Preprint: In proceedings of IEEE International Conference on Digital Health (ICDH) 2021

Secure Cloud EHR with Semantic Access Control, Searchable Encryption and Attribute Revocation

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*—To ensure a secure Cloud-based Electronic Health Record (EHR) system, we need to encrypt data and impose field-level access control to prevent malicious usage. Since the attributes of the Users will change with time, the encryption policies adopted may also vary. For large EHR systems, it is often necessary to search through the encrypted data in real-time and perform client-side computations without decrypting all patient records. This paper describes our novel cloud-based EHR system that uses Attribute Based Encryption (ABE) combined with Semantic Web technologies to facilitate differential access to an EHR, thereby ensuring only Users with valid attributes can access a particular field of the EHR. The system also includes searchable encryption using keyword index and search trapdoor, which allows querying EHR fields without decrypting the entire patient record. The attribute revocation feature is efficiently managed in our EHR by delegating the revision of the secret key and ciphertext to the Cloud Service Provider (CSP). Our methodology incorporates advanced security features that eliminate malicious use of EHR data and contributes significantly towards ensuring secure digital health systems on the Cloud.**

***Index Terms*—Attribute Revocation, Searchable Encryption, Electronic Health Record, Knowledge Graph (Ontology), Cloud Computing, Cloud Security**

I. Introduction

Healthcare organizations are increasingly adopting cloud-based technologies to maintain their digital records efficiently. A cloud-based Electronic Health Record (EHR) service allows centralizing patient data and using the advantages of elasticity and scalability of a cloud infrastructure [5], [34], [35], [40]. The cloud also offers a highly supportive atmosphere for efficiently handling the load [2]. Moreover, cloud services are usually a more economical solution than others when they develop a technology infrastructure to deploy their services. It has become much more popular due to the pay-as-you-go concept, which incentivizes customers to pay only for what they want and how much they use.

Although the Cloud offers many advantages, it also continues to pose unique risks to healthcare organizations in terms of data privacy and security. With the recognition of the security risks, all healthcare organizations must comply with Health Insurance Portability and Accountability Act (HIPAA) [14], [50] and Health Information Technology for Economic and Clinical Health (HITECH) [49] privacy guidelines set by regulatory authorities. Failing to comply with the acts can be financially devastating to an organization. Noncompliance carries a range of penalties depending on the degree of misconduct which can result in hefty fines. Violations can also lead to an arrest. As a result, an EHR solution must abide with all applicable laws and regulations and also allow an easy and seamless exchange of patient information.

A. Motivation

EHR system with semantic access control. Recently, Walid et al. [55] proposed a cloud-based EHR system that offers a semantically rich, policy-driven mechanism that employs Attribute-Based Access Control (ABAC) [22] to evaluate users' entrance to the system.

The architecture of their system was created using Semantic Web Technologies [7]. By referencing the HIPAA knowledge graph (ontology) developed in [23], they created a HIPAA-consistent knowledge graph. This way, the system systematically addressed the issues with the aforementioned compliance problem.

In particular, the system used a knowledge graph to derive user attributes and the EHR fields based on the type of request. The knowledge graph is queried using SPARQL, and it entails complete details of individuals in the organization and their associated unique attributes. The unique attribute control various access to different fields of an EHR. Thus, each individual has distinct access to a patient EHR.

The system also allowed the client to search through encrypted data based on keyword queries, without needing to download and decrypt all encrypted data from the cloud.

Disadvantages of the work [55]. Although their system has many good features, it still has a few disadvantages. In this work, we want to improve the system by resolving those disadvantages.

First of all, the system uses two different encryption schemes, one for searchability and the other for data encryption. It often becomes more cumbersome in large systems to have multiple schemes within a system, as more keys and policies may have to be maintained. A system would be better if it uses a single encryption scheme.

Second, although the system provides the search feature, the search time is quite slow. In the presence of Big data, it usually requires a lot of time and analysis to search through the data to locate specific patients with certain diseases or conditions. We want to speed up the search time.

Most importantly, the system doesn't provide revocation. It is essential that an attribute-based EHR system have an attribute revocation feature, since the organization policies and the user attributes keep changing with time. For example, a physician might have been promoted, so its attributes would change. Likewise, an employee might have moved to another department, or an employee leaves the organization, so its unique attributes must be revoked from the system. Often the organization policies also vary with time, which requires some attributes to be revoked.

B. Our Work

In this paper, we improve the EHR system in [55] by resolving the issues mentioned above. Below, we overview our system.

Underlying ABE encryption. We use the revocable, searchable ABE scheme introduced by Wang et al. [59]. The scheme provides both searchability and data encryption, which fits our purpose. The scheme also allows us to outsource the computation to the cloud, which we describe next.

Outsourcing the computation to cloud. In general, our system requires fewer computations as most functions are safely delegated to the cloud service provider (CSP). This is achieved by splitting each user's secret into two keys and having one key uploaded to the CSP while the other key is kept secret to the user.

For example, since each user has a dedicated private key stored within the CSP, partial decryption is delegated to the CSP. The output from the partial decryption still hides the plaintext from the CSP and outsiders. However, given the partial decryption from the CSP, the user can recover the plaintext record with much less effort.

Faster search time. Owing to the presence of Big Data, searching through encrypted data requires utmost attention. The search time can also be improved, thanks to the outsourcing framework.

In particular, to search for a keyword in the encrypted index for the EHR database, the user first creates a token connected with a keyword query. For privacy, the token hides the keyword from the CSP.

Once the CSP gets the search token, the CSP uses the token to run the search algorithm over the ciphertexts to see which ones have the privately linked keyword (s). When the indexes, keywords, and the user's attributes are set to meet the ciphertext access control policies, the cloud service retrieves the search results. The encoded version of the message containing the keyword is sent back to the recipient.

Our technique enables keyword search with substantially reduced network latency and client-side computing costs compared to prior work [55].

Attribute revocation. The knowledge graph records all users and patients in our framework, considering all attribute changes in the system along with the associated EHR fields. So, the knowledge graph functions to revoke unwanted user attributes and help to protect patient privacy.

The outsourcing framework also benefits when the system performs revocation. In particular, when a user attribute is revoked, the revocation takes place *only in the CSP level*. This works because a user secret is split into two keys, and revoking only one key would still disable the function of a user secret. The secret key that lies within the user remains firm during the entire revocation process, which greatly simplifies the revocation process.

Edge computing. The term edge computing [51] refers to the need to analyze data locally before sending it to the cloud, and we have followed this principle in our system. Inside the organizational periphery, which we refer to as the edge in our system, we enforce an access control mechanism on the records. All users are checked only within the organization's borders, preserving their anonymity. Within the organization limit, we have implemented a robust encryption approach that protects data integrity from privacy risks until moving it to the cloud. As a result, the frontier continues to be a formidable barrier to data protection.

Threat model. Cloud users, while storing their data on the cloud, usually categorize CSP on one of these threat models: the Honest-But-Curious (HBC) adversary model, where CSPs run the programs and algorithms correctly, but might look at the information passed between entities; or the malicious adversary model, where providers behave in an arbitrary manner that may be hostile to the cloud customer [42]. We have considered the HBC threat model for our approach since cloud users trust the functionality of their applications running on the cloud, but they may not fully trust the CSP whose dataset is stored in distant Cloud data centers. First of all, Clouds may be exposed to tainted workers who fail to adhere to data protection standards. Secondly, cloud applications may be subject to external cyberattacks, and cloud users may not be aware of the possible repercussions on their data security when such invasions occur. The users worry that the CSP might attempt to decrypt the data by analyzing it thoroughly or monitoring data traffic between users. We presume our framework to withstand a compromised user attempting to decode ciphertext with her decryption key and gain knowledge. We also assume our system to be resilient in the face of a user alliance trying to crack the ciphertext with decryption keys that no single member of the coalition can decode with her decryption key.

C. Organization

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 conclusions in Section VI.

II. Related Work

A. Electronic Health Record System

Digital health record systems are commonly employed to enhance hospital services, improve treatment efficacy, and reduce premiums [16], [25]. EHR records a patient's vital stats, diagnoses, medications, immunization history, laboratory and radiology reports, doctor notes, and other medical facts along with the patient's details. An EHR system provides several benefits such as accurate documentation, disease tracking, data sharing, statistical analysis, and so forth. Consequently, security and privacy concerns have hampered the spread of the EHR system, and they have received increasing focus in current years [35]–[37], [48]. Narayan et al. [45] recommended using ABE to protect the privacy of EHR data from outside threats, as well as the CSP. Fatos et al. [60] originally presented a multi-user fuzzy keyword search method that supported fine-grained permission restriction over encrypted data.

Unfortunately, most proposed solutions fall short in providing controlled access, encryption device, searchable encryption, and attribute revocation. Furthermore, the majority of the accessible application is licensed, making them costly to use. In this circumstance, our research effort tries to develop an open-source, low-cost EHR management system that can provide advanced data privacy and protection levels.

B. Regulatory Policies

Patient data is secured in the United States under some statutes; the most notable is the HIPAA Act. Electronic safe health information (ePHI) [11] is the name given to the information about one's health that is protected by these rules. The Health Information Technology for Economic and Clinical Health Act (HITECH) allows sharing ePHI while still requiring HIPAA privacy and protection laws to be applied more strictly and thoroughly [1]. These rules, on the other hand, make no mention of encryption principles or algorithms. Furthermore, data encryption in data access control and transfer is defined as addressable rather than mandatory. This left space for different definitions and then became a source of debate regarding sharing ePHI. Cloud-based EHR services in the United States are required to comply with these regulatory standards and ensure enhanced data protection combined with the seamless user experience that cloud services offer. This also requires implementing strict access control mechanisms to provide unauthorized access by any user is prohibited by their EHR.

C. Semantic Web Technologies

We have used Semantic Web technologies to develop our system's knowledge graph and the reasoning component of our system. These enable us to build the schema using W3C standardized languages that support our design requirements, including interoperability, sound semantics, Web integration, and the availability of tools and system components. Semantic Web tools enable data to be annotated with machine-understandable meta-data, allowing the automation of their retrieval and their usage of incorrect contexts. Semantic Web technologies include languages such as Resource Description Framework (RDF) [26] and Web Ontology Language (OWL) [43] for defining ontologies and describing meta-data using these ontologies as well as tools for reasoning over these descriptions.

Our most fundamental requirement is for a representation that supports interoperability at both the syntactic and semantic levels. OWL has well-defined semantics grounded in first-order logic and model theory, allowing programs to draw inferences with the assurance that the subsequent interpretation is sound. An important advantage for OWL over many other knowledge-representation systems is that it has well-defined subset profiles guaranteeing sound and complete reasoning with various levels of reasoning complexity and is designed to work with popular implementation technologies, such as OWL QL for databases and OWL RL for rule-based systems. A second design requirement is for a language that is designed to integrate well with the Web and Cloud, which is becoming the dominant technology for today's digital health systems. These technologies can be used to provide common semantics of service information and policies enabling all agents who understand essential Semantic Web technologies to communicate and use each other's data and Services effectively. OWL is built on basic Web standards and protocols and is evolving to remain compatible with them. It is possible to embed RDF and OWL knowledge in HTML pages, and several search engines (including Google) will find and process some embedded RDF.

D. Attribute-Based Encryption

ABE [17], introduced by Sahai and Waters, has been one way to ensure data security and eliminate risks. In ABE [17], The data is encoded using a set of attributes, and the private key is defined using a different set of attributes. Based on the threshold parameter, the ciphertext can only be deciphered if the two sets of attributes overlap. One of the EHR system security developments was known as ABE [3], [6], and [45]. It has been further divided into ciphertext-policy ABE (CP-ABE) [8] and key-policy ABE (KP-ABE) [4] due to lack of expressibility. The secret key is coupled with an attribute set in CP-ABE [8], and the ciphertext is paired with an access policy. In most cases, the policy is defined as a Boolean formula with a specific set of attributes. A secret key may decrypt a ciphertext if the attributes set match the access policy, while the whole scenario is reversed in the KP-ABE scheme.

CP-ABE [8] is considered more effective for authentication in the cloud because an individual ciphertext defines a policy that explicitly specifies attributes that data users must hold for the encryption process. Joshi et al. [24] developed attributed-based access control (ABAC) that is semantically enriched in accessing data leveraging CP-ABE [8]. Their model evaluates access categories based on user attributes and EHR fields. The national hub controls both EHR secure entry and distribution.

E. Attribute-Based Encryption With Attribute Revocation

Since the user's attribute can vary significantly over time, attribute revocation is crucial in ABE frameworks. Perretti et al. [46] were the first to implement attribute revocation, which they accomplished by a timed rekeying process. Each attribute had an expiration time in the system, so authority centers had to reprint revised keys regularly. The authority center had to cease releasing and modifying the current edition of the attribute to revoke an attribute in the scheme. Bethencourt et al. [8] later expanded Perretti's work where there was a single expiration time connected with the user's private key. Boldyreva et al. [9] proposed a revocable KP-ABE scheme that improved on their previous revocable IBE. Wang et al. [57], [56] presented two explicitly revocable CP-ABE frameworks based on bilinear and multilinear maps, accordingly. Several ABE systems involving instant attribute revocation were suggested in current history. Yu et al. [62] and Ibraimi proposed et al. [21] the CP-ABE scheme that employs a semi-trusted proxy server to execute instant attribute revocation. Their approach shifted the authority's responsibilities to the proxy server, significantly reducing the authority's burden. They have, nevertheless, been unable to obtain fine-grained access control. Furthermore, as the number of users increases rapidly, the proxy server's update work skyrockets. Li et al. [31] devised an effective CP-ABE scheme of user revocation with a lower computing expense. Several other schemes can be seen in [66], [47], [63], and [39].

Computational efficiency is another consideration in the latest ABE schemes. Outsourced decoding technologies can help decrease the user's computing load. Green et al. [18] first proposed an effective ABE scheme that facilitates outsourced decoding. The bulk of decryption activities are done by the CSP using the users' key. Zhou et al. in [65] suggested an optimized data management system centered on mobile devices, in which portions of the encryption and decryption processes were safely delegated to the CSP without sensitive data leakage. Li et al. [30] proposed an ABE scheme including complete verification for outsourced decryption, which addresses the problem of ensuring the accuracy of outsourced decryption for unauthorized individuals. The scheme implemented in our systems seems to be perfect compared to [20], [61], [54], and [41].

F. Keyword Search Over Encrypted Data

Fast and efficient searchability is required for any EHR system, particularly in the movement of evidence-based healthcare, because doctors have a limited time in which to make judgments. Dawes et al. [13] mentioned that time constraints are the most significant factor impeding computer systems in medical practice. Physicians indicated that response time is one of the obstacles to EHR system adoption in another study by Holden et al. [19]. Searchable encryption (SE) thus remains to be of the utmost important feature in EHR systems.

SE is an encryption technique that allows users to scan for keywords in cyphertext without revealing the keywords. Song et al. [52] first devised a realistic SE scheme focused on symmetric cryptography, establishing a significant standard for keyword search on encrypted data. Boneh et al. [10] later pioneered SE research into public-key cryptography. Following that, numerous SE schemes were developed to improve search performance, security issues, and search functionality [12], [15], [27], [32], [53]. Attribute-based keyword search, which combines ABE and SE properties, has seen many hypes in current history that can be seen in [28], [29], [33], [44], [58], [64], and [38].

III. Preliminaries

Let λ be the security parameter.

A. Revocable, Searchable ABE

In this section, we describe revocable, searchable attribute-based encryption scheme.

Syntax. Let $\mathcal{X}$ be the attribute universe. A revocable, searchable ABE consists of the following algorithms:

- $\mathsf{Setup}(1^\lambda, \mathcal{X}) \to (\mathsf{mpk}, \mathsf{msk}, \mathsf{msvk})$. The setup algorithm gets as input the security parameter λ, the attribute universe $\mathcal{X}$. It outputs the public parameter mpk, the master secret key msk, and the master secret version key. The master secret version key will be updated when users or attributes are revoked through algorithm $\mathsf{Update\text{-}msvk}$ described below.
- $\mathsf{KeyGen}(\mathsf{msk}, \mathsf{msvk}, x) \to (\mathsf{sk}_x^1, \mathsf{sk}_x^2)$. The key generation algorithm gets as input $\mathsf{msk}, \mathsf{msvk}$ and a set of attributes x. It outputs a pair of secret keys $(\mathsf{sk}_x^1, \mathsf{sk}_x^2)$.
  The first key sk_x^1 will be sent to the user, and the second key sk_x^2 will be stored on the cloud server.
- $\mathsf{Enc}(\mathsf{mpk}, \mathsf{msk}, f, m) \to \mathsf{ct}_f$. The encryption algorithm gets as input mpk, and a boolean formula f over $\mathcal{X}$, and a message m. It outputs a cipehrtext ct_f.
- $\mathsf{EncInd}(\mathsf{mpk}, W) \to I_W$. The encrypted index algorithm gets as input mpk, and a set of keywords W. It outputs an encrypted index I_W for W.
- $\mathsf{Token}(\mathsf{sk}_x^1, w) \to \mathsf{t}_w$. The token generation algorithm gets as input the user secret key sk_x^1 and a query keyword w. It outputs a token t_w.
- $\mathsf{Test}(\mathsf{sk}_x^2, I_W, \mathsf{t}_w) \to 0/1$. The test algorithm gets as input the clout secret key sk_x^2, the encrypted index I_W and the user generated token t_w. If the embedded keyword in 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 sk_x^2 when it receives the token t_w for the user; the encrypted index I_W is typically stored on the cloud server.
- $\mathsf{Decrypt\text{-}cloud}(\mathsf{sk}_x^2, \mathsf{ct}_f) \to \mathsf{pd}$. This algorithm gets as input the cloud secret key sk_x^2 and the ciphertext ct_f. If $f(x) = 1$, it outputs partial decryption pd; otherwise, it outputs an error.
- $\mathsf{Decrypt\text{-}user}(\mathsf{sk}_x^1, \mathsf{pd}) \to m$. Given the partial decryption, the user with sk_x^1 will recover the message m.
- $\mathsf{Update\text{-}msvk}(\mathsf{msvk}, x) \to \Delta_x$. 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 Δ_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.

- $\mathsf{Update\text{-}mpk}(\mathsf{mpk}, \Delta_x)$. This algorithm updates the master public key mpk using Δ_x.
- $\mathsf{Update\text{-}cloudkey}(\mathsf{sk}_x^2, \Delta_x)$. This algorithm updates the cloud secret key sk_x^2 using Δ_x.
- $\mathsf{Update\text{-}ct}(\mathsf{ct}, \Delta_x)$. This algorithm updates ciphertext ct using Δ_x.

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

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

In the above, all queries x that A makes to oracle $\mathsf{KeyGen}(\mathsf{msk}, \mathsf{msvk}, \cdot)$ should satisfy $f^*(x) \neq 1$. In addition, all queries m_0 and m_1 should have the same length.

A revocable, searchable ABE is said to be *revocation secure*, if for all polynomial adversary A, the probability $|\Pr[\mathsf{Expt}^{\mathrm{revoke}}(\lambda)] - 1/2|$ is negligible in λ.

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

```
$\mathsf{Expt}_A^{\mathrm{keyword}}(\lambda)$:
    $(\mathsf{mpk}, \mathsf{msk}, \mathsf{msvk}) \leftarrow \mathsf{Setup}(1^\lambda, \mathcal{X})$;
    $x \leftarrow A(\mathsf{mpk})$;
    $(\mathsf{sk}_x^1, \mathsf{sk}_x^2) \leftarrow \mathsf{KeyGen}(\mathsf{msk}, \mathsf{msvk}, x)$;
    $(W_0, W_1) \leftarrow A^{\mathsf{Token}(\mathsf{sk}_x^1, \cdot)}(\mathsf{mpk})$;
    $b \leftarrow_R \{0, 1\}$;
    $I_{W_b} \leftarrow \mathsf{EncInd}(\mathsf{mpk}, W_b)$;
    $b' \leftarrow A^{\mathsf{Token}(\mathsf{sk}_x^1, \cdot)}(I_{W_b})$
    If $b = b'$ output 1; otherwise output 0.
```

In the above, all queries w to $\mathsf{Token}(\mathsf{sk}_x^1, \cdot)$ should satisfy $w \notin \{W_0, W_1\}$.

A revocable, searchable ABE is said to be *keyword-search secure*, if for all polynomial adversary A, the probability $|\Pr[\mathsf{Expt}^{\mathrm{keyword}}(\lambda)] - 1/2|$ is negligible in λ.

The scheme we use. In this paper, we use the scheme in [59] that satisfies both revocation security and keyword-search security.

IV. System Architecture

The entire framework is based on the principles of Edge computing [51]. It is divided into two sections, with the organizational boundary comprising the Authentication Module and Data Processing Module as shown in Figure 1. Since organizations control these two modules, they are known as trustworthy bodies. All users are authenticated inside the organizational perimeters, preserving their anonymity. The other section concerns an untrustworthy CSP. Before uploading data to the cloud, a rigorous encryption approach is implemented within the organizational border to protect data integrity from privacy risks. An attacker may also be sabotaging the CSP. In our system, we assume a compromised CSP will behave in an honest-but-curious manner [42].

Our framework has a diverse set of users, authorities, and data owners from various medical fields. A single CSP stores the EHRs, encrypted index file, and user's secondary secret keys. The Authentication module performs a thorough check on any request to the framework. Each user is granted access rights based on attributes as determined by the organization's policies. Patients have read access to all fields of their EHR.

Use cases. Whether users choose to read, write, revoke an attribute, or browse through encrypted EHRs, our framework has multiple use cases. A user first asks for access to the EHR system. The Authentication Module reviews the application by looking over the user attributes in the user graph and ABAC rules defined according to the individual company policy. If the attributes follow the guidelines of the company, access is granted.

Whenever a user modifies an EHR, the framework uses the Data Processing Module to encrypt the updated details of the accessed fields. The Attribute Control Center in this module supplies the user attributes during the process. The Key Production Unit provides encryption keys for re-encryption. The EHR ontology housed with the CSP is then modified with the ciphertexts. A similar operation is performed during a read request.

During the search process, the user enters the search keyword as a request. The Key Production Unit provides the keys used for searching. Using the search keyword and hidden keys, the Token Origination Unit creates a trapdoor. The trapdoor is then sent to the CSP, where it is compared to the encrypted Indexes. The search operation retrieves encrypted EHRs if there is a match. The user may then choose any particular EHR to decrypt.

Attribute revocation is entirely handled in the Data Processing Module. The user gives revoked attributes to the Attribute Center, which it stores and supplies to Cryptography Unit. The Key Production Unit provides the master key. The ciphertext and the secondary secret that lies with the CSP are then updated to account for the changes.

In the following sections, we will go through each submodule in detail.

A. Authentication Module

Any login request passes through a comprehensive investigation in this module. The key policy behind the module is the ABAC. There are also several units within the module with critical functions. The user's login information is at first checked in the database. If it passes, the sub-modules begin to perform their functions.

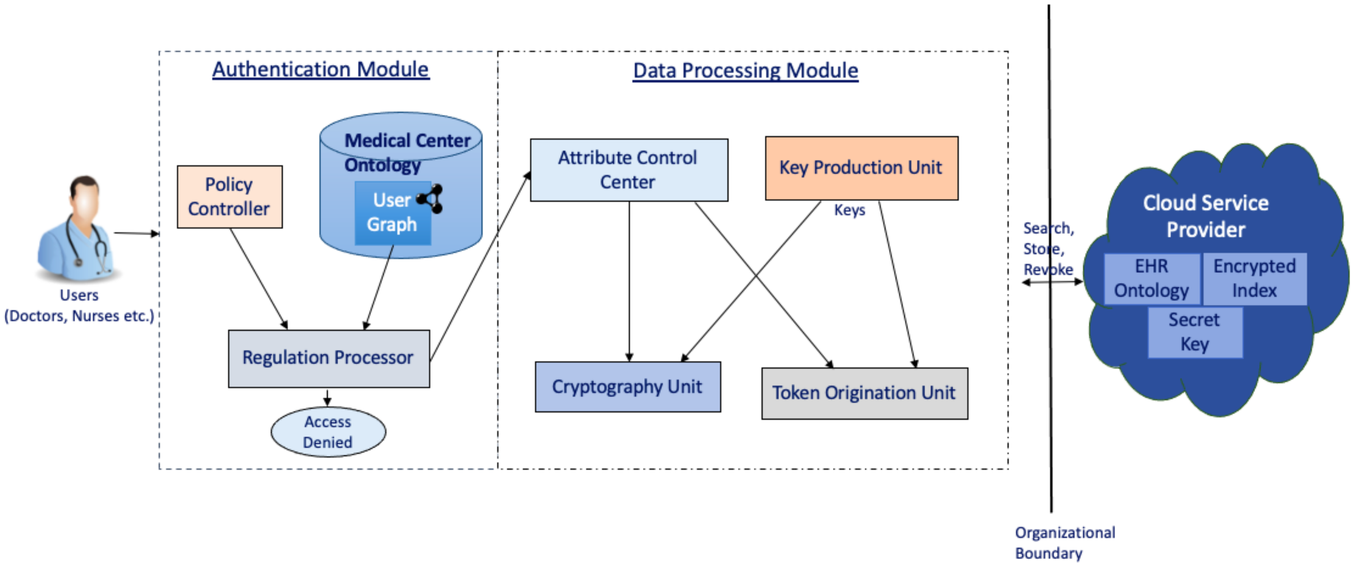

Fig. 1. System Architecture

The Medical Center Ontology store all attributes of every person belonging to the healthcare organization. This helps any user getting unnecessary information and only information to complete their job functions. It also preserves EHR access-related information involving the particular field level access for all users and patients of the EHR system. For example, a nurse is given access only to the data stored in the Lab Results and Doctor Notes fields of an EHR to prevent privacy leakage. The ontology is built considering the HIPAA regulations.

Policy Controller holds the organization's policies. It is often listed in terms of the attributes of the users. Different policies are set for various users that help to protect individual privacy.

Regulation Processor employs Semantic Web Rule Language (SWRL) to achieve access choices. It does it with the help of the Policy Controller and the output given by querying the Medical Center Ontology. The SWRL is unique for each individual in the EHR system. Regulation Processor also stores the user attributes and control policies for the data records and passes them to Cryptography Unit during the encryption-related process.

B. Data Processing Module

The Data Processing Module performs several critical functions in the EHR system. Several sub-modules within the module help data encryption, data decryption, search token generation, encrypted index creation, and attribute revocation. Since we have used a single scheme in all such operations compared to our previous work [55], it seems more straightforward and convenient.

The user attributes and the control policies stored within the Regulation Process are passed to the Attribute Control Center that supplies the user attributes for any operation. The Key Production Unit supplies the keys for encrypting or decrypting an EHR. Whenever a user modifies an EHR, the Attribute Control Center, and Key Production Unit function to encrypt the updated data. The EHR ontology housed with the CSP is then modified with the ciphertexts.

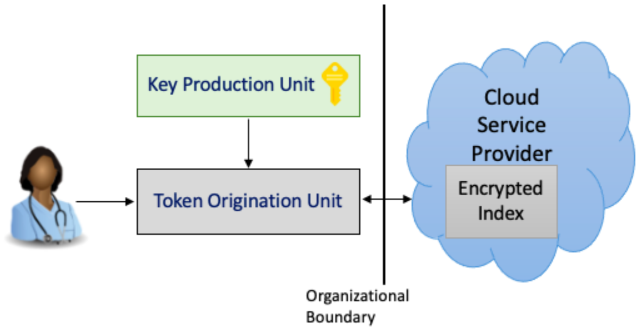

Fig. 2. Token generation process

During a search operation, the user gives the search keyword in the form of a query. The same Key Production Unit supplies the keys for the operation. A trapdoor is later generated, as shown in Figure 2. The trapdoor allows scanning through the encrypted index, and it is given to the CSP and compared to the encrypted indexes. The user gets the corresponding EHRs if there is a match. The user may then decrypt any particular EHR with the help of the Cryptography Unit and Key Production Unit. For example, a doctor may want to find patients with covid symptoms for immediate treatment. So, the doctor puts a search query, and by processing it with the secret key, a trapdoor is generated. The trapdoor scans through the encrypted indexes and returns all the corresponding patients.

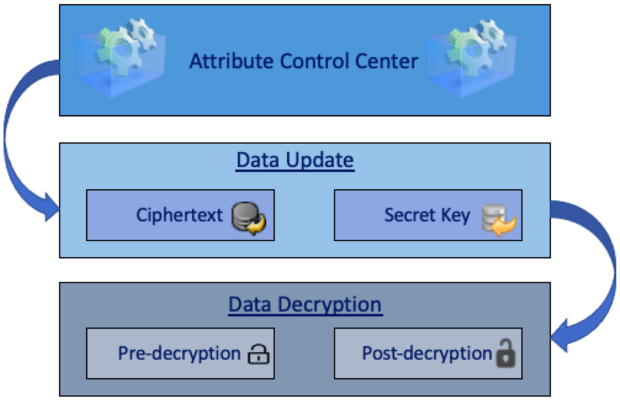

Fig. 3. Token generation process

Attribute Revocation functions are also carried out within the Data Processing Module. The process is depicted as shown in Figure 3. A user provides the attribute to be revoked, and the Attribute Control Center accepts it. These operations are most common when a user leaves an organization, gets a promotion, or organization policies change, so the last attributes need to be revoked. Key Production Unit supplies the master key during the process. The process is completed when the ciphertext and the secret user key that lies with the CSP are updated. The updated private key is then used for decrypting the EHRs at any later time.

C. Cloud Service Provider

The EHR ontology, cryptographic index, and secondary secret key are stored in the CSP. The CSP is located outsides the organization frontier. We consider it to be an honest-but-curious model [42]. The CSP not only runs the programs and algorithms correctly, but it can also examine the information shared in or out of the company. To address this, we enforce an authorization protocol on data inside the organizational periphery, which we refer to as the edge of our framework. As a result, users are checked only within the organization's boundaries, preserving their privacy. While transferring data to the cloud, we introduced a robust encryption mechanism at the organizational edge to shield data from privacy risks.

The EHR ontology that lies within the CSP defines patients, users, and patient records in the medical domain. The nodes of an ontology store the EHR records. The ontology was built previously in our lab following the HIPAA act [23], and it has been slightly modified to accommodate the enhancements. The encrypted index file holds the encrypted word token from each patient's EHR. It also preserves the unique patient id for locating. The dedicated secondary private key for users with several computational advantages in using the scheme resides within the CSP. We will discuss each of the files in detail in the next few subsections.

1) EHR Ontology: HIPAA act has been considered while designing the EHR ontology. It was built previously in our lab [23], and later it was updated to accommodate new enhancements: the knowledge graph stores users, patients, and EHR attributes. The EHR in our study has eight fields. Depending on the user and their attributes, different type of access are given. All such information is recorded within the ontology. The ontology further describes the roles and traits of health organization members and their various relationships.

2) Encrypted Index: The encrypted index file holds encrypted word tokens from each patient EHR along with the patient id. The file is needed for any search operation. Word tokens from each patient are extracted. The tokens are then preprocessed and then encrypted with the aid of the Cryptography Unit following the RSABE scheme [59]. The Key Production Unit supplies the public key during the process. The process is depicted in the Figure 4. The Attribute Control Center via the Regulation Processor obtains all the attributes from the Medical Center Ontology. All these functions are performed within the organization's limit. The file is then stored within the CSP.

V. Implementation

The open-source EHR software is built using the Python Django web-based framework. The application facilitates field-level ABE and access to patient EHR. Due to the presence of Big Data, it is often necessary to search through encrypted data within a limited time and computations; such functions are also available in the software. The user attributes keep changing with time, and it is also allowed in our open-source software. Model-View-Controller (MVC) architectural concepts have been followed while designing the application.

The EHR framework allows doctors to treat their patients securely. It also provides the necessary features that are needed for regular operation. The ABAC controls the field-level access to the EHR. The patient data, along with the encrypted index, are encrypted using the RSABE scheme. Searchable Encryption functions are also allowed using the same scheme in the system. The EHR ontology and the Medical Center Ontology have been built using the Protege [protege.stanford.edu]. Protege is a management framework for knowledge that is open-source. The ontologies are queried using SPARQL with Apache Jena library. The knowledge graph is modified using the SWRL rules. Thus, our application espouses ABE, searchable encryption, attribute revocation, and semantic web for smooth operation.

A. Dataset Description

There are more than eleven thousand patient records in our framework. A patient record has several fields, and we have considered the standard eight fields in all records in our system. There are almost thirty medical users with different attributes. The attribute defines the role and the type of access to the EHR system. All operations on the patient records are performed within the organizational limit before they are kept on the cloud.

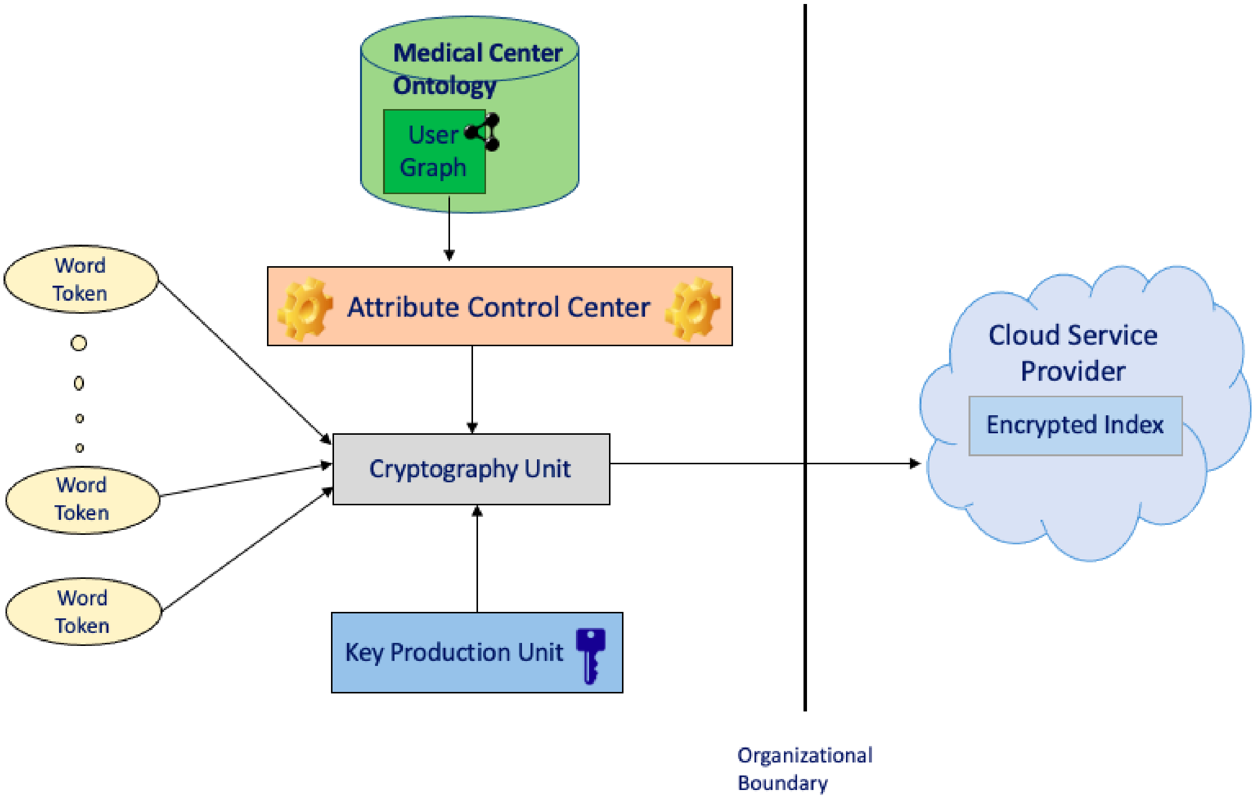

Fig. 4. Token generation process

B. Evaluation

To evaluate the EHR framework, we developed a proof of concept prototype that is described as follows. The EHR management software allows users to treat their patients supporting several critical functions. Suppose a doctor named Sarah submits an access request. The request is comprehensively evaluated in the Authentication Module; username and password are checked with the database; Policy Controller checks the policies, and Medical Center Ontology provides the unique attributes. The Regulation Processor then processes all these pieces of information. If Dr. Sarah plans to decrypt the EHR of Thomas, a patient in the EHR system, the request is processed in the Cryptography Unit by obtaining the Keys from the Key Production Unit. A similar operation is performed for encrypting an EHR. To search through encrypted EHRs, Sarah provides a search query that is processed by the Token Origination Unit by obtaining the secret keys from the Key Production Unit. To revoke the attribute of another Junior Doctor named Jennifer, Sarah submits the revoke attribute information to the Attribute Control Center. The request is processed, and later ciphertext and the secret key that lies with the CSP are updated.

To demonstrate the performance of our system, we calculated the time to produce tokens from the Token Origination Unit. It takes just 0.035 seconds on average, which is acceptable given the high security it enforces.

VI. Conclusion

In this paper, we developed an EHR system that supports field-level ABE, ABAC, searchable encryption, and attribute revocation by employing a knowledge graph that is HIPAA compliant. A knowledge graph produced by our framework accounts for all user roles and attributes in the healthcare organization. It records the attributes of the users along with EHR fields to provide meaningful access to the EHR system. Often in the presence of Big Data, doctors require to search through encrypted data within limited time and computation. Our framework allows search through the data within a short time. User attributes also keep changing with time because some users leave the organization, some may be promoted, or organization policies vary. So, all these changes require attributes to be revoked. Our framework also ingeniously addresses this issue by assigning ciphertext updates and secondary secret key updates to the cloud. The secret key that resides with the user remains firm. Usually, a system with many features has several keys that become an extra management load to the user. By using a single scheme for all the above operations, our framework seems to be more user-friendly. The knowledge graph, encrypted index file, and secondary secret key are stored in the CSP considering the HBC adversary model [42]. We have also assumed the principles of Edge Computing in our framework [51]. Users are verified within the organization's limit to protect privacy. All operations on the data were also performed within the organization frontier before moving it to the cloud to protect against privacy threats.

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.

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] Sanjay P Ahuja, Sindhu Mani, and Jesus Zambrano. A survey of the state of cloud computing in healthcare. *Network and Communication Technologies*, 1(2):12, 2012.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] Alexandra Boldyreva, Vipul Goyal, and Virendra Kumar. Identity-based encryption with efficient revocation. In *Proceedings of the 15th ACM conference on Computer and communications security*, pages 417–426, 2008.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Matthew Green, Susan Hohenberger, Brent Waters, et al. Outsourcing the decryption of abe ciphertexts. In *USENIX security symposium*, volume 2011, 2011.

[19] 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.

[20] 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.

[21] Luan Ibraimi, Milan Petkovic, Svetla Nikova, Pieter Hartel, and Willem Jonker. Mediated ciphertext-policy attribute-based encryption and its application. In *International Workshop on Information Security Applications*, pages 309–323. Springer, 2009.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] Ora Lassila, Ralph R Swick, et al. Resource description framework (rdf) model and syntax specification. 1998.

[27] 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.

[28] Jiguo Li, Xiaonan Lin, Yichen Zhang, and Jinguang Han. Ksf-oabe: Outsourced attribute-based encryption with keyword search function for cloud storage. *IEEE Transactions on Services Computing*, 10(5):715–725, 2016.

[29] 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.

[30] 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.

[31] Jiguo Li, Wei Yao, Yichen Zhang, Huiling Qian, and Jinguang Han. Flexible and fine-grained attribute-based data storage in cloud computing. *IEEE Transactions on Services Computing*, 10(5):785–796, 2016.

[32] 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.

[33] 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.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] Yuxi Li, Fucai Zhou, Yuhai Qin, Muqing Lin, and Zifeng Xu. Integrity-verifiable conjunctive keyword searchable encryption in cloud storage. *International Journal of Information Security*, 17(5):549–568, 2018.

[39] Zhenhua Liu, Shuhong Duan, Peilin Zhou, and Baocang Wang. Traceable-then-revocable ciphertext-policy attribute-based encryption scheme. *Future Generation Computer Systems*, 93:903–913, 2019.

[40] 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.

[41] Zhiquan 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.

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

[43] Deborah L McGuinness, Frank Van Harmelen, et al. Owl web ontology language overview. *W3C recommendation*, 10(10):2004, 2004.

[44] 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):1–12, 2016.

[45] 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.

[46] Matthew Pirretti, Patrick Traynor, Patrick McDaniel, and Brent Waters. Secure attribute-based systems. *Journal of Computer Security*, 18(5):799–837, 2010.

[47] Huiling Qian, Jiguo Li, Yichen Zhang, and Jinguang Han. Privacy-preserving personal health record using multi-authority attribute-based encryption with revocation. *International Journal of Information Security*, 14(6):487–497, 2015.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] Redwan Walid, Karuna Pande Joshi, SeungGeol Choi, and Daeyoung Leroy Kim. Cloud-based encrypted ehr system with semantically rich access control and searchable encryption. *UMBC Student Collection*, 2020.

[56] Hao Wang, Debiao He, Jian Shen, Zhihua Zheng, Xiaoyan Yang, and Man Ho Au. Fuzzy matching and direct revocation: a new cp-abe scheme from multilinear maps. *Soft Computing*, 22(7):2267–2274, 2018.

[57] Hao Wang, Zhihua Zheng, Lei Wu, and Ping Li. New directly revocable attribute-based encryption scheme and its application in cloud storage environment. *Cluster Computing*, 20(3):2385–2392, 2017.

[58] 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.

[59] 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.

[60] Fatos Xhafa, Jianfeng Wang, Xiaofeng Chen, Joseph K Liu, Jin Li, and Paul Krause. An efficient phr service system supporting fuzzy keyword search and fine-grained access control. *Soft computing*, 18(9):1795–1802, 2014.

[61] 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.

[62] Shucheng Yu, Cong Wang, Kui Ren, and Wenjing Lou. Attribute based data sharing with attribute revocation. In *Proceedings of the 5th ACM symposium on information, computer and communications security*, pages 261–270, 2010.

[63] Peng Zhang, Zehong Chen, Kaitai Liang, Shulan Wang, and Ting Wang. A cloud-based access control scheme with user revocation and attribute update. In *Australasian Conference on Information Security and Privacy*, pages 525–540. Springer, 2016.

[64] 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.

[65] 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 virtualiztion management (svm)*, pages 37–45. IEEE, 2012.

[66] Longhui Zu, Zhenhua Liu, and Juanjuan Li. New ciphertext-policy attribute-based encryption with efficient revocation. In *2014 IEEE International Conference on Computer and Information Technology*, pages 281–287. IEEE, 2014.