

# Semantically Rich Differential Access to Secure Cloud EHR

Redwan Walid

Department of Information Systems  
University of Maryland, Baltimore County  
Baltimore, MD, US  
rwalid1@umbc.edu

Karuna P. Joshi

Department of Information Systems  
University of Maryland, Baltimore County  
Baltimore, MD, US  
karuna.joshi@umbc.edu

Seung Geol Choi

Department of Computer Science  
United States Naval Academy  
Annapolis, MD, US  
choi@usna.edu

**Abstract**—Existing Cloud-based Electronic Health Record (EHR) services face challenges in handling heterogeneous data and maintaining performance with large records since they often use a relational database or only partially store information in a graph database. We have developed a novel approach that allows fine-grained field-level security for Cloud EHRs to protect patient privacy and data security. Our graph-based EHR has been developed by integrating Attribute-based Encryption (ABE) with ontology reasoning using Semantic Web technologies. The novelty of our approach lies in providing differential access to an EHR by using a comprehensive knowledge graph that stores all medical data as encrypted nodes, thereby handling heterogeneous patient data while preserving good performance. In this paper, we describe our system in detail, along with the results demonstrating that the system maintains consistent data retrieval performance with different data sizes and allows real-time updates on the data while supporting queries.

**Index Terms**—Attribute-Based Encryption (ABE), Attribute-Based Access Control (ABAC), Searchable Encryption (SE), Attribute Revocation, Electronic Health Record (EHR), Knowledge Graph (Ontology), Cloud Security, Cloud Computing

## I. INTRODUCTION

Medical organizations continually turn to cloud-based solutions to maintain their digital data as they allow centralizing patient data management and utilize cost-effective cloud-based storage services [4], [28], [29], [32]. Team members can collaborate more effectively in cloud settings since they can access the same infrastructure and work in parallel. Several research projects have been developed to emphasize safe, cloud-based EHR solutions [4], [29]. Despite the numerous benefits cloud services offer, they pose distinct threats to medical organizations regarding data privacy and security. Individuals whose health information gets unlawfully accessed suffer adverse consequences if their privacy is infringed. Given the security threats, all healthcare providers must adhere to the Health Insurance Portability and Accountability Act (HIPAA) [13], [40] and the Health Information Technology for Economic and Clinical Health (HITECH) [39] privacy requirements established by the Office for Civil Rights, U.S. Department of Health and Human Services. Thus, an EHR system must comply with all applicable regulations while allowing for a smooth and straightforward interchange of patient data. Failure to comply with the acts can have serious financial

consequences, including sanctions, detention, or significant fines.

While several Cloud-based EHR solutions have recently flooded the market, the majority of the proposed approaches lack guaranteeing an attribute-based access control and encryption mechanism. We have developed a cloud-based EHR management solution that guarantees secure encrypted access control and patient data security down to the field level. Healthcare providers can use our system to maintain EHRs securely at a very low operational cost.

This work extends our previous version [45] presented a cloud-based EHR system with a semantically rich, policy-driven method that assesses users' access to the system using Attribute-Based Access Control (ABAC) [18]. Semantic web technology [6] was used to design the system architecture. The system used a knowledge graph [20] containing details about each individual in the organization and their unique attributes. The knowledge graph was queried with SPARQL, and the results were utilized to extract user attributes and associated EHR fields based on the kind of request. A user's unique attributes control different access to distinct fields in an EHR. As a result, each person has unique access to a patient's EHR. The system also allowed the user to look through encrypted data using keyword searches rather than having to download and decode the data from the cloud. Moreover, the system allowed revoking user attributes by addressing their changes over time. The key features of this EHR include.

- The system handles data heterogeneity. Highly varied heterogeneous data sets exist in the medical domain that is needed to create reliable data models for better healthcare. This is an improvement over the previous version of our system [45] which was a combination of a graph database and flat files.
- The system allows flexible expansion of data schema. For instance, a doctor may get new certifications or degrees that can change his/her attributes. Moreover, all users and patients may also not have the same attribute. For example, a patient may have only one EHR field, like allergies, whereas another may have multiple EHR fields, like allergies, diagnosis, billing information, etc. It is apparent that medical users have additional attributes over time, and the system easily facilitates these changes.

- The data retrieval performance of our system remains unaffected by the number of EHR records. Data is growing daily, and providing fast patient service is essential.
- As with [45], the knowledge graph handles EHR functions such as ABAC, encryption/decryption, revocation, searchable encryption, and data storage.

The remainder of the paper is structured as follows – we discuss related work in Section II, preliminaries in section III, system architecture in Section IV, system implementation in Section V, and conclusion in Section VI.

## II. RELATED WORK

### A. Electronic Health Record System

An EHR system brings many benefits, including reliable recording, illness tracking, data exchange, statistical analysis, and more. Hospitals commonly use digital health record systems to improve services, increase clinical effectiveness, and lower deductibles [14], [22]. Security and privacy concerns have restricted the growth of the EHR system, and they have gotten a lot of attention in recent years [29], [30], [31], [38]. To safeguard the privacy of EHR data, Narayan et al. [36] suggested adopting ABE. Joshi et al. [21] presented a cloud-based EHR system that uses semantic web technologies to encrypt patient records using CP-ABE, and Wang et al. [47] proposed another solution that employs blockchain technology to secure the integrity and traceability of health data. Both systems lack searchable encryption and attribute revocation. Walid et al. [46] proposed an EHR system that allows ABE and searchable encryption using two schemes, so multiple keys must be managed. Moreover, their system does not address the attribute changes over time. Controlled access, searchable encryption, and attribute revocation are currently lacking in most cloud-based EHR systems. Furthermore, the bulk of the accessible applications uses a relational database or a flat file system, which has various drawbacks.

### B. Regulatory Policy

In the United States, patient data is protected by several laws, and the most important is the HIPAA Act [9]. The main goal of HIPAA is to preserve the privacy of individually identifiable health information. While the HITECH Act permits sharing electronic safe health information (ePHI), it also requires that HIPAA regulations be enforced [1]. There is no mention of encryption techniques in these regulations. Also, data encryption is described as addressable rather than required. This created room for multiple meanings, which became a point of contention when distributing ePHI.

### C. Semantic Web Technology

Our system's knowledge graph and reasoning were developed using semantic web technologies. These allow us to create the schema using W3C-defined languages that meet our design objectives, such as interoperability, sound semantics, and web integration. Semantic web technologies include languages for constructing ontologies and describing meta-data using these ontologies and tools for reasoning over these

descriptions, such as Resource Description Framework (RDF) [23] and Web Ontology Language (OWL) [35]. Data can be tagged with machine-understandable meta-data, enabling automated retrieval and use of appropriate contexts.

### D. Attribute-Based Encryption

ABE, introduced by Sahai and Waters [15], is considered one of the prominent security measures for EHR [2], [5], [36]. It uses a set of attributes to encode data and a distinct set of attributes to establish the private key. According to the threshold setting, the ciphertext can only be decoded if the two sets of attributes coincide. Due to a lack of expressibility, ABE has been separated into ciphertext-policy ABE (CP-ABE) [7] and key-policy ABE (KP-ABE) [3]. The ciphertext is paired with an access policy, and the secret key is associated with an attribute defined in CP-ABE. The policy is usually expressed as a Boolean expression with certain attributes. If the attributes match the access policy, a secret key can decipher a ciphertext, but in the KP-ABE scheme, the scenario is the opposite. Because each ciphertext sets a policy that specifies explicitly the properties that data users must possess for the encryption process, CP-ABE [7] is thought to be more effective for cloud authentication.

### E. Attribute-Based Encryption With Attribute Revocation

ABE architectures require an attribute revocation feature because the user's attribute might change considerably over time. Perretti et al. [37] were the first to use a timed rekeying mechanism to perform attribute revocation. The approach was later enhanced by Bethencourt et al. [7], who linked the user's private key to a single expiration period. One key factor in the most recent ABE algorithms is computational efficiency. Decoding technologies that are outsourced can help users save time and money on their computers. Green et al. [16] were the first to suggest such a scheme. The CSP, utilizing the users' key, performs most decryption operations. Zhou et al. in [52] proposed a mobile-centric data management system in which portions of the crypto operations were outsourced to the CSP without compromising sensitive data. Li et al. [25] presented an ABE system for outsourced decryption that includes thorough verification, addressing the issue of assuring the correctness of outsourced decryption for unauthorized personnel. In comparison to the scheme suggested in [17], [33], [44], [50], the scheme applied in our systems appears to be ideal.

### F. Searchable Encryption

SE is an encryption method that allows users to scan ciphertext for keywords without disclosing the keywords. The major obstacle inhibiting computer systems in medical practice, according to Dawes et al. [11], is time restrictions. Because doctors have limited time to make decisions, any EHR system must provide quick and fast searchability. Song et al. [42] developed the first practical SE technique based on symmetric cryptography, laying the groundwork for a substantial standard for keyword search on encrypted material.

Since then, many SE schemes were created to increase search efficiency, security, and functionality [8], [10], [26], [43]. In recent history, there have been several hypes around attribute-based keyword search, which combines ABE and SE features [24], [27], [48], [51].

### III. OVERVIEW OF OUR SYSTEM

*a) Multi-layer System:* We developed the EHR framework by employing a straightforward user-id/password-based authentication technique and establishing a policy-defined ABAC. All stakeholders, including caregivers, pharmacists, and patients, have access to the system. The system is split into four layers. Users can seek access to relevant EHR at layer 1. Users are authenticated at layer 2, where requested actions are assessed in light of access rules, policies, user attributes, and EHR fields. If the operation is allowed, the request is sent to layer 3, where adjustments are made to the EHR and encrypted using the user's attributes. The CSP is at layer 4 and acts like a data storage center for storing the Knowledge Graph, which details the relationships between different entities in the medical organization. It is outside the organizational border and considered untrusted, while layers 1 through 3 are inside.

*b) Encrypted data on the cloud:* As with [45], we use the revocable, searchable ABE technique proposed by Wang et al. [49]. The scheme allows for data encryption as well as searchable encryption. The data is encrypted using ABE, where the attributes are obtained by querying the knowledge graph.

*c) Knowledge Graph:* The EHR system functions such as ABAC, ABE, attribute revocation, searchable encryption, and storing data are integrated with the knowledge graph. HIPAA act has been considered while designing the graph. It holds user attributes, patient records, semantic web rule language (SWRL), medical entities, and object properties needed in an EHR system. The graph limits access to the system following the ABAC rules defined using SWRL. It provides attributes during any crypto operation and searchable encryption. Patient data is stored as encrypted nodes in the graph as a data properties to handle heterogeneity. There are other nodes for medical users like doctors, nurses, etc., and their attributes contain plaintext data that is not encrypted. Therefore, using dynamic SPARQL, each user can navigate the graph freely with the only restriction of decrypting the patient data.

*d) Threat Model and Edge Computing:* We adopt the honest-but-curious (HBC) threat model [34], where CSP runs the programs and algorithms correctly but may examine the information exchanged between entities. Our system uses the edge computing concept [41], which refers to the requirement to examine data locally before transmitting it to the cloud. The organizational perimeter is the edge, where a data access control mechanism is implemented. Users are only vetted inside the organization's boundaries, ensuring their anonymity. A powerful encryption technique within the edge is established that protects data integrity from privacy issues until it is moved to the cloud. As a result, the border remains a powerful data-protection barrier.

## IV. PRELIMINARIES

### A. Revocable, Searchable ABE

In this section, we describe revocable, searchable attribute-based encryption scheme [49] that we implement and use. Here, we will describe only the syntax of the scheme to show how this the scheme can be used for setting up the system initially, generating secret user keys, encrypting data, generating encrypted index file, generating tokens for searchable encryption, generating search results, decrypting data partially on the cloud, and fully on the user end, updating several keys like the master secret version key, master public key, secret key on the cloud, and ciphertext when the user attribute changes. For security definitions and proofs, we refer the readers to the original paper [49].

*a) Syntax:* Let  $\lambda$  be the security parameter. Let  $\mathcal{X}$  be the attribute universe. A revocable, searchable ABE consists of the following algorithms:

- $\text{Setup}(1^\lambda, \mathcal{X}) \rightarrow (\text{mpk}, \text{msk}, \text{msvk})$ . The setup algorithm gets as input the security parameter  $\lambda$ , the attribute universe  $\mathcal{X}$ . It outputs the public parameter  $\text{mpk}$ , the master secret key  $\text{msk}$ , and the master secret version key  $\text{msvk}$ .

The master secret version key will be updated when users or attributes are revoked through algorithm  $\text{Update-msvk}$  described below.

- $\text{KeyGen}(\text{msk}, \text{msvk}, x) \rightarrow (\text{sk}_x^1, \text{sk}_x^2)$ . The key generation algorithm gets as input  $\text{msk}$ ,  $\text{msvk}$  and a set of attributes  $x$ . It outputs a pair of secret keys  $(\text{sk}_x^1, \text{sk}_x^2)$ . The first key  $\text{sk}_x^1$  will be sent to the user, and the second key  $\text{sk}_x^2$  will be stored on the cloud server.
  - $\text{Enc}(\text{mpk}, \text{msk}, f, m) \rightarrow \text{ct}_f$ . The encryption algorithm gets as input  $\text{mpk}$ ,  $\text{msk}$ , and a boolean formula  $f$  over  $\mathcal{X}$ , and a message  $m$ . It outputs a ciphertext  $\text{ct}_f$ .
  - $\text{EncInd}(\text{mpk}, W) \rightarrow I_W$ . The encrypted index algorithm gets as input  $\text{mpk}$ , and a set of keywords  $W$ . It outputs an encrypted index  $I_W$  for  $W$ .
  - $\text{Token}(\text{sk}_x^1, w) \rightarrow \text{t}_w$ . The token generation algorithm gets as input the user secret key  $\text{sk}_x^1$  and a query keyword  $w$ . It outputs a token  $\text{t}_w$ .
  - $\text{Test}(\text{sk}_x^2, I_W, \text{t}_w) \rightarrow 0/1$ . The test algorithm gets as input the cloud secret key  $\text{sk}_x^2$ , the encrypted index  $I_W$  and the user generated token  $\text{t}_w$ . If the embedded keyword in  $\text{t}_w$  is contained in  $I_W$ , it outputs true; otherwise it outputs false.
- Note that this algorithm can be performed by the cloud that holds the key  $\text{sk}_x^2$  when it receives the token  $\text{t}_w$  for the user; the encrypted index  $I_W$  is typically stored on the cloud server.
- $\text{Decrypt-cloud}(\text{sk}_x^2, \text{ct}_f) \rightarrow \text{pd}$ . This algorithm gets as input the cloud secret key  $\text{sk}_x^2$  and the ciphertext  $\text{ct}_f$ . If  $f(x) = 1$ , it outputs partial decryption  $\text{pd}$ ; otherwise, it outputs an error.
  - $\text{Decrypt-user}(\text{sk}_x^1, \text{pd}) \rightarrow m$ . Given the partial decryption, the user with  $\text{sk}_x^1$  will recover the message  $m$ .

- Update-msvk(msvk,  $x$ )  $\rightarrow \Delta_x$ . This algorithm is run by the central authority to update the attribute  $x$  when a user with attribute  $x$  is revoked. The algorithm updates the master secret version key for the attribute  $x$ , and also outputs  $\Delta_x$  to be used for updating the master public key, the cloud secret key that is associated with attribute  $x$ , and ciphertexts associated with attribute  $x$ .
- Update-mpk(mpk,  $\Delta_x$ ). This algorithm updates the master public key mpk using  $\Delta_x$ .
- Update-cloudkey(sk $_x^2$ ,  $\Delta_x$ ). This algorithm updates the cloud secret key sk $_x^2$  using  $\Delta_x$ .
- Update-ct(ct,  $\Delta_x$ ). This algorithm updates ciphertext ct using  $\Delta_x$ .

## V. SYSTEM ARCHITECTURE

The system is based on edge computing ideas [41]. As illustrated in Figure 1, it is separated into two sections, with the organizational border consisting of the Authentication Module and Data Processing Module. Organizations are recognized as trustworthy bodies and regulate the Authentication and Data Processing Module. The other section is about a dubious CSP. The data is encrypted within the organizational limit before uploading data to the cloud. The CSP might potentially be sabotaged by an attacker. In our system, we expect a corrupted CSP to act in an HBC model [34].

The framework encompasses many healthcare industry users, authorities, and data owners. The knowledge graph, encrypted index file, and user's secondary secret keys are all stored in a single cloud server. Any request to the framework is thoroughly checked by the Authentication Module. The organization's policies decide which attributes are used to provide access permissions to each user. Patients own their data and have read-only access to all fields in their EHR.

*a) Use cases:* Our framework offers various use cases, including the ability to read, write, revoke an attribute, and browse encrypted EHRs. User first request access to the system. The Authentication Module examines the application by looking at the user attributes in the knowledge graph and the ABAC rules specified per the company's policies. Access is allowed if the attributes comply with the company's policies.

The framework employs the Data Processing Module to encrypt the updated information of the accessible fields whenever a user edits an EHR. During the procedure, the Attribute Control Center in the module provides the user attributes. The Key Production Unit provides encryption keys for re-encryption. The ciphertexts are later updated in the knowledge stored in the CSP. During a read request, a similar action is done.

The user inputs the search keyword as a query during the search procedure. The Key Production Unit creates the keys for the search. The Token Origination Unit uses the search phrase and concealed keys to produce a trapdoor. The trapdoor is transmitted to the CSP to be compared to the encrypted Indexes. If there is a match, the search procedure brings encrypted EHRs.

The Data Processing Module is solely responsible for attribute revocation. The user provides revoked attributes to the Attribute Center, which it keeps and sends to the Cryptography Unit. The Key Production Unit supplies the master key needed for operation. The ciphertext and the CSP's secondary secret are then updated to reflect the modifications.

We will go through each sub-module in depth in the sections that follow.

### A. Authentication Module

In this module, each login request undergoes a thorough analysis. The ABAC is the module's critical agenda. Within the module, there are multiple units. The database is initially verified for the user's login details. The sub-modules begin to conduct their duties if it passes.

Depending on requests, the Authentication Module interacts with the Document Processor Module and the knowledge graph in the cloud. The knowledge graph holds all users' attributes, access types, and permissions to linked EHR fields. To prevent privacy leaks, a nurse, for example, is only granted access to the data recorded in the Lab Results field of an EHR.

Policy Controller is where the organization's policies are kept. It is frequently described in terms of the users' attributes. Each user has their own set of attributes, which helps preserve their privacy.

Regulation Processor uses SWRL to accomplish access decisions. It accomplishes this with the assistance of the Policy Controller and the results obtained by querying the knowledge graph. Each person in the system has their own set of rules in SWRL. The user attributes and control policies for the data are also stored by the Regulation Processor and sent to the Cryptography Unit throughout the encryption process.

### B. Data Processing Module

Data cryptography, search token generation, encrypted index building, and attribute revocation are supported by sub-modules of the Data Processing Module. The Attribute Control Center gets the Regulation Processor's user attributes and control policies and assigns them to any activity. The Key Production Unit produces the EHR encryption or decryption keys. When a user changes an EHR, the Attribute Control Center and Key Production Unit encrypt the data. The ciphertexts are then uploaded to the knowledge graph as a new node.

The Key Production Unit provides the keys for building a trapdoor in the event of a search operation. The trapdoor searches the cloud's encrypted index for relevant EHRs. With the Cryptography Unit and Key Production Unit, the user may decrypt any EHR. A clinician, for example, could wish to locate patients with covid19 new variant symptoms so they can be treated quickly. As a result, the doctor enters a search query, processed with the secret key, to create a trapdoor. The trapdoor searches the encrypted indexes for all patient records and returns the relevant ones.

Within the Data Processing Module, Attribute Revocation functions are also performed. For example, when a user is promoted, the last attributes must be withdrawn, which

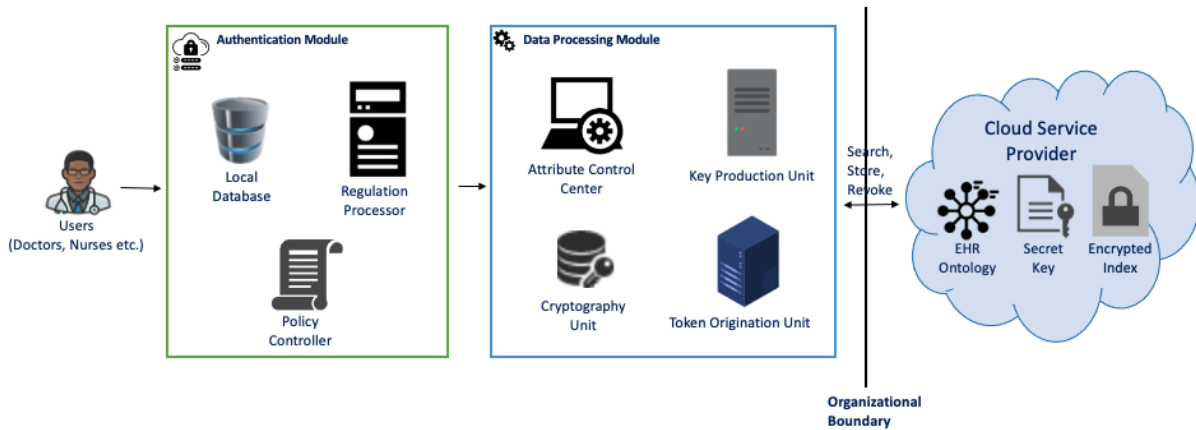


Fig. 1. System Architecture

necessitates these activities. During the procedure, the Key Production Unit provides the master key. The ciphertext and the private user key held by the CSP are changed. The newly updated private key is then used to decrypt EHRs at any moment in the future.

### C. Cloud Service Provider

The CSP contains the knowledge graph, cryptographic index, and secondary secret keys. The CSP is situated outside the organization's boundary and is assumed to be an HBC adversary model. The CSP can successfully run programs and algorithms and examine the information provided within and outside the company. We impose an authorization process on data within the organizational perimeter to deal with this, which we refer to as the framework's edge. Consequently, users are only verified inside the company's borders, protecting their privacy. We implemented the ABE encryption technique at the organizational edge to safeguard data from privacy issues when transmitting data to the cloud server.

1) *EHR Ontology*: The knowledge graph used in the framework to handle heterogeneity is shown in Figure 2. It was designed considering HIPAA regulations. As an information graph, the ontology records medical organization users like Doctors, Nurses, Medical Assistants, Pharmacists, and Patients. It also includes Certifications like MD (Doctor of Medicine), PharmD (Doctor of Pharmacy), etc. Hospital Wards like Oncology, Pediatric, and Specializations like Cardiology and Gynaecology are also included. Likewise, EHR fields like Prescriptions, Doctor Notes, Lab Results, etc., are stored in the ontology as data properties to store patient data. The EHR field access is controlled using ABAC to protect privacy. The Certification, Specialization, and Hospital Wards serve as the attributes of a user. Various users with specific attributes can have multiple kinds of access to the EHR fields.

The ontology in our system holds the encrypted patient data in the nodes. The nodes storing patient data are encrypted using the ABE scheme discussed earlier. This approach offers several advantages like managing fewer files, faster query

retrieval performance independent of data size, and flexible expansion of schemas. The ontology is queried using SPARQL, created dynamically by our system based on the user account.

2) *Encrypted Index*: The patient's encrypted EHR word tokens and unique patient IDs are stored in the encrypted index file. Any search procedure requires the file. Word tokens from the patient's EHR are retrieved. The tokens are then pre-processed before being encrypted using the RSABE technique [49] with the help of the Cryptography Unit. During the crypto procedure, the Key Production Unit provides the public key, and the EHR Ontology delivers the attributes to the Attribute Control Center. The file is produced within the organization's limit and saved in the CSP.

## VI. IMPLEMENTATION

Model-View-Controller (MVC) [12] architectural principles are used to create the open-source EHR software, which is constructed using the Python Django framework. The application allows field-level ABAC access to patient EHRs as well as data encryption using ABE. The RSABE scheme is used to encrypt the patient data and to create the encrypted index file. The application also enables a faster and more efficient search of encrypted data using a searchable encryption technique with the usefulness of the same scheme. The user attributes change over time, which is also permitted in the system.

The EHR system enables doctors to treat their patients safely. It includes the features that are needed for daily operations. Protege [protege.stanford.edu] was used to create the EHR ontology used in the system. It is an open-source knowledge graph management tool. The ontology is queried using SPARQL with the Apache Jena library, and the SWRL rules are used to modify it. Thus, the application uses ABE, searchable encryption, attribute revocation, graph database, and semantic web for seamless functioning.

### A. Dataset Description

We utilized the MIMIC-III [19] dataset to create our synthetic graph dataset of different sizes. We used data with 6000, 13000, 19000, 26000, and 32000 patient records for

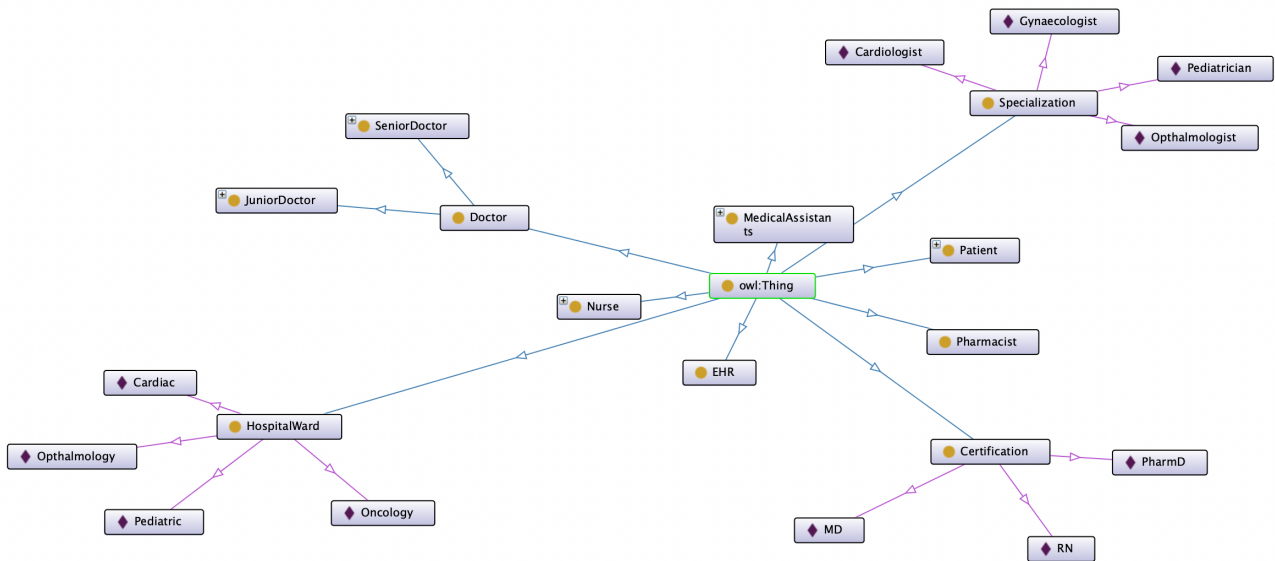


Fig. 2. EHR Ontology

our experiments. Each patient has several fields in their EHR based on their medical histories, such as Allergies, Billing Information, Prescription, etc. The nodes of the EHRs are encrypted within the organization’s perimeter and kept in the cloud server, as per the edge computing concept. We had 30 medical users like Doctors, Nurses, and Medical Assistants throughout the system. Each medical user has certification, specialization, and hospital ward attributes. Different users with distinct attributes have unique access to the EHR fields.

### B. Evaluation

We designed a proof of concept prototype to assess the EHR framework. Let’s say a Doctor named Rachel requests access. In the Authentication Module, the request is thoroughly assessed; the username and password are validated against the database; the Policy Controller examines the policies; the EHR Ontology offers the unique attributes of Doctor Rachel; and the Regulation Processor processes these pieces of information. If Doctor Rachel plans to retrieve the EHR of a patient named Andrew, the request is completed in the Cryptography Unit by getting the Keys from the Key Production Unit. Encrypting an EHR is accomplished similarly. To search through encrypted records, like to find the patient EHRs that have the word token virus, Doctor Rachel submits a search query to the Token Origination Unit that uses the secret keys obtained from the Key Production Unit. To remove an attribute from the system in case of any attribute changes, such as for Junior Doctor Jacky, who got new certification attributes, Doctor Rachel uploads the attributes to be revoked to the Attribute Control Center. The request is processed, and then the ciphertext and the CSP’s secret key are changed to accommodate the changes.

We evaluated the performance of queries with different data sizes to highlight the benefits of our proposed system

TABLE I  
QUERY PERFORMANCE BY OUR PROPOSED NOVEL SYSTEM

Data Size	Insert (and Encrypt)	Retrieve (and Decrypt)	Search	Delete	Revoke
6000	0.003376	0.010368	0.692188	0.002716	0.009746
13000	0.003302	0.009584	1.137799	0.002732	0.009722
19000	0.003264	0.009605	1.408645	0.002744	0.009637
26000	0.003132	0.009672	1.685153	0.002714	0.009594
32000	0.003584	0.009952	1.961842	0.002661	0.009665

compared to our old system [45]. The Table I shows the performance of our proposed system for various data sizes for different types of queries. Similarly, Table II shows the performance of the old system [45]. The performances on both tables are listed in seconds using an average of 10 queries.

The queries in different systems have slightly different operations that reflect their performances, as shown in the tables. The insert query in Table I means time to encrypt a patient EHR field followed by an insert SPARQL command to store into the knowledge graph. In contrast, the insert query in Table II means time to encrypt the same patient EHR field, create a folder, and dump the flat file. Thus, the data upload time is more in our proposed system than in the old system [45]. Moreover, the data size unaffected encryption performances in each system. The retrieve query in Table I denotes the time to decode an encrypted EHR field stored in the knowledge graph using the SPARQL query. However, in Table II, the retrieve operation represents the time to locate the

TABLE II  
QUERY PERFORMANCE BY OUR OLD SYSTEM [45]

Data Size	Insert (and Encrypt)	Retrieve (and Decrypt)	Search	Delete	Revoke
6000	0.002639	0.017153	0.695360	0.000344	0.015750
13000	0.002775	0.022646	1.157350	0.000229	0.016229
19000	0.002570	0.019395	1.427119	0.000190	0.015999
26000	0.002674	0.017417	1.692440	0.000191	0.016552
32000	0.002623	0.017676	1.977922	0.000204	0.016249

same patient’s EHR field and decode the data. Our proposed system shows improved data retrieval performance compared to the old system [45]. The performance is also independent of the data size, as depicted in Table I. The search operation performs better in the new system, as shown in Table I compared to the old system [45]. The tables show the time to search for patient EHRs with the keyword flu and then retrieve an EHR of one of the patients with flu. The delete query in Table I represents the time to remove a patient EHR field using a SPARQL query. On the other hand, the query in Table II means locating the folder containing the EHR field data and deleting it. Overall, our proposed system takes more time to delete the data than the old system [45]. However, the performance in each system stands unchanged by the data size. The revocation time shows better performance in our new system. In Table I, we recorded the time to revoke an attribute, update the ciphertext and secret key and finally retrieve a patient EHR field. In Table II, we recorded the time for the same operation, and overall it takes more time as the old system [45] is slow compared to our proposed system.

Our proposed system shows the great benefits of using a graph database based on our experiments. The data retrieval performance is fast, which is critical in the healthcare domain, and it is independent of the data volume. The query execution time remains almost the same. The reason is the graph database has no global index. Each vertex holds data about its neighbor nodes. So for a specific query, it does not need to handle irrelevant information. Only a particular node is addressed in the query to provide the results. Hence, the query performance remains solid.

## VII. CONCLUSION

This paper describes using a HIPAA-compliant knowledge graph to create an EHR system that supports field-level ABE, ABAC, searchable encryption, and attribute revocation. All user roles and attributes in the healthcare organization are represented in a knowledge graph. The graph keeps track of the users’ attributes and EHR fields to grant them restricted access to the EHR system. The patient data are stored as encrypted nodes in the knowledge graph, which provides many benefits. Doctors frequently need to scan through encrypted data in

a limited amount of time and computation in the presence of Big Data, which our framework can also handle. Because some users leave the organization, others get promoted, or the organization’s regulations change, user attributes change over time. As a result, all of these modifications need the revocation of attributes. Our system solves this problem by entrusting ciphertext and secondary secret key updates to the cloud. The secret key, which the user keeps, stays solid. Often, a system with a lot of functionality usually includes a lot of keys, which adds to the user’s administrative burden. Our framework appears more user-friendly by having a single scheme for all of the above actions. The knowledge graph, encrypted index file, and secondary secret key are all kept in the CSP considering the HBC adversary model. We also assumed Edge Computing ideas. Users are validated within the organization’s limits to ensure privacy. All activities on the data are completed inside the organization’s border before transporting them to the cloud to defend against privacy concerns.

We want to expand our research in many possible directions in the future. We plan to populate more data in the system on the scale of billions, and we expect our novel system to show better delete and encrypt performances. Additionally, we want to make our system more robust by having a backup server or redundant copies to avoid the central point of failure. We hope to broaden our future comparisons with other systems with distributed architectures like MapReduce. Furthermore, we would like to consider other threat models in the future.

## ACKNOWLEDGMENT

This work has been supported by Office of Naval Research under grants N00014-18-1-2453, N00014-19-WX-00568, and N00014-20-WX01704 and by NSF grant 1955319. We thank Dr. Michael A. Grasso for his vital feedback.

## 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 Madiseti. 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] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5):34–43, 2001.
- [7] 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.

- [8] 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.
- [9] Centers for Medicare & Medicaid Services. The Health Insurance Portability and Accountability Act of 1996 (HIPAA). Online at <http://www.cms.hhs.gov/hipaa/>, 1996.
- [10] 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.
- [11] 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.
- [12] John Deacon. Model-view-controller (mvc) architecture. *Online [Citado em: 10 de março de 2006.] http://www.jdl.co.uk/briefings/MVC.pdf*, 2009.
- [13] 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.
- [14] Allan H Goroll, Steven R Simon, Micky Tripathi, Carl Ascenzo, and David W Bates. Community-wide implementation of health information technology: the massachusetts ehealth collaborative experience. *Journal of the American Medical Informatics Association*, 16(1):132–139, 2009.
- [15] 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.
- [16] Matthew Green, Susan Hohenberger, Brent Waters, et al. Outsourcing the decryption of abc ciphertxts. In *USENIX security symposium*, volume 2011, 2011.
- [17] Junbeom Hur and Dong Kun Noh. Attribute-based access control with efficient revocation in data outsourcing systems. *IEEE Transactions on Parallel and Distributed Systems*, 22(7):1214–1221, 2010.
- [18] 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.
- [19] Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9, 2016.
- [20] 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.
- [21] 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.
- [22] Alex H Krist, Eric Peele, Steven H Woolf, Stephen F Rothemich, John F Loomis, Daniel R Longo, and Anton J Kuzel. Designing a patient-centered personal health record to promote preventive care. *BMC medical informatics and decision making*, 11(1):1–11, 2011.
- [23] Ora Lassila, Ralph R Swick, et al. Resource description framework (rdf) model and syntax specification. 1998.
- [24] Jiguo Li, Yuerong Shi, and Yichen Zhang. Searchable ciphertext-policy attribute-based encryption with revocation in cloud storage. *International Journal of Communication Systems*, 30(1):e2942, 2017.
- [25] Jiguo Li, Yao Wang, Yichen Zhang, and Jinguang Han. Full verifiability for outsourced decryption in attribute based encryption. *IEEE transactions on services computing*, 13(3):478–487, 2017.
- [26] 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.
- [27] Jingwei Li, Jin Li, Xiaofeng Chen, Chunfu Jia, and Zheli Liu. Efficient keyword search over encrypted data with fine-grained access control in hybrid cloud. In *International conference on network and system security*, pages 490–502. Springer, 2012.
- [28] 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.
- [29] 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.
- [30] Xiong Li, Maged Hamada Ibrahim, Saru Kumari, Arun Kumar Sangaiah, Vidushi Gupta, and Kim-Kwang Raymond Choo. Anonymous mutual authentication and key agreement scheme for wearable sensors in wireless body area networks. *Computer Networks*, 129:429–443, 2017.
- [31] Xiong Li, Jianwei Niu, Saru Kumari, Fan Wu, and Kim-Kwang Raymond Choo. A robust biometrics based three-factor authentication scheme for global mobility networks in smart city. *Future Generation Computer Systems*, 83:607–618, 2018.
- [32] 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.
- [33] Zhiqian Lv, Jialin Chi, Min Zhang, and Dengguo Feng. Efficiently attribute-based access control for mobile cloud storage system. In *2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, pages 292–299. IEEE, 2014.
- [34] Tim Mather, Subra Kumaraswamy, and Shahed Latif. *Cloud security and privacy: an enterprise perspective on risks and compliance*. ” O’Reilly Media, Inc.”, 2009.
- [35] Deborah L McGuinness, Frank Van Harmelen, et al. Owl web ontology language overview. *W3C recommendation*, 10(10):2004, 2004.
- [36] Shivaramkrishnan 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.
- [37] Matthew Pirretti, Patrick Traynor, Patrick McDaniel, and Brent Waters. Secure attribute-based systems. *Journal of Computer Security*, 18(5):799–837, 2010.
- [38] Bo Qin, Hua Deng, Qianhong Wu, Josep Domingo-Ferrer, David Naccache, and Yunya Zhou. Flexible attribute-based encryption applicable to secure e-healthcare records. *International Journal of Information Security*, 14(6):499–511, 2015.
- [39] 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.
- [40] 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.
- [41] Weisong Shi, Jie Cao, Quan Zhang, Youhui Li, and Lanyu Xu. Edge computing: Vision and challenges. *IEEE internet of things journal*, 3(5):637–646, 2016.
- [42] Dawn Xiaoding Song, David Wagner, and Adrian Perrig. Practical techniques for searches on encrypted data. In *Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000*, pages 44–55. IEEE, 2000.
- [43] 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.
- [44] Wenhai Sun, Shucheng Yu, Wenjing Lou, Y Thomas Hou, and Hui Li. Protecting your right: Verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud. *IEEE Transactions on Parallel and Distributed Systems*, 27(4):1187–1198, 2014.
- [45] Redwan Walid, Karuna P Joshi, and Seung Geol Choi. Secure cloud ehr with semantic access control, searchable encryption and attribute revocation. In *2021 IEEE International Conference on Digital Health (ICDH)*, pages 38–47. IEEE, 2021.
- [46] Redwan Walid, Karuna Pande Joshi, SeungGeol Choi, and Dae-young Leroy Kim. Cloud-based encrypted ehr system with semantically rich access control and searchable encryption. *UMBC Student Collection*, 2020.
- [47] Hao Wang and Yujiao Song. Secure cloud-based ehr system using attribute-based cryptosystem and blockchain. *Journal of medical systems*, 42(8):1–9, 2018.
- [48] Qinqin Wang, Yanqin Zhu, and Xizhao Luo. Multi-user searchable encryption with fine-grained access control without key sharing. In *2014 3rd International Conference on Advanced Computer Science Applications and Technologies*, pages 145–150. IEEE, 2014.



- [49] Shangping Wang, Duo Zhang, Yaling Zhang, and Lihua Liu. Efficiently revocable and searchable attribute-based encryption scheme for mobile cloud storage. *IEEE Access*, 6:30444–30457, 2018.
- [50] Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou. Achieving secure, scalable, and fine-grained data access control in cloud computing. In *2010 Proceedings IEEE INFOCOM*, pages 1–9. Ieee, 2010.
- [51] Fucai Zhou, Yuxi Li, Alex X Liu, Muqing Lin, and Zifeng Xu. Integrity preserving multi-keyword searchable encryption for cloud computing. In *International Conference on Provable Security*, pages 153–172. Springer, 2016.
- [52] Zhibin Zhou and Dijiang Huang. Efficient and secure data storage operations for mobile cloud computing. In *2012 8th international conference on network and service management (cnsm) and 2012 workshop on systems virtualization management (svm)*, pages 37–45. IEEE, 2012.