Risk based Hybrid Security Model for enhancing eHealth services of Cloud System

by

Azmat Ullah
18201207
Md. Tasnimul Hasan
18201058
Imam Hossain
18301276
Tonmoy Das
18301138

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
September 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Azmat Ullah 18201207	Md. Tasnimul Hasan 18201058
Imam Hossain 18301276	Tonmoy Das 18301138

Approval

The thesis/project titled "Risk based Hybrid Security Model for enhancing eHealth services of Cloud System" submitted by

1. Azmat Ullah (18201207)
2. Md. Tasnimul Hasan (18201058)
3. Imam Hossain (18301276)
4. Tonmoy Das (18301138)

Of Summer, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 28, 2022.

Examining Committee:

Supervisor:
(Member)

Dr. Muhammad Iqbal Hossain
Associate Professor
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD
Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

E-health is a relatively new healthcare innovation that incorporates electronic operations and communication. The Electronic Health Record (EHR) or Electronic Medical Record (EMR) in an e-health system contains all health data information, such as demographics, prescriptions, medical histories, laboratory reports, photographs, billing information, and any other sensitive patient information. Any unauthorized access has the potential to be devastating. In order to establish a highly secured model for EHR, our paper emphasizes the problem of the study and directions in cyber security. Cloud computing, on the other hand, provides excellent services to both patients and healthcare providers in terms of cost-effective data storage, processing, and updating, as well as better quality and enhanced efficiency. We aim to ensure that security and privacy are essential when accessing or sharing patient data among several stakeholders. In addition, we want to postulate an effective robust security method for EHR which will be a combination of data encryption and risk based access control which will give access to data by calculating and comparing risk scores of different parameters associated with the user's past activity. Also, explore methods to preserve the accuracy and secrecy of patients' data, as confidentiality and anonymity are considered to be important components when sharing or trying to access patient data between multiple stakeholders.

Keywords: eHealth; Encryption; Risk-based Access Control; Security;

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis has been completed without any major interruption.

Secondly, to our advisor Dr. Muhammad Iqbal Hossain Sir for his kind support and advice in our work. He helped us whenever we needed help.

And finally to our parents without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ABE Attribute-Based Encryption

$CP-ABE$ Ciphertext-Policy Attribute-Based Encryption

CSP Cloud Service Provider

EHR Electronic Health Records

EMR Electronic Medical Records

$KP-ABE$ Key-Policy Attribute-Based Encryption

OTP One-Time Password

PHR Personal Health Record

SSE Searchable Symmetric Encryption

Chapter 1

Introduction

1.1 An Introduction to E-Health

E-Health is a growing field at the intersection of health informatics, global health, and business that refers to healthcare services and information supplied or augmented via the internet and related technologies. In a broader sense, the phrase describes a state of thinking, a mindset, an approach, and a commitment to use information and communication technologies to improve health care at local, global, and regional levels. This new style of the invention has the potential to revolutionize health care by enhancing efficiency, widening and broadening accessibility, encouraging and connecting professionals and their patients, and empowering, devolving, and even partly deconstructing medicine in the procedure. The application of eHealth and smart health-care planning in emerging and developing countries has the ability to provide access to necessary remedies and preventive care, which can serve as the basis for significant economic growth. In industrialized countries, the use of eHealth has the potential to restructure the healthcare delivery business model while also increasing and enhancing the customized standard patient healthcare provided.

EHR (Electronic Health Record) is a systematized aggregate of a patient's electronic health information in an e-health system. These records involve all sensitive patient health information. Users typically have little to no knowledge of how their private data is handled. The existence of healthcare data in the cloud has piqued the interest of hackers, who attack systems in the hopes of obtaining confidential information and profiting. Each year, a large number of data breaches occur. In this regard, information systems' safety and confidentiality should be improved in order to make them more robust and of higher quality.

The provision of EHRs is a significant application of healthcare information technology (HIT) . The concept of EHRs is not new. EHRs have been around in some form or another for over five decades. Recent rapid improvements in technology, on the other hand, have information technology, particularly the creation and implementation of the broad use of mobile electronic devices such smartphones, mobile gadgets, and personal computers that are web-enabled PDAs (Personal Digital Assistants); have altered the ways in which records can be accessed and processed.

As a whole, eHealth standards promote clinical information sharing by enabling

syntactical compatibility between medical systems. They were not, however, created with the goal of safeguarding the patient's privacy. Furthermore, studies have concentrated their efforts on ensuring the security and privacy of data in motion, resulting in a multitude of solutions that address the safety of protocols for communication and communication routes. Protecting the security of data, on the other hand, is still in its early phases. Both monitoring the secure transmission and preservation of patient data (Safety) and ensuring access to user information (Privacy) are extremely difficult tasks.

Risk analysis in general is defined as a process of calculating the outcome that can cause any issue on the integrity of any organization. It is so significant for any organization to protect their data from any third party even erroneously. Now measuring risk in access control with the help of anomalies of the past behavior of the user activity and data sensitivity level dynamically the lessen the vulnerability of data breaching and misuse. Quantitative risk analysis includes considering different factors associated with the user, login session and the data. These parameters are flexible and upto organization's priority which can be added or removed according to requirement.

1.2 Research Problem

An EHR system's common problem includes the assurance of three things namely security, privacy and confidentiality [6]. If someone does not have authorization yet they are sharing sensitive details of health-related data then this can cause a data breach. In many circumstances, unavoidable systemic identification can lead to violation of privacy that occurs throughout the infrastructure of electronic health, as well as by the central technologies and authorities that monitor the behavior of healthcare staff and patients. The abuse of health record access can be done by healthcare providers either intentionally or accidentally.

As discussed in [24], concerns regarding health information privacy affected the willingness of patients to all health care professionals to share their medical data by means of cloud computing technology as per the study by Ermakova et al. People expect complete trust from their healthcare providers when it comes to privacy and data security. Information security, a branch of computer science, is concerned with safeguarding data from services supplied by a system from threats that try to compromise its confidentiality, integrity, and availability.

Cloud computing is a very popular way to save both time and money in many industries, including healthcare. Despite the advantages of eHealth clouds, there are still unresolved security and privacy issues that will require extensive research to fix [15]. As society is progressing, most things occur via clouds. Medical records hold sensitive and vital information regarding a patient's case. As a result, while storing and sharing this type of data, a high level of security is required. Furthermore, when communicating medical data on or over a network, the system must be protected from any assault that may occur.

A patient would not want their data out to the public and keeping this in mind, we

are trying to figure out if using a hybrid model for data security will be an efficient way to leap ahead. With this vision, our research is being conducted. Over the years, several kinds of research have been done to excel the security of eHealth data on the cloud but the studies have not been too consistent about the protection of the CIA triad over the cloud in the eHealth sector.

The question to answer in our research is:

How successful is the combination of attribute-based encryption which is searchable and access control based on risk and sensitivity calculation?

1.3 Research Objectives

The aim of our research is to develop an enhanced and elevated cloud system that will facilitate the sharing and integration of electronic health records by ensuring the integrity, confidentiality, authenticity, accountability, audit, nonrepudiation, and anonymity which are the security requirements in order to preserve medical data in a cloud environment [22]. We are planning to fulfill our goal by the following objectives putting in front:

1. To deeply understand encryption techniques and do risk analysis.
2. To deeply study the existing methods and techniques.
3. To identify the problems and drawbacks of existing methods and techniques.
4. To find out the solutions to these problems from literature, if any.
5. To propose and evaluate a model.
6. To offer recommendations on improving the model.

1.4 Research Structure

The entire paper is divided into six chapters. The first chapter give an introduction about the topic followed by research problem and research objective. In the second chapter, the related works are shown. The third chapter consists of the proposed model focusing on the encryption of data in the cloud followed by the risk factors. Chapter four contains an implemented idea with proper explanation. On top of that, chapter five contains the results of the findings. Lastly, chapter six draws to the conclusion and future aspects of this work.

Chapter 2

Literature Review

Cloud-based electronic health documents offer several benefits like the collection of patient information, the maintenance of their health records, and the organization of those records in order to improve effective communication between patients and care providers along with providing the facility to handle threats and attacks. However, the security and privacy of the data are the main concerns of our system.

There has been a significant amount of work done in the field of eHealth security and privacy threats. We studied many of these recent works that helped us propose our secure eHealth system architecture in this area.

2.1 Related Works

According to the paper [25], the state of the art in current eHealth cloud security and privacy research from five primary standpoints: data confidentiality, security control, and protocols, appropriate encryption, the requirements of security, and the recovery plans from disaster. The paper provides stakeholders with a clear overview of current security and privacy developments in eHealth, resulting in better knowledge, better designs, and making better decisions. For Examples- An Identity Management System (IMS), biometric-based IMS, Consolidated Identity Management (CIDM) system, Information Security Management System (ISMS), etc.

An attribute-based ciphertext policy encryption scheme is proposed in the research paper [10] which constitutes five algorithms- Initialize, Key-Gen, Encrypt, Decrypt, and Revoke. It is claimed that the ABE guarantees confidentiality, integrity, availability of data, and authentication. As a result, only legitimate authority can access EHR and change or delete EHR from the database in the cloud which aids the medical research field and ensures the integrity of the EHR.

After observing the failures of recent works in security requirements of anonymity, unlinkability, and vulnerability to impersonation attacks, the authors of the paper [19] proposed a secure EHR authorization system using elliptic curve encryption and public-key encryption with distributed EMR storage and sharing scheme. This system consists of four parties which are the doctor, the patient, the hospital's private cloud, and the public cloud. These four parties are connected by registering to the public cloud.

The public cloud has the key generation server which calculates the secret keys for every party by n elliptic curve followed by issuing the keys. In the case of visiting a doctor, the patient has to be authenticated by the doctor before sending the biomedical data. After receiving the index from the patient, the doctor then accesses the private cloud and attains the EMR. Following that, the doctor diagnoses the patient according to the patient's current condition and EMR. At last, the doctor sends this diagnosis report to the public cloud with encryption and signature mechanism.

Though this system is secured against impersonation attacks and fulfills the security requirements of patient anonymity and patient unlinkability, it has a computational cost 12TSign + 14TA + 6TS + 4TH which is higher compared to other schemes since it uses symmetric and asymmetric encryption for enhancing the security further.

This paper [5] discusses the implementation of a cloud application based on HL7. The application is then secured using risk-aware task-based access control. This security approach is safer than context-aware access controls in terms of preventing unauthorized access. Although, It has a delay of one second. This is implemented on AWS and is SOAP-based in nature. The evaluation of the delay time of the prototype implementation can be found. In the prototype java, aws with 64-bit Linux, 1 virtual core, and 1.7GB of ram were used. Task-based AIC is taking more time than other risk-based approaches as it is more efficient.

This paper [21] proposes an access control model of electronic health records in the cloud using Ciphertext-Policy Attribute-Based Encryption with the mention of its misuses. The CP-ABE method can give protection from attackers who are not verified in the organization. The verification will be done by the authority with the user details as attributes. Every verified user will be given a decryption key corresponding to that unique attribute. However, the threat remains from the users within. An existing user can share their decryption key with others that was provided by the organization. To solve this misuse, traceability, remote data auditing, and revocability were introduced. Firstly the data owner of the cloud generates permission keys for the verified users. When a user tries to do any action he/she has to enter the decryption key. Depending on whether it detects anomalies it will block the user or give access to data.

The paper [14] discusses Searchable Encryption (SE) and the cases where it can be used. As verified users need to query the encrypted data and need to keep the searched keywords data concealed from the cloud service providers SE is still needed. Searchable encryption works in a way where the data in the cloud will be encrypted and users can search over the encrypted data. There are four algorithmic phases in the SE which are setup, encryption, token generation, and query. Depending on who can do search actions there are two types of SE. Users who possess secret keys are able to make search tokens and this is part of the Searchable symmetric Encryption (SSE) scheme. Another scheme Searchable Asymmetric encryption (SAE) lets anyone who possesses a decryption public key to encrypt data however the search is limited to those who have the private key. The limitation of the SE scheme is that

as the security becomes enhanced the efficiency decreases and is not pragmatic for enlarged databases.

A non-cryptographic secure model is proposed to protect the privacy of EHR data in cloud systems [23]. The algorithm this approach uses hides patients' data by anonymization by means of signal processing. Although the implemented model uses only electrocardiograph (ECG) data, it can be utilized in any type of EHR data in the cloud. The author states the main reasons why the model does not use available cryptographic approaches such as ABE (attribute-based encryption), IBE (identity-based encryption), SKE (symmetric key encryption) due to being complex in the computation of those methods. Also, PKE (public key encryption) and blockchain are not used due to being slow-paced in operation and not having well-defined standards respectively. The algorithm uses two steps. Firstly it anonymizes data by algorithms of signal processing. Here, anonymization refers to removing all the personal data and making it obscure which can not be recognized without reconstruction. FFT (Fast Fourier Transform) is used for both anonymization and reconstruction of the data which outperforms other existing signal processing algorithms which are wavelet packet-based.

According to [9], using a newly constructed policy server at EUT and available for academic communities and enterprises creating eHealth applications, demonstrated a technique of demonstrating a requester's authority to access private eHealth data. This innovative technique, which allows eHealth individuals and organizations to manage access to clinical data of patients based on enforced privacy constraints, opens up a new window for the low-cost implementation of presumed "digital medical care" on a broad scale. The document also includes recommendations for what to do if the authorization procedures for personal data access do not work.

Two SEDSSE (Secure and Efficient Dynamic Searchable Symmetric Encryption) algorithms for medical cloud data have been proposed in this paper [13]. Firstly, they offer SEDSSE I, a Secure and Efficient Dynamic Searchable Symmetric Encryption method that combines the k-Nearest Neighbor (kNN) and Attribute-Based Encryption (ABE) techniques. Between the cloud server and search users, the suggested technique can accomplish forward privacy, backward privacy, and collusion resistance. Secondly, based on that scheme, they offer an upgraded technique called SEDSSE II to achieve the key non-sharing that hampers kNN-based searchable encryption approaches. Their suggested techniques have lower storage costs, lower search complexity, and updating complexity than existing DSSE schemes. Extensive testing shows that the suggested system is efficient in terms of storage overhead, index building, trapdoor generation, and query generation.

The framework that was proposed by Nathalie Baracaldo et al was the expansion of role-based access control (RBAC) which also includes the calculation of risk and trust in system users [3]. The proposed model [4] of Khalid et al. suggests enabling or disabling user roles based on session threshold. In normal risk-based, past behavior of the user is taken into account to measure the risk. Also defines normal behavior. The sensitivity of the data that a user wants to access is also considered for access. An action is only allowed when the benefit of the system is greater

than the risk. Now, in this Task-based AIC, the access control is on two factors. Firstly, the user is verified with the credential then the user is given permission to go further. Secondly, with the past history of the user, and sensitivity of data a score is calculated for risk. The score is then calculated with a predetermined risk threshold. If the score is greater then the user is now allowed to do anything further.

Quantified Risk adaptive access control (QRAAC) considers two parameters for granting access to data. One is doctors usage of data another one is the purpose of the data usage within a specific timeframe. The risk is then measured by utilization of information theory techniques for quantifying the uncertainty [16].

Risk based decision function gives an adaptability in access control using an appraisal factor for a user based on previously accessed record and sensitivity risk related to the data. Adjustment of user risks are done by the utilization of exponentially weighted moving average to recent behavior [11].

For dynamicity of calculating risk scores a method was developed which computes risk based on different variables and that computed score is forwarded into a measuring module which will decide actions to take based on the genetic algorithm [26].

User trust based on past activity was introduced in the sparse zone-based policy in risk calculation. The past activity parameter then further expanded with the role of users by the utilization of trust vector and cartesian product of different parameters like past login activity, user details, one user's appraisal to another [1]. Data document analysis with past behavior was proposed to add additional complexity of data in order to terminate inside vulnerabilities [12].

Chapter 3

Proposed Model

3.1 Approach of Proposed Model

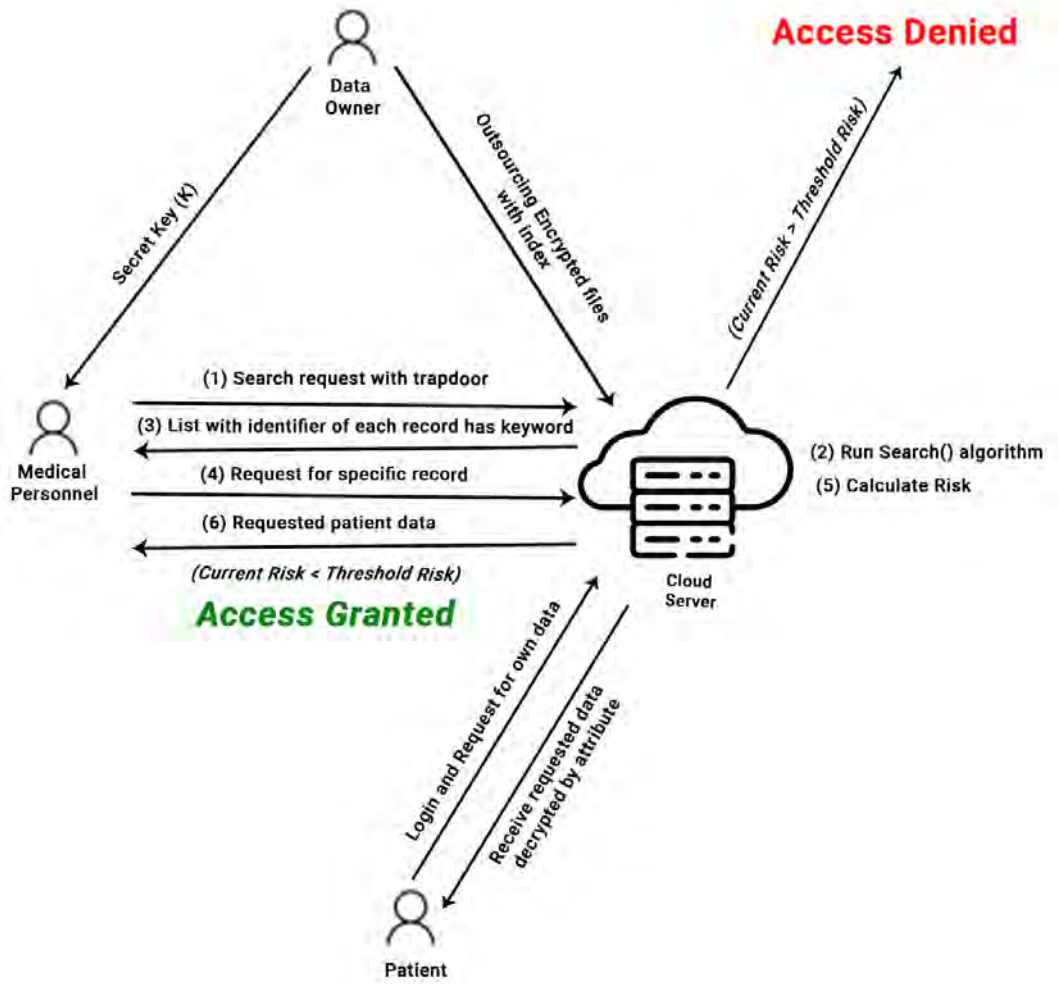

Figure 3.1: Our Proposed Model

Figure 3.1 demonstrates the overall idea where we propose a model for the security of electronic health records which will use attribute-based encryption while storing the data on the cloud storage with searchable functionality. Now the most crucial part of the security model is access control. Access control will be in two factored

ways starting with verification of the credentials. After getting verified, there will be risk analysis based on 9 parameters. The parameters are years of experience, designation, failed login ratio, referral index, working location index, working time index, appraisal factor, data sensitivity and probationary period. There will be two types of risk calculation while making a decision of granting a session: current risk and threshold risk as described in paper [5]. If current risk is less than threshold risk then data access session will be granted. Otherwise the session will be rejected.

This risk checking session will be only from the medical staff side as they have access to sensitive data. Patient side access control will consist of two factors of credential verification and further verification with OTP. Lastly, the patient will get their decryption by their attribute.

3.1.1 Primary Encryption

Primary Encryption

Third-party storage usage exposes data to a variety of privacy and security issues. To ease up, two sections of ABE are used which are Ciphertext policy attribute-based encryption (CP-ABE) and Key-Policy Attribute-Based Encryption (KP-ABE).

CP-ABE, which allows fine-grained access control, has been widely explored for the secure exchange of health data in cloud-based eHealth systems. It provides a viable solution in the cloud environment to privacy and security challenges. Potential aspects that CP-ABE can tackle are expressiveness, efficiency, user collusion resistance, and attribute/user revocation [7].

KP-ABE group of users consists of both patients and healthcare practitioners, as well as any entity with the right to keep data in the cloud. Someone can begin uploading files to the CSP after registering. As a result, we suppose that the user key can be saved on CSP's remote storage. The files must be transmitted and kept in encrypted form in a secure and private manner for prevention from both internal and external attacks. To accomplish this, the user contacts administration, who issues an ABE key based on a policy. Upon receiving their ABE secret key, the user can begin encrypting files. As a result, the user key gets encrypted using a set of properties that they set. Only users with keys that satisfy these criteria will be able to decrypt the data, hence the attributes can be thought of as access permissions [18].

The created ciphertext is transmitted to the CSP, who is unable to decrypt it because it lacks access to a valid private key, ensuring that the file's content stays private even if the CSP is acting maliciously.

Searchable Encryption

It is possible to look through ciphertext using searchable encryption without having to first decrypt the data. When a user wants to search with a keyword, he/she must first establish a trapdoor. The server will then get this trapdoor and search the keyword index using it. Finally, the server will send a list of documents that

contain the keyword the user looked for. By doing this, it ensures the further security of the user's private information and reduces the overhead associated with the transmission.

3.1.2 Risk Analysis

In the access control section, we consider risk analysis of authentication when anyone from administration attempts to access. We provide patients with access to their data through two-factor authentication, which includes logging in with a username and password.The verification process is then completed by sending an OTP via SMS. However, on the admin side, we will not only check the username and password, but we will also consider risk analysis. We are calculating risk for the admin staff rather than the patient because the admin staff has access to all types of sensitive data.

Risk is quantified using various parameters which are related to putting data at risk. Employees must be evaluated by a risk calculation module before being granted access to any medical data. First and foremost, the module is launched by authenticating the user. This module is made up of two ends. The first is the employee side, and the second is the cloud server side. The cloud server side monitors the employee side. After successfully completing the first step of authentication via username and password, an employee moves on to the second step. Once a specific file is requested by an employee, the risk value for the file and the file requester is computed. Following computation, it is decided whether or not to grant access to the data. Every authentication session is recorded for future risk value calculation.

Chapter 4

Methodology

The entire model has been divided into sub-parts consisting of encryption and risk factors which are discussed below.

4.1 Attribute Based Encryption

A significant percentage of medical data is saved in cloud-based platforms to improve system performance. This data is encrypted and stored in a variety of routes on cloud servers. In order to protect the confidentiality of medical data, we use symmetric cryptography because public key encryption methods are ineffective when encrypting very large amounts of data. The encryption key is concealed using attribute-based techniques where two structures of attribute-based encryption are used. The KP-ABE structure is used to regulate the level of access to healthcare and service providers like hospitals, labs, and healthcare related companies. Depending on the access policy, the CP-ABE structure is utilized for individuals and wards where patients frequently divulge their medical information.

A symmetric encryption key is encrypted and stored on a private location using attribute-based encryption: Given that PHRs can have EMR data, medical records, including results of tests, MRI reports, etc., it is not optimal to keep this much medical information about each person in one place. Using a symmetric encryption key, the PHR data is encrypted and stored arbitrarily on several cloud computing platforms. The symmetric encryption key and the file storage path are then encrypted using attribute-based encryption, depending on the application.

User attributes are used as the public key in attribute-based encryption, a sort of public key encryption. Aspect-based encryption may by default also include identity-based encryption because user identification is a particular attribute. The two attribute-based encryptions are key policy (KP-ABE) and cipher policy (CP-ABE). KP-ABE encryption is based on a set of attributes for the ciphertext and an access structure for the user's private key. Figure 4.1 gives a short run on the working mechanism. The user can successfully decrypt the ciphertext using this method if the attribute set matches the access structure. The user's private key is linked to the access structure to restrict which encrypted texts the user can decrypt as shown in figure below.

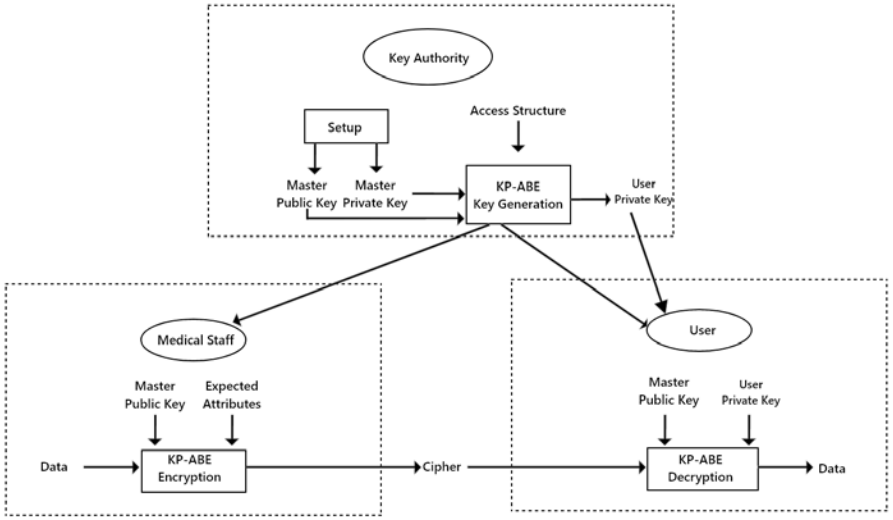

Figure 4.1: KP-ABE Scheme

In contrast to KP-ABE, CP-ABE encryption encrypts a message using a particular access policy and depends on the user's private key on a number of arbitrary parameters. By using this method, a user can only decrypt cipher-text if their characteristics match the specifications listed in the cipher-text. To limit who can decrypt the cipher-text, the access structure in CP-ABE is linked to the cipher-text. As a result, the ciphertext is labeled in KP-ABE, and it can be decoded if the labels correspond to the user's key access structure. In contrast, CP-ABE identifies the user's private key and adds an access structure to the ciphertext as per figure 4.2. If the access structure and the set of private key labels match, the ciphertext is available.

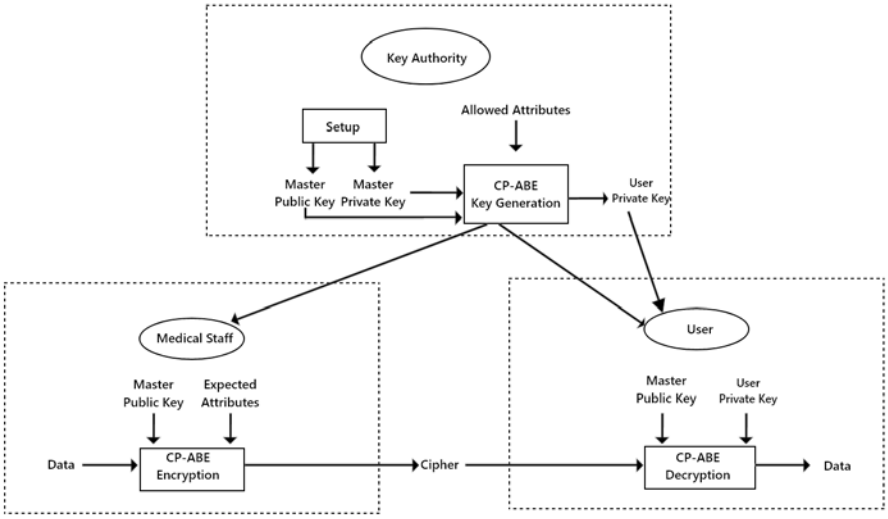

Figure 4.2: CP-ABE Scheme

Combining the ABE scheme to both CP-ABE and KP-ABE. As a result, the foundation of our technique is the notion that the entire set of attributes can be split

into m distinct, non-overlapping groupings. The hardest part of designing ABE is making sure that users who obtain essential components from several authorities do not plan attacks.

4.2 Searchable Encryption (SE)

SE refers to a server's capacity to search through ciphertext and get data without having to decrypt it. It secures the user's sensitive information by searching on the ciphertext and decreases the communication and computation overhead. An ideal SE performs both encryption function and keyword search on the ciphertext. Basic searchable encryption is made up of three components which are a data owner, a semi-trustworthy distant server, and a group of authorized data users with search access [17]. Each unit's functionality and capabilities are as follows:

- **Data owner:** Data owner is the unit that is trusted by all other participants in the system. It is responsible for generating the query private key and attribute private key which will be shared with authorized users for creating the trapdoor and decrypting the ciphertexts. Besides, this unit outsources a collection of files as well as some keywords, to be utilized in search operations later. The owner encrypts the files using system public parameters and access structure and then sends them to the distant server.
- **Authorized data users:** When an authorized user with access control wants to search files for the desired indexed keyword of interest, he or she must send a query of keyword message to the remote server, which serves as a trapdoor. As a result of this keyword message, many searchable techniques have evolved. Once the search is complete, the remote server returns the desired files containing the requested keywords to the authorized user. After that, user will decrypt the ciphertext using his attribute private key and obtain the plaintext.
- **Semi-honest distant server:** The search tasks on the ciphertexts are handled by the distant server unit. When a distant server receives a query of keyword request from an authorized user together with its trapdoor, it runs a search operation on the cipher and retrieves and sends the related files that contain the desired keyword to the user. We assume the server is somewhat trustworthy and curious. This indicates that, while the server follows the standards, it may investigate the requested data and derive extra information.

Figure 4.3 depicts the traditional model of Searchable Encryption consisting of three units. Data owner encrypts the data and outsource the data to the server along with encrypted index. Authorised data users are given a security key by data owner before making a keyword search. With the help of security key user make a trapdoor search and using this trapdoor server run a algorithm on index and return the resultant list of encrypted data.

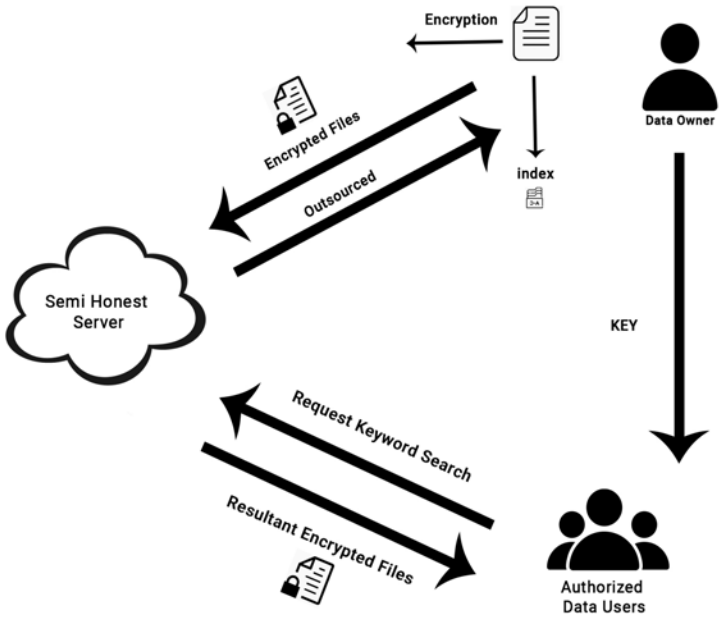

Figure 4.3: Traditional Model of SE

Functionalities and Techniques of SE Schemes

Various SE schemes have been proposed in the field of Cloud SE. The basic and common functionality of every scheme is outsourcing the encrypted data to the server while keeping the search option on them. Besides effective search on ciphertext, an SE scheme must address other search-related issues, such as query representation, user authorization, multi-user group access permissions, repudiation, etc [17]. Two common techniques of SE schemes are Searchable Symmetric Encryption (SSE) and Public Key Encryption With Keyword Search (PEKS). Besides these two, other techniques such as Hidden Vector Encryption (HVE), Identity-Based Encryption (IBE), etc. have been introduced in different SE schemes. These techniques are intended to keep client and server communication on the cloud secure and effective. Additionally, they facilitate single-user architecture and architecture for multiple users along with different kinds of query types like single-keyword search, multi-keyword search, ranked search, fuzzy multi-keyword search, etc.

Searchable Symmetric Encryption (SSE)

In our model, we have adopted Searchable Symmetric Encryption (SSE) since it offers multi-user settings and is appropriate for outsourcing sensitive data, such as EHR, to a distant server where a single data is shared by numerous recipients. By sending separated and hidden queries, Searchable Symmetric Encryption (SSE) enables users to upload data to the cloud with verifiable secrecy. The server can only discover the ciphertext through hidden query and isolation query, not the plain-

text. The query is being executed as a trapdoors-encrypted query, which is always generated using a secret key. The general SSE algorithm is given below:

- **Keygen(p):** This is a key generation process that the data owner controls. It accepts public security parameters, a set of attributes from users as input and produces two secret keys - query private key(K) and attribute private key(Ka). Both keys will be shared with authorized users to be utilized in trapdoor generation and decryption.
- **BuildIndex(p, D):** This is keyword index creation algorithm which is also run by data owner. The data owner must create an index table with a list of all the keywords for each document. After that, this keyword index table and system security parameters are sent to this algorithm as inputs and it returns a secure keyword index I.
- **Trapdoor(K, Ka, w):** This is a client-controlled keyword trapdoor generating algorithm. Using this trapdoor, users will send search request to the server. It accepts query secret key K, attribute private key Ka and a keyword w for the search as inputs and returns the trapdoor Tw for the given keyword w. It is mainly used to secure the search keyword when it is sent to the untrusted server so that server can not understand and infer any information from the keyword.
- **Search(I, Tw):** This is a server-based keyword search algorithm. Server uses the trapdoor as a token to execute the search operation on keyword index of each document to check if any document has the keyword or not. It generates a set of documents D(w) that contain the search term w after receiving as inputs a keyword index I and a trapdoor Tw. If keyword is not found in any document's keyword index, then it will return 0.

After the search operation, user can retrieve the specific document from the document list and decrypt the document using attribute private key(Ka) to the plaintext.

Mechanism of Searchable Symmetric Encryption (SSE)

In SSE method, the server stores the encrypted documents and indexes associated with the documents. Before storing the documents and associated index, two algorithms are run by the data owner. These are Encryption algorithm and BuildIndex algorithm. These algorithms together produce encrypted data and encrypted index. In our model, During encryption, data owner uses CP-ABE access policy to encrypt the document along with keyword index. In addition, the data owner runs the Keygen algorithm which generates the query private key and attribute private key and shares them with medical personnel. While searching, clients/medical personnel create the trapdoor(Tw) for any keyword w with a query private key(K) and send this trapdoor(Tw) to the server. The server then executes the Search(I, Tw) function for every record in the server taking keyword index as input to check if any record includes keyword w. After executing the function, the server eventually sends a list with unique identifiers of each record containing keyword w. Thus, the semi honest server can not infer any pertinent information from the index since only the client possesses the private master key which was used to produce the trapdoor.

4.3 Risk Analysis

Though the CP-ABE mechanism is contributing to access control using the role attribute of users, past-behavior and data sensitivity are not considered. Taking into account the fact that sensitive patient data could be misused by internal medical employees, risk analysis during accessing patient health records can be a solution to this issue. For this reason, we have further increased the access control part including the risk assessment after access request to specific record which will ensure more fine-grained access control.

In the access control part we are considering risk analysis of authentication when anyone from the administration side is trying to access any sensitive data. For patients, we are giving them access to their data by two factor authentication of login with their username or password. Then verification is done by sending OTP through SMS. However on the administration side, we are not only checking username and password but also we will be considering risk analysis. The reason we are not calculating risk for the patient but for the admin is because all types of sensitive data are exposed to the administration staff.

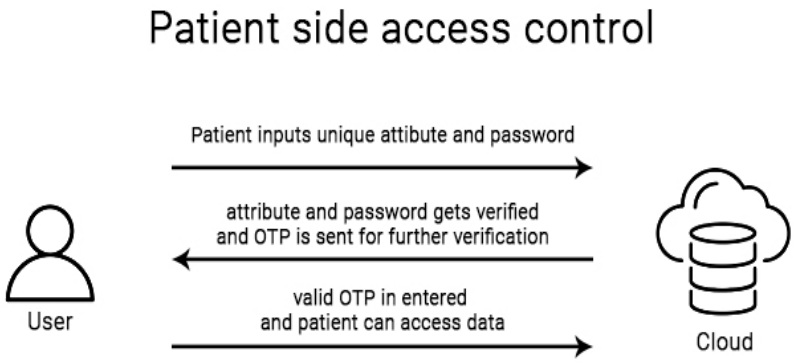

Figure 4.4: Access Control from the patient side

In figure 4.4, the access control mechanism from patient side is shown. When patient sends login request to the server with unique attribute and password, server first verifies the information that the user has provided. Upon verification, an OTP is sent to the user. If patient enters correct OTP, he/she will get the access to the data, otherwise the request will be terminated and the patient will go back to the login page.

Here risk is based on different quantized parameters that has the correlation posing threat to data privacy. To gain access to any medical data, employees have to be assessed by a risk calculation module. First of all, the module gets started by authentication of the user. This module consists of two ends: the employee side and the other the cloud server side. The employee side is tracked by the cloud server

side. After succeeding the first step of authentication by username and password, an employee faces the second step. Employee requests for a specific file and the risk value associated with the file and the file requester is calculated. After computation it is decided whether to give access to the data or not. After every session of authentication, it is recorded for future calculation of risk values. This proposed module has four steps:

- **Registration and login**: Every employee whether it's a doctor, medical staff or other has to be registered in order to login. After logging in risk is calculated using the risk parameters. Also, employees are required to maintain a similar location within workspace which has the influence of the risk parameter working location index.
- **Allowing access session**: To allow a session, parameters associated with risk are analyzed and necessary computations take place. Firstly, current risk values are checked against threshold risk after calculations. To grant a session, the current risk value has to be below the threshold value. If it crosses the threshold value, the session gets terminated.
- **Permission Provision**: When administration employees ask for any file access or permissions, we compare it against the risk value the employee has been assigned before. To get access, current risk must be lower than the threshold risk.
- **Session Termination**: If no risky behavior is found the session is continued otherwise it gets terminated. Also, there might be a temporary ban of medical staff accounts if there is a greater risk. To get out of the ban employees have to contact the technical admin.

Figure 4.5 on the following page demonstrates the risk based access control model from admin side. Firstly, any user from admin side sends access request with their ID and password. After verification, user can request for specific data to gain access. Then, this request goes to the system for calculation of the threshold risk followed by the current risk. After that, system compares both risk values and checks whether current risk is less than the threshold risk or not. If current risk value is within the threshold risk value, admin will gain access to that specific data. Otherwise, this access request will be denied and terminated by the system.

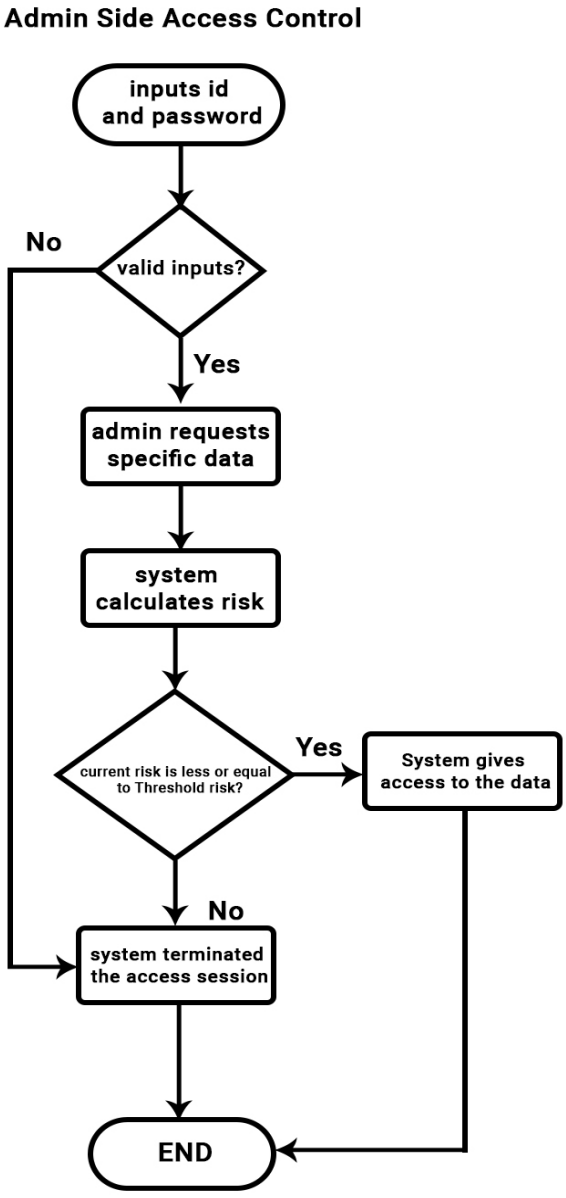

Figure 4.5: Risk based access control on admin side

4.3.1 Risk Parameters

One of the characteristics of our risk assessment is that it is score based. In the score based risk assessment, all of the parameters are quantized. Also there are few requirements of this score based requirements mentioned below [20]:

1. Parameters in the risk assessment should be normalized to a general scale
2. Parameters should avoid wide variability. If any parameter is in range of high variability the values should be reconsidered.
3. Parameter value should be adjustable to latest updates.
4. Suggestion of change in risk values should be given.

The risk parameters associated with our access control risk analysis are nine in total. Five of them are needed for calculating current risk and other four parameters are for threshold risk. Explanation of each parameter is explained here below:

- **Years of experience:** Any employee who spends more time on an organization is more reliable than any newcomer to the company. So, employees with most days or years of working will have the lowest risk value whereas employees with least working days have the highest risk value. In table 4.1, according to years of experience within the health organization risk values are segmented.

Years of Experience	Risk Value
1	0.7
2	0.6
3	0.5
4	0.4
5-10	0.3
10-15	0.2
Over 15	0.1

Table 4.1: Years of Experience & Risk value

- **Designation:** Not everyone should have the equivalent access to all the data. A receptionist who deals with doctor appointments in the health institution should not have access to cancer diagnostics data as it does not have relevance to the role. Also, if that particular role even gets access to the data. There will be a potential risk of data breaching. Any medical staff with higher rank will pose lower risk. On the other hand lower ranked staff will have the highest risk. The table 4.2 shows how we imposed the risk value of medical employees based on the salary and influence in service.

Designation	Risk Value
Healthcare Administrator	0.1
Senior Surgeons	0.2
Specialized Doctors	0.3
Junior Doctors	0.4
Medical Assistant	0.5
Nurse Practitioner	0.6
Medical Technician	0.7
Receptionist/Medical Biller	0.8

Table 4.2: Designation & Risk value

- **Failed Login Session:** It refers to the ratio of denied attempts and total data access attempts session after login. The lowest value will be 0 which means no failed session in the past and highest value will be 1 means failed all the sessions previously. After every login session of whether it is failed or granted the session log is stored in the cloud data based with the user for future measurement of risk with the parameter.

$$Failed\ Login\ Ratio\ = \frac{number\ of\ denied\ access\ attempts}{total\ access\ attempts} \tag{4.1}$$

- **Referral Index:** It refers to the value of reference or recommendation by someone senior. Let's say someone with higher rank referred to someone then the referral value would be higher. The table below shows the designation 8 or lower rank have the lowest referral index value of 0.1. For instance, matching with the table of "Risk values based on Designation", if a person is referred by a Receptionist who is of lowest rank (rank 8), the referral value is taken to be lowest of 0.1. On the other hand, if the person is referred by Healthcare Administrator (rank 1), the referral value is the highest of 0.8. Table 4.3, shows the designation 8 or lower rank have the lowest referral index value of 0.1. The referral value is considered as the referral index.

Designation Index	Referral Value
8	0.1
7	0.2
6	0.3
5	0.4
4	0.5
3	0.6
2	0.7
1	0.8

Table 4.3: Designation Index & Referral value

- **Working Location Index:** It is decided upon the ratio of access outside the medical service facility and total number of access within a specific time. Risk is proportional to higher value of the location index. The location of the role will be determined by IP address or MAC address. Firstly, the IP address

of a user or mac address of the organization devices will be stored in a cloud database. Anytime a role tries to login the stored mac address in the database will be checked against the device which requested data in the current session. New devices will be taken as outside the location of the health services facility which means potential vulnerability or risk.

$$Working\ Location\ Index = \frac{Number\ of\ access\ attempts\ outside\ the\ institution}{Total\ access\ attempts} \tag{4.2}$$

- **Working Time Index:** Every designation in the health organization has a fixed working time for a particular day. It's not rigid for all the roles. For example a role can have working hours for a particular day from morning to afternoon and another day afternoon to night. This parameter is calculated by the ratio of access attempts outside working time and total number of access attempts. A role that has working time from morning to afternoon if he tries to access outside, their time will be marked as outside working time in the cloud storage.

$$Working\ Time\ Index = \frac{Number\ of\ access\ attempts\ outside\ of\ working\ hours}{Total\ access\ attempts} \tag{4.3}$$

- **Appraisal factor:** An overseer can monitor the activities of employees and can set the appraisal factor based on their organizational or service performance and sincerity. Lower the performance higher the risk. The overseer can be immediately someone senior in rank or can be someone from the health management staff.

- **Probationary Period:** This is referring to remaining months to complete the probationary period as an employee. For someone who just joined the health services will have a default probationary period of 12 months. Anyone who does not have any probation month left still has the value of 1 due the being in the parameter as a denominator as a multiplied form. Otherwise the whole threshold risk value might occur in error.

- **Patient Data Sensitivity:** This sensitivity is set based on data hierarchy we consider risky from low to high. Any appointment related data does not pose much risk in even leakage so we can assign it to the lowest. The highest value we are considering is any data that is related to a patient with incurable disease and can lead to death. This data sensitivity can be flexible and changed upon health organizations' own defined magnitude of sensitivity. For evaluation purposes we can make the hierarchy of the data like shown in table 4.4. The data sensitivity is shown in a generalized way which defines level-7 to be the highest risk and level-1 to be the lowest of risk.

Data Sensitivity Level	Risk Value
Level-7	0.7
Level-6	0.6
Level-5	0.5
Level-4	0.4
Level-3	0.3
Level-2	0.2
Level-1	0.1

Table 4.4: Data Sensitivity Level & Risk value

The table 4.5, demonstrates the summarized value of all the nine parameters by showing only minimum and maximum values.

Risk Parameter	Minimum Value	Maximum Value
Years of Experience	0.1	0.7
Designation/Role	0.1	0.8
Failed Login Ratio	0	1
Referral Index	0.1	0.8
Working Location Index	0	1
Working Time Index	0	1
Appraisal Factor	0	1
Probationary Period	1	12
Patient Data Sensitivity	0.1	0.7

Table 4.5: Risk Parameter & Minimum Value & Maximum Value

4.3.2 Risk Calculation

As we have discussed previously two types of risk are being taken into account and compared. Based on comparison it is decided whether to give access or not.

Threshold Risk

In our case of risk calculation, threshold risk is the verge risk upto which point we would allow any risk. There are five parameters out of nine parameters which will be needed for this threshold risk. These risk parameters are mostly associated with the administrative staff such as their rank, years of experience, remaining probation period, the rank of their referral and the number of times they have failed to login, out of the total attempted login sessions. This threshold risk calculation formula as shown in 4.4 is derived using Naive Bayes classifier algorithm for trust model or risk based model [2].

The threshold risk formula is calculated as follows [8]:

$$Threshold\ Risk = \frac{Failed\,Login\,Ratio * designation\,risk * referral\,index}{Years\,of\,experience\,risk * remaining\,probationary\,period} \tag{4.4}$$

Firstly, computation of the threshold risk is done and we get a value. Following that, current risk computation is done and checked against the threshold risk. If the threshold risk is greater or equal to the current risk value, access to a data will provided otherwise the session will be terminated.

Current Risk

Then there is the current risk which is average of failed login ratio, working location index, ratio of during and outside work hour index, appraisal factor and data sensitivity associated with patient data. The current risk value is mostly associated This current risk is checked against the threshold risk to decide data access.

The current risk formula is calculated as follows-

$$Current\ Risk = \frac{FLR + WLI + WTI + AF + PDS}{5} \tag{4.5}$$

where, FLR = Failed Login Ratio
WLI = Working Location Index
WTI = Working Time Index
AF = Appraisal Factor
PDS = Patient Data Sensitivity

Chapter 5

Experimentation and Results

Our paper gives a security proposal to enhance existing security models by adding a layer of access control by evaluation of risk associated with the data requester and the data sensitivity. As our approach is model based rather than implementational we will be showing cases based analysis. There will be shown whether a data access session is granted or denied based on comparing current risk with the threshold risk. We can categorize risk the following way.

Threat Category	Case Description
Minor	1. A staff with medium level of years of experience requesting a data with low sensitivity 2. A staff with higher years of experience and higher rank requesting data with lower sensitivity
Moderate	3. A recently joined staff or lower ranked staff trying to access a mid level data in sensitivity. 4. A staff with mid level experience or rank trying to access a mid level data in sensitivity.
Severe	5. A low designation staff trying to access high sensitivity data. 6. A staff with less years of experience trying to access high level data.

Table 5.1: Threat Category & Case description

Now let's consider every category case mentioned in the above table. Firstly, we consider the minor category two cases.

Case 1: A staff with medium level of years of experience requesting a data with low sensitivity

Risk Parameter	Risk Value
Years of Experience	0.4
Designation/Role	0.6
Failed Login Ratio	0.45
Referral Index	0.5
Working Location Index	0
Working Time Index	0
Appraisal Factor	0
Probationary Period	1
Patient Data Sensitivity	0.1

Table 5.2: Parameter values of Case 1

From table 5.2, we set up different parameter values of the mentioned case-1. We set up years of experience risk value to be the mid value of 0.4 then patient data sensitivity 0.1 which is the minimum value. We set the working location index, appraisal factor, and working time index to 0 as for someone new. Also the referral index was set to 0.5 assuming that the individual was referred by someone in mid designation.

Calculating the risk values,

$$Threshold\ Risk = \frac{0.45 * 0.6 * 0.5}{0.4 * 1} = 0.3375 \tag{5.1}$$

$$Current\ Risk = \frac{0.1 + 0 + 0 + 0 + 0.45}{5} = 0.11 \tag{5.2}$$

Intended Outcome: Access granted

Outcome: The current risk is less than the threshold risk. So the system is granting access. Granting access to someone with mid level experience would not pose much of a threat to the data so access is acceptable.

Case 2: A staff with higher rank and years of experience requesting data with lower sensitivity

Risk Parameter	Risk Value
Years of Experience	0.1
Designation/Role	0.1
Failed Login Ratio	0.85
Referral Index	0.7
Working Location Index	0
Working Time Index	0
Appraisal Factor	0
Probationary Period	1
Patient Data Sensitivity	0.1

Table 5.3: Parameter values of Case 2

In the table 5.3 above, we set up values for case-2 so that it satisfies the criteria of higher years of experience, high designation and low data sensitivity. The risk value of years of experience was 0.1, designation 0.1 and data sensitivity 0.1. Also, the failed login ratio was set to 0.85.

Calculating the risk values,

$$Threshold\ Risk = \frac{0.85 * 0.1 * 0.7}{0.1 * 1} = 0.595 \tag{5.3}$$

$$Current\ Risk = \frac{0.1 + 0 + 0 + 0 + 0.85}{5} = 0.19 \tag{5.4}$$

Intended Outcome: Access granted

Outcome: The current risk is lower than the threshold risk. So the system is granting access. Definitely we do not want to stop someone higher rank to access a data with minimal. So, this case is working as intended.

Case 3: A recently joined and mid ranked staff trying to access a mid level data in sensitivity

Risk Parameter	Risk Value
Years of Experience	0.7
Designation/Role	0.5
Failed Login Ratio	0.25
Referral Index	0.4
Working Location Index	0
Working Time Index	0
Appraisal Factor	0
Probationary Period	12
Patient Data Sensitivity	0.4

Table 5.4: Parameter values of Case 3

In table 5.4, we set up values for case-3 so that it satisfies the criteria of recently joined, mid designation and mid data sensitivity. The risk value of years of experience was 0.7, designation 0.5 and data sensitivity 0.4 with a probation period of 12.

Calculating the risk values,

$$Threshold\ Risk = \frac{0.35 * 0.5 * 0.4}{0.8 * 12} = 0.00729 \tag{5.5}$$

$$Current\ Risk = \frac{0.35 + 0 + 0 + 0 + 0.4}{5} = 0.15 \tag{5.6}$$

Intended Outcome: Access denied

Outcome: The threshold risk is significantly so small compared to the current risk. In this case, the system will not grant access to someone with less experience and mid level role to mid sensitivity data.

Case 4: A staff with mid level experience or rank trying to access a mid level data in sensitivity

Risk Parameter	Risk Value
Years of Experience	0.4
Designation/Role	0.4
Failed Login Ratio	0.65
Referral Index	0.4
Working Location Index	0
Working Time Index	0
Appraisal Factor	0
Probationary Period	1
Patient Data Sensitivity	0.3

Table 5.5: Parameter values of Case 4

In table 5.5, we set up a value for case-4 such that years of experience, designation risk and referral index is value of 0.4 and patient data sensitivity is 0.3 so that it matches the case requirement.

Calculating the risk values,

$$Threshold\ Risk = \frac{0.65 * 0.4 * 0.4}{0.4 * 1} = 0.26 \tag{5.7}$$

$$Current\ Risk = \frac{0.3 + 0 + 0 + 0 + 0.65}{5} = 0.19 \tag{5.8}$$

Intended Outcome: Access granted

Outcome: The current risk is lower than the threshold risk. So the outcome will be granting access which is the same as the intended outcome. Giving access to data which have mid level sensitivity to a mid level experienced and ranked individual might not pose that much of a threat to data.

Case 5: A low designation but more years of experience staff trying to access high sensitivity data

Risk Parameter	Risk Value
Years of Experience	0.1
Designation/Role	0.8
Failed Login Ratio	0.25
Referral Index	0.2
Working Location Index	0.35
Working Time Index	0.25
Appraisal Factor	0.2
Probationary Period	1
Patient Data Sensitivity	0.7

Table 5.6: Parameter values of Case 5

In table 5.6, we set the values of years of experience to be minimum which is 0.1, designation 0.8, patient data sensitivity 0.7 to meet the case 5 requirements. For the working time index, working time index and appraisal factor we assumed values 0.35, 0.25, 0.2 respectively which denoted light or medium risk of those parameter values. As there is no probationary period left, the value is set to default 1 for that parameter.

Calculating the risk values,

$$Threshold\ Risk = \frac{0.25 * 0.8 * 0.1}{0.1 * 1} = 0.20 \tag{5.9}$$

$$Current\ Risk = \frac{0.85 + 0.35 + 0.25 + 0.3 + 0.7}{5} = 0.23 \tag{5.10}$$

Intended Outcome: Access denied

Outcome: The current risk is greater than the threshold risk. So the system is denying the access of a specific data. Granting access to someone in low designation but higher years of experience might pose a security threat. A receptionist with 15 years in the medical service cannot be given access to cancer related data. Our system is restricting access as we intended.

Case 6: A staff with less years of experience and low designation risk value trying to access high level data

Risk Parameter	Risk Value
Years of Experience	0.7
Designation/Role	0.8
Failed Login Ratio	0.45
Referral Index	0.1
Working Location Index	0
Working Time Index	0
Appraisal Factor	0
Probationary Period	10
Patient Data Sensitivity	0.7

Table 5.7: Parameter values of Case 6

In table 5.7, we set the values of years of experience, designation, patient data sensitivity and probationary period such that it satisfies the above mentioned condition of case 6. So, we set years of experience 0.7, designation risk value to be 0.8, probation period value 10, failed login ratio 0.45 and Patient data sensitivity 0.7.

Calculating the risk values,

$$Threshold\ Risk = \frac{0.45 * 0.8 * 0.1}{0.8 * 10} = 0.0045 \tag{5.11}$$

$$Current\ Risk = \frac{0.7 + 0 + 0 + 0 + 0.45}{5} = 0.23 \tag{5.12}$$

Intended Outcome: Access denied

Outcome: Again, the current risk is greater than the threshold risk. So the system is denying the access of a specific data. Granting access to a low ranked and low experienced individual poses a severe risk for data security. So in this case we are seeing an intended outcome.

Chapter 6

Performance Analysis

6.1 Analysis of Encryption

Attribute Based Encryption

For cloud systems, ABE is the right fit as a beginning encryption technique. It comes in helpful when we need to share a resource safely. The appropriate credentials allow access to a resource, which is not entirely public. The combination of using CP-ABE and KP-ABE keeps data more confidential, gives a secured access control and the whole system becomes scalable. CP-ABE performs highly efficiently which makes the privacy and accuracy even higher. The overall encryption scheme becomes quite robust.

Searchable Encryption

Traditional data utilization methods are based on plaintext keyword search. For example, when a user needs to search on encrypted files, he has to download the encrypted files first. Then, after doing the decryption he can search on the plaintext to retrieve the keyword. But, this method is tedious and brings a lot of computation and communication overhead. To overcome this issue, searchable encryption is introduced which ensures searching keyword on ciphertext securely, effectively, and efficiently.

Between the two main branches of searchable encryption, we choose Searchable Symmetric Encryption (SSE) since it supports multi-user access settings and it is suitable for a big organization application system where a massive amount of data is shared among numerous recipients. It uses hidden and isolation queries which ensures that the server can not learn anything about the plaintext except the ciphertext. Besides, another reason behind choosing SSE is that it has been widely used in the ehealth cloud over the years because of its high success rate compared to Searchable Asymmetric Encryption (SAE).

6.2 Analysis of Risk

The factors that have been used for calculating threshold risk are designation, failed login ratio, referral index and probationary period. Changes in any of those parameters affect the threshold risk value as a whole. In the graphs below we have shown how individual factors in the threshold risk formula change when a parameter changes. For each graph we made the factor we are trying to show relation with the threshold risk variable and made other factors rigid.

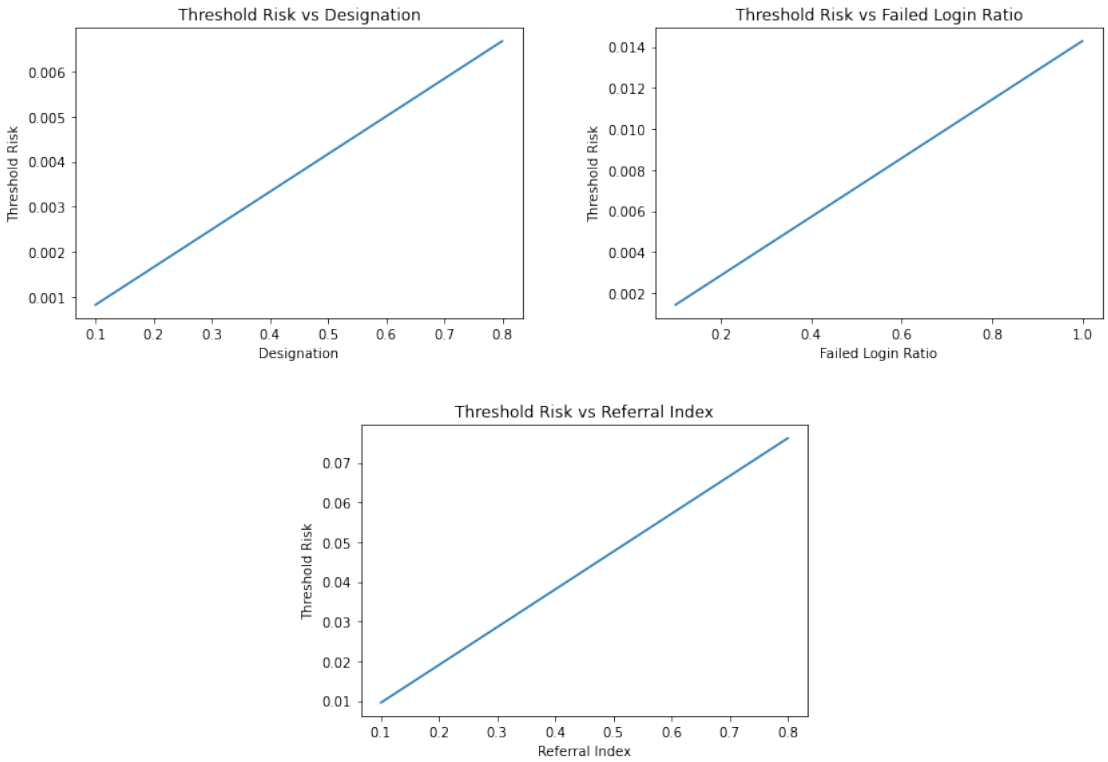

Figure 6.1: Linear relationship between Factors and Threshold risk

From the graph in 6.1 we see that the designation, referral index, and failed login ratio have a linear relationship with the graph. Which indicates any of the mentioned factors increase means increase in threshold risk. We should prioritize giving data access to higher designation. However, we are seeing threshold risk flexibility increasing when the designation risk is increasing. The reason is there are other parameters also being considered here. Threshold risk might increase with designation but years of experience, referral index, probation period will change the threshold value in a way that some new joined higher designation will not instantly get data. To get data access then other factors have to be in adequate value. Referral index also make the threshold risk flexible by increasing linearly. Let's say someone in mid level designation needs to access data of high level. If that individual is referred by higher than only he can access by additional referred index value. Also, failed login ratio does the same. However, this should be restricted to someone with higher years of experience and higher rank as they might have more failed login sessions in the organization for working there for years.

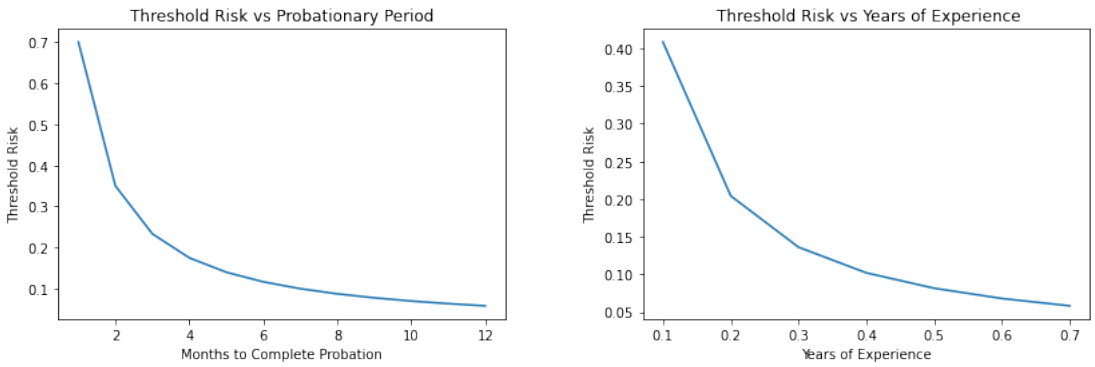

Figure 6.2: Inverse relationship between Factors and Threshold risk

On the other hand in figure 6.2, the risks of years of experience and time to complete the probationary period show an inverse relationship with the threshold risk. When someone with fewer years of experience or high probation time requests data, the threshold risk is low. The lesser value of threshold risk, the less chance of getting access granted. For example, if we see the probationary period graph, the highest probationary period is 12 and we get the threshold risk less than 0.1. That means there is not much scope for granting access. In a similar manner, with fewer years of experience, the threshold risk value decreases. In the graph, we see for the highest risk value of years of experience 0.7 the threshold risk value is around 0.05. This value is so small that for getting a successful data access session the value of current risk has to be less or equal to 0.05.

There might be cases where someone in a lower designation can be referred by someone with a higher designation. In these scenarios, the threshold risk will be linearly high which will result in someone with a lower rank accessing data with high sensitivity. To resolve this we can limit the referring criteria. In the new criteria of referring, only 2 or 1 rank higher employee can refer to a lower rank person.

Also, we will consider failed login ratio only for the higher rank and highly experienced staff due to the fact that high failed login ratio increases the threshold risk. As someone experienced in the organization with many years on the job will have a lot of failed login sessions in real life. So, the failed login ratio will not be taken into account for lower rank and fewer years of experience or might be considered inversely for lower designation.

Implementation Limitation

Our system needs updates on all the parameter values that are being considered for risk calculation. The designation risk parameter value will be given at the registration phase also the referral index. Designation increases with promotion. As days go by, years of experience, probationary period will be get up-to-date. Working time index, working location index, and failed login ratio will be derived from the previous logs of the employee history which will be a bit time-consuming while giving access compared to traditional access control. Furthermore, storing all those logs and values will require additional space in the cloud storage. To make it feasible in real life we have to evaluate the performance and latency of our system

by implementing on a small scale.

Chapter 7

Conclusion

eHealth data is vulnerable due to data being exposed to the cloud service holders and by the activity of users. Encryption solutions such as Attribute-based encryption was used to protect the data from outside attacks. If a user wants to access data he/she will be first verified then the data operation action will be held based upon the risk that will be calculated from previous behavior and sensitivity of the data. So even if any verified user tries to do some malicious data request the action will be denied. Our proposed model will be having attribute-based encryption with searchable encryption the access control of this model will be based on risk factor calculation.

Our future research will pay greater attention to the model's possible effects, and we will conduct more analysis to show that the suggested model will be more effective than existing models. There are a lot of potential areas that can be focused such as encryption, risk, and many more. The model will be explained in further detail and taking the model to be in a better stage than the current stage. Future research is required to confirm the types of findings that may be derived from this study.

Bibliography

[1] S. Chakraborty and I. Ray, "Trustbac: Integrating trust relationships into the rbac model for access control in open systems," in *Proceedings of the eleventh ACM symposium on Access control models and technologies*, 2006, pp. 49–58.

[2] W. Yuan, D. Guan, S. Lee, and Y. Lee, "A dynamic trust model based on naive bayes classifier for ubiquitous environments," in *High Performance Computing and Communications*, Springer Berlin Heidelberg, 2006, pp. 562–571. DOI: 10.1007/11847366_58. [Online]. Available: https://doi.org/10.1007/11847366_58.

[3] N. Baracaldo and J. Joshi, "A trust-and-risk aware rbac framework: Tackling insider threat," in *Proceedings of the 17th ACM symposium on Access Control Models and Technologies*, 2012, pp. 167–176.

[4] K. Z. Bijon, R. Krishnan, and R. Sandhu, "Risk-aware rbac sessions," in *International Conference on Information Systems Security*, Springer, 2012, pp. 59–74.

[5] M. Sharma, Y. Bai, S. Chung, and L. Dai, "Using risk in access control for cloud-assisted ehealth," in *2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems*, IEEE, 2012, pp. 1047–1052.

[6] H. O. Alanazi, A. Zaidan, B. Zaidan, M. Kiah, and S. Al-Bakri, "Meeting the security requirements of electronic medical records in the era of high-speed computing," *Journal of medical systems*, vol. 39, no. 1, pp. 1–13, 2015.

[7] M. Bahrami and M. Singhal, "A dynamic cloud computing platform for ehealth systems," in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, IEEE, 2015, pp. 435–438.

[8] L. H, N. S, Seemanthini, *et al.*, "Risk based access control in cloud computing," Oct. 2015, pp. 1502–1505. DOI: 10.1109/ICGCIoT.2015.7380704.

[9] A. Soceanu, M. Vasylenko, A. Egner, and T. Muntean, "Managing the privacy and security of ehealth data," in *2015 20th International Conference on Control Systems and Computer Science*, IEEE, 2015, pp. 439–446.

[10] H. Elmogazy and O. Bamasag, "Securing healthcare records in the cloud using attribute-based encryption.," *Comput. Inf. Sci.*, vol. 9, no. 4, pp. 60–67, 2016.

[11] D. Fall, T. Okuda, Y. Kadobayashi, and S. Yamaguchi, "Risk adaptive authorization mechanism (radam) for cloud computing," *Journal of Information Processing*, vol. 24, pp. 371–380, Mar. 2016. DOI: 10.2197/ipsjjip.24.371.

[12] Z. Lu and Y. Sagduyu, "Risk assessment based access control with text and behavior analysis for document management," in *MILCOM 2016-2016 IEEE Military Communications Conference*, IEEE, 2016, pp. 37–42.

[13] H. Li, Y. Yang, Y. Dai, S. Yu, and Y. Xiang, "Achieving secure and efficient dynamic searchable symmetric encryption over medical cloud data," *IEEE Transactions on Cloud Computing*, vol. 8, no. 2, pp. 484–494, 2017.

[14] R. Zhang, R. Xue, and L. Liu, "Searchable encryption for healthcare clouds: A survey," *IEEE Transactions on Services Computing*, vol. 11, no. 6, pp. 978–996, 2017.

[15] O. Ali, A. Shrestha, J. Soar, and S. F. Wamba, "Cloud computing-enabled healthcare opportunities, issues, and applications: A systematic review," *International Journal of Information Management*, vol. 43, pp. 146–158, 2018.

[16] S. Aqeeli, M. Al-Rodhaan, Y. Tian, and A. Al-Dhelaan, "Privacy preserving risk mitigation approach for healthcare domain," *E-Health Telecommunication Systems and Networks*, vol. 07, pp. 1–42, Jan. 2018. DOI: 10.4236/etsn.2018.71001.

[17] S. Buchade and P. Devale, "A study on searchable encryption schemes," *International Journal of Engineering & Technology*, vol. 7, no. 2, pp. 617–620, 2018.

[18] J. Zhang, J. Ma, Z. Ma, *et al.*, "Efficient hierarchical data access control for resource-limited users in cloud-based e-health," in *2019 International Conference on Networking and Network Applications (NaNA)*, IEEE, 2019, pp. 319–324.

[19] C.-L. Chen, P.-T. Huang, Y.-Y. Deng, H.-C. Chen, and Y.-C. Wang, "A secure electronic medical record authorization system for smart device application in cloud computing environments," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–31, 2020.

[20] M. Jayabalan, "Towards an approach of risk analysis in access control," in *2020 13th International Conference on Developments in eSystems Engineering (DeSE)*, 2020, pp. 287–292. DOI: 10.1109/DeSE51703.2020.9450772.

[21] S. K. S. Raja, A. Sathya, and L. Priya, "A hybrid data access control using aes and rsa for ensuring privacy in electronic healthcare records," in *2020 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, IEEE, 2020, pp. 1–5.

[22] O. Alabi, "A review on information security of cloud based electronic health record," in *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*, 2021.

[23] J. Jusak, S. S. Mahmoud, R. Laurens, M. Alsulami, and Q. Fang, "A new approach for secure cloud-based electronic health record and its experimental testbed," *IEEE Access*, vol. 10, pp. 1082–1095, 2021.

[24] I. Keshta and A. Odeh, "Security and privacy of electronic health records: Concerns and challenges," *Egyptian Informatics Journal*, vol. 22, no. 2, pp. 177–183, 2021.

[25] A. Sahi, D. Lai, and Y. Li, "A review of the state of the arts in privacy and security in the ehealth cloud," *Ieee Access*, 2021.

[26] M. Calvo and M. Beltrán, "A model for risk-based adaptive security controls," *Computers &amp Security*, vol. 115, p. 102 612, Apr. 2022. DOI: 10.1016/j.cose.2022.102612. [Online]. Available: https://doi.org/10.1016/j.cose.2022.102612.