and R. F. de Mello, "Multi-Keyword ranked search based on mapping set matching in cloud ciphertext storage system," Connection Science, pp. 1–18, Apr. 2020, doi: https://doi.org/10.1080/09540091.2020.1753175.
[26] H. R. Bosker, "Using fuzzy string matching for automated assessment of listener transcripts in speech intelligibility studies," Behav. Res. Methods, vol. 53, no. 5, pp. 1945–1953, 2021.
[27] M. Yamazaki, D. Li, T. Isshiki, and H. Kunieda, "SIFT-based algorithm for fingerprint authentication on smartphone," in 2015 6th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), 2015.
[28] W. Bian, P. Gope, Y. Cheng, and Q. Li, "Bio-AKA: An efficient fingerprint based two factor user authentication and key agreement scheme," Future Generation Computer Systems, vol. 109, pp. 45–55, Aug. 2020, doi: https://doi.org/10.1016/j.future.2020.03.034.
[29] S. Kamath K M, R. Srijith, P. Karen, and A. Sos S, "Fingerprint authentication using geometric features," in 2017 IEEE International Symposium on Technologies for Homeland Security (HST), 2017.
[30] M. Tahir, M. Sardaraz, Z. Mehmood, and S. Muhammad, "CryptoGA: a cryptosystem based on genetic algorithm for cloud data security," Cluster Comput., vol. 24, no. 2, pp. 739–752, 2021.
[31] M. Javurek, M. Turcanik, and B. Matej, "Model of Encryption System with Genetic Algorithm," 2019 Communication and Information Technologies (KIT), Oct. 2019, doi: https://doi.org/10.23919/kit.2019.8883476.
[32] J. kumar and V. Saxena, "Hybridization of Cryptography for Security of Cloud Data," International Journal of Future Generation Communication and Networking, vol. 13, no. 4, pp. 4007–4014, Jan. 2020.
[33] S. N. Sarode and G. R. Chillarge, "Efficient and Secure Multi-Keyword Ranked Search and Group Data Sharing for Encrypted Cloud Data," Journal of Scientific Research, vol. 66, no. 02, pp. 68–78, 2022, doi: https://doi.org/10.37398/jsr.2022.660210.

Authors' Profiles

Jaydip Kumar received his bachelor's degree in computer application from Indira Gandhi National Open University, India, in 2013. He received his master's degree in computer application from Indira Gandhi National Open University New Delhi, India, in 2015. He received Ph.D. in Computer Science from Babasaheb Bhimrao Ambedkar University Lucknow, India and published several outstanding research articles. His research interests are cloud computing, cloud data security and cryptography.

Hemant Kumar received MCA degree from Subharti University in 2018 and presently pursuing Ph.D. programme in Computer Science from Babasaheb Bhimrao Ambedkar University, Lucknow. He has published several research articles in the field of Software Testing and Machine Learning. His research interests are Software Engineering, Artificial Intelligence, Machine Learning and Cyber Security.

Karm Veer Singh received his Ph.D. degree from Indian Institute of Technology (BHU), Varanasi, India. He is working as Assistant Professor in Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India. His research interest includes Multimedia Information Retrieval, Quantum Neural Networks, Reliability, AI/ML in Medical Imaging, Pattern Recognition and Mathematical Modeling.

Prof. Vipin Saxena received his Ph.D. degree from Indian Institute of Technology, Roorkee, Uttarakhand, India. Presently, he is working as a Professor in the Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India. He has 30 years of teaching and 33 years of research experience and published more than 250 research articles in International and National Journals and Conferences, authored 05 books in the field of Computer Science and Scientific Computing, attended 62 International and National Conferences, and received three National Awards for meritorious research work in the field of Computer Science.

How to cite this paper: Jaydip Kumar, Hemant Kumar, Karam Veer Singh, Vipin Saxena, "Secure Data Storage and Retrieval over the Encrypted Cloud Computing", International Journal of Computer Network and Information Security(IJCNIS), Vol.16, No.4, pp.52-64, 2024. DOI:10.5815/ijcnis.2024.04.04