This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law. Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it’s important to you. Thank you.

Preprint: In proceedings of 7th International Workshop on Privacy and Security of Big Data (PSBD 2020), in conjunction with 2020 IEEE International Conference on Big Data (IEEE BigData 2020)

Cloud-based Encrypted EHR System with Semantically Rich Access Control and Searchable Encryption

Redwan Walid*, Karuna P. Joshi*, Seung Geol Choi†, Dae-young Kim*
*University of Maryland, Baltimore County
{rwalid1, karuna.joshi, leroy.kim}@umbc.edu
†United States Naval Academy
{choi}@usna.edu

Abstract—**Cloud-based electronic health records (EHR) systems provide important security controls by encrypting patient data. However, these records cannot be queried without decrypting the entire record. This incurs a huge amount of burden in network bandwidth and the client-side computation. As the volume of cloud-based EHRs reaches Big Data levels, it is essential to search over these encrypted patient records without decrypting them to ensure that the medical caregivers can efficiently access the EHRs. This is especially critical if the caregivers have access to only certain sections of the patient EHR and should not decrypt the whole record. In this paper, we present our novel approach that facilitates searchable encryption of large EHR systems using Attribute-based Encryption (ABE) and multi-keyword search techniques. Our framework outsources key search features to the cloud side. This way, our system can perform keyword searches on encrypted data with significantly reduced costs of network bandwidth and client-side computation.**

Index Terms—**Attribute Based Encryption, Attribute Based Access Control, Searchable Encryption, Electronic health Record, Knowledge Graph (Ontology), Cloud Computing, Cloud Security**

I. Introduction

A cloud-based Electronic Health Record (EHR) system helps a health organization record patient information in one place and take advantage of cloud-based storage services [4], [20], [21], [22]. Cloud storage provides a higher degree of efficiency and elasticity, as well as substantial cost savings. Medical centers have embraced cloud-based technology to manage and operate efficiently with their official records. Multiple research projects have been designed with such a focus on secure, cloud-based EHR approaches [21], [4].

On the other hand, storing EHR on third-party cloud services is associated with an increased risk of attack and data breaches leading to patient security issues. Governing bodies such as the Economic and Clinical Health Information Technology (HITECH) [26] and the Health Insurance Portability and Accountability Act (HIPAA) [11], [27] have recognized this issue. Therefore, incorporating an EHR system that conforms with all relevant laws and regulations facilitates a convenient and smooth sharing of patient data.

Motivation. Joshi et al. [17] presented a cloud-based electronic health records (EHR) system that provides important security guarantees. First, the health records are uploaded on the cloud server in an encrypted form to protect data privacy. Second, the system provides an access control mechanism so that only the right people can decrypt a given health record. Lastly, the system provides a tool for managing the access control policy easily but with rigor.

However, their system doesn't provide a key search. Owing to the large volume of data in the cloud, users may need to search through encrypted data. For example, physicians may search for a specific symptom or disease through encrypted EHRs on the cloud to choose patients requiring immediate treatment. Physicians may also need to find particular patients with a viral illness to prevent the community from the spread. When using the aforementioned system, due to lack of the keyword search feature, when a doctor wants to find a record with a specific illness system, the entire encrypted health records must be downloaded and then decrypted to find the right records. This incurs a huge amount of burden in-network bandwidth and the client-side computation.

A. Our Work

In this paper, we present our novel approach that facilitates searchable encryption of large EHR systems using Attribute based Encryption (ABE) and multi-keyword search techniques. Our framework outsources key search features to the cloud side. This way, our system can perform a keyword search with significantly reduced costs of network bandwidth and client-side computation. We follow the approach of Joshi et al. [17], and our framework uses an ABE [13] that provides owner-enforced fine-grained access control for encrypted data. In particular, we have implemented the ciphertext-policy ABE encryption scheme [6] in our system. Our system also implements a semantically rich, policy-driven framework that utilizes ABAC [15] to determine an individual's access to the system. We have used Semantic Web Technologies to incorporate this framework. We established a HIPAA-consistent knowledge graph (ontology) by referring to the HIPAA [16] knowledge graph built in our previous research. Our system derives user and EHR field attributes to implement safe access management and assignment-based encryption from the knowledge graph.

Keyword search. To support keyword search, we employ the attribute-based searchable encryption scheme [31], [24] that enables a user to create a small data (called a token) that is associated with a keyword she is interested in. For security, the association is secret; that is, the token itself doesn't reveal what keyword(s) it embeds.

Given this token, the cloud server can run the search algorithm over the ciphertexts to determine which ciphertexts have the secretly associated keyword(s). Only the encrypted containing the keyword will be sent back to the receiver. This way, our system provides keyword search with significantly reduced network bandwidth and client-side computation costs compared with the previous work [17].

We have implemented the m2-ABKS scheme [24] that allows users to search encrypted data in an efficient way.

Edge computing. Through conforming to edge computing [28] standards, we deliver a workaround. Edge computing relates to the concept of needing to run data processing before transferring it to the cloud. Likewise, we enact an access control function on data within the organizational perimeter, which we describe as the 'edge' in our framework. This means that users are verified inside the company boundaries, which maintains their identity's obscurity. We have also introduced a rigorous encryption method within the organizational frontier that safeguards data security against privacy threats before transferring it to the cloud. The institutional border, therefore, proves to be a potent edge in protecting data privacy.

Threat model. We require the system to be robust against any corrupted user who tries to gain information on the ciphertext that she can't decrypt with her decryption key. Further, we require the system to be robust even against a coalition of users who try to gain information about the ciphertext that no individual member of the coalition can decrypt with her own decryption key alone.

Cloud users, once placing their files in the cloud, typically classify cloud vendors through two major adversary models, the honest-but-curious adversary model, and the malicious adversary model [23]. We assume that the cloud server can be corrupted in an honest-but-curious manner. The cloud server runs the programs and algorithms accurately. Still, it could try to gain information on the encrypted data either by careful inspection or by observing the network communication between users.

B. Organization

The remainder of the paper is structured as follows – We discuss related work in Section II, preliminaries in section III, our system architecture in Section IV, and our system implementation in Section V. Future works and conclusions are summarized in Section VI and VII, respectively.

II. Related Work

A. Searchability

Given the short time frame is provided to the doctors when they make the decisions, fast and efficient searchability is needed for every health information system, especially in the trend of evidence-based medicine. Dawes and Sampson [10] reviewed ninety studies on information-seeking behavior in clinical practice. They pointed out that the lack of time is the primary factor that hinders the usability of information systems in clinical settings.

Holden [14] conducted twenty research interviews with physicians using EHR systems. Physicians reported that system response speed is one of the barriers to EHR uses. In this background, the demands of data encryption increased in EHRs' big transition to cloud-based storage, and usability decrease problem is raised because there is a lack of practical data management solutions for encrypted data, including searchability [29]. Thus, we addressed the searchability issues as the next step of our previous research.

B. Policy Compliance

In the United States, personal health information is protected by various laws, the most representative of which is the Health Insurance Portability and Accountability Act (HIPAA) since 1996. The health data protected by these laws are called electronic protected health information (ePHI) [8]. The Health Information Technology for Economic and Clinical Health Act (HITECH Act) facilitates the exchange of ePHI while allowing the broader and more rigorous application of HIPAA privacy and security rules [1].

However, these regulations do not specify encryption standards or algorithms. Also, encryption of data in data access control and transmission is designated as addressable, not required. This created room for various interpretations and became a source of contention when sharing ePHI.

C. Attribute Based Encryption

ABE [13] is one of the solutions to preserving data protection and attacks. It was identified as one of the EHR system security innovations [2], [5], [25]. ABE may be divided into key-policy ABE (KP-ABE) [3] on CP-ABE [6], depending on where the policy lies. In CP-ABE, the policy lies in a ciphertext, and each user receives a set of attributes according to her access credentials. The policy is usually described as a Boolean formula with a particular set of attributes as input. A secure CP-ABE scheme guarantees that a user can successfully decrypt a given ciphertext only when the user's attributes satisfy the ciphertext policy. Because an individual ciphertext can specify a policy that designates attributes that data users need to possess for decryption, CP-ABE is deemed more appropriate for access management in the cloud setting.

Joshi et al. [17] used CP-ABE to establish attributed-based access control (ABAC) that is semantically rich in data access. The model assesses classified access choices dependent on the user attributes and EHR's fields. The model does not enforce overhead access regulation on the patient. The central system manages both EHR safe access and allocation.

D. Searchable Encryption

Encryption has arisen as a decisive method for preserving data security and anonymity to protect personal privacy. The conception of valid and reliable information retrieval procedures over encrypted data are of paramount concern as encryption greatly obsoletes the conventional private information retrieval over plaintext.

The traditional situation, followed by several searchable encryption schemes [9], [19], [30], [18], is the outsourced cloud storage under which there are three parties: data owner, data user, and the cloud service provider (CSP). The data owner initially encrypts confidential data, building indexes, then outsourcing ciphertext and indexes to CSP. Data user needs creating a search token and apply it to CSP for the desired keyword. After this, CSP conducts a ciphertext scan and displays related data.

Boneh et al. [7] introduced the public key encryption with the keyword search system. Subsequently, several public key SE schemes were introduced that rely on a single keyword search [9], [19] or multi-keyword search [30], [18], [12]. Though, such systems are also only valid in single-user situations, which are, in fact, inefficient and unscalable.

In this paper, we implement the m2-ABKS [24] scheme concentrating on the multi-owner case to enable an attribute-based multi-keyword search using the CP-ABE scheme over encrypted files.

III. Preliminaries

In this section, we describe CP-ABE and attribute-based searchable encryption schemes.

A. CP-ABE

A ciphertext-policy ABE consists of four algorithms $(\mathsf{Setup}, \mathsf{Enc}, \mathsf{KeyGen}, \mathsf{Dec})$:

- $\mathsf{Setup}(1^\lambda, \mathcal{X}, \mathcal{M}) \to (\mathsf{mpk}, \mathsf{msk})$. The setup algorithm gets as input the security parameter λ, the attribute universe $\mathcal{X}$, and the message space $\mathcal{M}$. It outputs the public parameter mpk and the master secret key msk.
- $\mathsf{KeyGen}(\mathsf{mpk}, \mathsf{msk}, x) \to \mathsf{sk}_x$. The key generation algorithm gets as input msk and attributes x. It outputs a secret key sk_x.
- $\mathsf{Enc}(\mathsf{mpk}, f, m) \to \mathsf{ct}_f$. The encryption algorithm gets as input mpk, and a boolean formula f over $\mathcal{X}$, and a message $m \in \mathcal{M}$. It outputs a cipehrtext ct_f.
- $\mathsf{Dec}(\mathsf{mpk}, \mathsf{sk}_x, \mathsf{ct}_f) \to m$. The decryption algorithm gets as input sk_x and ct_f such that $f(x) = 1$. It outputs a message m.

Correctness. We require for all $x \in \mathcal{X}$ and all Boolean formula f over $\mathcal{X}$ such that $f(x) = 1$, and all $m \in \mathcal{M}$,

$$\Pr[\mathsf{Dec}(\mathsf{mpk}, \mathsf{sk}_x, \mathsf{Enc}(\mathsf{mpk}, f, m)) = m] = 1,$$

where the probability is taken over $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Setup}(1^\lambda, \mathcal{X}, \mathcal{M})$ and $\mathsf{sk}_x \leftarrow \mathsf{KeyGen}(\mathsf{mpk}, \mathsf{msk}, x)$, and the random coins of Enc.

Security. For a stateful adversary A and security parameter λ, we define an experiment $\mathsf{Expt}_A^{\text{ABE}}(\lambda)$ as follows:

```
$\mathsf{Expt}_A^{\text{ABE}}(\lambda)$:
    $f^* \leftarrow A(1^\lambda)$;
    $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Setup}(1^\lambda, \mathcal{X}, \mathcal{M})$;
    $(m_0, m_1) \leftarrow A^{\mathsf{KeyGen}(\mathsf{msk}, \cdot)}(\mathsf{mpk})$;
    $b \leftarrow_R \{0, 1\}$;
    $ct_{f^*} \leftarrow \mathsf{Enc}(\mathsf{mpk}, f^*, m_b)$;
    $b' \leftarrow A^{\mathsf{KeyGen}(\mathsf{msk}, \cdot)}(\mathsf{ct}_{f^*})$
    If $b = b'$ output 1; otherwise output 0.
```

In the above, all queries x that A makes to oracle $\mathsf{KeyGen}(\mathsf{msk}, \cdot)$ should satisfy $f^*(x) \neq 1$.

A CP-ABE scheme ABE is said to be *selectively secure* if for all polynomial adversary A, $|\Pr[\mathsf{Expt}_A^{\text{ABE}}(\lambda)] - 1/2|$ is negligible in λ.

In this work, we use a CP-ABE scheme from [6] that satisfies the above security requirement.

B. Attribute-based Searchable Encryption

Let $\mathcal{X}$ be the attribute universe and $W = \{w_1, \ldots, w_m\}$ be the keyword dictionary, $D = \{d_1, \ldots, d_p\}$ be the record set. An attribute-based searchable encryption consists of the following algorithms:

- $\mathsf{Setup}(1^\lambda, \mathcal{X}) \leftarrow (\mathsf{mpk}, \mathsf{msk})$. As with CP-ABE, the setup algorithm creates the public parameter mpk and the master secret key msk.
- $\mathsf{KeyGen}(\mathsf{mpk}, \mathsf{msk}, x) \to \mathsf{sk}_x$. As with CP-ABE, the key generation algorithm gets as input msk and attributes x. It outputs a secret key sk_x.
- $\mathsf{EncInd}(\mathsf{mpk}, f, W) \to I_f$. The encrypted index algorithm gets as input mpk, and a boolean formula f over $\mathcal{X}$, and a set of keywords W. It outputs an encrypted index I_f for W.
- $\mathsf{Token}(\mathsf{sk}_x, W') \to \mathsf{t}_x$. The token generation algorithm gets as input sk_x and a set of query keywords W'. It outputs a token t_x.
- $\mathsf{Search}(\mathsf{mpk}, I_f, \mathsf{t}_x) \to 0/1$. The search algorithm gets as input mpk, I_f and t_x. If $f(x) = 1$ and the embedded keywords in t_x and I_f match, it outputs true; otherwise it outputs false.
  Note that this algorithm can be performed without the master secret key msk or a decryptor's secret key sk_x. Therefore, a third party cloud server can run this algorithm.

Correctness. We require for all $x \in \mathcal{X}$ and all Boolean formula f over $\mathcal{X}$ such that $f(x) = 1$, and all set of keywords W,

$$\Pr[\mathsf{Search}(\mathsf{mpk}, \mathsf{EncInd}(\mathsf{mpk}, f, W), \mathsf{Token}(\mathsf{sk}_x, W)) = 1] = 1,$$

where the probability is taken over $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Setup}(1^\lambda, \mathcal{X}, \mathcal{M})$ and $\mathsf{sk}_x \leftarrow \mathsf{KeyGen}(\mathsf{mpk}, \mathsf{msk}, x)$, and the random coins of EncInd and Token.

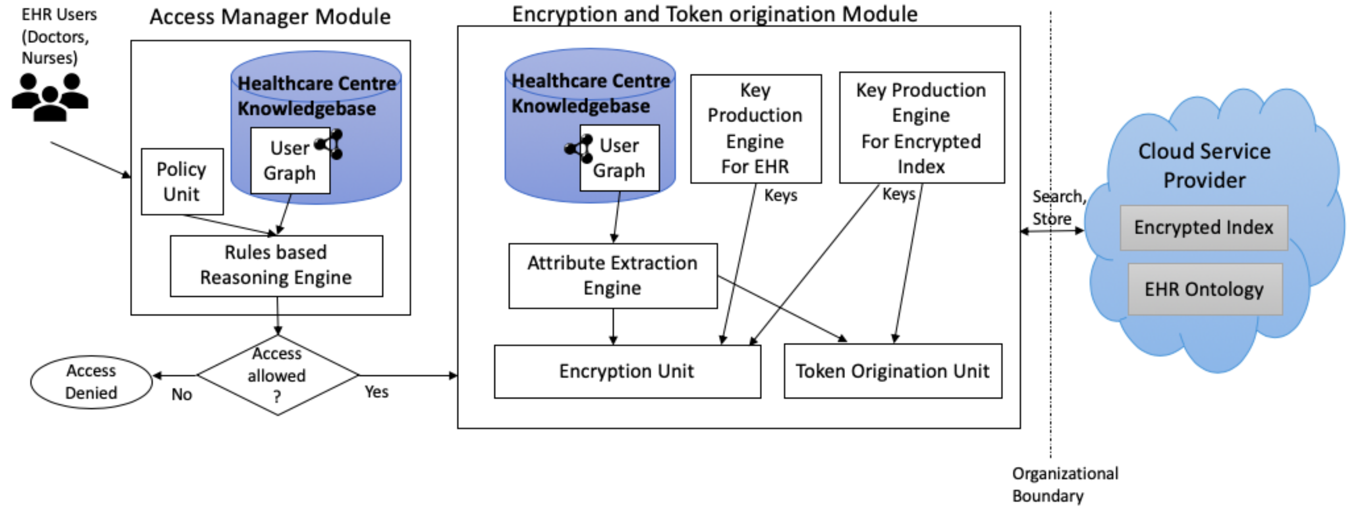

EHR Users (Doctors, Nurses)
Access Manager Module
Healthcare Centre Knowledgebase
User Graph
Policy Unit
Rules based Reasoning Engine
Access allowed ?
No
Yes
Access Denied
Encryption and Token origination Module
Healthcare Centre Knowledgebase
User Graph
Key Production Engine For EHR
Key Production Engine For Encrypted Index
Keys
Keys
Attribute Extraction Engine
Encryption Unit
Token Origination Unit
Search, Store
Cloud Service Provider
Encrypted Index
EHR Ontology
Organizational Boundary

Fig. 1. System Architecture

Security. For a stateful adversary A and security parameter λ, we define an experiment $\mathsf{Expt}_A^{\mathrm{ABKS}}(\lambda)$ as follows:

```
$\mathsf{Expt}_A^{\mathrm{ABKS}}(\lambda)$:
    $f^* \leftarrow A(1^\lambda)$;
    $(\mathsf{mpk}, \mathsf{msk}) \leftarrow \mathsf{Setup}(1^\lambda, \mathcal{X}, \mathcal{M})$;
    $\mathsf{sk}_{f^*} \leftarrow \mathsf{KeyGen}(\mathsf{msk}, f^*)$;
    $(W_0, W_1) \leftarrow A^{\mathsf{KeyGen}(\mathsf{msk},\cdot),\mathsf{Token}(\mathsf{sk}_{f^*},\cdot)}(\mathsf{mpk})$;
    $b \leftarrow_R \{0, 1\}$;
    $I_{f^*} \leftarrow \mathsf{EncInd}(\mathsf{mpk}, f^*, W_b)$;
    $b' \leftarrow A^{\mathsf{KeyGen}(\mathsf{msk},\cdot),\mathsf{Token}(\mathsf{sk}_{f^*},\cdot)}(\mathsf{ct}_{f^*})$
    If $b = b'$ output 1; otherwise output 0.
```

In the above, all queries x that A makes to oracle $\mathsf{KeyGen}(\mathsf{msk}, \cdot)$ should satisfy $f^*(x) \neq 1$. In addition, all queries W to Token should satisfy $W \notin \{W_0, W_1\}$.

An attribute-based searchable encryption ABKS is said to be *selectively secure*, if for all polynomial adversary A $|\Pr[\mathsf{Expt}_A^{\mathrm{ABKS}}(\lambda)] - 1/2|$ is negligible in λ.

In this paper, we use the m2-ABKS [24] scheme that satisfies the above security requirement.

IV. System Architecture

We describe our overall EHR system architecture (Figure 1) and various sub-modules in detail.

Participants and system components. Our system comprises multiple users, authorities, and data owners coming from a broad medical universe. A single cloud service provider (CSP) stores all EHR records and the encrypted indexes.

Any login request to the system passes a detailed check using the Access Manager module. Each user gets access privileges depending on their attributes checked with the organization policy. Patients are the data owners, so they have complete read access to their EHR records. An adversary can be undermining the cloud service provider as well. In this scenario, we believe a compromised supplier of cloud resources would act in an honest-but-curious manner [23].

The complete structure is split into two main parts, defining the secure edge as the operational border that contains the Access Manager module and Encryption and Token Origination module. Organizations regulate these modules, so they are regarded as trusted entities. The other part involves the untrusted provider of cloud services.

Use cases. Our system has several use cases depending on whether users want to read, write, or search through the encrypted EHRs.

At first, the user requests access to the EHR management system. The Access Manager module first assesses the submission by reviewing ABAC rules specified in compliance with the private organization policy. Access is authorized if the attributes match the organization's policies.

When a user edits a field, the system automatically encrypts revised information of the accessed EHR fields using the Encryption and Token Origination module. This unit derives user attributes from the central ontology housed with the CSP. The Key Production Engine for EHR provides encryption keys for encryption. The encrypted text then updates the ontology.

In the case of a search operation, the user provides the search keyword in the form of a query. The Key Production Engine for Encrypted Index provides the keys for searching. Token Origination unit generates a trapdoor using the keyword and keys. The trapdoor is later submitted to the CSP and matched against the encrypted Indexes. If there is a match, the user gets the corresponding EHR.

We will explain each of the sub-modules in-depth in the next few sections.

```
JuniorDoctor(?jd) ^ Certification(?c) ^
SeniorDoctor(?sd) ^ HospitalWard(?hw) ^
EHR(Prescription) ^ EHR(BillingInfo) ^
EHR(DoctorNotes) ^ EHR(ImmunizationDates) ^
worksIn(?sd, ?hw) ^ worksIn(?jd, ?hw) ^
isCertifiedBy(?jd, ?c) ^ reportsTo(?jd, ?sd) ^
canModifyPrescription(?sd, true) ^
canModifyBillingInfo(?sd, true) ^
canModifyDoctorNotes(?sd, true) ^
canModifyImmunizationDates(?sd, true) ->
canReadPrescription(?jd, true) ^
canReadBillingInfo(?jd, true) ^
canModifyDoctorNotes(?jd, true) ^
canModifyImmunizationDates(?jd, true)
```

Fig. 2. A SWRL rule for a junior physician with a partial access

A. Access Manager Module

ABAC is the fundamental principle behind this module. There are three sub-modules, Healthcare Centre Knowledge Base, Rule-Based Reasoning Engine, and the Policy Unit, within this module:

- The Healthcare Centre Knowledge Base sub-module uses an ontology to preserve all details of every individual belonging to the healthcare organization.
- Policy Unit holds the control policies.
- Rule-Based Engine utilizes Semantic Web Rule Language (SWRL) to leverage the security rules to execute access management decisions. Rule-Based Engine derives the semantics of the user and document from the Healthcare Centre Knowledge Base, which is present in the form of an ontology, to give access decisions.

Example: A junior physician. Figure 2 shows an example SWRL rule for a junior physician with access to EHR fields based on the access of a senior physician to whom the junior physician reports in the same hospital ward.

The Junior Doctor has partial access, i.e., modify access to two fields in the EHR, which are Allergies and Doctor Notes. The rest of the fields are only for reading. Often the access fields of a doctor vary within the various department. A junior doctor in a different department may have access to limited fields but may have full control.

Figure 3 shows an example SWRL rule for a junior doctor with full access to EHR fields based on the access of a senior physician to whom the junior physician reports in the same hospital ward.

B. Encryption and Token Origination Module

As shown in Figure 1, this module contains several sub-modules including the Healthcare Centre Knowledge Base, Attribute Extraction Engine, Encryption Unit, Key Production Engine for EHR, Key Production Engine for Encrypted Index, and Token Origination Unit.

The main principles underlying this module ABE [6] and SE [24]. Our system runs CP-ABE [6] and m2-ABKS [24] in parallel. We implemented CP-ABE scheme [6] for encrypting and decrypting EHRs. m2-ABKS scheme [24] allowed searching through the encrypted EHR in our framework.

```
JuniorDoctor(?jd) ^ Certification(?c) ^
SeniorDoctor(?sd) ^ HospitalWard(?hw) ^
EHR(Prescription) ^ EHR(BillingInfo) ^
EHR(Medication) ^ EHR(DoctorNotes) ^
EHR(Allergies) ^ EHR(Diagnoses) ^
worksIn(?sd, ?hw) ^ worksIn(?jd, ?hw) ^
isCertifiedBy(?jd, ?c) ^ reportsTo(?jd, ?sd) ^
canModifyPrescription(?sd, true) ^
canModifyBillingInfo(?sd, true) ^
canModifyMedication(?sd, true) ^
canModifyDoctorNotes(?sd, true) ^
canModifyAllergies(?sd, true) ^
canModifyDiagnoses(?sd, true) ->
canModifyPrescription(?jd, true) ^
canModifyBillingInfo(?jd, true) ^
canModifyMedication(?jd, true) ^
canModifyDoctorNotes(?sd, true) ^
canModifyAllergies(?sd, true) ^
canModifyDiagnoses(?jd, true)
```

Fig. 3. A SWRL rule for a junior physician with the full access

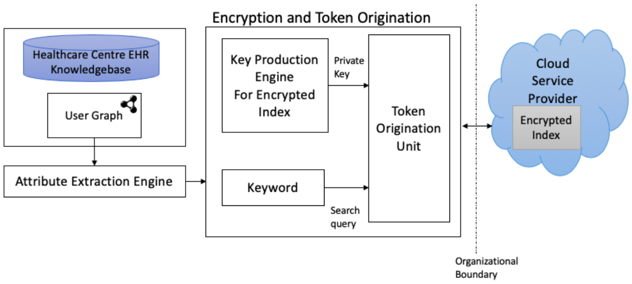

Encryption and Token Origination
Healthcare Centre EHR Knowledgebase
User Graph
Attribute Extraction Engine
Key Production Engine For Encrypted Index
Private Key
Token Origination Unit
Keyword
Search query
Cloud Service Provider
Encrypted Index
Organizational Boundary

Fig. 4. Token generation process

The attribute extraction engine queries the EHR ontology to obtain user attributes and EHR fields and authorization to read or write data. Data users' attributes combine with the master key and public key to produce separate private keys for EHR and Encrypted Index. Private keys for EHR operate as the encryption/decryption keys to protect the EHR, and Private keys for Encrypted Index allows searching through the encrypted data. After editing an EHR field, a new node is built in the ontology and the corresponding EHR field and stored within the cloud server.

There are separate Key Production Engines that provide distinct keys for encrypting, decrypting, or searching through encrypted EHRs. In the case of a search operation, the user provides the search keyword in the form of a query. The Key Production Engine for Encrypted Index provides the keys for searching.

The token Origination unit generates a trapdoor using the keyword and keys, as shown in Figure 4. The trapdoor is a query for searching through the encrypted index. The trapdoor is later submitted to the CSP and matched against the encrypted Indexes. If there is a match, the user gets the corresponding EHR. If a user wants to decrypt a particular EHR fetched by the CSP, it obtains the decryption keys from the Key Production Engine for EHR.

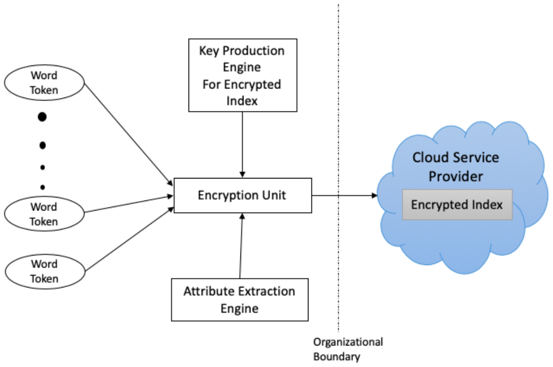


Fig. 5. Encrypted index generation process

C. EHR Ontology

The EHR ontology is designed following accordance with HIPAA [16] privacy and security rules. As an information graph, it records user and EHR attributes. The attributes of the EHR include Allergies, Billing Information, Prescription, Diagnosis, Doctor Notes, Prescription, Medication, Laboratory Results, and Dates of Immunization. Various users with specific designations in separate hospital wards can have multiple kinds of access to the EHR fields, which are stored within the ontology. The ontology outlines the functions and characteristics of members from medical organizations and their varying relations.

D. Encrypted Index

The encrypted index is a file stored within the cloud server. The file contains encrypted word tokens from each patient EHR and the unique patient id. Word token from each patient EHR extracted is encrypted using encrypted keys, and user attributes following the m2-ABKS scheme [24]. The Key Production Engine supplies the encryption keys for generating the Encrypted Index. User attributes obtained with the aid of the Attribute Extraction Engine from the EHR ontology. The procedure was repeated on all patients in the system, as shown in Figure 5. The encrypted word tokens and the individual patient ids for each patient form a data frame recorded in the encrypted index file that is stored in the CSP.

E. Cloud Service Provider

The cloud service provider reflects a central cloud computing platform. It remains beyond the organization's boundaries, containing the EHR ontology and the encrypted index. The EHR ontology includes all the descriptions of the multiple users and records as instances of groups established in the medical domain ontology. The ontology complies with HIPAA privacy and security rules [16]. The encrypted EHR data is stored on the cloud platform as nodes of the ontology. The encrypted index contains the encrypted token with its corresponding patient id from all patient instances.

Since we consider the honest-but-curious adversary model [23], CSP accurately runs the programs and algorithms but could look at the knowledge exchanged within and outside the organization. To tackle this, inside the organizational perimeter, we implement an access control mechanism on data, which we define as the 'edge' in our system. Therefore, users are verified inside the organization's limits, which preserves the anonymity of their identification. We have implemented a rigorous encryption process within the organizational frontier using the Encryption and Token Origination module to safeguard data protection against privacy threats before moving it to the cloud.

V. Implementation

The EHR Management Software is developed in Python using the Python Django web-based framework. It is an open-source web application that allows field-level attribute-based encryption and access control of patient EHR. The framework also enables search through encrypted EHR. The framework is based on the architectural concepts of the Model-View-Controller (MVC). EHR Management Software allows doctors, nurses, etc. to treat their patients by ensuring service securely. ABAC controls field-level access to patient EHR, ABE encrypts data to store securely within the cloud service provider, and SE allows search through the encrypted EHRs. The EHR ontology developed using the Protege [protege.stanford.edu]. The Stanford Center for Research in Biomedical Informatics developed Protege. It is an open-source editor and management framework with knowledge graphs. SPARQL with Apache Jena library queried the ontology using the open-source rdflib library. SWRL is used to modify the knowledge graph and extract inferences. The EHR Manager thus embraces the semantic web, CP-ABE, and SE to ensure a field-level EHR managing framework successfully.

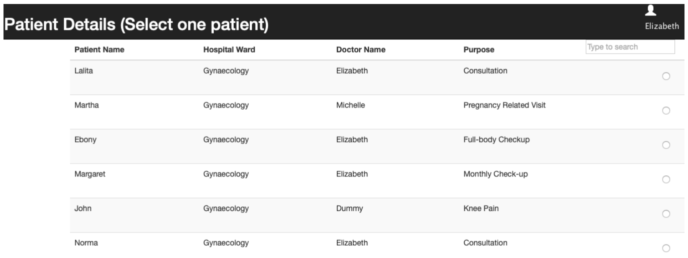
A. Dataset Description

There are 11200 patient instances in our system. Each patient has different fields like Allergies, Billing Information, Prescription, Diagnosis, Doctor Notes, Prescription, Medication, Lab Results, and Immunization Dates in their EHR that depends on the medical history. Following the edge computing idea [28], the EHRs are encrypted within the organization's limit and stored in the cloud server.

B. Proof of Concept

We developed a proof of concept detailed below for evaluating our system. Suppose Elizabeth is a doctor in our system. Elizabeth submits a login request that passes a detailed check using the Access Manager module. She qualifies the organization access policy, and the Access Manager Module grants the access decision. Elizabeth wants to search through the encrypted EHRs to find patients with coronavirus symptoms. So she wants to find patients having a fever. Figure 4 depicts the prototype interface of such a scenario.

We describe another prototype below. Lalita is a patient in our system, and Elizabeth is an authorized user, and her attributes match Lalita's EHR access policy. Elizabeth gets



Patient Details (Select one patient)

Patient Name	Hospital Ward	Doctor Name	Purpose	
Lalita	Gynaecology	Elizabeth	Consultation	○
Martha	Gynaecology	Michelle	Pregnancy Related Visit	○
Ebony	Gynaecology	Elizabeth	Full-body Checkup	○
Margaret	Gynaecology	Elizabeth	Monthly Check-up	○
John	Gynaecology	Dummy	Knee Pain	○
Norma	Gynaecology	Elizabeth	Consultation	○

Fig. 6. Interface of Dr. Elizabeth after login to the system.

the decryption keys from the Key Production Engine for EHR and decrypts the EHR using the Encryption Unit. Elizabeth receives different types of access, read or write, to various EHR fields. Figure 5 depicts the prototype interface of such a scenario.

We have measured the time it takes for a token to be produced from the Token Origination Unit to prove that our scheme is fast enough. On average, it just requires 1.36 seconds.

VI. Future works

We plan to use a single scheme for encrypting or decrypting EHR and searching through encrypted EHR, eliminating the encrypted index file. We intend to increase patient instances and perform extensive performance evaluation using standard benchmarks to show our system's usability and performance. The user attributes keep on changing with time, so we hope to include attribute revocation. Another dominant direction for us could be exploring the best practices for storing encrypted indexes and EHR.

VII. Conclusion

In this paper, we developed a framework that allows field-level attribute-based encryption and patient EHR access control. Owing to the large volume of data in the cloud, the doctors may need to search through encrypted data. So, our framework also enables search through encrypted EHR. ABAC controls field-level access to patient EHR, ABE encrypts data to store securely within the cloud service provider, and SE allows search through the encrypted EHRs. Our methodology created and produced a comprehensive knowledge graph that depicts the various stakeholders' roles and attributes in the medical organization and the specific relationships between them. We stored the ontology and the encrypted index within the CSP, considering the HBC adversary model [23]. We implement an access control mechanism on data inside the organizational perimeter, which we define as the 'edge' in our system. Therefore, users are authenticated inside the organization's limits, which preserves the anonymity of their identification. We have implemented a rigorous encryption process within the organizational frontier using the Encryption and Token Origination module to safeguard data protection against privacy threats before moving it to the cloud.

Acknowledgment

This work has been supported by Office of Naval Research under grants N00014-18-1-2453 and N00014-19-WX-00568.

References

[1] Health Information Technology for Economic and Clinical Health (HITECH) Act, Title XIII of Division A and Title IV of Division B of the American Recovery and Reinvestment Act of 2009 (ARRA), Pub. L. No. 111-5, 123 Stat. 226 (Feb. 17, 2009), codified at 42 U.S.C. §§300jj et seq.; §§17901 et seq.

[2] Joseph A Akinyele, Matthew W Pagano, Matthew D Green, Christoph U Lehmann, Zachary NJ Peterson, and Aviel D Rubin. Securing electronic medical records using attribute-based encryption on mobile devices. In *Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices*, pages 75–86, 2011.

[3] Nuttapong Attrapadung, Benoît Libert, and Elie De Panafieu. Expressive key-policy attribute-based encryption with constant-size ciphertexts. In *International Workshop on Public Key Cryptography*, pages 90–108. Springer, 2011.

[4] Arshdeep Bahga and Vijay K Madisetti. A cloud-based approach for interoperable electronic health records (ehrs). *IEEE Journal of Biomedical and Health Informatics*, 17(5):894–906, 2013.

[5] Josh Benaloh, Melissa Chase, Eric Horvitz, and Kristin Lauter. Patient controlled encryption: ensuring privacy of electronic medical records. In *Proceedings of the 2009 ACM workshop on Cloud computing security*, pages 103–114, 2009.

[6] John Bethencourt, Amit Sahai, and Brent Waters. Ciphertext-policy attribute-based encryption. In *2007 IEEE symposium on security and privacy (SP'07)*, pages 321–334. IEEE, 2007.

[7] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, and Giuseppe Persiano. Public key encryption with keyword search. In *International conference on the theory and applications of cryptographic techniques*, pages 506–522. Springer, 2004.

[8] Centers for Medicare & Medicaid Services. The Health Insurance Portability and Accountability Act of 1996 (HIPAA). Online at http://www.cms.hhs.gov/hipaa/, 1996.

[9] Reza Curtmola, Juan Garay, Seny Kamara, and Rafail Ostrovsky. Searchable symmetric encryption: improved definitions and efficient constructions. *Journal of Computer Security*, 19(5):895–934, 2011.

[10] Martin Dawes and Uchechukwu Sampson. Knowledge management in clinical practice: a systematic review of information seeking behavior in physicians. *International journal of medical informatics*, 71(1):9–15, 2003.

[11] Centers for Disease Control, Prevention, et al. Hipaa privacy rule and public health. guidance from cdc and the us department of health and human services. *MMWR: Morbidity and mortality weekly report*, 52(Suppl. 1):1–17, 2003.

[12] Zhangjie Fu, Xingming Sun, Qi Liu, Lu Zhou, and Jiangang Shu. Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing. *IEICE Transactions on Communications*, 98(1):190–200, 2015.

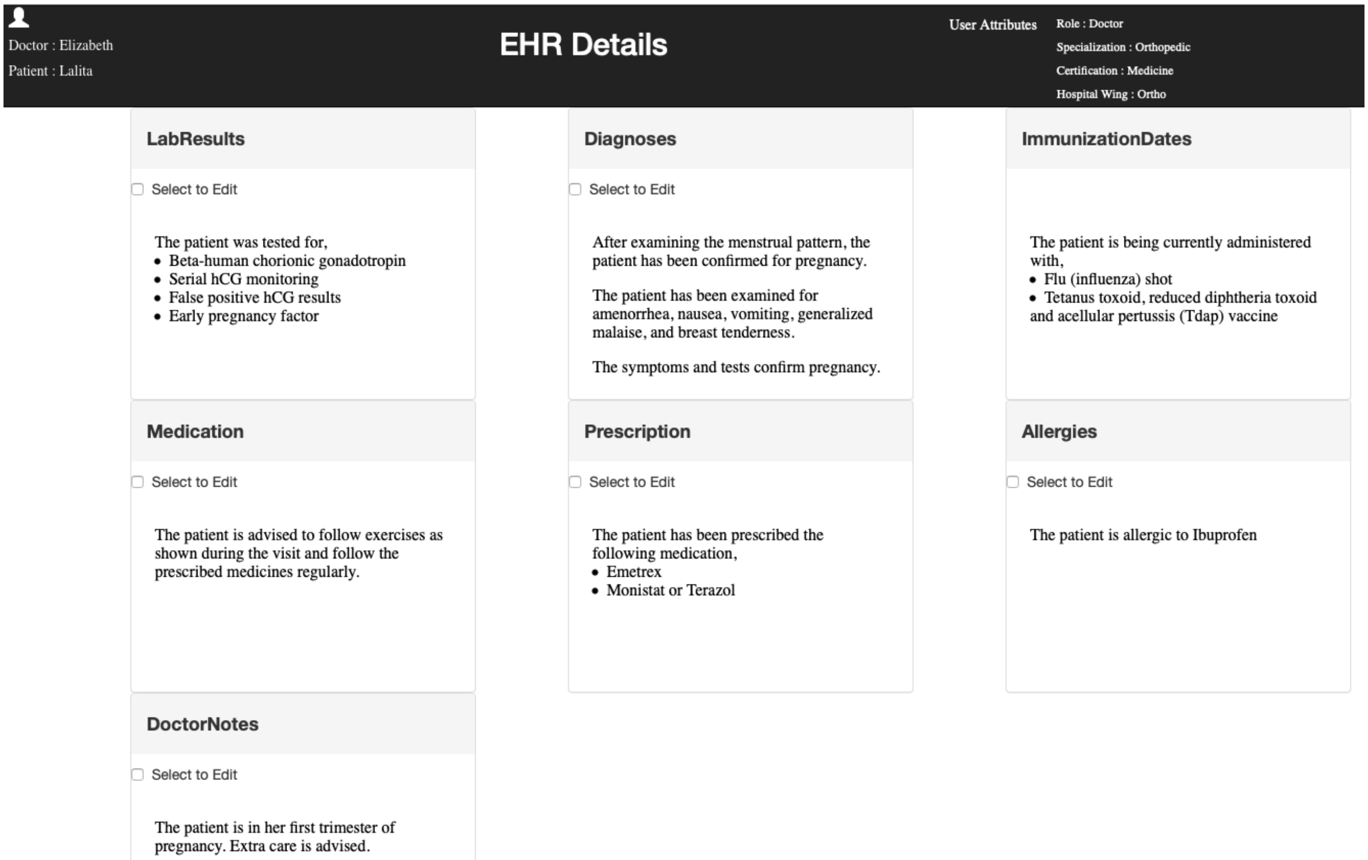
[13] Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM conference on Computer and communications security*, pages 89–98, 2006.

[14] Richard J Holden. What stands in the way of technology-mediated patient safety improvements? a study of facilitators and barriers to physicians' use of electronic health records. *Journal of patient safety*, 7(4):193, 2011.

[15] Xin Jin, Ram Krishnan, and Ravi Sandhu. A unified attribute-based access control model covering dac, mac and rbac. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 41–55. Springer, 2012.

[16] Karuna Pande Joshi, Yelena Yesha, Tim Finin, et al. An ontology for a hipaa compliant cloud service. In *4th International IBM Cloud Academy Conference ICACON 2016*, 2016.

[17] Maithilee Joshi, Karuna Joshi, and Tim Finin. Attribute based encryption for secure access to cloud based ehr systems. In *2018 IEEE 11th International Conference on Cloud Computing (CLOUD)*, pages 932–935. IEEE, 2018.



Doctor : Elizabeth
Patient : Lalita

EHR Details

User Attributes
Role : Doctor
Specialization : Orthopedic
Certification : Medicine
Hospital Wing : Ortho

LabResults

☐ Select to Edit

The patient was tested for,
- Beta-human chorionic gonadotropin
- Serial hCG monitoring
- False positive hCG results
- Early pregnancy factor

Diagnoses

☐ Select to Edit

After examining the menstrual pattern, the patient has been confirmed for pregnancy.

The patient has been examined for amenorrhea, nausea, vomiting, generalized malaise, and breast tenderness.

The symptoms and tests confirm pregnancy.

ImmunizationDates

The patient is being currently administered with,
- Flu (influenza) shot
- Tetanus toxoid, reduced diphtheria toxoid and acellular pertussis (Tdap) vaccine

Medication

☐ Select to Edit

The patient is advised to follow exercises as shown during the visit and follow the prescribed medicines regularly.

Prescription

☐ Select to Edit

The patient has been prescribed the following medication,
- Emetrex
- Monistat or Terazol

Allergies

☐ Select to Edit

The patient is allergic to Ibuprofen

DoctorNotes

☐ Select to Edit

The patient is in her first trimester of pregnancy. Extra care is advised.

Fig. 7. Dr. Elizabeth view to different types of access to various EHR fields.

[18] Hongwei Li, Dongxiao Liu, Yuanshun Dai, Tom H Luan, and Xuemin Sherman Shen. Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage. *IEEE Transactions on Emerging Topics in Computing*, 3(1):127–138, 2014.

[19] Jin Li, Qian Wang, Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou. Fuzzy keyword search over encrypted data in cloud computing. In *2010 Proceedings IEEE INFOCOM*, pages 1–5. IEEE, 2010.

[20] Ming Li, Shucheng Yu, Ning Cao, and Wenjing Lou. Authorized private keyword search over encrypted data in cloud computing. In *2011 31st International Conference on Distributed Computing Systems*, pages 383–392. IEEE, 2011.

[21] Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou. 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, 2012.

[22] Hans Löhr, Ahmad-Reza Sadeghi, and Marcel Winandy. Securing the e-health cloud. In *Proceedings of the 1st acm international health informatics symposium*, pages 220–229, 2010.

[23] Tim Mather, Subra Kumaraswamy, and Shahed Latif. *Cloud security and privacy: an enterprise perspective on risks and compliance*. " O'Reilly Media, Inc.", 2009.

[24] Yinbin Miao, Jianfeng Ma, Ximeng Liu, Fushan Wei, Zhiquan Liu, and Xu An Wang. m 2-abks: Attribute-based multi-keyword search over encrypted personal health records in multi-owner setting. *Journal of medical systems*, 40(11):246, 2016.

[25] Shivaramakrishnan Narayan, Martin Gagné, and Reihaneh Safavi-Naini. Privacy preserving ehr system using attribute-based infrastructure. In *Proceedings of the 2010 ACM workshop on Cloud computing security workshop*, pages 47–52, 2010.

[26] Rishi Kanth Saripalle. Fast health interoperability resources (fhir): Current status in the healthcare system. *International Journal of E-Health and Medical Communications (IJEHMC)*, 10(1):76–93, 2019.

[27] Matthew A Scholl, Kevin M Stine, Joan Hash, Pauline Bowen, L Arnold Johnson, Carla Dancy Smith, and Daniel I Steinberg. Sp 800-66 rev. 1. an introductory resource guide for implementing the health insurance portability and accountability act (hipaa) security rule, 2008.

[28] Weisong Shi, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. Edge computing: Vision and challenges. *IEEE internet of things journal*, 3(5):637–646, 2016.

[29] Dawn Xiaoding Song, D. Wagner, and A. Perrig. Practical techniques for searches on encrypted data. *Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000*, pages 44–55, 2000.

[30] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y Thomas Hou, and Hui Li. Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking. In *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security*, pages 71–82, 2013.

[31] H. Yin, J. Zhang, Y. Xiong, L. Ou, F. Li, S. Liao, and K. Li. Cp-abse: A ciphertext-policy attribute-based searchable encryption scheme. *IEEE Access*, 7:5682–5694, 2019.