A Hybrid Approach of Secret Sharing with Fragmentation and Encryption in Cloud Environment for Securing Outsourced Medical Database: A Revolutionary Approach

Dac-Nhuong Le[1], Bijeta Seth[2] and Surjeet Dalal[2]

[1]*Faculty of Information Technology, Haiphong University, Haiphong, Vietnam*
[2]*Department of Computer Science & Engineering, SRM University, Sonepat, Haryana, India*
E-mail: nhuongld@hus.edu.vn; bijetaoberoi@gmail.com; profsurjeetdalal@gmail.com

Received 17 March 2018; Accepted 05 August 2018;
Publication 12 September 2018

Abstract

Cloud Computing is observed as the greatest paradigm change in Information technology. Data outsourcing is an inventive representation with the intention of trustworthy storage and proficient query execution to customers. Data stored on the cloud is showing great attention. However, the security issues allied with data storage over the cloud is a chief daunting cause for potential adopters. Hence the focus is to find techniques that will offer more security. Many diseases fighting organizations are working together io implement cloud as a data sharing vehicle. It is obligatory to build up innovative solutions with the intention of amalgamate diverse approaches in order to generate flexible and adaptable systems, particularly for achieving elevated levels of utilization of developed algorithms. In this document, we suggest an innovative model based on fragmentation, secret sharing and encryption for medical databases which will divide the data amongst several cloud service providers. We develop a systematic structure exploiting the sensitive nature of information

Journal of Cyber Security and Mobility, Vol. 7_4, 379–408. *River Publishers*
doi: 10.13052/jcsm2245-1439.742

and results in enhanced security level. A database for medical system is represented as Entity association and Relational model. A cloud based model is proposed to offer secure patient centric right to access PHR in a competent way. The simulation results implemented in NetBeans Java for performance evaluation of existing cryptographic techniques are shown. Our security model is evaluated using CrypTool 1.4.30 considering the entropy of algorithms. The future work includes development of a computerized system retrieving, storing and maintaining data efficiently and quickly.

Keywords: Cloud Computing, Healthcare, Database Outsourcing, Secret Sharing, Fragmentation, Encryption.

1 Introduction

We are hasty approaching a new era in which we accumulate our data and perform expensive computation vaguely on remote servers-the CLOUD, in an admired idiom. The following Figure 1 mentions the cloud computing paradigm mentioning its characteristics, deployment models, service offerings and components. While the cloud offers several advantages like pay per usage, cost effectiveness, flexible, distributed computing environment and so on, it raises grave questions of confidentiality, since the data stored over the cloud can be harmed by unauthorized access of users [13, 15, 18, 23, 24].

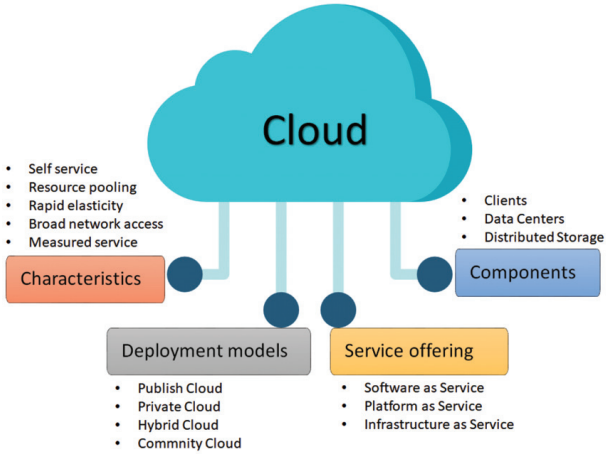

Figure 1 Cloud computing paradigm.

Traditional Healthcare Information systems were based on Client-Server architecture maintained in-house. The high capacity servers are accessed by client computers used by nurses, doctors, patients, and administrators. Existing IT solutions in healthcare suffer major challenges including hardware infrastructure costs, costs involved in networking, staff training, user licenses, security and backup and so forth [26]. Cloud computing is the prevalent impending revolution to the Healthcare organizations for clinically relevant services and enhanced patient outcomes and lessen the infrastructure management burden. Cloud entirely solves such Big-data issues by offering an inter-organizational medical sharing environment facilitating a huge and high capacity data center with numerous computers, storage devices, network devices and power systems [27] to Cloud providers. The health data often contains sensitive information like personal information, health records etc, it is necessary for the users to have assurance of data protection before storing the data onto the cloud. There are abundant applications of cloud based Personal Health Record systems and Bioinformatics research [28]. Despite several pros of Cloud computing, such as security, privacy, regulatory issues and governance make healthcare organizations reluctant to move their systems to cloud environment. This paper aims to provide and overtly diminish the set of latent adversaries for security by providing an architecture utilizing encrypted medical report and adopting secret sharing approach and applying it to Personal Health Records, and allocate the ensuing fragments to diverse independent Cloud providers, and Fragmentation technique is used for query implementation.

The paper proceeds in following way: Foremost in Segment 2, we present an overview of cloud scenario in healthcare: its need, benefits and challenges. Section 3 explains the motivational scenario. Section 4 explains techniques to make data confidential by using fragmentation, secret sharing and encryption methods. Section 5 describes the proposed cloud model and architecture overview. Section 6 has the Implementation, Discussion and Analysis. Finally, our conclusion and future scope are presented in Section 7.

2 Cloud Computing In Healthcare: Need, Benefits and Challenges

The environment in healthcare industry is changing gradually demanding for most effective medical services at low cost increasing competition level between different healthcare providers. Cloud computing aims to decipher clinical troubles faced these days and explain business deeds that have

plagued current healthcare providers. But still it is not widely adopted because healthcare industry still relies on paperwork. Health Information technology needs to be changed because then patients will be most benefited by this technology as they can look for preferred behaviour addressing their status of health and drive down cost and improve efficiencies. Cloud computing has several benefits like:

1) Cloud computing can act as a boon for healthcare actors like doctors, nurses, physicians, patients etc. by providing improved and quick access to data.
2) Cloud Computing can provide dynamic administration of infrastructure rather than time consuming manual entry of data by medical staff.
3) Improved services to patients as services can be given to patients remotely by an automated process where information can be made collected processed and delivered at any time by utilizing Cloud Computing.

2.1 Requirements of Healthcare Industry

Though cloud computing has entered various applications but it is not so much used in practice in healthcare industry where it is still underutilized. Certain requirements [1, 8, 10] that must be fulfilled by healthcare industry also like:

1) The system must be flexible to diverse departmental requirements and organizational sizes.
2) Open access to information and data sources should be encouraged.
3) Capital expense (CAPEX) to operational expense (OPEX) cost must be overseen in case of this migration from client server system to cloud model.
4) Portability is required to easily access remote data.
5) Security and privacy of data need to maintain.

2.2 Challenges in Implementation of Cloud Computing in Healthcare

Some of the significant barriers of cloud computing responsible for its slow adoption in healthcare are:

1) *Privacy and Security challenges*: Personal health records must be safely stored.
2) *Work-flow challenges*: From paper work and data entry system to cloud model, new training and new skill sets need to be planned.

3) *Reliability and performance*: Disaster recovery and performance must be taken into account.
4) *Integration and interoperability*: The Standard Development Organization (SDO) develops specification and principles to sustain healthcare information, exchange of information and systems integration which is difficult to maintain.
5) *Data portability and mobility of records*: The movement of data between healthcare organization and cloud vendors requires no disruption to data.
6) *Speed*: By using cloud computing, faster and accurate access to all information for healthcare service can be made.

2.3 Benefits of Cloud Computing for Healthcare

Cloud computing will be at the forefront of healthcare modernization as it provides several benefits [3–10, 21, 22, 30] like:

1) *Clinical Research*: Travelling long distance is not a problem for patients as experts can access patient information vaguely on request through internet.
2) *Electronic Medical Record* (EMR): Burdensome task is gradually offloading from hospital IT department by putting data online.
3) *Telemedicine*: Teleconsultation and Telesurgeries have increased manifold because of health record exchange, web conferencing and home monitoring through mobile technologies and intelligent medical devices.
4) *Health Information Exchange* (HIE): The delay in Cured by of patients by storing data in records on clouds has reduced significantly.
5) *Big data*: Cost of storing data has reduced as the cloud can store large data sets for EHRs, images related to radiology and genomic records for clinical drug trials.
6) *Type of Cloud*: The type of cloud public, private, hybrid which meets the needs of healthcare industry in a better way is still a question of debate.

Above points mention the advantages of adoption of cloud computing in healthcare industry.

3 Motivational Scenario

Database to be outsourced undergo from two chief challenges:

1) How service providers can guard outsourced databases from unauthorized access? Databases can be made secure by using encryption of data before outsourcing.

2) Protection of outsourced databases from storage service providers which are not fully trusted.

Subsequently, a well-built and protected database outsourcing technique that ensures protected storage and proficient query processing arises. An analytical framework is proposed to be developed using which an organization can store its data on the cloud in a same manner. It should be easy to use and should handle all operations within the trusted organization and send encrypted data to the cloud.

Our major contributions are shown as follows:

1) A novel proposal called **SecSFE** system permitting protected storage, authentication, and auditing of PHR on cloud through the upload and download algorithms is discussed.
2) We utilize the secret sharing approach to split and Blowfish encryption to outsource PHR to data-clouds-multilple independent clouds ensuring confidentiality and integrity of data. Fragmentation through Decision tree helps in efficient query transactions.
3) From the study, it is apparent that the projected idea can endure diverse attacks on the data saved on the cloud.

4 Techniques Used to Formulate Confidential Data by Means of Fragmentation, Secret Sharing and Encryption

4.1 Fragmentation

The segment describes Fragmentation which divides the attributes of relation in the trusted database management system satisfying all confidentiality constraints. Vertical fragmentation was anticipated by Navathe, Ceri, Wiederhold and Doa in 1984 and concentrated on transaction performance enhancement [19]. Ciriani et al. proposed an algorithm satisfying all confidentiality constraint and computed minimal fragmentation. S Sareen et al. in 2016 proposed an algorithm based on decision tree and fragmented attributes based on Decision tree algorithm [13, 14].

Consider a relational database $'D'$ with relation $R_1, R_2, \ldots, R_n$ and set attributes. A list of fragments obtained is denoted as $F = FR_1$, $FR_2, \ldots, FR_m$ satisfying:

- $\forall FR_i \in FR, i \in [1, 2 \ldots m]$, $F_i \leq A_f$ associates a single attribute in A_f with a fragment.
- $\forall a \in A_f$, there exists $FR_i \in FR : a \in FR_i$ guarantees that any attribute in A_f appears certainly in at least one fragment.
- $\forall FR_i, FR_j \in FR, i \neq j : FR_i \cap FR_j = \varphi$ guarantees unlikability between the different fragments.

Take in account a distributed style in which the data owner outsources his relational database $'D'$ having list of relational schemas $R = \{R_1, R_2, R_3, \ldots, R_N\}$ having attributes $a_1, a_2, \ldots, a_n$ on multiple clouds managed by different data storage providers. A set of confidentiality constraints $C_1, C_2, \ldots, C_n$ over R ensuring confidentiality can be one of the followings:

1) *Singleton constraint* (SC): A singleton set having sensitive attribute $a_{j,i}$ of relational schema R_i needs to be secluded with encryption.
2) *Association constraint* (AC): An association between subset of two or more attributes *AC* of relational schema $R_i = \{a_{1,i}, a_{2,i}, \ldots, a_{j,i}\}$ over the relational schema R_i have to be protected.
3) *Intertable constraint* (IC): It is represented as a link of relational schemas $IC = \{R_i, R_j\}$ of the relational database $'D'$ where R_i and R_j are allied with protected primary key or foreign key.

Creating a Decision Tree Algorithm shows in Algorithm 1.

Example illustrating fragmentation: Consider Table 1 as a relation patient with given attributes. The above algorithm can be used to split the Table 1 into three fragments fr_1, fr_2 and fr_3.

Table 2 defines the privacy and alliance constraints where c_0 represents highly sensitive information that must be protected. Constraints from c_1 to c_4 correspond to association constraints and constraints c_5, c_6 represent alliance of attributes that can collectively reveal the patients information.

Table 1 Relation Patient Stored in Plaintext

Id	Title	DateofBirth	Zip	Cured by	Illness
1	Dev	3/7/1981	131001	Medicine	Blood pressure
2	Neena	7/6/1970	131002	Surgery	Cancer
3	Inder	4/3/1967	131003	Medicine	Aids
4	John	12/9/1975	131003	Insulin	Diabetes
5	Meera	4/3/1989	131004	Surgery	Tumour

Algorithm 1 Creating a Decision Tree

BEGIN
Notations:
Make a starting root node array Arr[] storing all attributes excluding sensitive attributes.
Number of branches be denoted by m and initially set to zero.
Attributes and their branch index are stored in a decisional tree variable 2D array BT[][] that stores rows in form of branches and their corresponding child nodes as attributes in each row.
Steps:
For all $a_i \in Arr[]$ do perform following steps
For $m = 0$, construct a latest child node labelled with title as first attribute from array *Arr*[]
$m = m + 1$
Obtain the subsequently available attribute a_i from the array and perform the following steps:
For $i \leq m$ do
Char $= 1$;
If c_i is subset $(BT[][] \cup a_i)$ then
Char $= 0$;
Break;
Endif
Endfor
If *Char* $= 1$ then
Add the attribute a_i as a child node in the existing branch of Decisional hierarchy.
Else
Craft a new branch and insert the attribute a_i in it.
$m = m + 1$;
Endif
Endfor
Endfor
END

Table 2 Sample of Association Constraints and Confidentiality

Association Constraints
C0 = (Id)
C1 = (title, Illness)
C2 = (title, Cured by)
C4 = (title, zip)
C5 = (title, DateofBirth)
C6 = (DateofBirth, zip, Cured by)
C7 = (DateofBirth, zip, Illness)

4.2 Secret Sharing

This section provides an overview of sceret sharing scheme. Since the security of data is fully dependent on the secret keys used, a particular person shouldn't

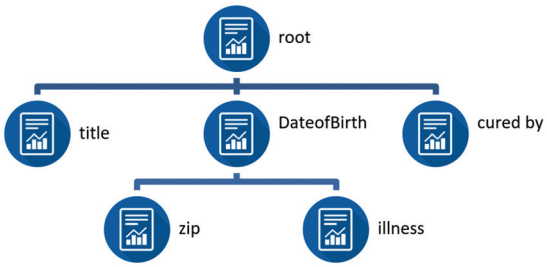

Figure 2 Decision tree based on fragmentation of attributes.

have full charge of the key. This has guided to the design of secret sharing schemes which is a cryptographic technique. Shamir and Blakely developed

Secret sharing schemes with the main aim of secure key management. The basic idea of any secret sharing scheme is to segregate the secret key or confidential information into different pieces and allocate them among different storage providers so that certain subsets of the persons can get together to recover the data with the help of this scheme [15–18].

1) *Important aspects of any secret sharing scheme*:

- *Secret*: any document containing secure keywords or encrypted data which is kept unknown for most of the members of the group.
- *Parties*: the devices used for storing the secret document or the key. or eg: computer,memory sticks etc.
- *Share*: the section of the key allocated to different persons whose combination can retrieve the original document.
- Dealer is the unit responsible for generating scheme parameters, producing the secret, defining initial share and sending them to the participants. The combiner is the entity accountable for pooling shares and reconstructing the shares.

2) *Important Schemes*: Shamir secret sharing [20] is based on secure order preserving technique based on polynomial interpolation and Lagrange's interpolation formula in which a single secret value s_j is shared among n different data centres out of which only k share are required to achieve the original values in order to reconstruct s_j. The security of the scheme lies on the fact that at least k points are required to uniquely reconstruct the polynomial of degree $k - 1$. Inverse of Shamir secret sharing algorithm has to be applied to retrieve back the original result. Any kind of structured query is being supported by Shamir secret sharing like range queries, exact match queries,

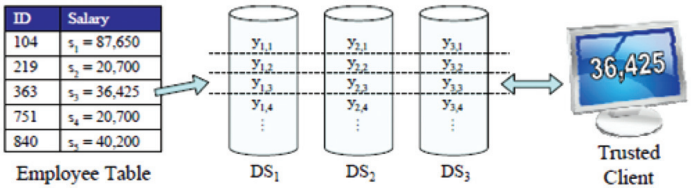

Figure 3 Employee table attributed over salary using Shamir Secret sharing.

aggregate queries etc. Example: having an employee table having id and salary as its attributes with x records and queries attributing over salary attribute. So, we divide salary among n data servers and split them into $s_1, s_2, s_3, \ldots, s_x$. Any trusted client only can retrieve the information as shown in the Figure 3:

The Threshold scheme: works at attribute level (*dividing attribute tuple by tuple*) and uses singleton constraint.

Database service providers $DSP_1, DSP_2, \ldots, DSP_n$ store fragments $fr_1, fr_2, \ldots, fr_n$ dividing our data a_s into n pieces $v_1, v_2, \ldots, v_n$ such that

1) Awareness of any k or more v_i pieces makes a_s effortlessly computable.
2) The awareness of any $k-1$ or fewer a_s pieces leaves a_s wholly uncertain.

This format is called threshold scheme. If $k = n$ then the participants have to construct the secret original datasets.

3) *Computation of shares and reconstruction of shares*: In secret sharing process, secret value is assigned to variable a_0 and random $k-1$ coefficients are chosen as $a_1, a_2, \ldots, a_{k-1}$. The polynomial of degree $k-1$ is generated as

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1} \tag{1}$$

Any secret information can be represented as $X(x_1, x_2, \ldots, x_n)$ as a set of n randomly chosen points and stored on Data storage providers (DSP). The secret value is equivalent to number of fragments and is computed by substituting the values of coefficients $a_1, a_2, \ldots, a_{k-1}$ as $f(x_i)$ and are stored on DSP_i. The variable id of relation patient table is separated into three shares and stored amongst three database service providers. For five tuples, five random polynomials are chosen with degree $2(k-1=3-1=2)$.

$$F(x) = a_2 x^2 + a_1 x + a_0 \quad \text{where } a_0 = a_s. \tag{2}$$

Let us take value of a_1 as $a_1 = (2, 1, 4, 3, 5)$ and $a_2 = (1, 2, 6, 3, 4)$ and $x = 1, 3, 4$ respectively. Thus, the value of polynomial becomes as shown

Table 3 Share Computations Using Polynomial Function

Id	Polynomial $F(x)$	X = 1 (share1)	X = 3 (share2)	X = 4 (share3)
41	$1x^2 + 2x + 1$	4	16	25
52	$2x^2 + 1x + 2$	5	23	38
63	$6x^2 + 4x + 3$	13	69	115
74	$3x^2 + 3x + 4$	10	40	64
85	$4x^2 + 5x + 5$	14	56	89

below for the corresponding values to id $1, 2, 3, 4, 5$ respectively. The values of the shares are computed by substituting the secret points in each polynomial.

$$\begin{cases} F1(x) = 1x^2 + 2x + 1 \\ F2(x) = 2x^2 + 1x + 2 \\ F3(x) = 6x^2 + 4x + 3 \\ F4(x) = 3x^2 + 3x + 4 \\ F5(x) = 4x^2 + 5x + 5 \end{cases} \tag{3}$$

The reconstruction of shares can be done only by trusted DBMS by using any set of k DSPs and retrieving atleast k shares. The coefficients of polynomials can be generated by using Newtons divided difference interpolation and the evaluate $a^s = a_0$ such as

$$\begin{aligned} F(x) \;=\; & yk + \Delta_d yk(x - x_k) + \Delta_d^2 yk(x - x_k)(x - x_{k+1}) \\ & + \cdots + \Delta_d^{n-k} yk(x - x_k)(x - x_{k+1}) \ldots (x - x_{n-1}) \end{aligned} \tag{4}$$

where $\Delta dyk, \Delta d^2 yk, \ldots, \Delta d^{n-k} yk$ are the first, second and k^{th} order divided difference.

In the taken example, to recreate the unique value of *id* say 1, it needs to retain the share from the corresponding values of id from different DSPs using three secret points (1, 4), (3, 16) and (4, 25) to generate value of $f(x) = x^2 + 2x + 1$ as shown in Table 7 and calculating second order polynomial by substituting values in equation from Table 7 as:

$$\begin{cases} F(x) = y_2 + \Delta_d y_2(x - x_2) + \Delta_d^2 y_2(x - x_2)(x - x_3) \\ F(x) = 16 + 9(x - 3) + 1(x - 3)(x - 4) \\ F(x) = x^2 + 2x + 1 \end{cases} \tag{5}$$

Likewise other values of sensitive attributes can be computed using Newton's divided difference tables and polynomials.

Table 4 Fragment 1 Stored on DSP_1

S.no	Id	Title
1	4	David
2	5	Neetu
3	13	Inder
4	10	John
5	14	Meera

Table 5 Fragment 2 Stored on DSP_2

S.no	Id	Dateofbirth	Zip
1	16	03/07/1981	131001
2	23	07/06/1970	131002
3	69	04/03/1967	131003
4	40	12/09/1975	131003
5	56	04/03/1989	131004

Table 6 Fragment 3 Stored on DSP_3

S.no	Id	Cured by	Illness
1	25	Medicine	Blood pressure
2	38	Surgery	Cancer
3	115	Medicine	Aids
4	64	Insulin	Diabetes
5	89	Surgery	Tumor

Table 7 Newtons Divided Difference Table for $id = 1$

id	x_i	y_i	$\Delta_d y_i$	$\Delta_d^2 y_i$
1	1	4	6	1
2	3	16	9	
3	4	25		

4.3 Encryption

Cryptography can be defined as the art of transforming plaintext (*readable text*) into ciphertext (*unreadable text*) which ensures data privacy and non alteration of data. Our paper focuses on providing protection to data from any modification or malfunctioning by untrusted third party. To provide security to data, it must be stored in an encrypted format. Some of the cryptographic approaches are discussed further.

- **Diffie Hellman** was developed in 1776 and considered to be the first public key algorithm. It securely exchanges information between two

parties over an untrusted network. **RSA** was developed by Rivest, Shamir, and Adleman in 1977 and is considered as one of the most popular and secure public key encryption method. Its encryption speed is average and is widely used for key distribution (*generally 1024,248 bits*) and digital signature process. It relies on the perception of "*integer factorization*" in which one way function method is used which is easy to compute one way but is difficult to compute in the reverse process. It suffers from Brute Force attack and Timing attack.

- **Advanced Encryption Standard** (AES) or Joan Daemen and Vincent Rijmen algorithm developed AES in 2001 and works in blocks of 128 bits up to 256 bits. It is based on substitution and permutation process. The key size has no maximum limit. The alike key is intended for encryption and decryption process.
- **Blowfish Algorithm** is considered to be the fastest symmetric block cipher used in place of DES/IDEA. It was developed in 1993 by Bruce Schenier. It is a 16 round Fiestal structure with a changeable key length (32–448 bits). All operations are done as addition on 32 bit words and *XOR*, and block size is of about bits. It is believed to be highly safe and no attack is found successful against it to be successful till date. Blowfish *S* boxes are key dependent.

Following Table 8 provides a comparison of the various algorithms studied.

5 Proposed Cloud Model

Security and privacy violation in clouds has been the primary factor restricting its widespread use. To assure the health record management over access to the Personal Health Record, it is further promising to encrypt the PHR before outsourcing. An automated process of collecting patients vital information and delivering this information to be stored, processed and distributed over cloud is proposed. In this section we are going to analyze our projected system model, which will allow information to be stored on multiple clouds transparently to the user. The foremost intend of our proposed structure is to offer secure patient-centric right to use personal health records and key management in a competent way.

Table 8 Comparison Table for Different Parameters

Algorithm	Year	Key Size	Block Size	Encryption Speed	Level of Security	Attacks
AES	2001	128, 192, 256 bits	128 bits	Faster	Excellent security	Key recovery, side channel attack
RSA	1977	1024, 248	$\leq log2(n)$ In practice, block size is 2^k bits with $2^k < n \leq 2^{k+1}$	Average	Adequate secure	Brute force attack, timing attack
Diffie Hellman	1976	Variable	Variable	Depend on key and block size	Good security	Man in the middle attack
Blowfish	1993	32–448	64 bits	Very fast	Highly secured	No attack found successful

5.1 SecSFE Components

The SecSFE is the proposed Secure Secret, Fragmented, and Encrypted system consisting of 4 steps:

1) Key Generation
2) Encryption/Decryption
3) Secret sharing
4) Fragmentation

Each time a patient visits the Healthcare Unit (HU), the authorized doctor (*client*) performs the signing and encrypting tasks of the PHR and sends it to the database. Next, the encrypted PHR is partitioned through to the secret sharing algorithm, and scatters each share over the Internet to different Cloud service providers. The blend of the three techniques helps to diminish the threat of information seepage in multi-clouds even further, in the visage of inquiring or hacked CSPs.

Furthermore, the availability of data stored in the cloud is increased. The recovery procedure is analogous to the storage course involving authentication of CSP's and reconstruction of encrypted PHRs. SecSFE involves:

1) *Secure Storage*: The subsequent steps are taken to amass the data provided by a data storage provider, namely, a doctor securely using encryption and TA:

- Define access rights for users wrt the files based on Role based access control.
- Data files are encrypted using Blowfish algorithm and secret keys generated.
- Signatures for encrypted PHRs are generated by SHA-1.
- An encrypted PHR and its allied signature are sent to the independent CSPs.
- The generated signature is stored in TPA database.

2) *Secure Verification*: In the proposed SecSFE scheme, if a user desires to access the PHR stored by Data storage provider, subsequent are the steps for uniqueness and access authentication:

- Request is sent by a client to DSP for a PHR.
- The Data storage provider verifies the access privileges of the clients by role based access control policy.
- The client is denied access on non-confirmation of access rights.
- On granting the access rights to the client, the power is conceded to the TPA to review the integrity of information.

3) *Secure Auditing*: In the proposed architecture shown in the Figure 4, if a client clears the access and verification phase, data integrity is verified by the TPA using the subsequent steps:

- The TPA requests the encrypted PHR from the CSP.
- The CSP sends the encrypted PHR to the TPA. The TPA generates the hash code using Bcrypt algorithm.
- The TPA compares the generated signature to the signature stored in its database.
- If the signatures match, the TPA confirms the Data Storage Provider (DSP) for data integrity.
- After verification of data integrity, the DSP requests secret key for decryption process generated from Blowfish algorithm and encrypted PHR from the DSP.
- The DSP sends the decrypted file to the client securely.
- The download operation completes from the role of multicloud proxy where it regenerates the splitted shares of PHR.

5.2 Architecture Overview

In the next section, we will provide an outline of a design in Figure 4 comprising the storage course from a solitary Healthcare unit (HU) including the data flow and security measures. The system will consist of four major components: *Data Storage Provider (DSP), Cloud, Data users, and Trusted Third Party Auditor*.

In the proposed system architecture, the patients are the entities whose personal health record will be uploaded in the Cloud. They are the data owners who are a trusted authority of her information. The information is stored in Personal Health Record. The users are database clients having the query to be executed. The users obtain their secret keys from trusted authority. Trusted Authority (TA) is responsible for user registration and authentication and stores user information on cloud data store. The trusted authority transforms the user queries and generates query implementation plan as a subset of queries and operations like join, delete. The query transformer generates query implementation plan considering the different attributes, relations and data stored in metadata repository. It contains information about data distribution among different fragments. After login authentication when user wants to upload any file he requests to cloud to upload a specified file. The Cloud performs key generation and distributes the keys to the Owners. The datasets are arbitrarily partitioned using horizontal and vertical partitioning. Then the

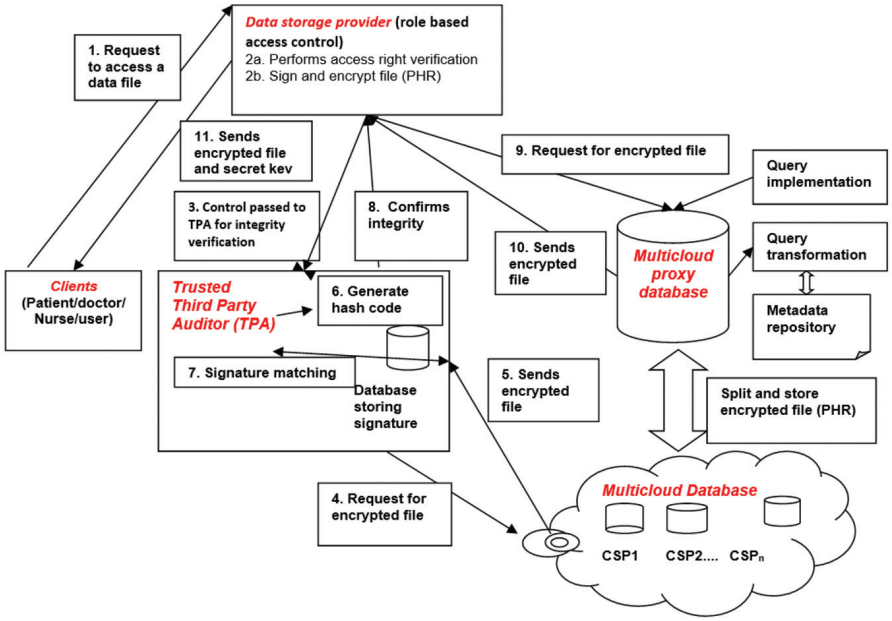

Figure 4 Architecture of proposed model.

partitioned dataset is encrypted using the encryption algorithm. Encrypted file and generated key are stored in the database. A key is sent to the user as an acknowledgment which is further used for downloading a file. When user wants to download his file, again he needs to specify a file name and key which is obtained in response while uploading a file. The cloud again decrypts the desired file with the help of key and sends back a decrypted file i.e. original file. Multiclouds represent different database storage providers in which information fragments are scattered. Database storage providers perform functions like overseeing the components of hardware and software, performance enhancement, backup and recovery, security enrichment. Multiclouds are proposed to be used so as to increase the security level and avoid vendor lock-in problem that exists in single cloud. They store the various database storage providers. Figure 4 shows the overall concept of the framework.

Domains using Personal Health record (PHR) [2] can be categorized as

- *Public domain*: including healthcare domain (*institutions, hospital, doctors, and nurses*) and Insurance domain.
- *Private domain*: including PHR owner, friend etc.

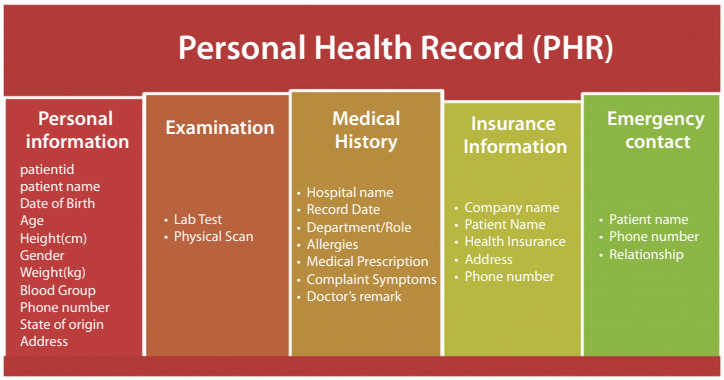

Figure 5 Attributes of PHR.

Personal health record can have following attributes personal health information, medical history, examination, emergency contact and insurance information as shown in Figure 5.

The client can select the operation to be performed as described by following algorithms and Encryption/decryption can be performed using Blowfish algorithm.

6 Implementation, Results and Discussion

Queries requested by users are converted into subqueries and next the best possible query execution plan is made based on the specified conditions.

Example showing precise match query implementation over three fragments: a simple query retrieving id, title, dateofbirth and illness of patients who have met surgery and have zip less than 131003 can be written as:

6.1 Exact Match Query Execution

Query Q: Exact match query execution

```
select tid, title, dateofbirth, illness
from patient
where zip <=131003 and Cured by = 'surgery';
```

Subqueries(Q1):

```
select tid, dateofbirth
from f2
where zip <= 131003;
```

Subqueries(Q2):

```
select tid, dateofbirth, illness
from ResQ1, f3
where resQ1.id = f3.id
and Cured by = 'surgery';
```

Subqueries(Q3):

```
select tid, name, dateofbirth, illness
from ResQ2, f1
where ResQ2.id = f3.id;
```

tid is the tuple identifier. But the final result is obtained from the following subqueries because the id attribute is stored on three fragments in form of three shares. The results obtained from the following subqueries are sent to different DSPs respectively storing different fragments respectively are shown in Table 9 below:

Table 9 Query Results

Tid	Title	Dateofbirth	Illness
2	Neetu	07/06/1970	cancer

6.2 Retrieving Shares from Three Fragments

Query: Retrieving shares from three fragments.

```
Select id
from f1
where ResQ3.tid = f1.tid;
```

Subqueries (Q1):

```
select id
from f2
where ResQ3.tid = f2.tid;
```

Subqueries (Q2):

```
select id
from f3
where ResQ3.tid = f3.tid;
```

Table 10 Newtons Divided Difference Table

Id	x_i	y_i	$\Delta_d y_i$	$\Delta_d^2 y_i$
1	1	5	9	2
2	3	23	15	
3	4	38		

The polynomial value is obtained by following equations:

$$\begin{aligned} F(x) &= y_2 + \Delta_d y_2(x - x_2) + \Delta_d^2 y_2(x - x_2)(x - x_3) \\ &= 23 + 15(x - 3) + 2(x - 3)(x - 4) \\ &= 23 + 15x - 45 + 2(x^2 - 7x + 12) \\ &= 2x^2 + x + 2 \end{aligned} \tag{6}$$

The final results are obtained in Table 11.

Table 11 Final Query Results

Tid	Id	Title	Dateofbirth	Illness
2	2	Neetu	07/06/1970	cancer

6.3 Implementation

The encryption module is implemented using JavaNet-Beans[1] IDE 8.0.2 on Windows 10 Platform. In the program, the secret key algorithms namely RSA, Diffie–Hellman, AES and Blowfish were implemented and their performance was compared by encrypting and decrypting input files in bytes of changeable contents and sizes.

6.4 Discussion and Analysis

In the end, the results were obtained which concluded that Blowfish is considered to be most secure and fastest algorithm. The Table 12 shows the Encryption/Decryption times and key sizes respectively in milliseconds.

1) *Security Analysis*: Fine sources of randomness are crucial in cryptography, and entropy is frequently employed to measure randomness. Low entropy

Table 12 Comparison Table for Different Parameters

Algorithm	Encryption Time (ms)	Decryption Time (ms)	Encryption Key Size	Decryption Key Size	File Size (Bytes)
AES	2525	3362	375	973	5351
RSA	4020	3702	375	32	5351
Diffie Hellman	2137	3362	973	973	5351
Blowfish	1399	2420	40	40	5351

[1]https://netbeans.org/

Algorithm 2 Uploading Operation

Data: A file as a plaintext.
Output: Shares N_i stored in cloud storage S_i.
BEGIN
For (each file) do
1. Generate 64 base encoding;
2. Generate key pairs (*secret key, public key*) for each file;
3. Encrypt and compress each file;
4. Determine number of shares N_i
5. Determine number of thresholds M_i
6. For each file do apply secret sharing scheme to split file into shares N_i where $i = 1, 2, \ldots, N$
   For each shares N_i do
      Hash and sign each share N_i
   End for
   For (all shares N_i and signatures of each share) do
      Upload all shares N_i and signatures into cloud storage S_i;
   End for
End for
End for
END

Algorithm 3 Downloading Operation

Data: N_i shares stored in cloud storage S_i.
Output: Original file decrypted as plaintext.
BEGIN
For all (cloud storage S_i and all shares N_i) do
1. Verify digital signatures of all shares;
2. If (*threshold shares M_i is corrupted*) then
   Redownload that threshold share and signature;
   Apply secret sharing scheme to reconstruct the original file;
   Decrypt and decompress the downloaded file;
End if
End for
END

signifies that the source probably isn't actually random. In particular, if we have a random variable X that takes on values $x_1, \ldots, x_n$ with probabilities $p(x_1), \ldots, p(x_n)$ respectively, then the entropy of X is:

$$H(X) = -\sum_{i=1}^{n} p(x_i) log p(x_i) \tag{7}$$

This value is maximized when all of the probabilities are the same. If we have 2^n different symbols, that maximum value will be n bits of entropy per symbol. That's the hypothetical highest level of entropy that we can achieve.

2) *Numerical Analysis*: The security has been analyzed using CrypTool 1.4.30[2]. The Entropy of the mentioned algorithms is considered as an evaluation parameter. The entropy of a document is an index of its information content and is calculated in bits per character. Higher the entropy value signifies higher the hardness of the key. From the evaluation Table 13, we observed that the entropy of proposed Paillier is higher than other algorithms.

The bar graph for considering parameters among RSA, EIGamal, Paillier and proposed Paillier has been shown in Figure 6 and for entropy in Figure 7.

Table 13 Comparison Table for Entropy

Algorithm	Entropy
RSA	1.5
DES	2.53
Diffie Hellman	3.34
Blowfish	4.49

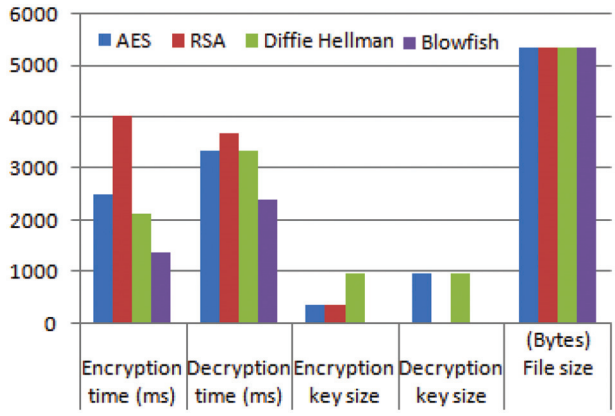

Figure 6 Bar graph for various discussed algorithms.

[2]https://www.cryptool.org/

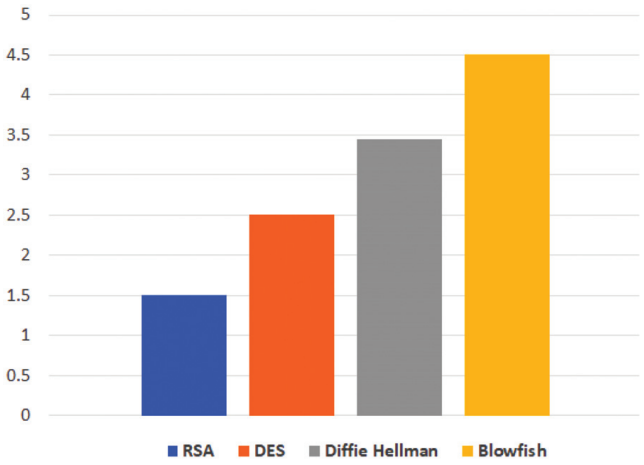

Figure 7 Bar graph for entropy of different algorithms.

7 Conclusion and Future Work

Cloud computing has emerged out as an ideal data sharing medium to share patient data. The concept of fragmentation, secret sharing and encryption aims to safeguard the privacy of outsourced information and users queries. Sensitive attributes can be protected by splitting secret values and distributing them with different fragments on multiple database service providers. Next, encryption enhances the security level further. The projected system is a novel patient-centric framework with a set of mechanisms for information access management to PHRs kept in cloud severs. The proposed system strives to lend trustworthy and scalable data storage and key management at a much reduced cost. The privacy is assured by means of confidentiality constraints recitation the sensitiveness of attributes and their relationships. Our main intend is to construct the system for users in an easy manner and devoid of installing any applications on the customer site. In our prospect work, we plan to implement the proposed architecture to provide complete computerized data storage systems for hospitals so that information be able to be stored, maintained, restructured, retrieved efficiently in addition to securely.

References

[1] Rolim, C. O., Koch, F. L., Westphall, C. B., Werner, J., Fracalossi, A., and Salvador, G. S. (2010). A cloud computing solution for patient's data collection in health care institutions. In *Second International Conference on eHealth, Telemedicine, and Social Medicine, 2010. ETELEMED'10.* (pp. 95–99). IEEE. DOI: 10.1109/eTELEMED 2010.19

[2] Li, M., Yu, S., Zheng, Y., Ren, K., and Lou, W. (2013). Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. *IEEE Transactions on Parallel and Distributed Systems,* 24(1), 131–143.

[3] Nkosi, M. T., and Mekuria, F. (2010). Cloud computing for enhanced mobile health applications. In *IEEE Second International Conference on Cloud Computing Technology and Science (CloudCom),* 629–633. IEEE.

[4] Ikuomola, A. J., and Arowolo, O. O. (2014). Securing patient privacy in e-health cloud using homomorphic encryption and access control. *International Journal of Computer Networks and Communications Security,* 2(1), 15–21.

[5] Soubhagya, B. (2013). A homomorphic encryption technique for scalable and secure sharing of personal health record in cloud computing. *International Journal of Computer Applications*, 67(11).

[6] Deng, M., Petkovic, M., Nalin, M., and Baroni, I. (2011). A Home Healthcare System in the Cloud–Addressing Security and Privacy Challenges. In *IEEE International Conference on Cloud Computing (CLOUD)*, (pp. 549–556). IEEE. DOI: 0.1109/CLOUD.2011.108

[7] Sundararaman, K., Parthasarathi, J., Rao, S. V., and Rao, G. A. (2008). Hridaya A tele-medicine initiative for cardiovascular disease through convergence of grid, Web 2.0 and SaaS. In *Second International Conference on Pervasive Computing Technologies for Healthcare, 2008. PervasiveHealth* (pp. 15–18). IEEE. DOI: 10.1109/PCTHEALTH.2008.471015

[8] Zafar, Z., Islam, S., Aslam, M. S., and Sohaib, M. (2014). Cloud computing services for the healthcare industry. *Int. J. Multidiscip. Sci. Eng.,* 5, 25–29.

[9] Benaloh, J., Chase, M., Horvitz, E., and Lauter, K. (2009). Patient controlled encryption: ensuring privacy of electronic medical records. In *Proceedings of the 2009 ACM workshop on Cloud computing security* (pp. 103–114). ACM.

[10] Reddy, G. N., and Reddy, G. J. (2014). Study of Cloud Computing in HealthCare Industry. *arXiv preprint arXiv:1402.1841*.

[11] Huda, M. N., Sonehara, N., and Yamada, S. (2009). A privacy management architecture for patient-controlled personal health record system. *J. Engineering Science and Technology,* 4(2), 154–170.

[12] Li, M., Yu, S., Ren, K., and Lou, W. (2010). Securing personal health records in cloud computing: Patient-centric and fine-grained data access control in multi-owner settings. In *International conference on security and privacy in communication systems* (pp. 89–106). Springer, Berlin, Heidelberg.

[13] Sareen, S., Sood, S. K., and Gupta, S. K. (2016). Towards the design of a secure data outsourcing using fragmentation and secret sharing scheme. *Information Security Journal: A Global Perspective,* 25(1–3), 39–53.

[14] Bkakria, A., Cuppens, F., Cuppens-Boulahia, N., Fernandez, J. M., and Gross-Amblard, D. (2013). Preserving Multi-relational Outsourced Databases Confidentiality using Fragmentation and Encryption. *JoWUA*, 4(2), 39–62.

[15] Pundkar, S. N., and Shekokar, N. (2016). Cloud computing security in multi-clouds using Shamir's secret sharing scheme. In *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT),* (pp. 392–395). IEEE.

[16] Ren, K., Wang, C., and Wang, Q. (2012). Security challenges for the public cloud. *IEEE Internet Computing*, 16(1), 69–73. DOI: 1089-7801/12/$31.00 IEEE Computer Society.

[17] Arun, V., Padma, S. K., and Shyam, V. (2015). Mobile admittance of Health Information with privacy and analysis in Telemedicine. In *International Conference on Trends in Automation, Communications and Computing Technology (I-TACT-15)*, (pp. 1–6). IEEE. DOI: 978-1-4673-6667-0/15/$31.00

[18] Hossain, M. A., Hossain, M. B., Uddin, M. S., and Imtiaz, S. M. (2016). Performance Analysis of Different Cryptography Algorithms. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(3).

[19] Navathe, S., Ceri, S., Wiederhold, G., and Dou, J. (1984). Vertical partitioning algorithms for database design. *ACM Transactions on Database Systems (TODS),* 9(4), 680–710. DOI: 10.1145/1994.2209

[20] Shamir, A. (1979). How to share a secret. *Communications of the ACM*, 22(11), 612–613. DOI: 1145/359168.359176

[21] Huang, J., Sharaf, M., and Huang, C. T. (2012). A hierarchical framework for secure and scalable ehr sharing and access control in multi-cloud. In *41st International Conference on Parallel Processing Workshops (ICPPW),* (pp. 279–287). IEEE. DOI: 10.1109/ICPPW.2012.42

[22] Liu, C. H., Chen, T. L., Lin, H. Y., Lin, F. Q., Liu, C. M., Wu, E. P., and Chen, T. S. (2013). Secure PHR Access Control Scheme in Cloud Computing. *International Journal of Information and Electronics Engineering*, 3(3), 329. DOI: 10.7763/IJIEE.2013.V3.328

[23] Jain, A., and Soni, B. K. (2017). Secure Modern Healthcare System Based on Internet of Things and Secret Sharing of IoT Healthcare Data. *International Journal of Advanced Networking and Applications*, 8(6), 3283.

[24] Gupta, P., Koushal, V., Narayan, C., and Anand, A. (2017). Building Genetic Database at Medical Institutes: Implement Patient Cost Audit and Improve Biomedical Research. Annals of neurosciences, 24(1), 3–4.

[25] Van, V. N., Long, N. Q., and Le, D. N. (2016). Performance analysis of network virtualization in cloud computing infrastructures on openstack. In *Innovations in Computer Science and Engineering* (pp. 95–103). Springer, Singapore.

[26] Bamiah, M., Brohi, S., and Chuprat, S. (2012). A study on significance of adopting cloud computing paradigm in healthcare sector. In *International Conference on Cloud Computing Technologies, Applications and Management (ICCCTAM)*, (pp. 65–68). IEEE.

[27] Mastelic, T., Oleksiak, A., Claussen, H., Brandic, I., Pierson, J. M., and Vasilakos, A. V. (2015). Cloud computing: Survey on energy efficiency. *Acm computing surveys (csur),* 47(2), 33.

[28] Sandhu, R., Gill, H. K., and Sood, S. K. (2016). Smart monitoring and controlling of Pandemic Influenza A (H1N1) using Social Network Analysis and cloud computing. *Journal of Computational Science,* 12, 11–22.

[29] Van, V. N., Long, N. Q., Nguyen, G. N., and Le, D. N. (2016). A performance analysis of openstack open-source solution for iaas cloud computing. In *Proceedings of the Second International Conference on Computer and Communication Technologies* (pp. 141–150). Springer, New Delhi.

[30] Le, D. N., Kumar, R., Nguyen, G. N., and Chatterjee, J. M. (2018). *Cloud Computing and Virtualization*. John Wiley & Sons.

Biographies

Dac-Nhuong Le has a M.Sc. and Ph.D. in computer science from Vietnam National University, Vietnam in 2009 and 2015, respectively. He is Deputy-Head of Faculty of Information Technology, Haiphong University, Vietnam.

He has a total academic teaching experience of 12 years with more than 50 publications in reputed international conferences, journals and online book chapter contributions (Indexed By: SCI, SCIE, SSCI, ESCI, Scopus, ACM, DBLP). His areas of research include: evaluation computing and approximate algorithms, network communication, security and vulnerability, network performance analysis and simulation, cloud computing, IoT and image processing in biomedical.

His core work in network security, soft computing and IoT and image processing in biomedical. Recently, he has been the technique program committee, the technique reviews, the track chair for international conferences: FICTA 2014, CSI 2014, IC4SD 2015, ICICT 2015, INDIA 2015, IC3T 2015, INDIA 2016, FICTA 2016, ICDECT 2016, IUKM 2016, INDIA 2017, CISC 2017, FICTA 2018, ICICC 2018, CITAM 2018 under Springer-ASIC/LNAI Series. Presently, he is serving in the editorial board of international journals and he authored 8 computer science books by Springer, Wiley, Chapman and Hall/CRC Press, Lambert Publication, Scholar Press, and VSRD Academic Publishing.

Bijeta Seth is pursuing a Ph.D. from the Department of Computer Science & Engineering at SRM University, received her M.Tech Degree from the Department of Computer Science & Engineering at Seth Jai Prakash Mukand Lal Institute of Engineering & Technology, Radaur in 2011. She has completed her B.Tech from Department of Computer Science and Engineering at Haryana Engineering College, Jagadhri in 2005. Her research interests are in the areas of Cloud Computing and Artificial Intelligence and she has published many papers in the national and international journals and confereces.

Surjeet Dalal received his Ph.D. Degree in 2014 from Suresh Gyan Vihar University Jaipur (Rajasthan) and M.Tech Degree in 2010 from PDM College of Engineering, Bahadurgarh Haryana. He has completed B.Tech (Computer Science & Engineering) from Jind institute of Engineering & Technology Jind (Haryana) in 2005. He has more than nine years of teaching experience in various colleges under Kurukshetra University Kurukshetra. His current research area is Artificial Intelligence, Multi-agent system, Case-based reasoning and Cloud Computing. He has presented more than twenty papers in the national/international conferences. He has published more than thirty papers in the national and international journals. He has guided many M.Tech students for their thesis work under Kurukshetra University, Kurukshetra.

He is the reviewer of many national/international journal of repute in India and Abroad. He is the professional member of various professional and research committees. He is the professional member of CSI India, IEEE New York, IETE Chandigarh and ISTE-AICTE New Delhi.