Privacy-Aware Cloud Architecture for Collaborative Use of Patients' Health Information

Fadi Alhaddadin * and Jairo Gutierrez

Posted Date: 6 May 2023

doi: 10.20944/preprints202305.0416.v1

Keywords: Cloud Architecture; Data Privacy; Data Confidentiality; Information Sharing; Health Information; Patient Records

Article

Privacy-Aware Cloud Architecture for Collaborative Use of Patients' Health Information

Fadi Alhaddadin [1]* and Jairo Gutierrez [2]

[1] CUC-Ulster University; fadi.alhaddadin@cuc-ulster.edu.qa
[2] Auckland University of Technology, jairo.gutierrez@aut.ac.nz
* Correspondence: fadi.alhaddadin@cuc-ulster.edu.qa, Tel.: +974 40 198 189

Abstract: Cloud computing appears to be the dreamed vision of the healthcare industry; it refers to means of storing and accessing data and programs over the internet instead of the computer's hard drive. However, the adoption of cloud computing requires solving several issues and information privacy is a major one. This work proposes a cloud architecture design for the healthcare information system. The proposed architecture enables for storing and sharing information in a privacy-preserving manner. Patients' information in the proposed architecture is divided into four categories identified in the case study data analysis. User identity management protocol (U-IDM) is employed for controlling the access to patients' information, and patients have means of control over who can access their information. A scenario-based instantiation validated the proposed architecture's privacy-preserving patient data exchange. The instantiation proved that the proposed architecture allows sharing healthcare information without violating the privacy of patients.

Keywords: cloud architecture; data privacy; data confidentiality; information sharing; health information; patient records

1. Introduction

Recently, the healthcare sector has shown a growing interest in information technologies to facilitate new methods of collecting, managing, and analyzing health-related information. In fact, the healthcare sector is under pressure to embrace many new technologies that are available on the market such as the Internet of Things (IoT) [1], and as consequence to such embracement, the amount of healthcare information is rapidly growing in details and diversity, driven by record keeping, compliance, regulatory requirements, and of course, patient care [2]. Such information generates special value when it is exchanged and collaboratively used among different parties involved in the healthcare area [3]. Several researchers consider immediate access to previously generated medical records during healthcare service delivery as highly important, it leads to effective ways of preventing and managing illnesses, as well as the discovery of new drugs and therapies [4].

In the healthcare domain, patients usually acquire medical care from a wide range of caregivers based on their proximity, quality of care received, cultural attitudes and bedside manner. Medical care may be received from various caregivers such as hospitals, pharmacy, laboratory, physician groups, nurses, school clinics, and public health places. This has led to the fragmentation of patients' information in heterogeneous systems. The majority of this collected information is stored in heterogeneous distributed health information systems which are mainly proprietary [5]. As a consequence, health-related information stored in these different systems cannot be easily accessed to present a clear and complete picture of an individual patient when needed. For example, when a patient visits a healthcare provider such as general practitioner, he or she often requires additional medical services or attention over a period of time whether it is specialized medical examination such as magnetic resonance imaging scans, or a routine medical examination such as cholesterol test and blood sugar checks.

The concept of sharing information in the healthcare domain helps to better understand the health needs and therefore improve the quality of care provided to patients [6]. For that, the seamless exchange of multimedia clinical information is considered as a fundamental requirement. Different

 doi:10.20944/preprints202305.0416.v1

technological approaches can be adopted for enabling the communication and sharing of health records segments [7]. However, there are a number of challenges that need to be overcome before obtaining the best of what sharing information in the healthcare can offer, of which interoperability and data privacy are major ones. Interoperability is defined as the ability for two or more systems of components to exchange information and use the information that has been exchanged [8]. It is the ability to share and use information across multiple system technologies seamlessly. Interoperability is a fundamental requirement for the health care system to derive the societal benefits promised by the adoption of electronic healthcare records [7].

1.1. Information Privacy

Information privacy is the desire of individuals to control or have some influence over data about themselves [9]. It is, in other words, the right of individuals to determine how and to what extent information they communicate to others is used. Healthcare data includes sensitive records that should not be made available to unauthorized people to protect the privacy of patients. Information privacy protection is very essential to build users' trust in order to reach the full potential of information sharing in the healthcare domain [10] [11]. Therefore, it is a mandatory step to adhere to legal frameworks such as the Health Insurance Portability and Accountability Act (HIPPA) [12] and the Data Protection Act [13]. Such frameworks clearly specify the responsibilities of organizations with regards to the privacy protection of personal health information. However, complying with these frameworks is both challenging and costly for healthcare organizations [14]. The main privacy challenge remains in the management of this collected data which is still largely unaddressed. There are many policy-related issues such as privacy policies that must be addressed to realize the full potential of sharing healthcare information [15] [16]. Sharing healthcare information using healthcare information systems based on privacy preservation rarely handles healthcare information sharing among healthcare-related entities at different places [17]; therefore, there is a need to address such collaboration based on privacy preservation.

1.2. Research Problem

Due to the diversity and complexity of the existing healthcare structure, in which patients' health information is distributed to multiple entities such as hospitals, healthcare centers and cloud servers, an appropriate architecture is one of the most important design issues for sharing healthcare information in a privacy-preserving manner. A centralized architecture design would not be convenient due to the lack of interoperability of most healthcare information systems. Currently, there are no policies for healthcare data standardization and normalization for proper data governance, it is also determined that there is no existing single data standardization structure that can effectively share and interpret patient data within heterogeneous systems [18] [19].

Despite the use of information technology solutions in the healthcare industry, there are various challenges encountered such as the high infrastructure management costs, dynamic needs for computational resources, scalability multi-tenancy and increased demand for collaboration [20]. The advancement in the healthcare industry requires modernizing healthcare information systems to facilitate collaboration and coordination among parties involved in the healthcare domain at lower costs. In healthcare, the availability of information regardless of the location of the patient and the clinician is a key driver towards patients' satisfaction and healthcare service betterment. For that, there is a stressing need for having a decentralized design of the architecture for healthcare information systems that allows for asynchronous interactions among parties involved in the healthcare domain with respect to privacy regulation [21] [22].

1.3. Cloud computing in healthcare

Cloud computing appears to be the dreamed vision of the healthcare industry; it refers to means of storing and accessing data and programs over the internet instead of the computer's hard drive [23]. Cloud computing matches the need of healthcare information sharing directly to various

 doi:10.20944/preprints202305.0416.v1

healthcare-related parties over the internet, regardless of their location and the amount of data being shared. The use of cloud computing in the healthcare sector has increasingly been highlighted as having great potential in facilitating data-driven innovations [24]. It offers functionality for managing information in a distributed, ubiquitous, and on-demand network access to a shared pool of configurable computing resources [25]. Resources in cloud computing can be rapidly provisioned and released with minimal management effort supporting several platforms, systems, and applications [26]. Cloud computing is an attractive paradigm of computing for the healthcare domain, due to the elasticity of resources and reduction of the operational costs. This allows for new ways of developing, delivering, and using healthcare services [27]. It offers practical solutions in the healthcare domain and sharing information is one of them [28].

Due to the rapidly growing applications of e-health systems which have been hampered by the conventional healthcare information system's lack of interoperability, cloud computing is found to be the best option for the global e-healthcare systems that are in place [29]. However, the adoption of cloud computing in the healthcare domain faces privacy-related challenges [30] [31]. Such challenges are caused by the fact of having medical data and information that is classified as confidential, stored in cloud servers, a virtual world where information can be easily hacked [32].

From the consumers' perspective, privacy when storing and sharing health-related information on the cloud is a primary concern, because data is stored in different places. Such concern prohibits the adoption of cloud computing in the healthcare domain [31]. The authors in [33] conducted a review of security and privacy-preserving challenges in e-health solutions. The review included various privacy-preserving approaches to ensure the privacy and security of electronic health records (EHRs) in the cloud. The review revealed a number of crucial privacy challenges which must be addressed before obtaining the full potential of cloud computing in the healthcare domain. The authors wrote "Studies must focus on efficient, comprehensive security mechanisms for EHR and also explore techniques to maintain the integrity and confidentiality of patients' information".

The main issue in the adoption of cloud computing in the healthcare domain is keeping sensitive information in the hand of a third party [31]. The owners of the information (patients' records) demand high levels of security and privacy on their information. Although data is usually encrypted, the owners require having control over their data to perform operations such as updating records. In the normal process, data transferred to the cloud goes through traditional encryption methods for security reasons, however, the data holder needs to decrypt the data whenever an operation is required on it. The data user provides the private key to the cloud provider to decrypt data in order to execute any required calculations. The decryption of the data at the cloud provider side causes privacy and confidentiality issues. Moreover, a patient's record may include information that might not always be needed for all different instances of medical treatments, for example, a patient who has a certain sexual disease might not want a practitioner at an emergency practice to access and read information related to such disease when it is not needed in that particular treatment instance, therefore, accessing such unneeded information may also cause breach of the patient's privacy.

This paper proposes a cloud architectural design for storing and collaboratively using patients' health information with respect to the privacy and confidentiality of it. The proposed architecture adopts three main approaches to protect the privacy and confidentiality of information; (1) grouping patients' health information according to the need of them in different instances, (2) adopting the searchable symmetric encryption (SSE) mechanism, and (3) exploiting the user identity management protocol (U-IDM). The rest of this paper is organized as follows: Section 2 presents the related work. Section 3 presents the proposed cloud architectural design. This section includes instantiation, implementation, and testing of the system. Section 4 includes discussion, and finally, section 5 presents conclusion and future research directions.

2. Related Work

Patients, families, and a diverse team of frequently highly specialized healthcare workers all contribute to the delivery of healthcare. To deliver exceptional treatment, all of these team members must be engaged in a collaborative and coordinative manner. Moreover, Information about patients'

 doi:10.20944/preprints202305.0416.v1

health generates special value when it is exchanged and collaboratively used among different parties involved in the healthcare area [6]. The definition of the term "collaboration" in the field of healthcare includes the concept of sensibly sharing a collective perspective that includes information, norms, social expectation, activity goals, and meaning. It is the communication that occurs among healthcare practitioners when sharing information and skills regarding patient care [34]

Researchers and individual users have paid a lot of attention to healthcare information among all the shared information [35]. In the digital healthcare era, electronic health records (EHR) have captured the processes of disease occurrence, development, and treatment which makes it of great value and essential tool to use for medical services [36]. It is of the utmost importance to harness medical information scattered across healthcare institutions to support in-depth data analysis and achieve personalized healthcare [37]. Therefore, a comprehensive and integrated healthcare infrastructure is required to facilitate the sharing of information among various healthcare institutions and domains.

The seamless exchange of vital information among healthcare practitioners played a significant role in reducing medical errors and facilitated better integration of health-related records [38]. However, to realize the full potential of collected medical data, health-related information technology systems and products are required to share information seamlessly among each other, but unfortunately, the vast majority of medical devices, electronic health records, and other systems lack interoperability [39]. Patients' health records are often stored in a non-standard, non-coded, structured, and non-structured form hindering the exchange of information among health information systems [40]. The heterogeneity is currently a major challenge in the healthcare industry to achieve interoperability especially among proprietary applications provided by different vendors [39]. For instance, a hospital may use one or more applications to share clinical and administrative information, and each application may support multiple communication interfaces and protocols that must be modified and maintained.

To achieve interoperability among different systems in the healthcare domain, several efforts have been put forth by various desperate parties. The work in [41] provided a review of some proposed cloud architectures for healthcare, along with issues in both technologies and the crucial reasons for moving forward with a cloud-based e-healthcare system. The issues of security and privacy were highlighted as a barrier to the adoption of cloud computing in the healthcare domain.

An approach towards achieving interoperability between information technology systems is Unified Modelling Language (UML) [42]. UML, or Unified Modelling Language, is a standardized modelling language used in software engineering to visually represent software systems. UML has been studied extensively for its potential to facilitate communication and interoperability between information systems. Many UML-based approaches have been proposed to facilitate the communication of information systems such as [43], [44]. UML can be used to represent the structure, behavior, and interactions of different systems and components, providing a common language and framework for communication among stakeholders. Similarly, the work in [45] proposed a new approach to specifying data integration toward interoperability based on data models such as entity-relationship (ER) and UML. The authors draw attention to a critical problem that results from the incompatibility of data models, such as the use of proprietary terminology, data structures, data formats, and semantics by different software systems. Data must be shared between software systems, and frequently, challenging data conversions or transformations are necessary. Process modelling is also difficult due to the complexity of the healthcare information systems and requirements, which accounts for the slow adoption of process modelling standards [46].

Enterprise Service Bus (ESB) [47] is another approach that serves as a platform for integrating different applications and services within an enterprise. EBS is an architectural pattern whereby a centralized software component performs integrations between applications. In a service-oriented architecture (SOA) [48], EBS implements a communication mechanism between software programs that interact with one another by acting as a central hub or mediator, enabling communication and information exchange between different systems and services.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

In [49], the authors conducted a study to understand and provide ongoing research topics, challenges, and future directions concerning ESB applications. While ESB is seen as a powerful approach for enterprise integration and data exchange, its adoption in the healthcare domain requires meeting difficult requirements related to accessibility, in light of the fragmentation of patients' information in heterogeneous proprietary systems.

Blockchain is another approach that is widely used in healthcare information systems due to its decentralization and security features [50]. The adoption of blockchain technology becoming a widespread trend in distributed computing. Many researchers considered the use of if for sharing information across healthcare information systems [51]. An example of suggested blockchain-based application is found in [52]. The author proposed a distributed smart and secure healthcare system using blockchain and edge computing for sharing medical information across different institutions. The proposed system shares data through separating medical data processing, access control, and data sharing into local and blockchain networks. The work also presents a data-sharing security algorithm based on the value of the shared data.

The proposed system has a significant drawback related to quality of service and information privacy. In terms of quality of service, it is difficult to guarantee consistent service quality because various validation times depend on the security level, while the move of data through edge nodes in the local network may lead to centralization and privacy leads especially because data may not be encrypted.

Another blockchain-based work for sharing healthcare information is presented in [53]. The authors presented attribute-based encryption system for authorization and dynamic authentication of medical on-demand service in remote medical systems. Blockchain in the system was exploited along with distributed database technologies to protect the integrity of information. However, the approach suffered from a limitation related to centralization and security.

The authors in [54] conducted a review of blockchain-based secure sharing of healthcare data. The review included an evaluation of the development of blockchain in healthcare from various perspectives. It also analyzed the approaches of blockchain from different application scenarios. The results show that blockchain technology has an advantage in the field of healthcare, but the technology is suffering from issues, including low throughput and low scalability, which limits its adoption in the healthcare industry. Users can store information on a decentralized platform using unforgeable ledgers. Digital encryption can ensure data security and individual privacy. This technology has the potential to reduce operating costs, and to increase synergies while preserving the integrity of data [55]. However, many issues related to its adoption in different industries remain unaddressed [54].

3. Proposed Architecture

Considering the complexity of the existing healthcare structure where patients' health information is distributed to multiple entities such as hospitals, healthcare centers and cloud servers, a centralized architectural design of information systems for the healthcare domain would not be suitable, especially when interoperability remains a challenging obstacle among the vast majority of healthcare information systems. A non-centralized architectural design would be the most suitable option for the healthcare sector so that disparate entities can collaborate through sharing information related to patients and their health.

In this work, a new cloud-based architecture is proposed for storing and sharing healthcare information in a privacy-preserving manner. There are two sources of information that informed the design of the proposed cloud architecture: case study findings and literature review [56]. The characteristics of the proposed cloud architecture are as the following:

- **Just-enough information disclosure**: Disclosing only the right information according to the context in which information is required.
- **Accessible location of information**: Storing patients' information in once place for easy access whenever information is required.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

- **Unified platform**: Accessing information through a unified platform is a key characteristic toward improving healthcare services.
- **Adherence to the legal privacy-related frameworks**: The architecture should adhere to privacy-related regulations and policies such as HIPPA and the information privacy act when using information.
- **Patients control**: Patients should have a means of control over who can access their information.
- **Cloud provider blindness**: The cloud provider should not be able to read or access patients' information that is stored on the cloud.

3.1. Architecture fundamental aspects

There are two fundamental aspects of the proposed architecture design that enable it to store and share patient health-related information in a privacy-preserving manner. The first fundamental aspect is structuring patient information into categories. This aims to eliminate the exposure of information that is not needed during instances of medical treatments.

Structuring patient information also contributes towards allowing patients to have means of control over who can access their information while it is stored on the cloud. The second fundamental aspect is the use of a searchable symmetric encryption scheme (SSE) [57] which enables to search through encrypted information without decrypting it. The objective of the searchable encryption scheme is to store patient information on the cloud without the ability of the cloud provider to learn the content of the stored information.

3.1.1. Structuring Patients Information

A fundamental aspect of the proposed cloud architectural design is the accommodation of patients' health information under four main categories which were identified in the case study findings [56]. These categories are Information that is constantly required in every patient's visit (All_V), Information that is required in patients' emergency visits (Em_V), Information that is required in out-patients' clinical visits (OutP_V), and information required for research purposes (R). This paper focuses on information categories that are used for medical treatment purposes; therefore, the (R) category is explained in other work.

The main goal of structuring patients' health information is twofold; firstly, to limit the exposure of information in instances when it is not needed. Secondly, limiting the exposure of information leads to better means of privacy protection that patients desire to have for their health information. The proposed system design stores patients' information in three groups referred to as documents. Each document has identifying tags and contains files.

Each file has the name of the patient, name of document, and a sub-tag used by the application system to identify and locate it. The system's identifying tags are used to technically facilitate access to documents and do not indicate the content of documents.

For example, and for simplification purposes, the tags used for the documents are 1, 2, and 3. All patients registered in the proposed system have their information organized into doc-1, doc-2, and doc-3. In the practical implementation of the proposed system, information stored in each document is subject to change according to the medical treatment changing needs. The information categories comprise information contained in different documents, therefore, accessing a category of information is a result of accessing one document or more. For example, when a user has the right to access information about a patient in an emergency setting (Em_V), doc-1 and doc-2 are released to the user, while a combination of document doc-1, doc-2 and doc-3 are released for users who have access to all information related to patients' health (OutP_V) category. Table 1 illustrates the information categories and their comprising documents.

Table 1. Information categories and their comprising documents

Doc-1	Doc-2	Doc-3
Full Name	Drug Allergies	All the information that is not
Date of Birth	Discharge, Summaries	contained in doc-1 and doc-2

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

Gender	Blood Type
Ethnicity	Laboratory Results
Significant Conditions	Next Kin
NHI Number	
Current Medication	

3.1.2. Searchable Symmetric Encryption (SSE)

Searchable symmetric encryption is a corner stone of the proposed system architecture. The main objective of the proposed system architecture is to store patients' information on the cloud without the ability of the cloud provider to read it. Achieving this is considered easy but not practical without a mechanism that enables to search through encrypted information without decrypting it.

The proposed system employs a searchable symmetric encryption (SSE) approach. The SSE approach enables outsourcing data storage while preserving the ability to selectively search over it. There are three models for searching on encrypted data identified in the literature namely searching on public-key encrypted data [58], single-database private information retrieval (PIR) [59] and finally searching on private-key encrypted data [57] which is the approach employed in the proposed cloud architecture. For consistency purposes, the private key is denoted by secret key (S_k) throughout the paper.

In the secret-key-encrypted data model, the data is encrypted by the user and is organized in an arbitrary way prior to encrypting it. The data is stored on a server in encrypted form and decrypting it can only happen by the S_k. In this model, the initial work for the user is large when data is large, while subsequent work such as accessing the data is small. The user work is large because data pre-processing requires performing a number of processes to facilitate searchability on it while it is encrypted. Structuring data as part of the pre-processing allows for efficient access to relevant data. In this proposed system, Information is partitioned into portions denoted by documents as explained earlier. For every patient, there is a root secret key (S_{kR}) that is used to encrypt 3 secret keys (Sk). Secret keys are used to encrypt patients' documents (doc-1, doc-2, and doc-3). Each document is encrypted with its corresponding S_k. Indexes and trapdoors -explained further in this section- are generated to identify and decrypt documents respectively. An important property of the secret-key-encryption approach is that anyone who can decrypt information for a document can also decrypt any file in that document. This means, anyone who has access to a document can have access to all files within that document.

The main goal of employing the SSE approach is to store patients' health information on the cloud in a searchable manner and only authorized parties can access it. Moreover, the cloud provider can never learn anything about the information stored, it receives encrypted information to store and releases it without decrypting it.

The decryption of each document under the secret root key requires the secret key for it which is released upon authenticated and authorized user requests. The cloud provider is not informed about the content of any document; therefore, the challenge remains in identifying encrypted document/s without decrypting them. The searching capability of the SSE approach is achieved using a secure index mechanism [60]. The secure index is a structure of data that stores document collections while supporting efficient keyword search, for example, given a keyword (w), the index returns a pointer to the documents that contain it. The secure index works by searching for a string exact match in encrypted documents. Every document contains a collection of encrypted strings, and a string is chosen to be the searching keyword for the document that contains it. The selected keyword is computed using the secret key by which the entire document is encrypted. The resulting ciphertext is then used to search for an exact match in documents. For example, a keyword in a document is "Basic-Information". This keyword is computed using the secret key of the encrypted document and the resulting ciphertext is e.g. "JK^78Uo8361KL$#VWL". The combination of keyword and its corresponding ciphertext is then used to identify the document which contains the keyword "Basic-Information". However, a keyword may appear in different documents, therefore, a number of keywords and their corresponding ciphertexts are put together in an encrypted index and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

corresponding trapdoor to assure the accuracy of document identification. Alternatively, a document's unique name can be used to achieve the same outcome accurately such as doc 1. Table 2 demonstrates an example of an encrypted index generated for a document listed under a secret root key and its corresponding trapdoor.

Table 2. Encrypted index and Trapdoor for a document

Doc-1 Encrypted Index		Doc-1 Trapdoor	
Basic Info	JK^78Uo8361KL$#VWL	**Sk**	JK^78Uo8361KL$#VWL
Significant	RM*#%H)GIDU784K2%		RM*#%H)GIDU784K2%
Medication	B&0*9QOVPI(068B%#O		B&0*9QOVPI(068B%#O
Doc-1	APV*89&@JE)<I@DO$		APV*89&@JE)<I@DO$

To achieve the properties of the SSE approach, authors in [57] proposed the below five algorithms which are the Key Generation algorithm (**KeyGen**), Key derivation algorithm (**KeyDer**), Index Generation algorithm (**IndexGen**), Trapdoor Generation algorithm (**Trap**), and a Search algorithm (**Search**). Below is the description of these algorithms:

KeyGen Algorithm: A probabilistic algorithm that sets up the searchable encryption scheme. It is responsible for generating a secret root key for patient's documents as a collection. It takes a security parameter k and generates a secret root key (SkR) for the patient SkR. This key is used for wrapping and unwrapping the secret keys of all documents that belong to the patient.

$$KeyGen\ (1^k) \rightarrow S_{KR}$$

KeyDer Algorithm: Employed for generating a secret key (S_k) for each document listed under the secret root key (S_{kR}). It takes the document name and secret root key (SkR) as input and generates a secret key (S_k) for the document. This secret key will be used to encrypt and decrypt the information contained in its corresponding document.

$$KeyDer\left(sk_{(i_1 \ldots . i_{n-1})}, (i_1 \ldots . i_n)\right) \rightarrow sk_{(i_1 \ldots . i_n)}$$
$$Fsk_{(i_1 \ldots . i_{n-1})}$$

IndexGen Algorithm: Responsible for generating an encrypted index (I) for every document. It takes a number of keywords in a document such as the name of the document or its title and encrypts them using the document secret key (S_k). The output of the IndexGen algorithm is an encrypted searchable index I for every document to be used for searching it.

$$IndexGen\left(sk_{(i_1 \ldots . i_n)}, (i_1 \ldots . i_n), word_{w_1 \ldots . w_n}\right) \rightarrow sk_{(i_1 \ldots . i_n)}, I$$

$$Enc\ (sk_{(i_1 \ldots . i_n)}, I) \rightarrow C_1$$

Trap Algorithm: Responsible for generating trapdoors for documents. It takes the secret key of a document and keywords' ciphertexts as input and outputs a corresponding trapdoor (T) which is used for decrypting the document.

$$Trap\left(sk_{(i_1 \ldots . i_n)}, (i_1 \ldots . i_n), word_{(w_1 \ldots . w_n)}\right) \rightarrow T$$

Search Algorithm: Uses the decrypted index and the trapdoor for one document to find it. It takes the decrypted index and the trapdoor as inputs and identifies the encrypted document as an output.

$$Search\ (C_1, T_1) \rightarrow \text{Encrypted ciphertexts}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

Similarly, in the proposed system, the process of preparing patients' information for storage involves five steps:

1. Generating a secret root key (S_{KR}) for the patient.
2. Generating a secret key (S_K) for every document of patient information and choosing a keyword of each document.
3. Keywords are encrypted using their corresponding secret keys and the resulting ciphertexts are listed to form an encrypted index.
4. Trapdoors are then created which involves combining the secret keys with the ciphertexts. Trapdoors will be used to identify and decrypt documents.
5. Documents are encrypted using their corresponding secret keys.

By following the above five steps, it becomes feasible to search for patients' documents while they are encrypted without having to perform decryption operations on them. Further explanation of how information is obtained from the cloud and decrypted is provided in the following section.

3.2. Architectural Design and Components

The proposed architecture comprises five architectural components that are required for storing healthcare information on the cloud and collaboratively use it in a privacy-preserving manner. These components are Requesting Agent, User Application, Cloud Service Registry, Secret Key Server, and Cloud Service Provider. Each component is responsible to accomplish certain tasks as a contribution to achieving the main objectives of the proposed architecture. Figure 1 illustrates the architectural design and the relationship of its comprising components.

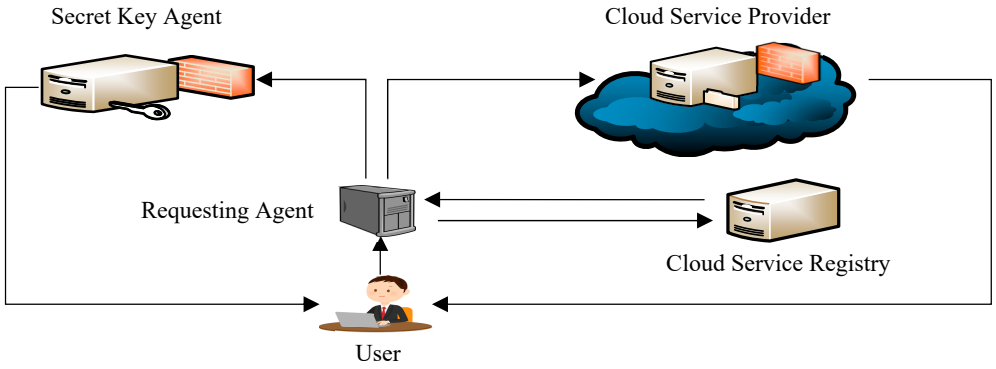

Figure 1. Proposed Architectural Design

Requesting Agent

The Requesting Agent (RA) is a server that is responsible for receiving requests from users and forwarding them to both the Cloud Service Provider and the Secret Key Agent after authenticating users. It is the point of contact through which users send requests to store or access information stored on the cloud. Users are authenticated and their access rights are identified before requests are forwarded by the RA. In other words, it plays the role of the gate person who does not allow unauthorized users to access the system. The RA has a limited communication channel with the users, one-way communication channel with both the Cloud Service Provider (CSP) and the Secret Key Agent (SKA), and a two-way communication channel with the Cloud Service Registry (CSR) for users' authorization. The RA receives requests from users and only responds with information that is limited to confirmation of authentication.

The RA stores the required information for identifying patients who are registered in the system. Information stored on the RA is important for facilitating secure access to patients' information that is stored on the cloud. Every patient has a unique code referred to as system ID which is generated by the RA and used for searching purposes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

When a user requests to access patient information, the patient system ID is used to identify the patient's information that is stored on the cloud. The main role of the RA in the proposed design includes receiving requests from users, authenticating users, forwarding user requests to both the CSP and the SKA. The information stored on the RA for every patient is organized into 3 sections, every section contains information that is important to facilitate access to patients' information in a secure and privacy-preserving manner. Table 3 illustrates the information is stored in the RA.

Table 3. Information stored on the Requesting Agent for every patient.

Full Name	D.O. Birth	NHI Number
ID	S_{KR}	doc-1 Index doc-2 Index doc-3 Index
User ID User ID User ID	Email Phone Number	

Section 1 includes information that is required to identify patients on the system. Users request to access patients' information by using patients' basic identification information such as name, date of birth, and NHI number. Section 2 includes the patient's system ID, S_{KR}, and indexes. The system ID is required to identify patients' information that is stored on both; the cloud and the Secret Key Agent (SKA). The S_{KR} is required to decrypt the trapdoors that are stored on the SKA, and the indexes are needed to identify the documents stored on the CSP. Section 3 includes information that is required by the Cloud Service Registry (CSR) for authorizing users to access patients' information. It includes a list of users who have permanent consent to access the patient's information, as well as the patient's contact details for requesting and obtaining patient consent. There are two types of consents that patients grant to users for accessing their information: permanent consents which are granted by patients to their local GPs or pre-determined medical institutions/practitioners, and temporary consents which patients grant to medical practitioners/institutions for casual incidents or clinical visits. Patients optionally grant permanent consent to users to access their information. Temporary consent is granted to a user who does not have permanent consent and requires access to a patient's information. There are two methods of requesting and obtaining temporary consents in the proposed system: mobile phone in a form of text message, or via email confirmation. More information about patient consent is provided further in the paper.

Standard User Application

The proposed system architecture requires having a standard application that is installed and run locally on users' machines. Accessing patients' information stored on the cloud can only happen through a standard user application (UA). Having a unified platform to access patients' information was identified in the case study findings as a desired characteristic of healthcare information systems, therefore, the proposed architecture design employs a standard UA through which users can access information that is stored on the cloud.

The UA plays a key role in the proposed system architecture; it facilitates means of standardization to the process of storing, accessing, categorizing, and structuring information. There are three main functions that UA is responsible for which are storing, accessing, and updating patient information on the cloud. These functions are performed using buttons that are available on the UA interface, these buttons are ENROL, REQUEST, UPDATE, and RESEARCH. Further explanation about these functions is presented further in the paper.

There is a number of characteristics that UA has which enable it to store, access, and update patients' information on the system. The following are the main characteristics of the UA employed in the proposed system design:

- Standard presentation and categorization of information

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

Categorizing information is part of the UA's functionalities. The application organizes patient information files into three documents (doc-1, doc-2, and doc-3) before it is stored on the cloud. The UA has a standard user interface for all users. Information is accessible when it appears in predetermined fields on the user interface. Information is presented in their associated fields only when it is decrypted. Information fields remain blank when their corresponding files are not decrypted. For example, a field on the application interface is predetermined for information related to patient mental health, this field remains blank when the logged-in user is not authorized to access the document in which mental health file exists.

- Information pre-processing, encrypting, and decrypting

The properties of the searchable symmetric encryption (SSE) approach employed in the proposed architectural design are achieved by operations performed by the UA. The pre-processing operations together with encryption/decryption operations are all performed by the UA. The UA is responsible to pre-process the information by organizing them and encrypting them following the SSE approach before it is sent for storage. It is also responsible for requesting to access information and decrypt it when it is received.

- Characteristics related to accessing patients' information for research purposes

The UA has important characteristics that are related to accessing patients' information for research purposes in a privacy-preserving manner. Entering kiosk mode, disabling certain functionalities such as copy-paste functionality, allowing and prohibiting communication channels. are all important characteristics of the UA. These characteristics aim to ensure the privacy of patients' information when used for research purposes. Further details about the characteristics related to using patients' information for research purposes are provided further in this chapter.

Cloud Service Registry

The proposed cloud architecture in this research employs the concept of the user identity management protocol for the cloud computing paradigm (U-IDM) proposed in [61]. U-IDM was initially proposed for cloud computing customers and cloud service providers. The main objectives of U-IDM were to achieve a set of global security objectives in cloud computing environments which include user authentication, authorization and accounting. It aimed to protect customers and cloud provider's infrastructures by preventing unauthorized users to gain access to services or facilities delivered by cloud providers.

The main component of the U-IDM paradigm is the Cloud Service Registry (CSR). The CSR plays a vital role in the proposed architecture. CSR provisions access information according to users' privileges in a form of service level agreements (SLAs). Services in the context of the proposed architecture include the provision of access to patients' information that is stored on the cloud. There are three information categories that require access rights from the CSR which are All_V, Em_V, and OutP_V. As discussed earlier, each information category contains one or more documents. The CSR grants access to information categories by providing access to the documents that form these categories. Repeating the example of the Em_V category, it is a combination of doc-1 and doc-2 Therefore, granting access to the Em_V category requires the CSR to include the name of documents or their identifying tags with the user authentication confirmation. The CSR stores the names of categories and their comprising documents' tags. A list of registered users is stored on the CSR. Each user has a record of information related to the information that they can access. Users are listed under the name of their organizations. Searching for a user requires knowing the organization he/she belongs to. Table 4shows an example of users' lists who are affiliated to an organization.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

Table 4. Example of users list in stored on the CSR

ID#	Role	Access Privilege
29930894	Nurse	doc-1 ǀ doc-2
29930804	Doctor	doc-1 ǀ doc-2 ǀ doc-3
29930832	Receptionist	doc-1
29930930	Doctor	doc-1 ǀ doc-2 ǀ doc-3

Nevertheless, an important task of the CSR in the proposed architecture is to obtain patients' consent for accessing their information. The CSR does not authorize users to access patients' information without having patients' consent. As mentioned earlier, when a user requests to access patient information, the RA authenticates the user and forwards the request to the CSR for authorization. Part of the information included in the RA's forwarded request includes a list of permanently authorized users to access the patient's information. This list enables the CSR to find out whether the user is granted permanent consent to access the patient's information or not. If the user is not included in the list, the CSR promptly sends a request for temporary consent to the patient, and the patient can promptly grant consent or reject.

Secret Key Agent

The Secret Key Agent (SKA) resides in a server that stores the required information for decrypting information stored on the cloud. As explained earlier, for every patient, there are 3 secret keys (S_k) listed under a secret root key (S_{kR}) which are used to decrypt 3 documents. All secret keys are stored together with trapdoors for all documents related to one patient (under one S_{kR}). The main functionality of the SKA is to receive requests from the RA and send the required trapdoors directly to the user. SKA has a one-way communication channel with the RA which is to receive requests, and a one-way communication channel with users to send secret keys, encrypted indexes, and trapdoors.

Cloud Service Provider

The cloud service provider (CSP) holds information related to patients' health. The main goal of the proposed architecture is to store all patients' information in one place which is the cloud. The CSP serves by storing and releasing encrypted information related to patients upon users' requests. The CSP has a one-way communication channel with the RA, and a one-way communication channel with users. It receives requests from authenticated and authorized users through the RA and releases the required information in its encrypted form to users. Information stored on the cloud is contained in encrypted documents. The CSP cannot learn anything about the content of the documents stored. The cloud receives encrypted documents to store and release them to users without performing any decryption process on the documents. The CSP employs a string match algorithm that aims to identify documents. Every patient has 3 encrypted documents that are labelled by the patient's system ID. Further explanation about the role of the string match algorithm is provided in the following section.

3.3. System Instantiation

The proposed architecture design aims to store healthcare data on the cloud and access it for legitimate purposes while protecting privacy. This section uses a scenario to show how the suggested architecture accomplishes this goal. The instantiation is provided in two parts: the first shows a patient enrolling in the system and storing their information on the cloud, while the second section shows how this information can be accessed for medical treatment.

Storing patient information on the cloud - Scenario

Let's assume that Bob visits a system authorized doctor and requests to enroll in the system. The doctor enters Bob's information through the user application (UA). The doctor clicks on the **ENROL**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

button. The process of storing Bob's information on the cloud involves 3 stages: information preparation, authentication and authorization, and storage.

Stage 1: Information Preparation (Searchable Symmetric Encryption): Prior to forwarding the information to the RA, Bob's information undergoes a number of pre-processing algorithmic operations performed by the UA as preparation for storage. The operations aim to encrypt Bob's information for storing it on the cloud in a searchable manner.

1. A random secret root key (SkR) is generated for Bob's information using KeyGen algorithm.
2. A number of keywords from each document are selected, and a secret key (S_k) for encrypting them is generated using KeyDer algorithm.
3. The IndexGen algorithm encrypts selected keywords for every document using their corresponding S_k. The goal in this step is to create an encrypted index for each document to identify it while encrypted.
4. After encrypted indexes are generated for all documents, the ciphertexts of keywords with their corresponding S_k for each document is grouped to be the documents' trapdoors.
5. The last step in the information preparation process involves encrypting the patient's documents and their corresponding trapdoors. Each document is encrypted using the S_k that is included in its corresponding trapdoor, and trapdoors are encrypted using the S_{kR} which was generated in the first step.

Stage 2: Authentication and Authorization: When Bob's information is pre-processed, it is forwarded to the Requesting Agent (RA) in a form of Request of Enroll (ROE). Table 5 presents information contained in the request of enroll (ROE).

Table 5. Information included in the ROE

Section 1	Section 2		Section 3
Organization ID Practitioner ID Patient Name	S_{KR}		Encrypted information
Patient DOB	doc-1 Index	doc-1 Trapdoor	doc-1
Patient NHI	doc-2 Index	doc-1 Trapdoor	doc-2
Consent Method: Phone#	doc-3 Index	doc-1 Trapdoor	doc-3

The ROE includes three sections that include different information as the following: Section 1 includes information that is required to identify a) the doctor (organization id and user id), b) Bob and practitioners who have permanent consent to access Bob's information, and c) method of obtaining Bob's consent to access his information. Section 2 includes information required to identify and decrypt Bob's information, and section 3 includes Bob's encrypted Bob's information (3 encrypted documents).

When the RA receives the request, it authenticates the doctor and forwards his information (contained in the first section of the ROE) and Bob's phone number to the cloud service registry (CSR) for authorization. The CSR then sends a text message to Bob requesting consent to store his information on the cloud, the content of the message includes:

Please reply **YES** to authorize **(doctor name**) from (**organization name**) to enroll you and store your health information on the system. Upon receiving a **YES** reply from Bob, the CSR sends a confirmation of authorization to the RA.

Stage 3: Information Storage

When the RA receives confirmation of authorization from the CSR, it does the following actions:

1. Generates a unique code for the patient referred to as (System ID).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

2. Sends Bob's encrypted information labelled by Bob's system ID to the cloud service provider (CSP).
3. Sends the encrypted trapdoors to the secret key agent (SKA) for storage. Information sent to the SKA is also labelled by Bob's system ID.
4. The information sent to both CSP and SKA is deleted from the RA.
5. The RA stores the following information:

- Bob's identification information
- Bob's system ID,
- Bob's SkR,
- Document indexes
- Names of users who have permanent consent to access Bob's information (if Bob has provided any)
- Information required for obtaining Bob's temporary consent

Below is the state-of-the-art of Bob's information while stored in the cloud:

1. Bob's information is stored in encrypted form and labelled by Bob's system-generated ID. The cloud provider is not able to learn the content of the information.
2. The trapdoors are encrypted using Bob's SkR and stored on the SKA labelled by Bob's system ID. The SKA is unable to learn the content of the trapdoors without having Bob's SkR that is stored on the RA.
3. The RA is the only entity in the system that can identify Bob in the system and his SkR. The RA stores all the information that is required to access Bob's information as presented in Figure 2.

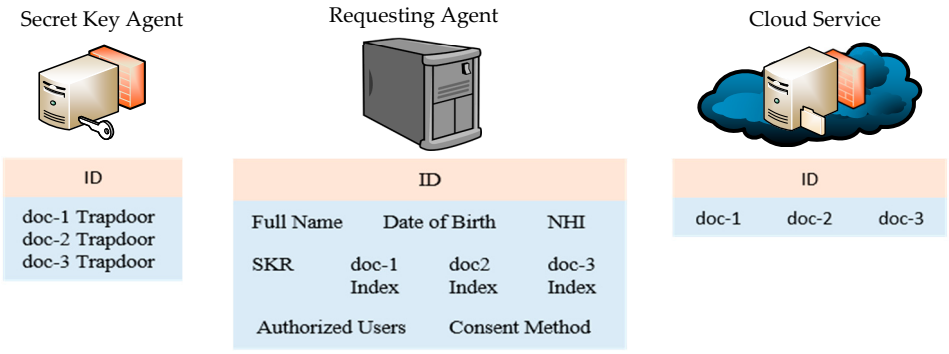

Figure 2. Bob's information stored in the proposed architecture

Therefore, accessing Bob's information can only happen through collaborative interactions among CSP, SKA, and the RA. Compromising 1 or 2 of these architectural components will be fruitless to any disparate party in terms of accessing Bob's information.

Having discussed the process of storing Bob's information on the cloud, the following subsection presents the process of accessing Bob's information for genuine reasons such as providing healthcare to a patient. For this, the scenario presented in the following subsection involves the same patient (Bob) requiring healthcare assistance by a different medical practitioner who also has access to the system.

- Accessing stored patient information – Scenario

Bob visits a hospital for urgent medical treatment. He walks into the emergency department and meets one of the nurses in charge. The nurse requires accessing Bob's information and updating his records to include information about Bob's visit, medical condition, and other information related to his visit.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

3.3.1. Protocol to access information stored in the cloud

The process of accessing Bob's information comprises 4 stages as the following:

Stage 1: Generating user request

The nurse enters Bob's basic information into the user application and clicks on the **REQUEST** button to generate a user request that is forwarded to the RA. The user request includes information about both Bob and the nurse.

Stage 2: Authentication and Authorization

When the RA receives the request from the user (nurse), it authenticates the user and forwards the request to the CSR for authorization. For this, the RA does the following actions:

1. It searches for Bob's information using his basic information and finds his System ID.
2. It sends a request of authorization to the CSR. The request includes the following information:

- Information that is required to identify the nurse (user ID);
- List of users who have permanent consent to access Bob's information;
- Bob's mobile number for requesting his consent if required in this particular instance.

When the CSR receives the request from the RA, it does the following actions:

1. It searches for the nurse information to identify her access rights to patient information. This happens by searching through the list of users that is stored locally on the CSR.
2. It checks if the nurse is permanently consented to access Bob's information using the list of users who have permanent consent to access Bob's information.
3. The CSR finds out that the nurse is allowed access doc-1 and doc-2 (Em_V) of patients' information, but she is not permanently consented to access Bob's information, therefore, Bob's consent is required.
4. The SCR sends a request of consent to Bob in the form of a text message. The content of the message includes:

*Please reply **YES** to temporarily authorize (**nurse name**) at (**organization name**) to access your health information.*

Upon receiving a YES from Bob, the nurse becomes temporarily authorized to access Bob's information. The CSR sends a confirmation of authorization to the RA. The confirmation of authorization includes the information category that the nurse can access (doc-1 and doc-2), and confirmation of obtaining Bob's consent to access his information. The nurse is then added temporarily to the list of authorized users (stored on the RA) as a temporarily authorized user. However, any authorization granted by the CSR remains valid for 1 hour, after that it is automatically deleted from the list of authorized users.

Stage 3: Releasing Information

Upon receiving confirmation of authorization from the CSR, the RA forwards requests to both, the CSP and the SKA to send Bob's information to the nurse. As illustrated in Figure 3, the request to the CSP includes:

- Bob's system ID
- Indexes of doc-1 and doc-2
- The nurse's application address

While the information included in the request to the SKA includes:

- Bob's system ID and S_{kR}
- Trapdoor-1 and Trapdoor-2 tags
- The nurse's application address

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

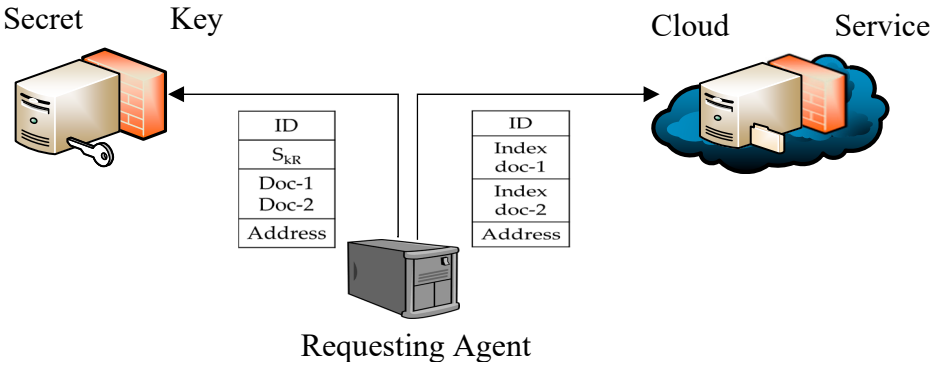

Figure 3. Requesting information from CSP and SKA

As presented in Figure 4, when the RA requests are received by the CSP and the SKA, they do the following actions:

The CSP:

1. Searches for Bob's information using Bob's system ID
2. Searches for the doc-1 and doc-2 using their indexes.
3. Sends the identified documents (doc-1 and doc-2) to the nurse using her application physical address.

The SKA:

1. Searches for the encrypted trapdoors using Bob's system ID
2. Decrypts the trapdoors using Bob's SkR
3. Sends trapdoors for doc-1 and doc-2 to the nurse application using her application physical address.
4. Re-encrypts the trapdoors using the same Sk_R and drops the SkR (deletes it).

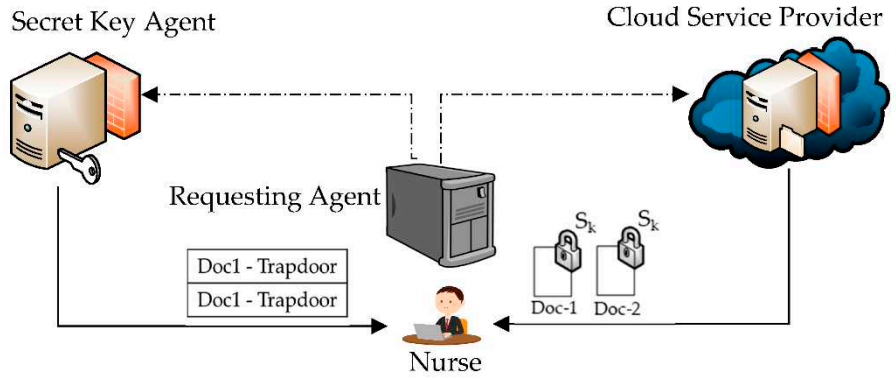

Figure 4 Releasing information to the nurse application

Sage 4: Decrypting Information

When the information from CSP and SKA is received by the nurse's application, doc-1 and doc-2 are identified and decrypted using their corresponding trapdoors. When information is decrypted, files in each document appears in their predetermined fields on the nurse's UA. Fields that belong to the files contained in doc-3 remain blank. The nurse application stores the trapdoors temporarily to be used for re-encrypting the information which is further explained in the following subsection.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

3.3.2. Updating patient information

Assuming that the nurse has made an update on Bob's information such as information related to current medication. The nurse clicks on the **UPDATE** button on her UA interface. As illustrated in Figure 5:

Stage 1: The nurse's UA encrypts doc-1 and doc-2 using their secret keys obtained from the trapdoors. The encrypted information (doc-1 and doc-2) is forwarded to the RA.

Stage 2: The RA receives the request from the nurse and does the following:

1. It searches for Bob's information to identify him.
2. It searches through the list of authorized users to access Bob's information and finds the nurse listed as temporarily authorized users to access doc-1 and doc-2 of patients' information.
3. It forwards the encrypted information, indexes for doc-1 and doc-2, and Bob's system ID to the CSP.

Stage 3: When the CSP receives the information from the RA, it does the following actions:

1. It searches for Bob's encrypted documents using indexes and system ID.
2. It identifies the documents using the indexes and replaces them by the new ones
3. It deletes the indexes received from the RA.

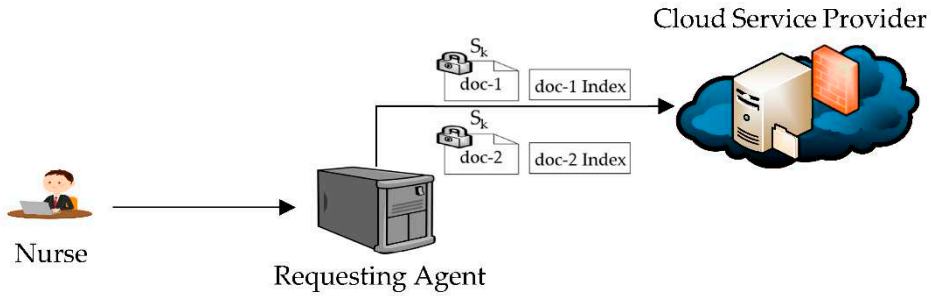

Figure 5. Updating patient information

3.4. Architecture Implementation

Having discussed the components and protocol of the proposed architecture, the architecture is further implemented and adapted to data sharing use. The proposed architecture was built using Amazon Web Services (AWS) which provides cost-effective cloud computing solutions [62]. AWS Software Development Kit (SDK) was used with Java language to implement and test the proposed architecture design and validate its concept. The implementation of the proposed architecture aimed to elaborate on how the proposed architecture enables for collaborative use of patients' information in a privacy-preserving manner. The main objective of the implementation was twofold: firstly, to elaborate on how patients' information can be collaboratively shared and used in the proposed architecture with assurance to its privacy protection, and secondly to illustrate on how patients' information is protected from a number of privacy-related threats including confidentiality and unauthorized access.

The elaboration is presented in two parts: the first part presents a scenario that involves a patient who walks into a hospital for urgent medical treatment and is seen by a nurse. The goal of this part is to show how a user (nurse) can access a patient's information according to certain access rights without questioning the privacy of the information. The elaboration also aims to validate the concept of information separation in real cloud-based application contexts.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

The second part of the elaboration presents the results of tests that have been performed on the implemented architecture. The architecture was tested in terms of its ability to preserve the privacy of information while it is stored on the cloud. Four tests were performed which covered the following aspects:

- The ability of the cloud provider to access and read the information that is stored in the cloud
- The ability of unauthorized users to access patients' information stored in the cloud

The following section presents information about the implementation of the cloud architecture in terms of its architectural components and the privacy-preserving techniques. The implementation aims to mirror the proposed architecture design in terms of the employed components and the privacy-preserving techniques for the goal of validating the concept of the proposed architectural design.

3.4.1. Implementation setup

For the implementation of the proposed architecture, a virtual private cloud (VPC) was created. The VPC represents the entire proposed system architecture, and it is completely isolated from the internet. The VPC has two subnets: the first subnet does not allow access from the internet; it is only accessed locally. The second subnet enables internet access to allow for communication with the user application.

The public subnet has limited access and could be communicated with through the internet; however, only registered IP addresses can communicate with it. The EC2 represents the RA in the proposed design. Access to it happens only through a particular port. Moreover, the database in the implemented design only accepts SQL traffic. Patients' documents and their associated trapdoors are stored in two different places that are not accessible through the internet as seen in Figure 6. Access to them can only happen locally.

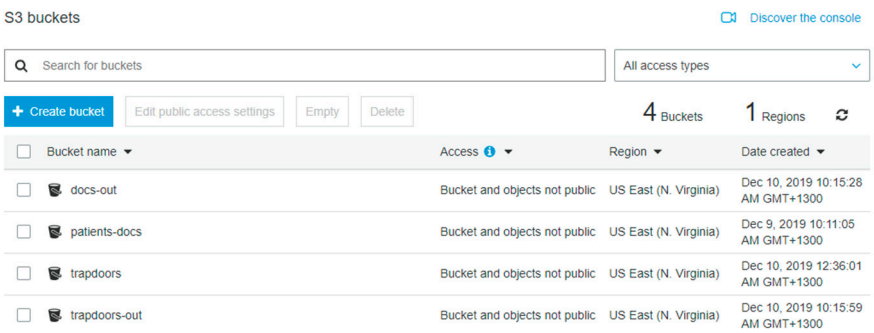

Figure 6. Separation of information stored on the cloud

Access to patients' documents can only happen by requests from the CSR, while the trapdoors can only be accessed through the SKA. Documents stored on the cloud are all encrypted.

Accessing Patient information – Scenario

For the goal of testing the architecture, dummy information about 3 patients was used. Each patient had three documents in the system. Documents were all encrypted and stored in the cloud database. Encrypted documents were stored on one database (CSP) and their trapdoors were stored on a different database (SKA). The illustration involved a patient (Bob) who walked into the hospital for urgent medical treatment. The nurse wishes to access Bob's information to update information regarding Bob's visit and current medication.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

The nurse needs to login to the system for authentication purposes. The user is required to have username and password which are entered through the standard user application. Access to the system only happens through the user application. When the RA receives the user credentials (username and password), it searches for the user information in the list of registered users. The RA in the implemented architecture had a database that contains names of registered users, this database was used for authenticating users. When the user is found, authentication is confirmed.

User authorization

When the user is authenticated by the RA, another window pops up on the user application for entering Bob's basic information. Bob's information is used by the RA to identify him in the system. When information is entered by the nurse and forwarded to the RA, the RA searches for the patient in the list of the registered patients in the system. The RA in the implemented system had another database that contains information about all patients enrolled in the system. When the patient is found, the CSR is called for authorization. The RA sends Bob's information to the CSR along with the users' information.

The CSR searches for the user in the list and finds out that the user is a nurse and is allowed to access Doc-1 and Doc-2 of patients' information. Table 6 is the table used by CSR to authorize users. The CSR confirms to the RA that the user is allowed to access Doc-1 and Doc-2 of Bob's information. The assumption made here was that Bob's has received a text message from the CSR and has granted consent for the nurse to access his information.

Table 6. The users' table used by CSR to authorize users

SysId#	UserId	Role	Access Privilege
29930894	7777	Receptionist	doc-1
29930804	8888	Nurse	doc-1 ǀ doc-2
29930832	9999	Specialist	doc-1 ǀ doc-2 ǀ doc-3
29930930	6666	Receptionist	doc-1

Releasing Information

When the RA receives confirmation of authorization from the CSR to access Doc-1 and Doc-2 of Bob's information, it does the following:

1. It sends Bob's system ID and indexes of Doc-1 and Doc-2 to the CSP.
2. It sends Bob's system ID and trapdoor tags to the SKA.

Decrypting Information

In response to the requests received from the RA, the CSP searches for Doc-1 and Doc-2 using their indexes and send them to the user. And the SKA does the same for the trapdoors and sends them to the user. The user then has two encrypted documents and two trapdoors. The user application associates trapdoors to their corresponding documents using the string exact match mechanism. And the secret keys in the trapdoors are then used to decrypt the documents.

3.4.2. Testing the architecture

 doi:10.20944/preprints202305.0416.v1

One of the main requirements of storing healthcare information on the cloud is the protection from unauthorized users. The cloud architecture was tested in terms of its ability to prevent unauthorized cloud users from accessing the information. Four tests were performed against the implemented system, summary of the tests and results is presented in Table 7.

Table 7. Summary of architecture test results

Test	Description	Result	Pass/Fail
Unauthorized System Access	Un authorized user made request to download a trapdoor that is stored on the cloud	Access Denied	Pass
Getting access to documents stored on the cloud	A document downloaded without having the secret key.	Document was viewed in its encrypted form	Pass
Unauthorized operations	Attempt made to access the server or the database by sending queries	Access Denied	Pass
Unknown users	Unregistered user attempted to log in to the system	Login failed	Pass

4. Discussion

The adoption of cloud computing for the healthcare information systems is a major improvement in the healthcare domain, due to the significant benefits that cloud computing technology offers. The proposed architecture enables for storing and sharing patient information in a privacy-preserving manner. The proposed cloud architecture will serve the healthcare domain by storing all patients' health information in one place (cloud) so that genuine users can access it regardless of their locations. System instantiation of how the designed architecture works in terms of sharing healthcare information was presented in a scenario-based fashion. The feasibility and usability of the proposed architecture were confirmed, and the validity of the architecture in terms of preserving the privacy of information was successfully tested and proven.

4.1. Privacy-preservation

Information in the proposed architecture is encrypted before it is sent to the cloud for storage. The cloud stores encrypted data without decryption details (secret keys). This separation of information (encrypted information and secret keys) makes is difficult for cloud provider to learn the content of the information through decrypting it. Exploiting the searchable symmetric encryption scheme (SSE) ensures that cloud providers can fulfill users' requests by releasing the needed information without decrypting it. The decryption of information can only happen by genuine parties.

Dividing patients' information into many divisions according to the need of it overcomes the issue of unnecessary disclosure of information. The designed architecture categorizes patients' information into four categories of which three are for medical treatment purposes namely All_V, Em_V, and OutP_V, and one is for research purposes (R). Medical practitioners do not always require accessing the entire information about a patient every time medical treatment is needed. For example, information about a sexual disease may not be needed in urgent medical treatments such as car accidents or minor incidents such as skin wounds and cuts. This was derived from the findings of the case study data analysis and supported by the literature.

Nevertheless, the designed architecture requires obtaining patients' consent whenever accessing their information is required. The user identity management protocol (U-IDM) preserves the confidentiality of the information that is stored on the cloud and grants patients a means of control over who can access their information.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

4.2. Security Analysis

The proposed system design prioritizes privacy and security when storing and sharing healthcare data on the cloud. The proposed system design includes several security approaches on different levels of the architecture.

User Application Level

The user application in the real implementation of the system design allows only certain operations to be performed by the user. Users in the proposed system design are given accessibility to the system that is controlled by the enabled features and functions of the user application. For example, a nurse's log-in credentials enable certain functions on the user application to access the system, meaning that a nurse cannot perform operations to modify the way information is stored on the system. Moreover, the encryption and decryption processes are not controlled by the user; they are performed internally by the user application. The user application in the real implementation of the system may have the characteristic of hiding all information that is related to the encryption and decryption of information.

Access control

The proposed system design employs the concept of user identity management (U-IDM), which is the Cloud Service Registry (CSR). The CSR is a component that is not located at the user side, meaning that users cannot attempt to add or modify access rights to the system. Moreover, accessing the system can happen through requests that are sent from the user application to the Requesting Agent (RA), who authenticates and authorizes users before their requests are processed further in the system.

The collaboration between the RA and the CSR is the only way to forward users' requests to access the system. Therefore, there are three security stations in the system that the user must go through to access the system, as presented in Figure 7.

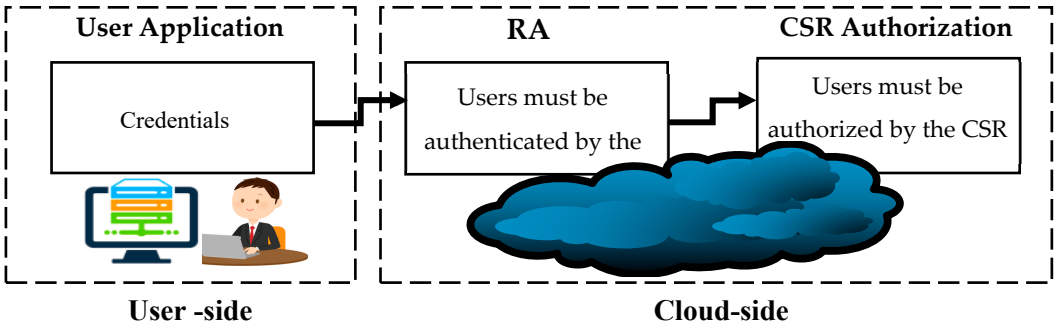

Figure 7. Security stations in the proposed system

The first station is at the user side which involves entering the users' credentials to access the system. Users' credentials in the real implementation can be through information entered by the system (user ID and Password) or other forms of credentials such as figure prints. The second station is at the cloud side which involves authenticating the user. The user must be authenticated by the RA component before the authorization process takes place. The third station involves the authorization process. Once a user is authenticated by the RA, the authorization process happens by the CSR. The CSR is not accessible by the user; it only communicates with the RA component.

The separation of information

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2023 doi:10.20944/preprints202305.0416.v1

The separation of information while it is stored in the system, makes it difficult to perform any unauthorized actions that could lead to reading the information that is stored on the cloud, especially because the cloud provider cannot learn the content of the information that is stored on the cloud. The proposed system design provides means of security to the information in its simplest implementation due to: (1) storing encrypted information and decryption keys on different components of the architecture; and (2) requiring the collaboration of three different components of the architecture to gain access to the information that is stored in the system. The information required to identify patients and their information is stored on the RA. Therefore, compromising any of these three components (CSP, SKA, or RA) will be fruitless to any party in terms of reading the information stored on the cloud.

4.3. Architecture Limitation

In practical implementation of the proposed architecture, there are some bandwidth-related issues that need to be addressed before obtaining the best of what the proposed architecture can offer in terms of information sharing. Bandwidth can be a significant issue when adopting cloud computing in healthcare, especially when it comes to accessing large amounts of data, such as medical images or electronic health records (EHRs) [63]. These issue maybe:

1. Network congestion: Healthcare organizations may experience network congestion when multiple users are accessing cloud-based services simultaneously. This can cause slowdowns, latency, or even complete loss of service.
2. Geographic location: Healthcare providers located in remote or rural areas may not have access to high-speed internet, which can make accessing cloud-based services difficult.
3. Data-intensive applications: Certain healthcare applications, such as EHRs or medical imaging systems, generate large amounts of data that need to be transferred over the internet. This can be a bandwidth-intensive process, which can lead to slowdowns and performance issues.
4. Security: High-bandwidth applications may require additional security measures to protect patient data. This can further slowdown the data transfer process and add to the bandwidth requirements.
5. Cost: Higher bandwidth requirements can result in increased costs for healthcare organizations, which may make cloud adoption less feasible.

To mitigate these bandwidth issues, healthcare organizations can take several steps to improve their network infrastructure and better support the bandwidth requirements enabling them to benefit from the advantage of what cloud computing offers in term of sharing information. These steps include:

1. Assessing their current bandwidth requirements and planning for future needs. This can help ensure that their network infrastructure can handle the bandwidth requirements of cloud-based applications.
2. Investing in high-speed internet connections and upgrading network infrastructure, including routers and switches.
3. Considering cloud providers that offer content delivery networks (CDNs) to minimize latency and speed up data transfer times.
4. Implementing data compression and deduplication techniques to reduce the amount of data that needs to be transferred over the internet.
5. Prioritizing network traffic to ensure that bandwidth-intensive applications are given priority over less critical applications.

5. Conclusion

In conclusion, information about patients' health generates special value when it is exchanged and collaboratively used among different parties involved in the healthcare domain. Cloud computing technology appears to be the dreamed vision of the healthcare industry because it matches the need for healthcare information sharing directly to various healthcare-related parties over the internet, regardless of their location and the amount of information being shared. However, the adoption of cloud computing in the healthcare domain has always been hindered due to many challenges in which information privacy is a major one. In this work, a cloud architecture was proposed for healthcare information systems to collaboratively share and use information in a privacy-preserving manner. The proposed architecture was implemented and tested in terms of its ability to share information in privacy-preserving manner. Potential challenges that may arise in the practical implementation of the proposed architecture were highlighted and recommended set of actions were provided in response to these challenges. The research findings and outcomes provide multiple directions for extending and expanding upon the scope and focus of the present research.

Firstly, getting feedback from medical practitioners on the prototype of the designed architecture is an important direction of future research. The proposed architecture has satisfied the need for sharing healthcare information in a privacy-preserving manner, however, getting feedback from the medical practitioners and experts from the healthcare domain may further validate and improve the design of the architecture to best serve the domain.

Secondly, patients' information in the present research has been categorized into four categories of which three were for medical treatment purposes, however, a research direction would refine these categories to further limit the exposure of information when it is needed for medical treatment purposes. This direction would require deeper knowledge in the medical field to allow feeding the research with more technical data related to what and when patients' health information is needed.

Moreover, the proposed architecture allows for manually enrolling patients and storing their information on the cloud, however, healthcare data is today collected using various advanced methods such as mobile devices, wearable sensors, and home wireless networks which can automatically transmit and receive data. Researchers have proven that utilizing the data collected in these methods contributes significantly to healthcare service betterment. Therefore, a research direction can be to expand the proposed architecture design to accommodate patient-generated information that is collected by these data collecting methods.

Author Contributions: Fadi Alhaddadin - System architecture design, system implementation and validation. Jairo Gutierrez provided expertise that greatly assisted and guided the research. Both authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. M. Gupta, M. Thirumalaisamy, S. Shamsher, A. Pandey, D. Muthiah and N. Suvarna, "Patient Health Monitoring using Feed Forward Neural Network with Cloud Based Internet of Things," in 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022.
2. B. W. Mamlin and W. M. Tierney, "The Promise of Information and Communication Technology in Healthcare: Extracting Value From the Chaos," The American Journal of The Medical Sciences, vol. 351, no. 1, pp. 59-68, January 2016.
3. A. L. Neves, A. W. Carter, L. Freise, L. Laranjo, A. Darzi and E. K. Mayer, "Impact of sharing electronic health records with patients on the quality and safety of care: a systematic review and narrative synthesis protocol," BMJ Open, vol. 8, no. 8, pp. 1-8, 2017.

4. S. Kalkman, J. v. Delden, A. Banerjee, B. Tyl, M. Mostert and G. v. Thiel, "Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence," Journal of Medical Ethics, pp. 3-13, 2019.
5. E. Kim, S. M. Rubinstein, K. T. Nead, A. P. Wojcieszynski, P. E. Gabriel and J. L.Warner, "The Evolving Use of Electronic Health Records (EHR) for Research," Seminars in Radiation Oncology, pp. 354-361, 2019.
6. T. Kitamura, K. Kiyohara, T. Matsuyama, T. Hatakeyama, T. Shimamoto, J. Izawa, C. Nishiyama and T. Iwami, "Is Survival After Out-of-Hospital Cardiac Arrests Worse During Days of National Academic Meetings in Japan? A Population-Based Study," Journal of Epidemiology , vol. 26, no. 3, pp. 155-162, 5 March 2016.
7. C. S. Gray, J. Barnsley, D. Gagnon, L. Belzile, T. Kenealy, J. Shaw, N. Sheridan, P. W. Nji and W. P. Wodchis, "Using information communication technology in models of integrated community-based primary health care: learning from the iCOACH case studies," Implementation Science, pp. 1-14, 2018.
8. W. Oude, L. v. Velsen, M. Huygens and H. Hermens, "Requirements for and Barriers towards Interoperable eHealth Technology in Primary Care," IEEE Internet Computing , pp. 10-19, 2015.
9. F. Bélanger and R. E. Crossler, "Privacy in the digital age: a review of information privacy research in information systems," MIS Quarterly, vol. 35, no. 4, pp. 1017-1042, December 2011.
10. T. White, E. Blok and V. D. Calhoun, "Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed," Special Issue: The ENIGMA Consortium: the first 10 years, pp. 278-291, 2022.
11. F. N. Wirth, T. Meurers, M. Johns and F. Prasser, "Privacy-preserving data sharing infrastructures for medical research: systematization and comparison," BMC Medical Informatics Decision Making, pp. 1-13, 2021.
12. Public Law , "HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT OF 1996," Public Law 104-191, 104th Congress, 1996.
13. G. Gunasekara and E. Dillon, "Data Protection Litigation in New Zealand: Processes and Outcomes," Victoria University of Wellington Law Review (VUWLR) , vol. 39, 2008.
14. A. Gkoulalas-Divanis and G. Loukides, "Introduction to Medical Data Privacy," in Medical Data Privacy Handbook, Switzerland, Springer International Publishing, 2015, pp. 1-14.
15. S. Meng, S. Fan, Q. Li, X. Wang, J. Zhang, X. Xu, L. Qi and M. Z. A. Bhuiyan, "Privacy-Aware Factorization-Based Hybrid Recommendation Method for Healthcare Services," IEEE Transactions on Industrial Informatics, pp. 5637 - 5647, 2022.
16. G. Gürsoy, T. Li, S. Liu, E. Ni, C. M. Brannon and M. B. Gerstein, "Functional genomics data: privacy risk assessment and technological mitigation," Nature Reviews Genetics, p. 245–258, 2022.
17. M. Tanriverdi, "A Systematic Review of Privacy Preserving Healthcare Data Sharing on Blockchain," Journal of Cybersecurity and Information Management (JCIM) , pp. 31-37, 2020.
18. S. M. Blackman, "Towards a Conceptual Framework for Persistent Use: A Technical Plan to Achieve Semantic Interoperability within Electronic Health Record Systems," in Proceedings of the 50th Hawaii International Conference on System Sciences, 2017.
19. J. Shahid, R. Ahmad, A. K. Kiani, T. Ahmad, S. Saeed and A. M. Almuhaideb, "Data Protection and Privacy of the Internet of Healthcare Things (IoHTs)," Applied Sciences, pp. 1-22, 2022.
20. Priyanga.P and MuthuKumar.V.P, "Cloud computing for healthcare organisation," International Journal of Multidisciplinary Research and Development , pp. 487-493, 2015.
21. V. Casola, A. Castiglione, K.-K. R. Choo and C. Esposito, "Healthcare-Related Data in the Cloud: Challenges and Opportunities," IEEE Cloud Computing, 2016.
22. A. Svensson, "Challenges in Using IT Systems for Collaboration in Healthcare Services," Int J Environ Res Public Health, pp. 1-12, 2019 .
23. E. Griffith, "What Is Cloud Computing?," 03 May 2016. [Online]. Available: http://au.pcmag.com/networking-communications-software-products/29902/feature/what-is-cloud-computing. [Accessed 21 March 2017].
24. K. Cresswell, A. D. Hernández, R. Williams and A. Sheikh, "Key Challenges and Opportunities for Cloud Technology in Health Care: Semistructured Interview Study," JMIR Human Factors, 2022.
25. P. Mell and T. Grance, "The NIST Definition of Cloud Computing," in National Institute of Standards and Technology , 2011.
26. C. Doukas, T. Pliakas and I. Maglogiannis, "Mobile healthcare information management utilizing Cloud Computing and Android OS," in Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, 2010.
27. L. Griebel, H.-U. Prokosch, F. Köpcke, D. Toddenroth, J. Christoph, I. E. Ines Leb and M. Sedlmayr, "A scoping review of cloud computing in healthcare," BMC Medical Informatics and Decision Making, pp. 1-16, 2015.
28. R. Zhang and L. Liu, "Security Models and Requirements for Healthcare Application Clouds," in IEEE 3rd International Conference on Cloud Computing (CLOUD) ,, 2010.

29. M. Sharma and R. Sehrawat, "A hybrid multi-criteria decision-making method for cloud adoption: Evidence from the healthcare sector," Technology in Society, 2020.
30. B. Yüksel, A. Küpçü and Ö. Özkasap, "Research issues for privacy and security of electronic health services," Future Generation Computer Systems, pp. 1-17, 2017.
31. H. Raj, M. Kumar, P. Kumar, A. Singh and O. P. Verma, "Issues and Challenges Related to Privacy and Security in Healthcare Using IoT, Fog, and Cloud Computing," in Advanced Healthcare Systems: Empowering Physicians with IoT-Enabled Technologies, Scrivener Publishing LLC, 2022.
32. H. Aziz and A. Guled, "Cloud Computing and Healthcare Services," Journal of Biosensors & Bioelectronics, 2016.
33. S. Chenthara, K. Ahmed, H. Wang and F. Whittaker, "Security and Privacy-Preserving Challenges of e-Health Solutions in Cloud Computing," IEEE Access, pp. 74361-74382, 2019.
34. C. R. Weir, K. W. Hammond, P. J. Embi, E. N. Efthimiadis, S. M. Thielke and A. N. Hedeen, "An exploration of the impact of computerized patient documentation on clinical collaboration," International Journal of Medical Informatics, vol. 80, no. 8, pp. 62-71, August 2011.
35. B. E. Dixon, P. J. Embi and D. A. Haggstrom, "Information technologies that facilitate care coordination: provider and patient perspectives," Translational Behavioral Medicine, vol. 8, no. 3, p. 522–525, 2018.
36. M. M. Bertagnolli, B. Anderson, A. Quina and S. Piantadosi, "The electronic health record as a clinical trials tool: Opportunities and challenges," Clinical Trials, vol. 17, no. 3, pp. 237-242, 2020.
37. X. Liu, Z. Wang, C. Jin, F. Li and G. Li, "A Blockchain-Based Medical Data Sharing and Protection Scheme," IEEE Access, pp. 118943-118953, 2019.
38. G. Cordovano and S. N. Shah, "Requesting Medical Records," Journal of Ahima, 2020.
39. D. Gupta, S. Malik and A. Rana, "Adopting Semantic Interoperability for Improved Healthcare," SSRN, 2022.
40. N. Rajkumar, M. Muzoora and S. Thun, "Dentistry and Interoperability," Journal of Dental Research, pp. 1-5, 2022.
41. L. Devadass, S. S. Sekaran and R. Thinakaran, "Cloud Computing in Healthcare," International Journal of Students' Research In Technology & Management, vol. 5, no. 1, pp. 25-31, 2017.
42. T. Quatrani, "Introduction to the Unified Modeling Language," IBM, 2003.
43. T. Górski, "UML Profile for Messaging Patterns in Service-Oriented Architecture, Microservices, and Internet of Things," Applied Sciences, vol. 12, no. 24, 2022.
44. K. Thramboulidis and F. Christoulakis, "UML4IoT—A UML-based approach to exploit IoT in cyber-physical manufacturing systems," Computers in Industry, pp. 259-272, 2016.
45. R. J. Petrasch and R. R. Petrasch, "Data Integration and Interoperability: Towards a Model-Driven and Pattern-Oriented Approach," Modelling, pp. 105-126, 2022.
46. L. Pufahl, F. Zerbato, B. Weber and I. Weber, "BPMN in healthcare: Challenges and best practices," Information Systems, 2022.
47. M.-T. Schmidt, B. Hutchison, P. Lambros and R. Phippen, "The Enterprise Service Bus: Making service-oriented architecture real," IBM Systems Journal, pp. 781-797, 2005.
48. N. Niknejad, W. Ismail, I. Ghani, B. Nazari, M. Bahari and A. R. B. C. Hussin, "Understanding Service-Oriented Architecture (SOA): A systematic literature review and directions for further investigation," Information Systems, vol. 91, 2020.
49. O. Aziz, M. S. Farooq, A. Abid, R. Saher and N. Aslam, "Research Trends in Enterprise Service Bus (ESB) Applications: A Systematic Mapping Study," IEEE Access, pp. 31180 - 31197, 2020.
50. C. C. Agbo, Q. H. Mahmoud and J. M. Eklund, "Blockchain Technology in Healthcare: A Systematic Review," Healthcare, vol. 7, no. 2, 2019.
51. H. Jin, Y. Luo, P. Li and J. Mathew, "A Review of Secure and Privacy-Preserving Medical Data Sharing," IEEE Access, vol. 7, pp. 61656-61669, 2019.
52. A. A. Abdellatif, A. Z. Al-Marridi, A. Mohamed, A. Erbad, C. F. Chiasserini and A. Refaey, "ssHealth: Toward Secure, Blockchain-Enabled Healthcare Systems," IEEE Network, vol. 34, no. 4, pp. 312 - 319, 2020.
53. R. Guo, H. Shi, D. Zheng, C. Jing, C. Zhuang and Z. Wang, "Flexible and Efficient Blockchain-Based ABE Scheme With Multi-Authority for Medical on Demand in Telemedicine System," IEEE Access, vol. 7, pp. 88012 - 88025, 2019.
54. P. Xi, X. Zhang, L. Wang, W. Liu and S. Peng, "A Review of Blockchain-Based Secure Sharing of Healthcare Data," Applied sciences, 2022.
55. J. Qu, "Blockchain in medical informatics," Journal of Industrial Information Integration, vol. 25, 2022.
56. F. Alhaddadin, J. A. Gutiérrez and W. Liu, "Privacy-aware cloud-based architecture for sharing healthcare information," Auckland University of Technology, Auckland, New Zealand, 2020.
57. R. Curtmola, J. Garay, S. Kamara and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," in 13th ACM Conference on Computer and Communications Security, 2006.

58. D. Boneh, G. D. Crescenzo, R. Ostrovsky and G. Persiano, "Public Key Encryption with keyword Search," in International Conference on the Theory and Applications of Cryptographic Techniques, Berlin, Heidelberg, 2004.
59. Y.-C. Chang, "Single Database Private Information Retrieval with Logarithmic Communication," in The 9th Australasian Conference on Information Security and Privacy, Sydney, Australia, 2004.
60. E. Goh, "Secure Indexes," IACR ePrint Cryptography Archive, 2003.
61. S. Eludiora, O. Abiona, A. Oluwatope, A. Oluwaranti, C. Onime and L. Kehinde, "A User Identity Management Protocol for Cloud Computing Paradigm," Int. J. Communications, Network and System Sciences, vol. 4, pp. 152-163, 2011.
62. Amazon, "aws," 2019. [Online]. Available: https://aws.amazon.com.
63. J. S. Marwaha, A. B. Landman, G. A. Brat, T. Dunn and W. J. Gordon, "Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation," npj Digital Medicine, 2022.