International Journal of Information Management Data Insights 4 (2024) 100211

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

Leveraging semantic context to establish access controls for secure cloud-based electronic health records

Redwan Walid [a,*], Karuna Pande Joshi [a], Seung Geol Choi [b]

[a] Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, 21250, USA
[b] Department of Computer Science, United States Naval Academy, Annapolis, MD, 21402, USA

ARTICLE INFO

Keywords:
Electronic Health Record (EHR)
Attribute-Based Encryption (ABE)
Attribute-Based Access Control (ABAC)
Searchable Encryption (SE)
Attribute Revocation
Knowledge Graph
Cloud Security
Cloud Computing

ABSTRACT

With the continuous growth of cloud-based Electronic Health Record (EHR) systems and medical data, medical organizations are particularly concerned about storing patient data to provide fast services while adhering to privacy and security concerns. Existing EHR systems often face challenges in handling heterogeneous data and maintaining good performance with data growth. These systems mostly use relational databases or partially store data in a knowledge graph, making it challenging to handle big data and allowing flexible schema expansion. Hence, there is a need to address these problems. This paper provides a solution by proposing a novel graph-based EHR system integrating Attribute-Based Encryption and Semantic Web Technologies to ensure fine-grained EHR field-level security of patient records. Our approach leverages semantic context to query through a knowledge graph that stores encrypted medical data in the nodes, making it possible to handle heterogeneous data while ensuring optimal performance and preserving patient privacy.

1. Introduction

Healthcare providers are increasingly embracing cloud computing, which provides numerous advantages for managing digital data. Cloud storage enables healthcare providers to collect patient data centrally and enhance efficiency compared to local storage (Bahga and Madisetti (2013), Li et al. (2011), Li, Yu, et al. (2012), Löhr et al. (2010)). Its increasing popularity is partly due to its greater flexibility compared to local storage (Ahuja et al. (2012)). Additionally, the pay-as-you-go model has made cloud storage an attractive solution. Cloud environment enhances collaboration among team members as they can access the same infrastructure and work simultaneously. This helps to improve workflow processes and productivity.

Although cloud computing provides numerous advantages, it also poses severe risks to the privacy and security of healthcare providers, as listed in Table 1. Moreover, risk management in health technologies has shown great importance (Reveilhac and Blanchard (2022)). One of the significant issues lies in the vulnerability of patient data to breaches and unauthorized access. The sensitive nature of medical data attracts cybercriminals seeking to use personal information for financial gain or identity theft. Security and privacy violations may have severe consequences for patients and medical organizations. Both patients and medical organizations may suffer financial losses due to the misuse of compromised data. Moreover, patients may distrust the healthcare system due to disclosing their private information, making it more difficult to make correct diagnoses and develop effective treatment plans. In addition to financial losses, healthcare providers found in violation of privacy and security laws such as the Health Insurance Portability and Accountability Act (HIPAA) (Cohen and Mello (2018), Scholl et al. (2008)) and the Health Information Technology for Economic and Clinical Health Act (HITECH) (Saripalle (2019), Burde (2011)) may suffer significant fines and legal consequences, damaging their reputation and possibly threatening their ability to continue offering services. Therefore, Electronic Health Recording (EHR) systems must comply with all relevant rules and regulations while ensuring a smooth and straightforward exchange of patient data.

While various cloud-based attribute-controlled EHR systems have been proposed lately, most systems face challenges such as maintaining steady performance with data growth and handling heterogeneous data. These systems (Walid et al. (2020, 2021), Joshi et al. (2021, 2018a)) use relational databases or a combination of a knowledge graph and flat files to store patient data that have fixed schema, which results in several problems. First of all, these systems fail to handle heterogeneous medical data. Healthcare data is generated by distinct sources and

* Corresponding author. Present address: Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD, 21250, USA.
E-mail address: rwalid1@umbc.edu (R. Walid).

https://doi.org/10.1016/j.jjimei.2023.100211

Available online 11 January 2024

Table 1
Security and privacy issues in cloud computing for healthcare providers.

Issue	Description
Data Security	Protecting sensitive patient data from unwanted access and breaches is essential
Data Encryption	Ensuring data is encrypted both in transit and at rest
Compliance	Adhering to healthcare regulations (e.g., HIPAA and HITECH) to avoid legal and financial penalties
Identity Management	Implementing robust authentication and access control mechanisms to prevent unauthorized access
Data Portability	Ensuring smooth data transfer between cloud providers or from cloud systems back to local systems
Service Level Agreements	Defining security, uptime, and data availability agreements
Data Backups	Regularly taking backups of patient data to prevent data loss
Incident Response	Having a well-defined plan for responding to data breaches or security incidents

diagnostic tools with different formats, coding schemes, and terminologies, which the current systems fail to handle and cannot thus provide precise clinical assessments. Secondly, the current EHR systems cannot maintain steady performance with data growth. Due to the unstable performance scalability of the systems, physicians find it challenging to retrieve patient information quickly and effectively. As a result, the service delivery takes longer to complete. Finally, the systems do not leverage semantic context to allow reasoning through the medical data to preserve privacy and data security. This problem could lead to poor data utilization and privacy and security violations. However, a recent system proposed by Walid et al. (2023) addresses some problems using a complete graph-based approach but has a few limitations. First, they used Semantic Web Rule Language (SWRL) rules in their system to control user access. SWRL has problems such as limited tool support and development resources, resulting in fewer available tools and libraries. SWRL also suffers scalability issues when dealing with large datasets or complex rule sets, leading to performance degradation. Second, the knowledge graph (ontology) used in their system lacked critical data and object properties needed in a comprehensive healthcare environment. Hence, this paper addresses the following research questions.

- How can we use a comprehensive knowledge graph in the EHR system to handle heterogeneous medical data?
- How does storing encrypted data in the nodes of a knowledge graph help to maintain stable data retrieval performance?
- How can the queries made by the users leverage semantic context?
- How does using SPARQL in the EHR system offer more benefits and overcome the limitations of using SWRL?

1.1. EHR System

Multi-layer System. Our EHR system comprises multiple layers involving different entities with various functions. It was designed by implementing a user-id/password-based authentication and setting a policy-defined Attribute-Based Access Control (ABAC) approach (Hu et al. (2015), Yuan and Tong (2005)). The system is divided into four layers, as shown in Fig. 1. Layers 1 through 3 are inside the organizational order, and layer 4 is outside. Users seek access to the system at layer 1 and are authenticated at layer 2. If the operation is allowed, the request is forwarded to layer 3, where modifications are made to the patient EHR and then encrypted using the user's attributes. The CSP at layer 4 is regarded as untrustworthy (Shi and Dustdar (2016)). It works like a data warehouse for keeping the knowledge graph, patient data, and encrypted index file.

Encryption Scheme. Encryption allows healthcare providers to maintain the confidentiality, integrity, and accessibility of patient data while complying with healthcare regulations and standards (Carroll et al. (2011), Salomon (2003)). We use the Revocable, Searchable Attribute-Based Encryption (RSABE) scheme proposed by Wang, Zhang, et al. (2018) to encrypt patient data in our system. The scheme uses Cyphertext Policy - Attribute Based Encryption (CP-ABE) (Bethencourt et al. (2007)), where the policy described in terms of user attributes is used to encrypt data.

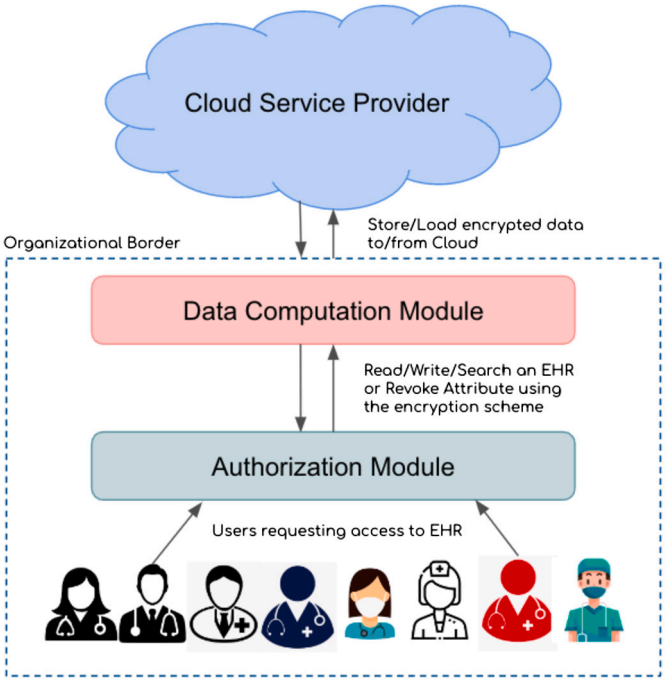

Fig. 1. Multiple layered EHR system.

Knowledge Graph. A knowledge graph is a knowledge base that integrates data using a graph-structured data model (Fensel et al. (2020)). It has been widely used for sharing and reusing data and knowledge in various research areas (Antunes et al. (2022)). The system functions like access control, encryption, user attribute revocation, searchable encryption, and storing data are incorporated with the knowledge graph. HIPAA Act was taken into account when creating the knowledge graph. The knowledge graph contains medical users and their attributes, patient data, other entities, and object properties required in a comprehensive healthcare environment. The knowledge graph restricts access to the system, adhering to the ABAC rules established by the healthcare provider, and leverages semantic context while allowing SPARQL (SPARQL Protocol and RDF Query Language) queries through it. SPARQL is a query language used to retrieve and manipulate data stored in RDF format on the Semantic Web. Several algorithms and tools have also been developed to allow end users to extract useful insights from the knowledge graphs (Niazmand et al. (2022), Jain et al. (2021)).

Using a comprehensive graph-based approach in our system has various advantages (Chen et al. (2020), Hasan et al. (2020)). When searching relevant data, graph-based systems excel, no matter how big or small. It never needs to load unnecessary data for a specific query. Due to the absence of a global index, each vertex holds data about its near neighbors. Hence, the graph's data retrieval performance remains stable as the data set grows. In addition to permitting queries, graph-based systems update Big Data in real time. Similarly, they enable schema growth while serving queries; vertices and nodes can be added and withdrawn as needed.

Searchable Encryption. Searchable encryption enables secure and efficient searches on encrypted data without revealing sensitive data (Wang et al. (2016), Bösch et al. (2014)). We used the same RSABE scheme to enable searchable encryption. To search through the encrypted data, the user first creates a token linked to a keyword query. The token hides the keyword from the CSP for privacy reasons. When the CSP receives the search token, it runs the search algorithm across the ciphertexts to determine which records contain the privately linked keyword(s). The CSP sends the search results when user attributes, encrypted indexes, and keywords fulfill the ciphertext access control policy.

Cloud Outsourcing. Outsourcing computations to the cloud can benefit medical organizations from cost savings and a reduced burden on their IT infrastructure while ensuring that sensitive patient data remains secure (Motahari-Nezhad et al. (2009), Ahmadi and Aslani (2018), Sadiku et al. (2014)). Our system assigns several computations to the CSP, relieving the user's workload. This is done by splitting the user's secret key into two keys, one kept by the user and the other uploaded to the CSP.

Attribute Revocation. Attribute revocation is critical in medical organizations where patient data or organization policy may need to be restricted or updated based on changing conditions and circumstances (Yu et al. (2010), Li, Yao, et al. (2017)). All users and their attributes in our system are recorded in the knowledge graph, which considers all attribute changes along with the corresponding EHR fields. As a result, the knowledge graph helps preserve patient privacy by revoking undesirable user attributes in the encryption policy string.

Edge Computing. Our system uses the edge computing principle (Shi et al. (2016), Cao et al. (2020)), which involves analyzing data at its source before sending it to the cloud. The organizational boundary in the system serves as the edge and incorporates a data access control technique. This technique ensures that only authorized users with the right attributes defined by the organization policy have access. Also, an encryption technique is implemented at the edge to safeguard data integrity and address privacy concerns before the data is transferred to the cloud.

Threat Model. We incorporated the Honest-But-Curious (HBC) threat model into our system. Cloud users commonly categorize CSPs as either following the HBC adversary model or the malicious adversary model (Mather et al. (2009)). In the HBC model, CSPs execute algorithms and programs successfully but can inspect the data sent among parties. On the other hand, in the malicious adversary model, CSPs may behave unpredictably and potentially harm the cloud user.

Additionally, we assume our system will remain resilient in the event of user collaboration aimed at decrypting ciphertext using decryption keys that no individual coalition member can decrypt.

The remainder of the paper is structured as follows – we discuss related work in Section 2, preliminaries in Section 3, system architecture in Section 4, system implementation in Section 5, discussion in Section 6, and conclusion in Section 7.

2. Related Work

2.1. Electronic Health Record

The EHR is a digital version of a patient's medical information, including diagnosis, treatments, prescriptions, immunization history, test results, doctor's notes, etc. EHR systems are frequently used by hospitals to enhance services, boost clinical efficacy, and reduce deductibles (Goroll et al. (2009), Krist et al. (2011)). Privacy and security issues have constrained the development of the EHR system and have drawn a lot of interest recently (Li, Yu, et al., 2012, Li, Ibrahim, et al., 2017, Li, Niu, et al., 2018, Qin et al., 2015). Narayan et al. (2010) advised implementing ABE to protect the confidentiality of EHR data from outside dangers and the CSP. Joshi et al. (2018b) proposed a cloud-based EHR system that employs CP-ABE to encrypt patient data and uses semantic web technologies to enable semantic reasoning; nevertheless, it is devoid of searchable encryption and attribute revocation. Another framework, presented by Wang and Song (2018), uses blockchain to protect the accuracy and traceability of health data but is deficient in controlled access, searchable encryption, and attribute revocation. Walid et al. (2020) suggested an EHR system that uses multiple encryption techniques to provide ABE and searchable encryption, requiring several keys to be handled. Additionally, their system employs flat files to hold patient data and ignores user attribute changes over time. Nowadays, the majority of cloud-based EHR systems lack controlled access, searchable encryption, user attribute revocation, and cloud outsourcing. Additionally, a flat file or a relational database system, which both have their share of problems, is used by most accessible systems.

2.2. Legal Mandates

Patient data is protected by a number of laws, the most significant of which is the HIPAA Act (Centers for Medicare & Medicaid Services (1996)). HIPAA's fundamental objective is to safeguard the confidentiality of personally identifiable health information while ensuring its appropriate use and dissemination to support high-quality medical treatment. HITECH is another act that promotes the use of health information technology in medical organizations, and it also requires that HIPAA Act regulations be enforced more severely and rigorously (Burde (2011)). Nevertheless, these acts do not refer to any encryption methods and data encryption is regarded as an option rather than a need. This allowed different interpretations, which became a contentious issue when electronic health records started evolving.

2.3. Semantic Web Technology

Semantic web technologies contain languages like Resource Description Framework (RDF) (Lassila et al. (1998)) and Web Ontology Language (OWL) (McGuinness et al. (2004)) for building ontologies, defining meta-data using these ontologies and tools for reasoning over these descriptions. Data can be tagged to enable automatic retrieval and utilization in the appropriate settings with the aid of machine-understandable meta-data. Since OWL has well-defined semantics based on first-order logic and model theory, programs confidently conclude that the interpretation is accurate. Compared to other knowledge-representation systems, OWL offers several advantages: it is designed to work with standard implementation technologies, such as OWL QL for databases and OWL RL for rule-based systems; it also has well-defined subset profiles that guarantee sound and complete reasoning at various levels of reasoning sophistication. OWL overtook other technologies as the standard for contemporary electronic health systems by enabling seamless connections to the Web and Cloud. Thus, we used semantic web technologies to build the knowledge graph and reasoning portion of our system.

2.4. Attribute-Based Encryption

ABE is an encryption scheme that allows access to encrypted data based on attributes assigned to users rather than specific identities, enabling more flexible and fine-grained access control. It was developed by Sahai and Waters (Goyal et al., 2006) and is regarded as one of the key EHR security solutions. It eliminates vulnerabilities and guarantees data security. It employs a specific set of attributes to create the private key and a different set of attributes to encrypt data. The ciphertext can only be deciphered if the two sets of attributes match. ABE has been further divided into ciphertext-policy ABE (CP-ABE) (Bethencourt et al. (2007)) and key-policy ABE (KP-ABE) (Attrapadung et al. (2011)) due to lack of clarity. In CP-ABE, the ciphertext is attached to an access policy, and the secret key is attached to the user attributes. A Boolean expression with user attributes typically expresses the access

policy. A secret key can decode a ciphertext if its attributes and access policy attributes match. However, the situation is reversed in the KP-ABE scheme, where the access policy is attached to users' secret keys. CP-ABE is considered more efficient for cloud authorization since each ciphertext establishes an access policy that lists the attributes that data users must possess for the encryption process.

2.5. Attribute Revocation

Attribute revocation involves invalidating or removing assigned attributes from a user or policy string in an access control system, limiting their access to resources or data based on those attributes. As user attributes change over time, ABE systems need attribute revocation. Pirretti et al. (2010) were the first to conduct attribute revocation using a timed rekeying process. Authority centers had to regularly recreate altered keys since each attribute in the system had an expiration date. To revoke an attribute, the authority center had to stop issuing new versions of the attribute and making changes to the existing ones. Bethencourt et al. (2007) proposed another approach by associating the user's private key with a single expiration time. Boldyreva et al. (2008) proposed a revocable KP-ABE system. Wang et al. proposed revocable CP-ABE schemes based on bilinear and multilinear maps (Wang et al. (2017), Wang, He, et al. (2018)). Other CP-ABE approaches, described by Yu et al. (2010) and Ibraimi et al. (2009), performed instantaneous attribute revocation via a semi-trusted proxy server. The workload of the authority was significantly reduced as the proxy server took over the authority's duties. However, they have not been able to obtain access control with more precise granularity. Additionally, the proxy server's update effort drastically rises as the number of users grows. Li, Yao, et al. (2016) developed a CP-ABE technique for user revocation that is less resource-intensive and inexpensive.

2.6. Searchable Encryption

SE is an encryption technique that enables users to search for keywords in cyphertext without revealing the keywords by adding an extra layer of protection. According to Dawes and Sampson (2003), the main barrier to using computers in healthcare systems is time constraints. Any EHR system has to have rapid and efficient searchability since physicians only have limited time to make judgments. A different poll by Holden (2011) mentions that physicians said that reaction time is one of the obstacles to using an EHR system. Consequently, SE is a vital part of EHR systems that can help reduce service delays in clinical settings. The first practical SE method based on symmetric encryption was created by Song et al. (2000). Later, Boneh et al. (2004) contributed to SE research on public-key cryptography. Afterward, several SE techniques were developed to improve search functionality, security, and efficiency (Curtmola et al. (2011), Li et al. (2010), Sun et al. (2013), Li et al. (2014), Fu et al. (2015)). The popularity of attribute-based keyword searches combining ABE and SE functions has increased significantly in recent years (Li, Shi, et al. (2017), Li, Li, et al. (2012), Wang et al. (2014), Zhou et al. (2016), Miao et al. (2016), Li, Lin, et al. (2016), Li, Zhou, et al. (2018)).

3. Initial Groundwork

3.1. Revocable, Searchable ABE Scheme

The RSABE encryption scheme (Wang, Zhang, et al. (2018)) that we used in our EHR system is explained in this section. Due to the unique blend of features that match our system's requirements, we decided to use the RSABE scheme based on our research on Google Scholar, ACM digital library, etc. In dynamic contexts where user attributes may change regularly, RSABE's effective attribute revocation guarantees that data confidentiality and access control are maintained. The support for keyword search over encrypted data enables smooth data retrieval. The scheme's outsourced decryption method also lessens the computational load on the user end and improves system performance. The scheme's proven security guarantees confidence in protecting sensitive data in the cloud. Overall, the effectiveness, security, and adaptability of RSABE make it the best option for boosting the functionality and security of the data in our EHR system.

3.1.1. Syntax

Assume that the security parameter is λ and $\mathcal{X}$ be the attribute universe. A revocable, searchable ABE is composed of these subsequent algorithms:

- $\mathsf{Setup}(1^{\lambda}, \mathcal{X}) \rightarrow (\mathsf{mpk}, \mathsf{msk}, \mathsf{msvk})$. The security parameter λ and the attribute universe $\mathcal{X}$ are given as input to the setup algorithm, which produces the master public parameter mpk, the master secret key msk, and the master secret version key msvk.
  When the user attribute is revoked in the encryption policy string, the master secret version key is updated using the Update-msvk algorithm described below.
- $\mathsf{KeyGen}(\mathsf{msk}, \mathsf{msvk}, x) \rightarrow (\mathsf{sk}_x^1, \mathsf{sk}_x^2)$. The master secret key msk, the master secret version key msvk, and a set of attributes x are given as input to the key generation algorithm, which produces a pair of secret keys $(\mathsf{sk}_x^1, \mathsf{sk}_x^2)$.
  The first key sk_x^1 is stored locally by the user, and the second key sk_x^2 is stored on the cloud server.
- $\mathsf{Enc}(\mathsf{mpk}, \mathsf{msk}, f, m) \rightarrow \mathsf{ct}_f$. The master public parameter mpk, the master secret key msk, a boolean formula f over the attribute universe $\mathcal{X}$, and a message m are given as input to the encryption algorithm, which produces a ciphertext ct_f.
- $\mathsf{EncInd}(\mathsf{mpk}, W) \rightarrow I_W$. The master public parameter mpk and a set of keywords W are given as input to the encrypted index algorithm, which produces an encrypted index I_W for the set of keywords W.
- $\mathsf{Token}(\mathsf{sk}_x^1, w) \rightarrow \mathsf{t}_w$. The user secret key sk_x^1 and a query keyword w are given as input to the token generation algorithm, which produces a token t_w.
- $\mathsf{Test}(\mathsf{sk}_x^2, I_W, \mathsf{t}_w) \rightarrow 0/1$. The cloud secret key sk_x^2, the encrypted index I_W, and the user generated token t_w are given as input to the test algorithm. It produces true if the embedded keyword in the user generated token t_w is present in the encrypted index I_W and false otherwise.
  The encrypted index I_W is usually stored on the cloud server, and the cloud server can execute the test algorithm when it receives the token t_w from the user.
- $\mathsf{Decrypt\text{-}cloud}(\mathsf{sk}_x^2, \mathsf{ct}_f) \rightarrow \mathsf{pd}$. The cloud secret key sk_x^2 and the ciphertext ct_f are given as input to this algorithm. It produces partial decryption pd if $f(x) = 1$; else, it produces an error.
- $\mathsf{Decrypt\text{-}user}(\mathsf{sk}_x^1, \mathsf{pd}) \rightarrow m$. The partial decryption pd and the user secret key sk_x^1 is given to this algorithm, which produces the message m.
- $\mathsf{Update\text{-}msvk}(\mathsf{msvk}, x) \rightarrow \Delta_x$. When the attribute x is revoked, the central authority uses this algorithm to update the master secret version key for the attribute x. The algorithm produces Δ_x, which is used to update the master public key, the cloud secret key that is associated with attribute x, and ciphertexts associated with attribute x.
- $\mathsf{Update\text{-}mpk}(\mathsf{mpk}, \Delta_x)$. The master public key mpk is updated by this algorithm using Δ_x.
- $\mathsf{Update\text{-}cloudkey}(\mathsf{sk}_x^2, \Delta_x)$. The cloud secret key sk_x^2 is updated by this algorithm using Δ_x.
- $\mathsf{Update\text{-}ct}(\mathsf{ct}, \Delta_x)$. The ciphertext ct is updated by this algorithm using Δ_x.

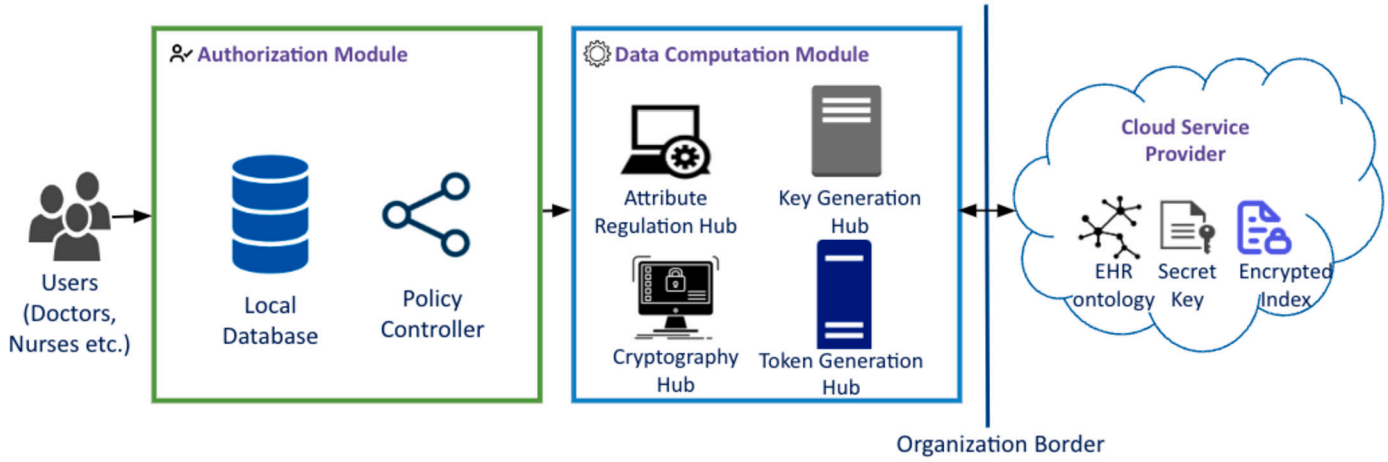

Fig. 2. System architecture.

3.1.2. Revocation Security

Revocation security in ABE makes sure that even if a user has previously been allowed access, they are no longer able to decode freshly encrypted data if their attributes are revoked or updated.

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.

3.1.3. Keyword Search Security

Keyword search security in ABE ensures that authorized users can perform efficient and privacy-preserving searches over encrypted data using keywords, without compromising data confidentiality.

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.

4. System Architecture

The system is designed on the principles of edge computing (Shi et al. (2016)). As shown in Fig. 2, it is divided into two parts: the Authorization Module and Data Computation Module within the organizational boundary and the untrusted CSP outside the organizational boundary. The organization governs the Authorization Module and Data Computation Module. All users are validated inside the organization's border, maintaining their anonymity. Before uploading data to the cloud, an ABE encryption technique is applied to the data within the organization's boundary to protect data integrity from privacy and security concerns.

The framework comprises numerous users, authorities, and data owners within the medical organization. A single CSP stores the user's secondary secret keys, the knowledge graph, and an encrypted index file. The Authorization Module carefully examines each request made to the framework. ABAC policy, defined by the organization, is used to control access to the system, and ABE is used to encrypt data. As the data owner, patients have read access to their EHR.

The framework supports several use cases, such as browsing through encrypted data, reading and writing data, and revoking user attributes in the encryption policy string. A user must submit an access request before being granted access to the system. The Authorization Module analyzes the application by examining the user attributes in the knowledge graph and the ABAC rules established following the organization's regulations. If the attributes adhere to the company's policies, access is permitted.

The system uses the various components within the Data Computation Module to encrypt the revised data of the accessible fields whenever a user modifies an EHR. The Attribute Regulation Hub in the Data Computation Module offers the user attributes during the process. Re-encryption keys are given by the Key Generation Hub. The EHR ontology stored in the cloud that holds patient data is later updated with the new ciphertext. The same steps are taken during a read request.

During the search process, the user enters a query with a search keyword. The Key Generation Hub produces private keys used in the search. A trapdoor is created by the Token Generation Hub using the search keyword and private keys. The trapdoor is then sent to the CSP and compared to the encrypted indices. If a match happens, the process retrieves encrypted EHRs. The user can then decrypt any EHR he wants.

The attribute revocation functionality is performed within the Data Computation Module. The Attribute Regulation Hub receives the attribute that needs to be revoked from the policy string, records it, and then transmits it to the Cryptography Unit. The Key Generation Hub gives the master key required to complete the process. The secondary secret stored in the cloud and the ciphertext are later revised to incorporate the changes.

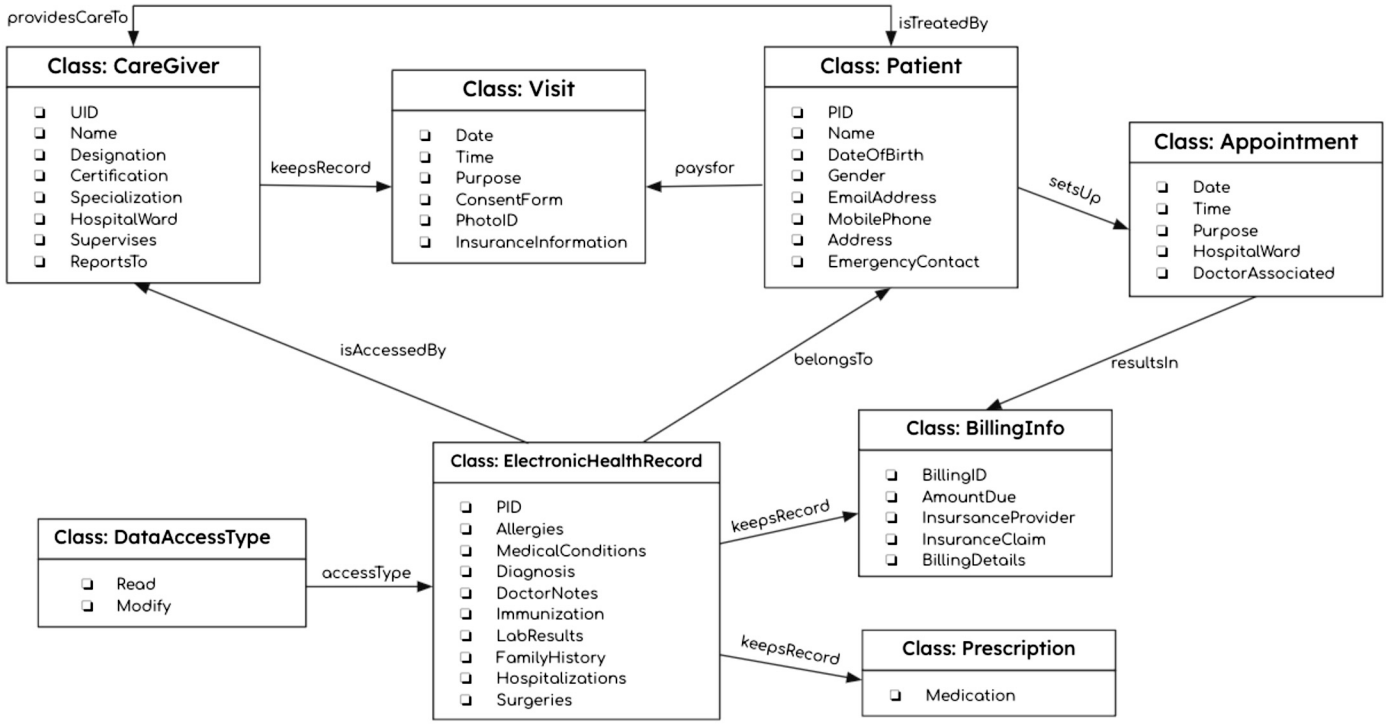

Fig. 3. EHR ontology.

We describe the different modules in our system in the following few subsections.

4.1. Authorization Module

Every login request is carefully examined in this module. Database authentication and the ABAC are the essential goals of the module. First, the user's login information is checked against the database, and then the sub-modules start carrying out their tasks if it succeeds.

The Authorization Module communicates with the Data Computation Module and the cloud-stored ontology in response to queries. For example, a junior physician, based on his attributes, is only permitted access to the information stored in the Doctor Notes field of an EHR to prevent privacy leaks.

The policy of the organization that is described in terms of the user attributes is kept in the Policy Controller. Because each user has unique attributes, the privacy of the data is maintained in the EHR system.

4.2. Data Computation Module

Different hubs of the Data Computation Module allow various system functions such as data cryptography, search token generation, encrypted index building, and attribute revocation. The Policy Controller sends the user attributes to the Attribute Regulation Hub, which then allocates them to the requested system function. The Key Generation Hub generates the encryption keys during the process.

When an EHR is modified, the Attribute Regulation Hub and Key Generation Hub help encrypt the revised data. The ciphertexts are then added as a new node to the EHR ontology in the cloud.

In the event of a search operation, the Key Generation Hub gives the private keys needed to construct a trapdoor to scan through the encrypted index that pulls the encrypted EHRs from the cloud. Later, the user can choose to decrypt any EHR using the Cryptography Hub and Key Generation Hub.

Attribute Revocation in the encryption policy string is also carried out in the Data Computation Module. These actions are required because the user attributes keep changing with time, so the organization's policy also changes. The Key Generation Hub supplies the master key during the process. Later, the private user keys and the ciphertext held by the CSP are modified. The EHRs can then be decrypted at any time using the freshly modified private key.

4.3. Cloud Service Provider

The EHR ontology, the secondary secret keys, and the encrypted index are kept in the CSP. It is considered untrustworthy because it is outside the organization's perimeter. Following the HBC threat model, it runs programs and algorithms correctly but might examine the information exchanged between parties. To address this, we implement an authorization process on data within the organizational boundary. Thus, user privacy is protected because authentication only occurs inside the organization's perimeter. Moreover, we enforced an ABE encryption method at the organization's border to protect data from privacy concerns before storing the data on the cloud.

4.3.1. EHR Ontology

The EHR ontology used to handle heterogeneity in the framework is shown in Fig. 3. It was developed following HIPAA regulations. It covers diverse medical entities, their properties, and their relationships. The ontology stores medical organization users like physicians, nurses, patients, etc. It also records the visit and appointment details of the patients. The purposes of the patient's visit are stored as data properties in the ontology. It records medical user attributes as data properties. User certifications are stored in the ontology, and their examples are MD (Doctor of Medicine), CMA (Certified Medical Assistant), PharmD (Doctor of Pharmacy), CPhT (Certified Pharmacy Technician), etc. Hospital Wards such as Cardiac and Ophthalmology and Specializations such as Gynaecology and Ophthalmology are also stored in ontology as data properties. The Certification, Specialization, and Hospital Wards serve as the attributes of a user. The ontology keeps track of each patient's billing information using data properties like Amount Due, Insurance Provider, Billing Details, etc. Likewise, patients' EHR encrypted fields like Doctor Notes, Diagnosis, Allergies, and Surgeries are stored in the ontology as data properties.

The encrypted patient data are stored in the nodes of the ontology. Using the RSABE scheme discussed earlier, the data is encrypted within the organization boundary. Numerous benefits come with this technique, including fewer files to manage, quicker query results, steady performance regardless of data volume, and flexible schema expansion. Our system writes dynamic SPARQL queries for the ontology based on the user account and type of request. SPARQL is a standardized and interoperable approach for examining and analyzing data inside the Semantic Web ecosystem due to its interaction with other Semantic Web technologies, including RDF, RDFS, and OWL. Compared to SWRL, SPARQL has several advantages. Its simplicity and convenience of use enable users to write queries efficiently. Besides, it can manage large-scale RDF datasets, offering scalable and effective methods for manipulating and retrieving data. In comparison, SWRL can be more challenging to deal with and may have scalability issues even if it is an effective tool for expressing complex rules. Thus, SPARQL is well-known for its simplicity, speed, and broad accessibility.

The Semantic Web's vision, knowledge integration, and the enormous potential of linked data on the web are all made possible by SPARQL. SPARQL and SQL are similar regarding query syntax, such as SELECT statements, FROM clauses, and WHERE clauses used to indicate data retrieval. Both languages enable joining and filtering techniques. In contrast to SQL, which is geared toward relational databases, SPARQL is intended for querying data in RDF format. Despite their shared characteristics, SPARQL and SQL serve various data models and have diverse use cases within their respective domains. The following statements show example SPARQL queries to update the node of the Allergies field of the EHR of a patient with ID 100. To update into an EHR field, SPARQL needs a delete query before the insert query.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX cc: <http://www.semanticweb.org/researchuser/ontologies
     /2023/1/encnode#>


DELETE
    {cc:100 cc:Allergies ?object}
WHERE
    {cc:100 cc:Allergies ?object .}
```

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX cc: <http://www.semanticweb.org/researchuser/ontologies
     /2023/1/encnode#>


INSERT
    {cc:100 cc:Allergies "The patient is allergic to Ibuprofen."}
WHERE
    {cc:100 cc:Allergies ?object .}
```

4.3.2. Encrypted Index

The encrypted index file is needed to search through the encrypted patient EHRs. To create the file, the patient's EHRs are initially preprocessed using several steps such as tokenization, lowercasing, removing punctuations, etc. Then, the Key Generation Hub supplies the public key, and the EHR Ontology supplies the attributes to the Attribute Regulation Hub. Later, the word tokens are encrypted using the RSABE scheme. The encrypted word tokens and the corresponding patient IDs form a data frame and are stored in the encrypted index file. The entire process is done within the organization boundary, and the file is stored in the CSP.

5. Implementation

The EHR system was developed utilizing the Python Django framework and Model-View-Controller (MVC) design principles (Deacon (2009)). The system enables data encryption to patient health records using ABE and field-level ABAC. The encrypted index file is generated, and the patient data are encrypted using the RSABE encryption technique. The system uses a searchable encryption approach for quicker and more effective search over encrypted data. The system also considers user attributes change over time and revokes unwanted ones in the encryption policy string.

The system allows physicians to treat patients quickly. It has every functionality required for day-to-day operations. The EHR ontology was built using Protege [protege.stanford.edu], an open-source knowledge graph management tool. With the Apache Jena library, SPARQL is used to query the ontology. Hence, the system uses ABE, searchable encryption, attribute revocation, knowledge graph, and semantic web for smooth operation.

5.1. Dataset Description

We generated synthetic graph datasets of various sizes using the MIMIC-III dataset (Johnson et al. (2016)). Our experiments used data with 40,000, 80,000, 120,000, 160,000, and 200,000 patient records. Based on the medical history, each patient has multiple fields in their EHR, including Immunization, Surgeries, Diagnosis, Precriptions, etc. Following the edge computing principle, the EHRs are encrypted within the organization boundary, inserted into nodes of the knowledge graph, and stored in the cloud server. Our system has 30 medical users, such as physicians, nurses, etc. Each medical user has several attributes, such as specialization, certification, and hospital ward attributes, determining their access to the system.

5.2. Evaluation

To evaluate the EHR system, we created a proof-of-concept. Consider that Richard, a physician, wants to use the system. The request is examined in the authorization module, the database verifies the username and password, the Policy Controller checks the organization's policies, and the EHR Ontology provides the attributes. If Richard wishes to access the EHR of any patient, the request is fulfilled in the Cryptography Hub by obtaining the private keys from the Key Generation Hub. Similar steps are taken to encrypt an EHR. To search through encrypted data, Richard sends a search request to the Token Generation Hub to generate the search token needed to complete the process. To revoke a user attribute from the encryption policy string, Richard uploads the revoked attribute to the Attribute Regulation Hub. Later, the ciphertext and the secret keys in the cloud are updated to complete the entire process.

We assessed the performance of queries with varying data sizes in order to demonstrate the advantages of our proposed system over that suggested by Walid et al. (2021). Both systems used the same RSABE scheme for all system functions and the same dataset of different sizes. The right section in Table 2 shows the performance of our proposed system for various data sizes for different types of queries where we stored all patient data in the encrypted nodes of a knowledge graph. The same table in the left section shows the performance of the system proposed by Walid et al. (2021), where they use SWRL rules in the system and a partial knowledge graph-based approach, i.e., encrypted patient data reside as flat files in their system. The query performances on the tables are recorded in seconds using an average of ten queries.

The SPARQL queries in the two systems have distinct processes that affect their performances, as shown in Table 2. The retrieve query in our proposed system represents the time to decrypt an encrypted EHR field stored as a node in the knowledge graph using a SPARQL query. Meanwhile, the retrieve process in the system by Walid et al. means the time to find the same patient's EHR field stored as a flat file inside a directory and decrypt the data. Compared to the system by Walid et al. (2021), the performance of our proposed system for data retrieval is better and unaffected by the data sizes. Since both systems rely on the same encrypted index file that was created using the same scheme and data,

Table 2
Comparison of performance between our novel system and system proposed by Walid et al.

Data size	Walid et al. system's performance			Our novel system's performance		
	Decrypt (and Retrieve) (s)	Search (s)	Revoke (s)	Decrypt (and Retrieve) (s)	Search (s)	Revoke (s)
40,000	0.0172066	0.1847897	0.0193267	0.0103081	0.1847897	0.0106127
80,000	0.0174678	0.3376495	0.0184021	0.0106438	0.3376495	0.0102164
120,000	0.0173397	0.4931780	0.0176004	0.0103354	0.4931780	0.0105571
160,000	0.0182211	0.6392195	0.0175358	0.0102828	0.6392195	0.0105158
200,000	0.0186612	0.7903189	0.0167967	0.0108574	0.7903189	0.0100075

Table 3
Percentage decrease in query time for different data sizes.

Use Case	40,000	80,000	120,000	160,000	200,000
Decrypt (and Retrieve)	40.09%	39.07%	40.40%	43.57%	41.82%
Revoke	45.09%	44.48%	40.02%	40.03%	40.42%

their search performances are identical. Nonetheless, our proposed system would perform better if a user wants to decrypt an EHR from the search result. In addition, our new system outperforms the Walid et al. system regarding revocation performance. We noted the time to revoke an attribute in the encryption policy string, update the ciphertext and secret key, and decrypt a patient EHR field.

Based on our experiments, our proposed system shows many benefits of using a graph-based approach. We have shown the percentage decrease in time for our system in Table 3 for different sizes based on the Decrypt and Revoke use cases run by the SPARQL queries. The data retrieval time is almost reduced by 40% on average for different data sizes, and query execution time stays nearly constant. The graph-based system has no global index; each vertex keeps data about its neighbor nodes. Therefore, it does not need to process unrelated data for a particular query. As a result, the query performances remain the same. The revoke query also took 40% less time on average in our system for different data sizes.

6. Discussion

We developed a novel EHR system that addresses problems associated with the current cloud-based attribute-controlled EHR systems. We explain our contributions and practical implications in the new few sub-sections.

6.1. Contributions to Literature

Our system is a novel version of the current cloud-based attribute-controlled EHR systems. Current systems available in the literature (Walid et al. (2023, 2020, 2021), Joshi et al. (2021, 2018a, 2017), Bahga and Madisetti (2013), Li, Yu, et al. (2012), Narayan et al. (2010), Dixit et al. (2022, 2019)) use relational databases or a combination of a knowledge graph and flat files to store patient data. These systems have fixed schema, leading to problems like inflexibility in adapting to changes, difficulty scaling with data growth, poor data retrieval performance, etc. Moreover, these systems cannot handle heterogeneous medical data, resulting in fragmented patient records, compromised clinical decision-making, etc. Also, these systems use SWRL rules, which have limited tool support and development resources and suffer from scalability issues when handling large datasets. Furthermore, these systems do not take advantage of cloud outsourcing and may have performance bottlenecks during peak usage, higher infrastructure costs, increased latency for critical tasks, etc.

Handling heterogeneous data in an EHR system enables comprehensive and interoperable data management, accommodating various sources and growing healthcare practices. Our system handles data heterogeneity by using a graph-based approach. This fosters integrated patient records, improves data retrieval performance, enhances clinical decision making, and mitigates problems related to fragmentation in medical information.

Flexible data schema enables the EHR system to adapt to evolving healthcare requirements, enhance data quality for accurate patient records, and support evolving medical standards and guidelines. Our system allows flexible expansion of data schema. For example, medical users like doctors and nurses can get degrees and certifications over time, which change their attributes, and our system addresses all these changes. Moreover, all users and patients may also have different attributes. For example, a patient may have only one EHR field, like allergies, whereas another may have multiple EHR fields, like allergies, diagnosis, prescriptions, etc.

Efficiently querying patient data enhances system performance and scalability. The simplicity, speed, and wide accessibility of SPARQL make it a better choice compared to SWRL in our EHR system. It enables users to write queries quickly. It can also handle large RDF datasets, offering scalable and effective methods for manipulating and retrieving data.

Delegating partial computing to the cloud enables EHR systems to improve scalability, performance, and cost-efficiency. It cuts the infrastructure cost, enhances fault tolerance, data security, and compliance. Moreover, it reduces client-side computations and allows medical organizations to offload the burden of managing complex infrastructure and resources to the cloud server. Our system requires fewer computations as some functions are safely delegated to the cloud. For example, our system assigns partial decryption tasks to the cloud with the help of the secondary secret keys stored in the cloud.

6.2. Implications for Practice

Our novel system can be used by any healthcare organization that plans to use an attribute-based and cloud-based system. It can help medical organizations to comply with HIPAA and HITECH regulations. The system would use user attributes to control access to an EHR at the field level instead of the traditional document level. The system would also address user attribute changes and leverage semantic context in the queries to limit unnecessary data access. The physicians can search over encrypted data without compromising privacy and security issues. The physicians could also provide faster and more scalable services to the patients.

7. Conclusion

This paper proposes an EHR system that uses a knowledge graph to manage system features such as ABE, ABAC, searchable encryption, and attribute revocation. The system handles heterogeneous data and enables flexible database schema expansion. The graph in the system stores the patient data as encrypted nodes, providing advantages such as constant data retrieval performance. The system allows physicians to search through encrypted data using searchable encryption. The graph records user attribute changes, which keep changing with time. Thus, the organization's policies change sometimes, so the system revokes user attributes from the policy string. A highly functioning system frequently has several keys, which adds to the administrative load. Our system appears more simple by using a single scheme for all features.

Our system also uses less local computing by sending tasks like partial decryption to the cloud server. Considering the HBC adversary model, the secondary secret keys, the encrypted index file, and the knowledge graph are stored in the cloud server. Based on the principles of edge computing, data computations are done locally before being stored on the cloud.

7.1. Future Work

In the future, we want to expand our research in several ways. Integration with existing systems, historical databases, and interoperability standards are frequent requirements when deploying a new EHR system in healthcare organizations. We plan to address potential difficulties that may arise in putting the novel system into practice and integrating it with the current healthcare system. For any EHR system, it is essential to ensure that it is user-friendly and intuitive for healthcare professionals to navigate and utilize effectively. We plan to design the user interface and run a user-centric evaluation to address usability challenges. Although we have demonstrated the system performance with the MIMIC-III dataset, we plan to assess the system scalability and performance with larger datasets and real-world deployment scenarios. We plan to get feedback and validate our system with medical practitioners.

CRediT authorship contribution statement

Redwan Walid: Conceptualization, Data curation, Methodology, Project administration, Software, Writing – original draft, Writing – review & editing. **Karuna Pande Joshi:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Resources. **Seung Geol Choi:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation.

Declaration of competing interest

The authors of this article have no conflict of interest to declare.

References

Ahmadi, M., & Aslani, N. (2018). Capabilities and advantages of cloud computing in the implementation of electronic health record. *Acta Informatica Medica*, *26*(1), 24.

Ahuja, S. P., Mani, S., & Zambrano, J. (2012). A survey of the state of cloud computing in healthcare. *Network and Communication Technologies*, *1*(2), 12.

Antunes, A. L., Cardoso, E., & Barateiro, J. (2022). Incorporation of ontologies in data warehouse/business intelligence systems-a systematic literature review. *International Journal of Information Management Data Insights*, *2*(2), Article 100131.

Attrapadung, N., Libert, B., & De Panafieu, E. (2011). Expressive key-policy attribute-based encryption with constant-size ciphertexts. In *International workshop on public key cryptography* (pp. 90–108). Springer.

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

Bethencourt, J., Sahai, A., & Waters, B. (2007). Ciphertext-policy attribute-based encryption. In *2007 IEEE symposium on security and privacy (SP'07)* (pp. 321–334). IEEE.

Boldyreva, A., Goyal, V., & Kumar, V. (2008). Identity-based encryption with efficient revocation. In *Proceedings of the 15th ACM conference on computer and communications security* (pp. 417–426).

Boneh, D., Di Crescenzo, G., Ostrovsky, R., & Persiano, G. (2004). Public key encryption with keyword search. In *International conference on the theory and applications of cryptographic techniques* (pp. 506–522). Springer.

Bösch, C., Hartel, P., Jonker, W., & Peter, A. (2014). A survey of provably secure searchable encryption. *ACM Computing Surveys (CSUR)*, *47*(2), 1–51.

Burde, H. (2011). The hitech act: An overview. *AMA Journal of Ethics*, *13*(3), 172–175.

Cao, K., Liu, Y., Meng, G., & Sun, Q. (2020). An overview on edge computing research. *IEEE Access*, *8*, 85714–85728.

Carroll, M., Van Der Merwe, A., & Kotze, P. (2011). Secure cloud computing: Benefits, risks and controls. In *2011 information security for South Africa* (pp. 1–9). IEEE.

Centers for Medicare & Medicaid Services (1996). The health insurance portability and accountability act of 1996 (HIPAA). Retrieved from http://www.cms.hhs.gov/hipaa/.

Chen, Z., Wang, Y., Zhao, B., Cheng, J., Zhao, X., & Duan, Z. (2020). Knowledge graph completion: A review. *IEEE Access*, *8*, 192435–192456.

Cohen, I. G., & Mello, M. M. (2018). Hipaa and protecting health information in the 21st century. *JAMA*, *320*(3), 231–232.

Curtmola, R., Garay, J., Kamara, S., & Ostrovsky, R. (2011). Searchable symmetric encryption: Improved definitions and efficient constructions. *Journal of Computer Security*, *19*(5), 895–934.

Dawes, M., & Sampson, U. (2003). Knowledge management in clinical practice: A systematic review of information seeking behavior in physicians. *International Journal of Medical Informatics*, *71*(1), 9–15.

Deacon, J. (2009). Model-view-controller (mvc) architecture. [Citado em: 10 de março de 2006]. Retrieved from http://www.jdl.co.uk/briefings/MVC.pdf.

Dixit, S., Joshi, K. P., & Choi, S. G. (2019). Multi authority access control in a cloud ehr system with ma-abe. In *2019 IEEE international conference on edge computing (EDGE)* (pp. 107–109). IEEE.

Dixit, S., Joshi, K. P., Choi, S. G., & Elluri, L. (2022). Semantically rich access control in cloud ehr systems based on ma-abe. In *2022 IEEE 8th intl conference on big data security on cloud (BigDataSecurity), IEEE intl conference on high performance and smart computing, (HPSC) and IEEE intl conference on intelligent data and security (IDS)* (pp. 1–10). IEEE.

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., Wahler, A., Fensel, D., et al. (2020). Introduction: What is a knowledge graph?. In *Knowledge graphs: Methodology, tools and selected use cases* (pp. 1–10).

Fu, Z., Sun, X., Liu, Q., Zhou, L., & Shu, J. (2015). Achieving efficient cloud search services: Multi-keyword ranked search over encrypted cloud data supporting parallel computing. *IEICE Transactions on Communications*, *98*(1), 190–200.

Goroll, A. H., Simon, S. R., Tripathi, M., Ascenzo, C., & Bates, D. W. (2009). Community-wide implementation of health information technology: The Massachusetts ehealth collaborative experience. *Journal of the American Medical Informatics Association*, *16*(1), 132–139.

Goyal, V., Pandey, O., Sahai, A., & Waters, B. (2006). Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM conference on computer and communications security* (pp. 89–98).

Hasan, S. S., Rivera, D., Wu, X.-C., Durbin, E. B., Christian, J. B., & Tourassi, G. (2020). Knowledge graph-enabled cancer data analytics. *IEEE Journal of Biomedical and Health Informatics*, *24*(7), 1952–1967.

Holden, R. J. (2011). 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.

Hu, V. C., Kuhn, D. R., Ferraiolo, D. F., & Voas, J. (2015). Attribute-based access control. *Computer*, *48*(2), 85–88.

Ibraimi, L., Petkovic, M., Nikova, S., Hartel, P., & Jonker, W. (2009). Mediated ciphertext-policy attribute-based encryption and its application. In *International workshop on information security applications* (pp. 309–323). Springer.

Jain, S., Seeja, K., & Jindal, R. (2021). A fuzzy ontology framework in information retrieval using semantic query expansion. *International Journal of Information Management Data Insights*, *1*(1), Article 100009.

Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Anthony Celi, L., & Mark, R. G. (2016). Mimic-iii, a freely accessible critical care database. *Scientific Data*, *3*(1), 1–9.

Joshi, M., Joshi, K., & Finin, T. (2018a). Attribute based encryption for secure access to cloud based ehr systems. In *2018 IEEE 11th international conference on cloud computing (CLOUD)* (pp. 932–935).

Joshi, M., Joshi, K., & Finin, T. (2018b). Attribute based encryption for secure access to cloud based ehr systems. In *2018 IEEE 11th international conference on cloud computing (CLOUD)* (pp. 932–935). IEEE.

Joshi, M., Joshi, K. P., & Finin, T. (2021). Delegated authorization framework for ehr services using attribute-based encryption. *IEEE Transactions on Services Computing*, *14*(6), 1612–1623.

Joshi, M., Mittal, S., Joshi, K. P., & Finin, T. (2017). Semantically rich, oblivious access control using abac for secure cloud storage. In *2017 IEEE international conference on edge computing (EDGE)* (pp. 142–149). IEEE.

Krist, A. H., Peele, E., Woolf, S. H., Rothemich, S. F., Loomis, J. F., Longo, D. R., & Kuzel, A. J. (2011). Designing a patient-centered personal health record to promote preventive care. *BMC Medical Informatics and Decision Making*, *11*(1), 1–11.

Lassila, O., Swick, R.R. et al. (1998). Resource description framework (rdf) model and syntax specification.

Li, H., Liu, D., Dai, Y., Luan, T. H., & Shen, X. S. (2014). 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.

Li, J., Li, J., Chen, X., Jia, C., & Liu, Z. (2012). Efficient keyword search over encrypted data with fine-grained access control in hybrid cloud. In *International conference on network and system security* (pp. 490–502). Springer.

Li, J., Lin, X., Zhang, Y., & Han, J. (2016). Ksf-oabe: Outsourced attribute-based encryption with keyword search function for cloud storage. *IEEE Transactions on Services Computing*, *10*(5), 715–725.

Li, J., Shi, Y., & Zhang, Y. (2017). Searchable ciphertext-policy attribute-based encryption with revocation in cloud storage. *International Journal of Communication Systems*, *30*(1), Article e2942.

Li, J., Wang, Q., Wang, C., Cao, N., Ren, K., & Lou, W. (2010). Fuzzy keyword search over encrypted data in cloud computing. In *2010 proceedings IEEE INFOCOM* (pp. 1–5). IEEE.

Li, J., Yao, W., Han, J., Zhang, Y., & Shen, J. (2017). User collusion avoidance cp-abe with efficient attribute revocation for cloud storage. *IEEE Systems Journal*, *12*(2), 1767–1777.

Li, J., Yao, W., Zhang, Y., Qian, H., & Han, J. (2016). Flexible and fine-grained attribute-based data storage in cloud computing. *IEEE Transactions on Services Computing*, *10*(5), 785–796.

Li, M., Yu, S., Cao, N., & Lou, W. (2011). Authorized private keyword search over encrypted data in cloud computing. In *2011 31st international conference on distributed computing systems* (pp. 383–392). IEEE.

Li, M., Yu, S., Zheng, Y., Ren, K., & Lou, W. (2012). 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.

Li, X., Ibrahim, M. H., Kumari, S., Sangaiah, A. K., Gupta, V., & Choo, K.-K. R. (2017). Anonymous mutual authentication and key agreement scheme for wearable sensors in wireless body area networks. *Computer Networks*, *129*, 429–443.

Li, X., Niu, J., Kumari, S., Wu, F., & Choo, K.-K. R. (2018). A robust biometrics based three-factor authentication scheme for global mobility networks in smart city. *Future Generation Computer Systems*, *83*, 607–618.

Li, Y., Zhou, F., Qin, Y., Lin, M., & Xu, Z. (2018). Integrity-verifiable conjunctive keyword searchable encryption in cloud storage. *International Journal of Information Security*, *17*(5), 549–568.

Löhr, H., Sadeghi, A.-R., & Winandy, M. (2010). Securing the e-health cloud. In *Proceedings of the 1st acm international health informatics symposium* (pp. 220–229).

Mather, T., Kumaraswamy, S., & Latif, S. (2009). *Cloud security and privacy: An enterprise perspective on risks and compliance*. O'Reilly Media, Inc.

McGuinness, D. L., Van Harmelen, F., et al. (2004). Owl web ontology language overview. *W3C Recommendation*, *10*(10).

Miao, Y., Ma, J., Liu, X., Wei, F., Liu, Z., & Wang, X. A. (2016). 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.

Motahari-Nezhad, H. R., Stephenson, B., & Singhal, S. (2009). Outsourcing business to cloud computing services: Opportunities and challenges. *IEEE Internet Computing*, *10*(4), 1–17.

Narayan, S., Gagné, M., & Safavi-Naini, R. (2010). Privacy preserving ehr system using attribute-based infrastructure. In *Proceedings of the 2010 ACM workshop on cloud computing security workshop* (pp. 47–52).

Niazmand, E., Sejdiu, G., Graux, D., & Vidal, M.-E. (2022). Efficient semantic summary graphs for querying large knowledge graphs. *International Journal of Information Management Data Insights*, *2*(1), Article 100082.

Pirretti, M., Traynor, P., McDaniel, P., & Waters, B. (2010). Secure attribute-based systems. *Journal of Computer Security*, *18*(5), 799–837.

Qin, B., Deng, H., Wu, Q., Domingo-Ferrer, J., Naccache, D., & Zhou, Y. (2015). Flexible attribute-based encryption applicable to secure e-healthcare records. *International Journal of Information Security*, *14*(6), 499–511.

Reveilhac, M., & Blanchard, A. (2022). The framing of health technologies on social media by major actors: Prominent health issues and COVID-related public concerns. *International Journal of Information Management Data Insights*, *2*(1), Article 100068.

Sadiku, M. N., Musa, S. M., & Momoh, O. D. (2014). Cloud computing: Opportunities and challenges. *IEEE Potentials*, *33*(1), 34–36.

Salomon, D. (2003). *Data privacy and security: Encryption and information hiding*. Springer Science & Business Media.

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

Scholl, M.A., Stine, K.M., Hash, J., Bowen, P., Johnson, L.A., Smith, C.D., & Steinberg, D.I. (2008). Sp 800-66 rev. 1. An introductory resource guide for implementing the health insurance portability and accountability act (hipaa) security rule.

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, *3*(5), 637–646.

Shi, W., & Dustdar, S. (2016). The promise of edge computing. *Computer*, *49*(5), 78–81.

Song, D. X., Wagner, D., & Perrig, A. (2000). Practical techniques for searches on encrypted data. In *Proceeding 2000 IEEE symposium on security and privacy. S&P 2000* (pp. 44–55). IEEE.

Sun, W., Wang, B., Cao, N., Li, M., Lou, W., Hou, Y. T., & Li, H. (2013). 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* (pp. 71–82).

Walid, R., Joshi, K. P., & Choi, S. G. (2021). Secure cloud ehr with semantic access control, searchable encryption and attribute revocation. In *2021 IEEE international conference on digital health (ICDH)* (pp. 38–47).

Walid, R., Joshi, K. P., & Geol Choi, S. (2023). Semantically rich differential access to secure cloud ehr. In *2023 IEEE 9th intl conference on big data security on cloud (BigDataSecurity), IEEE intl conference on high performance and smart computing, (HPSC) and IEEE intl conference on intelligent data and security (IDS)* (pp. 1–9).

Walid, R., Joshi, K. P., Geol Choi, S., & Kim, D-y. (2020). Cloud-based encrypted ehr system with semantically rich access control and searchable encryption. In *2020 IEEE international conference on big data (Big Data)* (pp. 4075–4082).

Wang, H., He, D., Shen, J., Zheng, Z., Yang, X., & Au, M. H. (2018). Fuzzy matching and direct revocation: A new cp-abe scheme from multilinear maps. *Soft Computing*, *22*(7), 2267–2274.

Wang, H., & Song, Y. (2018). Secure cloud-based ehr system using attribute-based cryptosystem and blockchain. *Journal of Medical Systems*, *42*(8), 1–9.

Wang, H., Zheng, Z., Wu, L., & Li, P. (2017). New directly revocable attribute-based encryption scheme and its application in cloud storage environment. *Cluster Computing*, *20*(3), 2385–2392.

Wang, Q., Zhu, Y., & Luo, X. (2014). Multi-user searchable encryption with fine-grained access control without key sharing. In *2014 3rd international conference on advanced computer science applications and technologies* (pp. 145–150). IEEE.

Wang, S., Zhang, D., Zhang, Y., & Liu, L. (2018). Efficiently revocable and searchable attribute-based encryption scheme for mobile cloud storage. *IEEE Access*, *6*, 30444–30457.

Wang, Y., Wang, J., & Chen, X. (2016). Secure searchable encryption: A survey. *Journal of Communications and Information Networks*, *1*, 52–65.

Yu, S., Wang, C., Ren, K., & Lou, W. (2010). Attribute based data sharing with attribute revocation. In *Proceedings of the 5th ACM symposium on information, computer and communications security* (pp. 261–270).

Yuan, E., & Tong, J. (2005). Attributed based access control (abac) for web services. In *IEEE international conference on web services (ICWS'05)*. IEEE.

Zhou, F., Li, Y., Liu, A. X., Lin, M., & Xu, Z. (2016). Integrity preserving multi-keyword searchable encryption for cloud computing. In *International conference on provable security* (pp. 153–172). Springer.