Towards Secure and Privacy Preserving e-Health Data Exchanges Through Consent Based Access Control

by

Abel Bradley Saed Bacchus

A thesis submitted in partial fulfillment
of the requirements for the degree of

Masters of Science

in

Computer Science

University of Ontario Institute of Technology

September 2017

Abstract

How we administer healthcare continues to evolve alongside the advancement of information technology. As we become more connected, the Internet of Things and our want to share information in a timely manner encourage us to redefine and enhance how we exchange health information. A fully integrated, universal health record system in Canada remains a distant goal. It requires thoughtful legislation, sufficient resources and the best of our technological and security implementation before realization. Nevertheless, we need such a system and are steadily working towards it.

While there are a number of obstacles in attempting a universal health record system, this thesis presents a solution for secure health information exchanges. A valuable component in establishing a complete framework for all health information exchanges. We present two protocols. Consent based access control (CBAC) and a fairness aware privacy preservation protocol (FAPP). These two protocols grant patients control in how their sensitive health information is used and provides avenues for certain third parties to collect patient information without compromising security and privacy.

Acknowledgements

I would like to thank my supervisor, Dr. Xiaodong Lin and co-supervisor Ying Zhu, for their patience and encouragement. I would also like to thank Dr. Aiqing Zhang for her beneficial discussion and help. Finally, I would like to thank my family for their love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Balancing the privacy and security of sensitive health records with our need to share and process information rapidly is a challenging process. Patients demand privacy. They want satisfaction in knowing that their health information is used for specific purposes and only by those given express permission. Healthcare providers aspire to provide exceptional and timely care. This often calls for access to sensitive patient details. This is also perplexing for non-critical data requesters like insurance institutions. In this report we introduce new protocols to structure how information is shared among entities while maintaining patient confidentiality.

The advancement of information technology and the Internet of Things (IoT) is redefining modern day interactions. Healthcare has experienced a fundamental overhaul of processes at all levels [16]. Improvements in healthcare have introduced more effective means to collect and analyze data. Furthermore, information can now be accessed and shared at an unprecedented rate. New IT tools serve well to educate and empower patients, allowing them to have more input into whom and how their information is shared. In these innovative times, we adapt new policies and practices to ensure data moves as securely and efficiently

as possible. Global trends in healthcare point towards change in how care is administered. Patients hope to exercise more choice. They wish to see providers on their own terms, at their own convenience and require personalized care with easy and timely access to their own records. All communities and demographics face their own challenges. Hospitals and medical organizations must implement complex multi-organizational networks and processes to address those challenges efficiently [19].

Healthcare is a complicated issue. Meaningful change requires thoughtful legislation, subject to budgetary and privacy restrictions. A secure healthcare exchange framework is a major requirement towards successful centralized healthcare records. Therefore, we outline a framework for securely exchanging healthcare information with patient consent and preserving privacy.

1.1 Background and Motivation

Healthcare records and information exchanges are steadily trending towards becoming entirely electronic. Many practices, however, still store records in paper format and requisitions are handled through fax and phone. Nevertheless, as older practitioners retire and we adapt newer technologies we should expect to see a larger shift towards electronic health exchange. As more of our information moves online, we can expect more concern for privacy preservation. Accordingly, several standards and bills have been introduced to address privacy concerns and standardize electronic information. These range from PIPEDA, HIPAA to HL7. However, rules and standardization are insufficient in meeting growing cybersecurity demands. In May, 2017, a wave of cyber attacks infected over 45,000 computers across 74 countries. The attacks ultimately resulted in the closure of 16 hospitals

who were unable to retrieve basic medical files [11]. Ransomware perpetrators requested a total of 300 Bitcoin, equivalent to approximately $760,000 USD at that time.

There have been several attempts to create security mechanisms for electronic health records. Unfortunately, due to the nature of e-Health records a universally accepted mechanism for exchanges has not come to fruition. Securing medical information is challenging. Specifically, balancing record ownership and control between medical providers and patients is difficult. This hampers the steady implementation of newer technologies. Current policies allow for a number of permitted data custodians with owners being allowed to grant, block or revoke access. These privileges should allow record owners to also retroactively annul access to any requesting parties. As we become more immersed in the Internet of Things, ensuring that patients control access to their own information is paramount. Granting them the freedom to control their own information would enhance patient centered care and improve patient mobility in socialized healthcare systems. Currently, finding, retaining and moving between primary care physicians is difficult and at times costly.

We define consent as access to records granted by a patient. This consent is exercised to give data requesters permission to read and edit records. Consent should be revocable after issuance. Informed consent is crucial in ensuring that patients are aware of associated risks in the future and should also be a cornerstone in healthcare information sharing.

Canada is one of the foremost healthcare providers in the world. Being publicly funded, Canadian healthcare is socialized and administration is handled by provinces or territories based on the guidelines set by the federal government. All Canadian citizens are entitled to basic health coverage regardless of medical history.

Canada also boasts state of the art treatment with one of the highest life expectancies and lowest infant mortality rates in the world [12]. Unfortunately, a growing elderly population introduces challenges. Longer lifespans also mean that healthcare providers pay more per individual over his/ her lifetime. Furthermore, Canada's population continues to grow through immigration as others look to Canada for social and economic security. According to [49], Canada in 2016 spent $228 billion CAD on healthcare, representing nearly 11% of Canada's gross domestic product. This exceeds the previous healthcare bill of $219 billion [13]. Furthermore, as evident in Table 1.1, Canada's expenditure on health has increased year after year. With Canada's rising immigration, and aging population it is likely that there will be an increase in future expenditure. However, rising healthcare costs are not sustainable and we should expect to see changes to mitigate costs in the coming years.

Ontario has attempted to implement a centralized healthcare system in prior years. Work continues to progress on eHealth Ontario [17], an initiative to connect all electronic Health record systems. Unfortunately, eHealth Ontario has been plagued with scandal. The Ontario Auditor General, Jim McCarter, called the project a "$1 Billion waste [30]." The scandal resulted in the Health Minister at the time, David Caplan, resigning. These failings have reduced public confidence in a workable centralized health exchange system. However, much of the scandal concerned misappropriation of funds [30]. There is still an interest in more efficient health information exchanges and a connected Canadian healthcare system. The auditor general went on to say that the implementation of electronic health records in every province could save them $6 billion. While Canadians benefit from universal healthcare, a universal Electronic Health Record (EHR) System would improve care and reduce costs [5]. A RAND study found that the USA could to save approximately $81 billion USD annually by moving to a universal EHR system.

Year	Percentage of GDP Spent on Health	Total Cost Per Year	Difference from Previous Year	Percent increase from previous year
2000	8.3	91 229.7	N/A	N/A
2001	8.7	98 714.1	7848.4	8.20
2002	8.9	105 721.3	7007.2	7.10
2003	9.0	113 063.6	7342.3	6.94
2004	9.1	121 100.5	8036.9	7.11
2005	9.1	128 444.9	7344.4	6.06
2006	9.2	137 388.3	8943.4	6.96
2007	9.3	146 313.8	8925.5	6.50
2008	9.5	156 450.9	10137.1	6.93
2009	10.6	167 734.9	11284	7.21
2010	10.6	175 558.2	7823.3	4.66
2011	10.2	180 880.0	5321.8	3.03
2012	10.2	186 344.4	5464.4	3.02
2013	10.1	191 940.4	5596	3.00
2014	10.0	198 054.2	6113.8	3.19
2015	10.3 *	203 666.9 *	5612.7	2.83
2016	10.3 *	209 481.0 *	5814.1	2.85

Table 1.1: **EXPENDITURE ON HEALTHCARE PER YEAR [32]**

Currency Measured in Millions of Canadian Dollars
* Provisional Value

The Canadian healthcare system, as it is, faces a number of dilemmas. The Conference board of Canada, in their 2012 Summit on Sustainable Health and Health Care, discussed at length a number of those issues in [29]. Canada was given an overall grade of B on its report card and trailed behind 9 other OECD countries. Japan, Switzerland and Italy held the top positions with each of them receiving an A grade. While clinical procedures have evolved, the current healthcare system is not efficiently configured to maximize efficiency. Constraints that inhibit progress include:

- Aging physical structures
- Old service delivery models

- A lack of provider incentives
- Labor contracts
- Stagnant information flow

Healthcare model of the 1960s. [29] Argues that the current Canadian healthcare system is locked in a model from the 1960's. Healthcare systems in Canada were designed to protect patients financially should they face health calamities. More specifically, if patients are met with disastrous health diagnoses or are faced with intensive surgery, associated costs would not become burdensome. In much of these cases, patients received acute care primarily through hospitals. However, much has changed since our healthcare system was initiated. For one, much treatment is delivered through a variety of non-hospital sites. These include homes, walk-ins and community clinics. The Canadian healthcare model was not originally designed for such healthcare delivery methods. Furthermore, advances in electronic records have not been readily adapted. Parties that use EMRs utilize different platforms. Oftentimes, there are difficulties in communicating among these different systems. Thus, there is a lack of cohesion. According to [29], new healthcare strategies must:

1. Fix Healthcare System gateways
2. Invest and use new technology
3. Enhance the current compensation model and related contracts
4. Focus on the health and wellness of Canadians
5. Build a more transparent and accountable healthcare system
6. Empower patients through consent

Current Issues. Much can be done to improve our healthcare system. A few of the current issues facing Canadian healthcare are summarized below:

1. Financial complications: These include wastage and inefficient usage of funds.

2. Communication Shortfalls: The inability of medical practitioners to send and receive records efficiently.

3. Lack of patient control: Patients have limited control over access to their records. They should be able to grant, block and retroactively revoke access to specific parties with ease.

4. Lack of consideration and inclusion for third party non-medical entities that may require access. Organizations, such as medical insurers, may need to verify medical information without direct access. Currently, access may result in these institutions receiving excessive information.

1.2 Objectives and Contributions

Our main contributions are as follows:

1. We propose a framework for secure and centralized healthcare information exchange. Original records are encrypted and offered to a data center. Others cannot access data without its owner's consent. Additionally, we propose methods to partially expose information to non-essential entities.

2. We propose a consent based access control mechanism for health information exchanges [54]. Data requesters must first negotiate with owners to access health information. Once an agreement is made, a consent token is given so that the intended party can access information.

3. We propose a conditional proxy re-encryption algorithm, which allows data centers to re-encrypt information without exposing plaintext. Conditions are integrated into the re-encryption key to enhance collusion resistance. Mutual authentication and contextual privacy are also achieved by using the public keys of receivers in the encryption algorithm.

4. Finally, we propose a method to preserve privacy in health insurance quotes by adopting k-anonymity techniques and propose a fairness-aware and privacy-preserving insurance application protocol [55].

1.3 Thesis Organization

The remainder of this thesish will organized as follows. Chapter 2 discusses relevant works pertaining to securing and preserving privacy in patient records during data exchanges. Chapter 3 introduces the security primitives used in this thesis. These serve as the basis for our proposed schemes and are fundamental in cryptographic exchanges. We then describe our framework for Consent Bases Access Control in Chapter 4. Next, we introduce a Fairness Aware Privacy Protocol in Chapter 5. Protocols introduced in this thesis will include sections that describe protocol architecture, functionality, security analyses and performance evaluations. Finally, we conclude and discuss our future work in Chapter 6.

Chapter 2

Related Work

2.1 Security and Privacy Preservation in e-health Systems

In the age of information, privacy preservation remains an integral part in the development of electronic frameworks. We require privacy in the most mundane sites and online tools. Measures must be made to ensure that our data remains in the right hands. With more sensitive information, the need for privacy preservation rises significantly. This is not lost on researchers who continue to develop frameworks for healthcare [24], [27]. Electronic Patient Records or EPRs, contain sensitive details on ones health. Therefore, data confidentiality is essential in the development of acceptable systems. An individual's medical records tend to accumulate much information over their lifetime. These records can include digital and rendered images, diets, medication, sexual preferences and psychological profiles [28]. Many diagnoses and treatments carry with them great social stigma and discrimination [6].

There currently exists several threats to health information systems. Rindfleisch

in [37], comments on our apprehensions to adapt newer security tools and our inability to effectively balance policies with evolving technologies. [39] has listed a number of threats and categorized them as either power failure, acts of human error, technical obsolescence or hardware and software errors. Regarding health information, [37] categorizes threats to health information privacy as organizational or systemic threats. Organizational threats include accidental disclosure, insider curiosity, unauthorized network intrusions and physical data breaches by internal or external parties [5]. On the other hand, Systemic threats refer to those threats to patient privacy where those with legal access to information can abuse patient information. This may be in the form of an insurance company denying coverage or claims based on medical records or work places denying employment to those with pre-existing conditions.

In this section we present background information, related work and a literature review on access control and privacy preserving techniques for electronic health record systems.

2.1.1 E-Consent

Informed consent is legally required when performing medical procedures. E-Consent is a model approach which allows patients more control over who can access their information [15]. Generally, patients provide blanket consent for access to their information. This allows organizations to request information in the future when such requests are unwarranted. E-Consent is a broad term, specific forms can be categorized as follows [15]:

General consent grants blanket permission to access information. This may become problematic if patients move, change physicians, or only perform a singular

procedure at a site. Furthermore, there is no limit to what data can be requested.

General consent with specific denials grants consent as above but can restrict access to particular information and deny disclosure to specific parties or for specified purposes.

General denial with specific consent in contrast to the above, blocks all access with the exception to particular information, or for specific parties or purposes.

General denial patient provides blanket denial. For every instance of care, a new consent would be required from the patient.

E-Consent systems provide the following [15]:

1. Verify an individual's identity. Check whether patients and data requesting parties are who they say they are.
2. Check that an individual's affiliation with requesting parties is valid, that they are members of permitted health organizations and that they are permitted to represent those health organization based on their clinical roles.
3. Verify health organization existence and identities.
4. Recognize and register defined purposes behind consent requests.
5. Record whether consent has been allowed or denied by the patient or authorized agent presently or in the past.
6. Retrieve any consent instruction associated with clinical data.

7. Match level consent to data access. This is generally done by levels. This would define what data is released based on level of consent. For example: data bound by an instance of care, or data bound to an entire patient record. This defines access based on the level of consent.

8. Record access control details or sets of complex consent instructions.

9. Verify the existence of data requesters and their roles or association with an organization.

10. Record the delegation of consent by one party to another.

2.1.2 Access Control Mechanisms

Access Control Mechanisms are procedures created to protect information and prevent unauthorized access to sensitive data assets. These mechanisms are crucial in curtailing unwarranted access to patient records.

Discretionary Access Control (DAC) grants creators or owners of objects to administer access rights. Owners can also grant and revoke access to other users or groups. DACs often rely on access control lists for implementation. The DAC model is based on resource ownership where identity plays a key role in access control.

Mandatory Access Control (MAC) is a traditional model where permissions are assigned by an administrator. Access to resources can only be granted or revoked by administrators or users with elevated privileges.

Role Based Access Control (RBAC) permissions are assigned based on roles [41], [18]. Users are appointed roles with set privileges. Roles are assigned based on

credentials, qualifications or responsibilities as established by organizations. Privileges can be revoked, updated or assigned easily. RBAC provides hierarchical features. More privileged employees inherit lower privileged rights. RBAC was originally developed by NIST and distributed as INCITS 359-2004 by the International Committee for Information Technology Standards in 2004 [21]. Since its inception, RBAC has become a de facto standard for access control. It has experienced several iterations and has formed the foundation for a number of new access control frameworks. RBAC is an open ended concept and is open to interpretation. Implementations may be simple or complex. RBAC does not have a sole definitive model as it may have too little or too many constraints. NIST has organized their model for RBAC into a four step sequence with each step increasing in complexity and capabilities [40]. Figures 2.1-2.6 are simple representations of RBAC protocol architectures. Variations of RBAC implementations are detailed below:

1. **Flat RBAC**: This is the simplest RBAC model and embodies only the essential RBAC aspects and is based on traditional group access control. Users are assigned roles and roles are assigned permissions. Users then acquire their permissions through the roles that they are assigned. User-to-role and permission-to-role assignment have a many to many relationship. Therefore, a single user can be given many roles and a role can be issued to a number of users. User-review mandates that user and role assignment relations can be traced. Lastly, Users must also be allowed to simultaneously exercise multiple roles which prevents users from being restricted by roles being activated one at a time.

2. **Hierarchical RBAC**: This category requires the addition of role hierarchies. We can understand a hierarchy as the mathematical partial orderings which

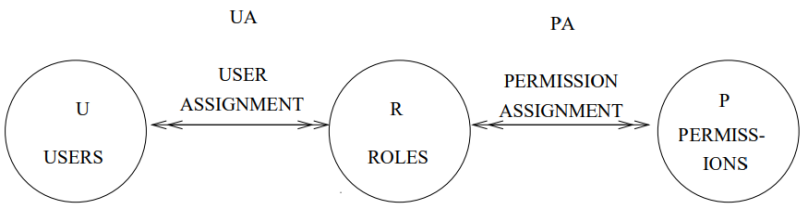

Figure 2.1: Flat RBAC [40]

define relations among roles. Partial orders are relations which are transitive, reflexive and anti-symmetric. Senior roles acquire permissions from their juniors. NIST defines two sub categories for Hierarchical RBAC as:

- General Hierarchical RBAC: An arbitrary partial order is imposed. This serves as the role hierarchy.
- Restricted Hierarchical RBAC: In this implementation, restrictions are used to control hierarchy structures. These may be in the form of trees, inverted trees or customized structures.

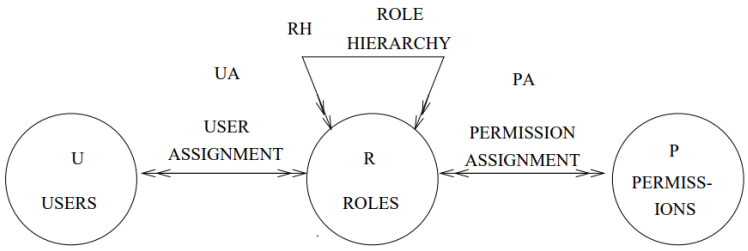

Figure 2.2: Hierarchical RBAC [40]

Role hierarchies may be implemented as either inheritance or activation hierarchies. In Inheritance hierarchies, senior roles often inherit from junior roles, this is called permission inheritance. Cases where senior roles do not automatically activate junior roles fall under the activation interpretation of hierarchies. These are referred to as activation hierarchies.

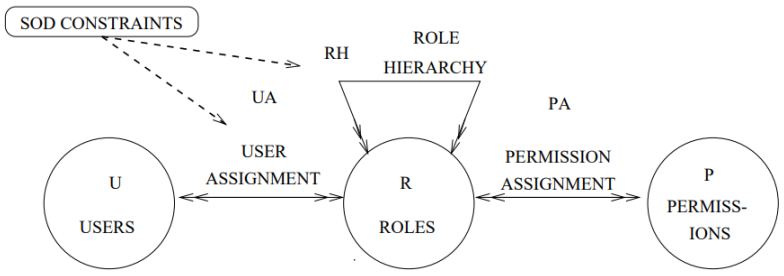

Figure 2.3: Constrained RBAC - Statatic SOD [40]

3. **Constrained RBAC**: At this level constraints are added to RBAC. Constrains can either be associated with user-role assignment as in Figure 2.3 or the activation of roles within user sessions as in Figure 2.4. Constraints are also inherited within role hierarchies. Separation requirements are used to implement conflict-of-interest policies. Conflict of interest can arise in an RBAC system when a user gains authority for permissions with conflicting roles. These are used to ensure users do not overstep their authority. Constrained RBAC is separated into two categories based on separation of duty (SOD). Separation of duty refers to how tasks and privileges are assigned among roles to prevent users from gaining excessive authority. SOD is used to mitigate fraud and damage. IT administration aggregates responsibilities and authority for duty amongst several users, thereby requiring the involvement of multiple users prior to a fraudulent or damaging activity. RBAC follows the principle of least privilege. Least privilege refers to the administrative practice of minimizing user permissions so that users have only sufficient permissions to perform their functions.

   The NIST model for Constrained RBAC allows for both static and dynamic separation of duty. Static separation of duty (SSD) addresses conflict of interest by constraining how roles are assigned to users. More specifically, if a user

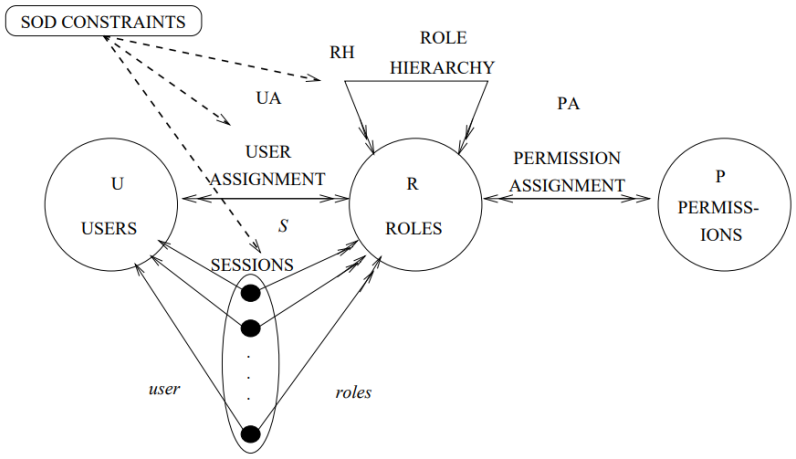

Figure 2.4: Constrained RBAC - Dynamic SOD [40]

is authorized as a member of one role, they may be prohibited from having other specific roles. With SSD, inheritance is limited to prevent conflict of interest. Dynamic Separation of Duty (DSD) allows RBAC administrators to use organization specific policies. This allows users to gain otherwise conflicting roles during enrollment. However, NIST does not permit a user to assume multiple roles which would conflict simultaneously. DSD generally allows for greater operational flexibility.

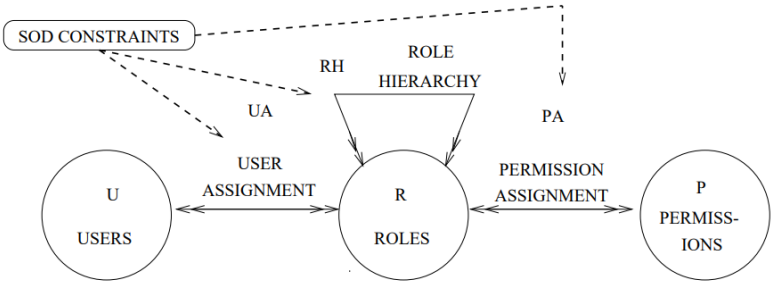

Figure 2.5: Symmetric RBAC - Static SOD [40]

4. **Symmetric RBAC**: Symmetric RBAC includes a permission-role review. Especially in growing organizations, permission schemes evolve. Older permission-role assignments may no longer be relevant or become inappropriate as sit-

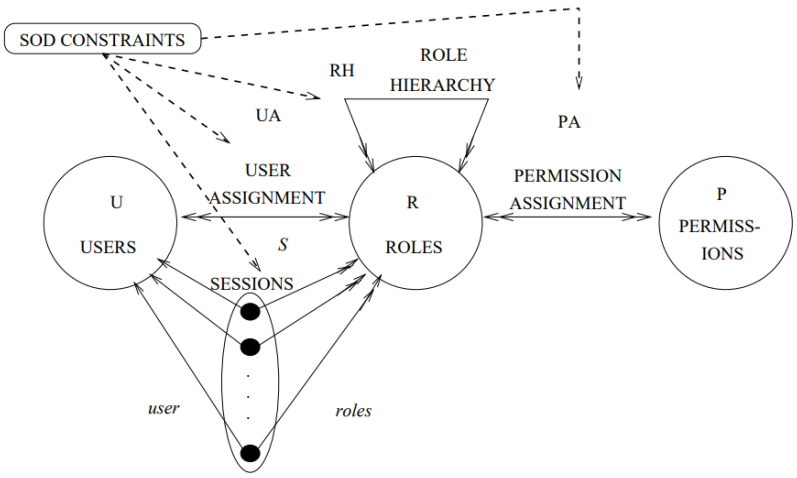

Figure 2.6: Symmetric RBAC - Dynamic SOD [40]

uations change. This becomes more tedious as users and roles are spread across different administrative boundaries. Effectively maintaining permission assignments necessitates identifying and reviewing the assignment of permissions to roles. This is done to ensure the principle of least privilege. An organizational review of permission assignments can be useful in a number of situations. When a user, leaves an organization or department, switches jobs or a certain set of permissions become obsolete, permission-role reviews will be useful. Deleting all of a user's data and accounts after they have left or having administration revoke all associated permissions would still leave a system with data garbage. Older permissions or roles could still haunt an employee when they switch jobs within an organization, however, deleting those permissions would inhibit their ability to work. Moreover, the complexity of permission-role reviews may not be necessary for some smaller organizations or organizations with limited roles. Therefore, a separate level is necessary in RBAC to ensure permission-role integrity throughout organizations.

A summary of RBAC at different levels can be found in Table 2.1.

Level	Name	RBAC Functional Capabilities
1	**Flat RBAC** (see Figure 2.1)	• Users must acquire permissions through roles • There must be support for many to many user role assignment • There mus be support for many to many permission-role assignment • There must be suport for a user-role assignment review • Users must be able to use permissions from multiple roles simultaneously
2	**Hierarchical RBAC** (see Figure 2.2)	Flat RBAC + • Must include support for role hierarchy • **level 2a** Support for arbitrary hierarchies • **level 2b** Support for limited hierarchies
3	**Constrained RBAC** (see Figure 2.3 & 2.4)	Hierarchical RBAC + • There must be enforcement of separation of duties • **level 3a** Support for arbitrary hierarchies • **level 3b** Support for limited hierarchies
4	**Symmetric RBAC** (see Figure 2.5 & 2.6)	Constrained RBAC + • There must be support for permission-role review, the performance comparable to user-role review • **level 3a** Support for arbitrary hierarchies • **level 3b** Support for limited hierarchies

Table 2.1: **SUMMARY OF RBAC VARIATIONS ORGANIZED BY LEVEL [40]**

Team Based Access Control (TMAC) is an approach to applying role-based access control in collaborative environments. Teams are abstractions which encapsulate collections of users according to roles.

2.1.2.1 Experimental Access Control Methods

Ni and Trombetta introduce a Privacy Aware version of RBAC [31]. Privacy Aware Role Based Access Control or P-RBAC is an extension of the RBAC model. It adds support for the expression of privacy related policies. P-RBAC represents a family of conceptual RBAC models; these are illustrated in figure 2.7. Each component of the P-RBAC family adds rich feature sets if needed. Core P-RBAC is the foundation of all P-RBAC implementations. Hierarchical P-RBAC adds Role, Data and Purpose hierarchies to the Core. Conditional P-RBAC allows for Permission Assignment Sets and Boolean Expressions. Universal P-RBAC combines the features from Conditional and Hierarchical P-RBAC.

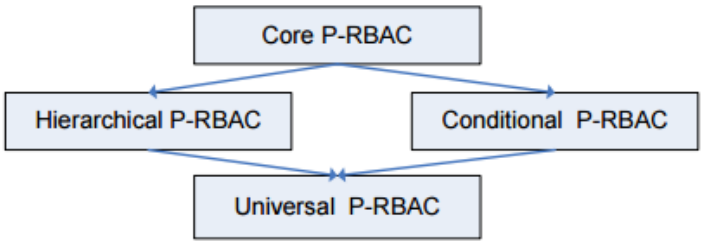

Figure 2.7: The family of conceptual P-RBAC models [31]

Russello et al. in [38] present a framework for consent based workflows. Their framework was designed for healthcare systems and can enforce consent based access control as well as the need to know principle. Furthermore, they attempt to release end users, in this case healthcare professionals, from the responsibility of security related configurations. This allows medical professionals to prioritize healthcare related duties.

2.2 Privacy and e-Health Communication Standards

There are a number of standards that outline the minimum security requirements for Electronic Health Records. ISO TC 215 [23] and ISO 18308 [8] provide technical specifications for EHRs and EHR architectures. Other standards include DICOM which standardizes imaged based information, The Continuity of Care Record, HISA (EN 12967), CONTSYS (EN 13940) and ANSI X12 (EDI) [43]. OpenEHR and Health Layer 7 or HL7 standardizes communications between physical and electronic record systems [42]. These standards provide flexible guidelines in how information is exchanged.

2.2.1 Mobile Healthcare

Works are ongoing to create more mobile and ubiquitous healthcare. There are massive adoptions worldwide to employ more wireless infrastructure to include emerging wireless applications [51]. The motivation for such moves are to enhancement of healthcare convenience and access. Much of the work revolves around remote monitoring and making conventional technologies more mobile and accessible.

2.2.2 Employing Consent and Access Control in Health Frameworks

The concept of adding consent to medical frameworks is not a new one. While consent is not usually built into to access control models, it remains a vital part of the medical processes. Attempts have been made implement and integrate consent and access control, however, these have not seen widespread adoption. In early 1996, Anderson proposed the British Medical Association Security Policy in [4]. It

later became a basis for access control policy for the British NHS. In this section we look at the implementation of consent and access control mechanisms in the healthcare field.

Cassandra [9] is a language for expressing access control policies on large scale distributed systems. Cassandra is role based and was designed specifically for Electronic Health Record systems (EHRs) in the UK. Cassandra supports credential based access control and remote policies can be queried. Cassandra provides dynamic RBAC, role revocation, distributed trust management and negotiation. In simpler terms, Cassandra is a framework for policy language for access control management in large systems. With Cassandra, consent is applied through a two step appointment mechanism. Patients issue consent roles while clinicians issue consent request roles. Mutual acceptance result in the creation of consented relationships.

The model in [33] describes a decentralized approach to electronic consent and health information access control. Part of the design includes the development of several transfer protocols for the transference of health and consent information. By default, data is protected through transfer protocols according to a patient's specified consent conditions. This model introduces the concept of placeholders which allow for preset consent configurations for records. The model for a healthcare system includes independent but cooperating health facilities where there are no centralized storage for health data and no centralized patient registrar. Consent is managed by local eConsent systems which allow the electronic recording of patient consent for access to health information and uses these consent conditions to manage access. All transfers between facilities are initiated by destination facilities, transfer commencement depends on the acceptance of consent conditions.

2.3 Public Health Information Management

Health information is distributed across business-to-business (B2B) healthcare networks. Electronic record keeping is not legally mandated. Many smaller organizations still house paper based record systems. Therefore, much information must be transmitted via fax or one time deliveries where physical copies are replicated and shared.

Typically national healthcare systems are structured hierarchically, Canada has a national authority on health which governs provincial authorities. Local authorities, clinics, hospitals and family doctors report to and are funded through provincial authorities. However, each of these entities may subscribe to different electronic medical record systems. These differ from national disease and drug registries. Consequently, it is difficult to standardize communication and data storage across a healthcare network due to differences between record systems.

2.4 Electronic Record Services

To manage health records organization or individuals use electronic health records (EHR) or electronic medical record (EMR) services. EHRs are longitudinal health record systems for patient information. These are available to specialty health clinics, health registries, hospitals and cooperating hospitals. EMRs are smaller scale patient record systems. These are typically used by smaller clinics, private physicians and physician groups. Popular solutions include Telus Health provided by Telus, Abelmed and OSCAR an open-sourced solution. Personal health records (PHR) are health recording systems used by patients to manage their own health information. These have seen a rise in popularity. Google and Microsoft have launched their own PHR platforms with Google Health and Microsoft HealthVault.

2.5 Data Integration

Consolidating data from multiple sources often poses challenges. Datasets are organized differently and carry diverse requirements. Consolidated data may conflict and even summary information may require added identification. After several years of research on public health, [20] addresses a few of these issues and provides three new protocols for privacy preserving data integration primarily for public surveillance on health information. These include an anonymized aggregation protocol and multi-party secure computational protocol. Hu also employs a semi-trusted party in his scheme. This reduces the need for a fully trusted third party.

Chapter 3

Security Primitives

Developing a new protocol presents unique challenges. Reliable protocols need to be founded on sound cryptographic principles based on proven concepts. To that end, this section explains the underlying cryptographic techniques used as the foundation of our framework.

3.1 Bilinear Pairing and Intractable Problems

Definition 1: Bilinear Pairing [34] Let g and h be two generators of two multiplicative cyclic groups $\mathbb{G}_1$ and $\mathbb{G}_2$ with the same prime order q. A mapping $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ is called an admissible bilinear pairing if it satisfies the following properties:

1. Bilinear: For all $V, Q \in \mathbb{G}_1$ and $a, b \in Z_q^*$, where $\hat{e}(V^a, Q^b) = \hat{e}(V, Q)^{ab}$.
2. Symmetric: $\hat{e}(V, Q) = \hat{e}(Q, V)$.
3. Non-degenerate: $\hat{e}(V, Q) \neq 1_G, where V, Q \neq 1_G$.
4. Computable: $\hat{e}$ is efficiently computable.

Definition 2 Discrete Logarithm (DL) Assumption [34] Let g be a generator of a multiplicative cyclic group $\mathbb{G}$ with the prime order q. On input $X \in G$, there is no probabilistic polynomial time algorithm that outputs a value $x \in Z_q^*$ such that $g^x = X$ with non-negligible probability.

Definition 3: Computational Diffie-Hellman (CDH) Assumption [34] Let g be a generator of a multiplicative cyclic group $\mathbb{G}$ with the prime order q. On input g^x, $g^y \in G$, there is no probabilistic polynomial time algorithm that outputs a $g^{xy} \in \mathbb{G}$ with non-negligible probability.

Decisional Bilinear Diffie-Hellman (DBDH) Assumption [10] Let $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$ be a bilinear map, where $\mathbb{G}_1$ and $\mathbb{G}_2$ are two multiplicative cyclic groups with the same prime order q. Let g and h be the generators of $\mathbb{G}_1$ and $\mathbb{G}_2$, respectively. On input (g, g^a, g^b, Q) for $a, b \in Z_q^*$ and $Q \in \mathbb{G}_2$, there is no probabilistic polynomial time algorithm to decide whether $Q = \hat{e}(g, g)^{a/b}$ with non-negligible probability.

3.2 Signcryption

Signcryption schemes have seen extensive acceptance and have been adopted into many applications. The combination of signing and encryption into a singular scheme has created greater efficiency in how to conceal and verify information and identities. Al-Rayami and Paterson [3] introduced the notion of certificateless private key cryptography (CLPKC) in 2003. This formed the basis of modern signcryption schemes, which simultaneously achieve signing and encryption. Signcryption private keys consist of a partial key and secret value generated by a key generation center and a user respectively. Certificateless signcryption schemes have seen

much growth. Ongoing research focuses on enhancing efficiency while maintaining security goals.

Barreto et al. in [7] was the first to propose a CLSC scheme without bilinear pairing. Shi et al. in [46] have demonstrated an efficient and secure certificateless signcryption scheme without bilinear pairing. Several others have since proposed other versions of CLSC schemes without bilinear pairing [22], [26]. These schemes however, have been either less secure or inefficient.

Below we relate a generic signcryption algorithm [3], [22] & [26]:

Setup (k) Setup is performed by the KGC to generate the master secret key msk and public parameters $params$.

1. Generate large primes p and q. The length of q is k and $p = 2q + 1$
2. Select a generator g with the order q.
3. Select a random point. $x \in Z_q^*$ and $y = g^x \bmod p$.
4. Choose hash functions (n is the size of the message to be signcrypted):

$$H_1 : \{0,1\}^* \times Z_p^* \rightarrow Z_q^*$$

$$H_2 : \{0,1\}^* \times Z_p^* \times Z_p^* \times Z_p^* \rightarrow Z_q^*$$

$$H_3 : Z_p^* \times Z_p^* \rightarrow \{0,1\}^n$$

$$H_4 : \{0,1\}^* \times Z_p^* \times Z_p^* \times \{0,1\}^n \times Z_p^* \rightarrow Z_q^*$$

5. Return:

   Parameters $params = (p, q, g, y, H_1, H_2, H_3)$

   Master Secret Key $msk = (p, q, g, x, H_1, H_2, H_3)$

For simplicity we assume that $params$ is publicly available and accessible to all functions.

Set Secret Value Given the publicly available parameters, a user u uses this algorithm to generate his/ her public and private keys.

1. Generate secret key sk_u. sk_u is randomly selected where $sk_u \in Z_q^*$
2. Calculate the public key pk_u where $pk_u = g^{sk_u}$.
3. Return the secret and public key pair (sk_u, pk_u).

Partial Private Key Extract (ID_u, msk) This operation is performed by the KGC. A user with identity ID_u and the KGC generate the partial public and private keys $P_u and D_u$.

1. Select a random number s_u where $s_u \in Z_q^*$.
2. Calculate P_u: $P_u = g^{s_u}$.
3. Calculate D_u: $D_u = (s_u + xH_1(ID_u, P_u))$ mod q.
4. Return partial private key D_u and partial public key P_u - (D_u, P_u).

Set Private Key (D_u, sk_u) This algorithm is run by the user to generate the full private/ secret key.

1. Set the full secret key SK_u = (sk_u, D_u)
2. Return SK_u.

Set Public Key (P_u, pk_u) This algorithm is used by the user to set the full public key PK_u.

1. Set $PK_u = (pk_u, P_u)$
2. Return PK_U

Signcrypt $(M, ID_A, ID_B, SK_A, PK_A, PK_B)$ This algorithm is run by a sender A with an identity ID_A using their full secret key SK_A. Receiver B has an identity ID_B and a full public key PK_u. Sender A signcrypts a plaintext message "M" to send to receiver B. To create a ciphertext sender A:

1. Selects a random number r_A, where $r_A \in Z_q^*$.
2. Calculate $c_1 = g^{r_A} \bmod p$.
3. Compute $k_A = H_2(ID_A, pk_A, P_A, y)$
4. Compute $k_B = H_2(ID_B, pk_B, P_B, y)$
5. Compute $h_B = H_1(ID_B, P_B)$
6. Compute $\xi = (pk_B^{k_B} P_B y^{h_B})_A^r \bmod p$
7. Compute $c_2 = H_3(\xi) \oplus M$
8. Compute $h = H_4(ID_A, pk_A, P_A, c_1, c_2, \xi, M)$
9. Compute $c_3 = [(k_A pk_A + D_A)/(r_A + h)] \bmod q$
10. Return $C = (c_1, c_2, c_3)$

Unsigncrypt $(C, ID_A, ID_B, PK_A, SK_B)$ This function is performed by the receiver B with an identity ID_B. B unsigncrypts the ciphertext C from sender with ID_A with public key PK_A using its secret key SK_B. Using the following steps, the receiver will decrypt a ciphertext C and receive a resulting decryption δ which can be either corresponding plaintext or a rejection message.

1. Compute $h_A = H1(ID_A, P_A)$
2. Compute $h_B = H_1(ID_B, P_B)$
3. Compute $k_A = H_2(ID_A, pk_A, P_A, y)$
4. Compute $k_B = H_2(ID_B, pk_B, P_B, y)$
5. Compute $\xi = (c_1)^{k_B sk_B + SK_B} \bmod p$
6. Compute $M = c_2 \oplus H_3(\xi)$
7. Compute $h = H_4(ID_A, pk_A, P_A, c_1, c_2, \xi, M)$
8. If $(c_1 g^h)^{c_3} == pk_A^{k_A} PK_A y^{h_A} \bmod p$

   Return M

   Otherwise/ else Return "Rejection Message."

3.3 Conditional proxy re-encryption

Proxy re-encryption (PRE) is a primitive used to transform ciphertext into a subsequent ciphertext. This is usually performed by a semi-trusted entity without ever accessing the original plaintext [48]. This primitive is suitable for a number of applications, especially where file management and encryption are concerned. Conditional proxy re-encryption (C-PRE) differs from the norm in that ciphertexts are

encrypted with an additional condition. Functions for system setup and key generation follow the norm. Algorithms for re-encryption key generation, first level encryption, re-encryption and decryption, however, require an additional condition c [44]. The added condition c must be equivalent in all functions. Otherwise, decrypting ciphertext will prove unsuccessful.

Generally, C-PRE schemes are comprised of the following algorithms [50], [44] [1]:

Global Setup (1^λ): This is a preliminary algorithm that utilizes the security parameter λ and creates a set of global parameters $params$. $params$ includes variables to describe an elliptic curve and variables pertinent to key generation and creating hash values. For simplicity we assume that all algorithms have access to the contents of $params$.

Results: $params$

Key Generation (msk, ID_i): This generates a private key for some user with a public identity of ID_i. msk represents the master secret key. A secret key sk_{ID_i} is generated for the corresponding user, where $ID_i \in \{0,1\}^*$.

Results: Secret/ Private key for ID rk_{ID_i}

Re-encryption Key Generation (sk_{ID_1}, w, ID_1, ID_2): This algorithm generates re-encryption keys. This is usually run by the user with ID_1 for transmission to an entity ID_2 with its secret key SK_{ID_1} and a condition w.

Results: Partial re-encryption key: $rk_{ID_1 \xrightarrow{w} ID_2}$.

Encryption (ID, m, w): This algorithm is used to perform the initial encryption. It accepts an identity ID. m represents a plaintext message where $m \in M$ (M denotes message space) and w is a condition. A ciphertext CT is outputted upon

[1] In this conditional proxy re-encryption scheme we use identities over the conventional public/ private key methods.

completion.

Results: Ciphertext CT.

Re-encryption ($CT_{ID_1}, rk_{ID_1 \xrightarrow{w} ID_2}$) Re-encryption is performed by a trusted third party or proxy. The algorithm uses a ciphertext CT_{ID_1} associated with the user identity ID_1 and a condition w. $rk_{ID_1 \xrightarrow{w} ID_2}$ is the re-encryption between ID_1 and ID_2 using the same condition w. This algorithm creates a new ciphertext CT_{ID_2} under ID_2.

Results: Re-encrypted Ciphertext CT_{ID_2}

Decryption (CT, sk_{ID}) This algorithm uses the secret key sk_{ID} of an entity ID and a ciphertext CT and returns a plaintext message m or an error symbol $\perp$.

Results: Corresponding plaintext - m or error message - $\perp$.

Chapter 4

Consent Based Access Control

This chapter proposes a consent-based access control (CBAC) mechanism for health record systems. After obtaining consent from patients, a healthcare organization can gain access to their data, which is encrypted by a healthcare provider. This is achieved by a cryptographic primitive: conditional proxy re-encryption. By doing so, a patient's medical data is protected against access from unauthorized parties. This includes the data centers where information is housed. Additionally, the proposed scheme achieves collusion resistance. Furthermore, mutual authentication and contextual privacy are attained. Performance evaluation demonstrates that the proposed CBAC scheme can achieve security and privacy preservation with high computational efficiency.

4.1 Introduction

Currently, healthcare providers globally are migrating towards electronic medical records. This provides healthcare organizations a convenient and reliable way to share and access health information. As we move healthcare to the digital world,

privacy preservation in data has become imperative. In the past, a number of security and privacy preserving mechanisms have been proposed for health record systems [25], [53], [24], [27] and [1]. However, due to the nature of the healthcare industry, securing patients' medical data is challenging.

These challenges plague technological implementation and adoption. One such challenge is the problem of data ownership and control. Healthcare information systems allow for a number of permitted data custodians. However, due to the sensitivity of records, the owner retains a right to grant and revoke access to requesting parties. These privileges extend beyond the release of information for retroactive revocation. In other words, it is critical for patients to control access to their data. Access of data must be associated with proper consent. Furthermore, a patient has the right to retroactively withdraw or revoke consent at anytime. From this, we conclude that consent is fundamental in healthcare information sharing. Unfortunately, developing a new consent-based access control scheme to manage access for sensitive health information poses unique challenges. Healthcare providers and insurance companies must be able to request patient files. Prior to an entity receiving any records, patients must permit any requests. This may be in the form of allowing partial access to a subset of requested records or the full record. Receivers must be able to verify the received information. Finally, patients should be able to retroactively deny access to records if they deem it necessary.

Based on the above observations, we propose a novel consent-based access control scheme for health record systems. A user's health information is encrypted by data providers before transmission to a data center for storage. A data requester negotiates with a user to obtain consent tokens for accessing health data. Meanwhile, the user sends a consent notification with a re-encryption key to the data center. The data center re-encrypts the requested encrypted data with the re-encryption

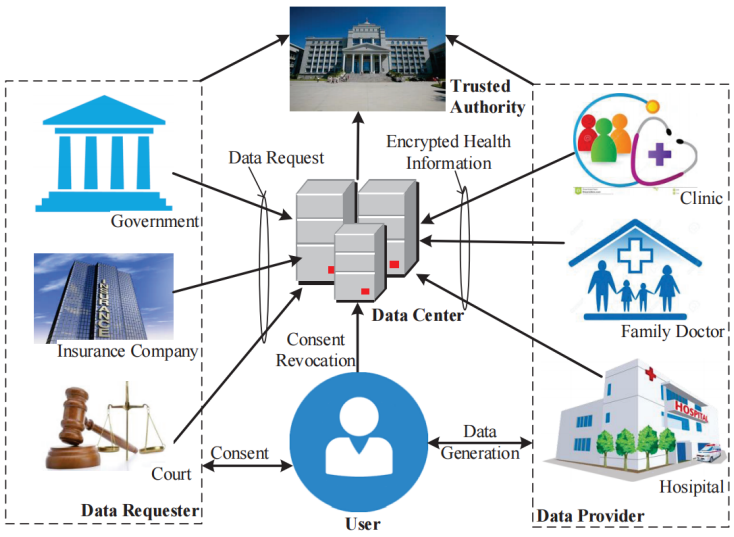

Figure 4.1: System architecture of the electronic health information record system

key without accessing the plaintext and sends the ciphertext to the data requester. The data requester is able to decrypt the ciphertext with his private key and the consent token from the user. In this way, a user's health information is exchanged among the data provider, data center and data requester without leaking a user's data.

4.2 CBAC System Model

In this section, we present the design goals and framework for electronic health record systems.

4.2.1 CBAC System Architecture

The proposed electronic health information record system is composed of a trusted authority (TA), data provider (d), data center (c), data requester (s) and users (u), as shown in Figure 4.1. Their functions are described as follows.

4.2.1.1 Trusted Authority

These are trusted entities which manage the system. All other entities, including the data provider, data center, data requester and users, must register to join the system. The TA generates public/private key pairs for those that register.

4.2.1.2 Data Provider

Medical institutions, such as clinics, hospitals or family doctors, act as data providers in the system. When a user arrives at the medical institution for health care or treatment, the institution will record his/her health status and diagnosis; which are components of the user's health information. Next, the institution encrypts the user's health information and sends it to a data center. We assume that the data provider is trustworthy.

4.2.1.3 Data Center

The data center is a semi-trusted authority, which records health information provided by data providers. Notably, all data providers send health information to the data center as ciphertexts. Moreover, the data center is unable to decrypt ciphertexts without consent. We assume that the data center is honest but curious about stored information. Therefore, it is only semi-trusted.

4.2.1.4 Data Requesters

Insurance companies, governments, courts, or other entities may need to access a user's health information under certain circumstances, e.g., health insurance companies require customer records for claims. Data requesters may take the form of other data providers such as other clinics which require patient information. The

requester should first obtain consent from users. It then sends a data request with an associated consent token to the data center. The data provider may query the data center to access the user's health information.

4.2.1.5 Users

When a user arrives at a medical institution, i.e., the clinic or the hospital, for health care or treatment, their health information will be generated and recorded. In some situations, the user may form an agreement with a data requester allowing it to access some of their health information by sending a consent token to the data requester.

4.2.2 CBAC Design Goals

Based on the above system model, the design goals of the scheme are as follows:

4.2.2.1 Data confidentiality and integrity

Data cannot be accessed by other entities without user consent. Additionally, the data should be protected from modification.

4.2.2.2 Contextual privacy

Data requesters are allowed to request a user's health information from the data center. Data Providers and Data Centers remain oblivious as to the contents of the data transaction. In simpler terms, the data provider is unaware of the request while the data center is unaware of whose information was requested. Furthermore, anyone eavesdropping on the data transmission cannot derive which information was received or requested by the data requester.

4.2.2.3 Consent revocation

Users have the privilege to terminate consent at any time prior to the expiration of that consent. This is done through a consent revocation token.

4.2.2.4 Mutual authentication

The data provider and the data center should be able to authenticate each other. Additionally, the user can only provide access to their own data. Consequently, the data center should be able to authenticate the consent provider. This is achieved through signcryption.

4.2.2.5 Collusion resistance

After receiving the re-encryption key intended for a data requester from a user, the data center can re-encrypt the user's data required by the requester. However, the data requester cannot decrypt a user's other data which exceed the consent given by colluding with the data center.

4.3 CBAC Protocol Description

In this section, we provide an overview and describe the proposed CBAC protocol and consent-based access control mechanism in detail.

4.3.1 An overview of the proposed CBAC protocol

When a user with identity ID_u arrives at a healthcare provider (data provider), for treatment, they will consent for records to be created. Interactions between them will generate an original health record for the user. The data provider will store

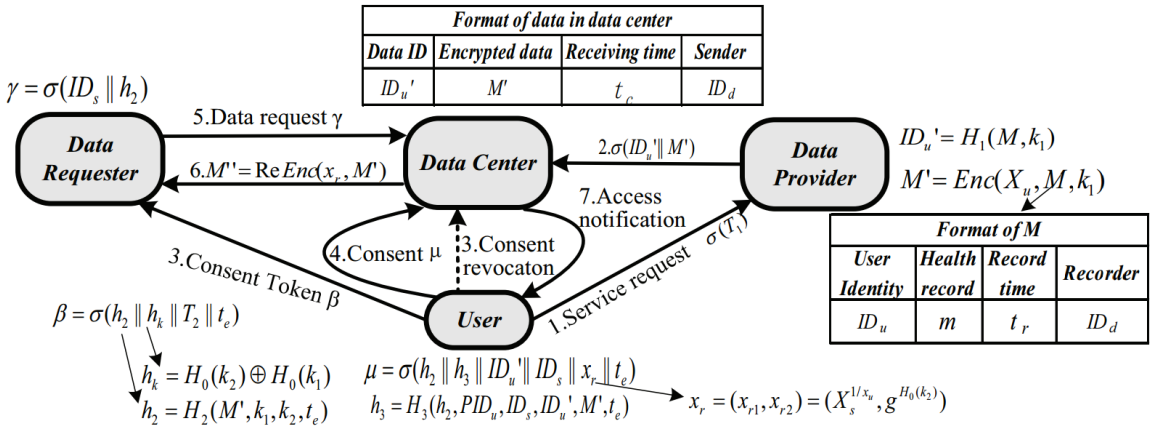

Figure 4.2: Proposed Protocol

the user's heath information M in the format as shown in Figure 4.2. Then, the data provider sends the data $(ID'_u||M')$ to the data center, where ID'_u is the record ID and M' is the an encryption of M. Upon receiving the record from the data provider, the data center stores the message in the format as shown in Figure 4.2.

If an entity wishes to gain access to a patient's files they will query the user for consent. Once an agreement is achieved between the user and the requester, the user will send a consent token to the data requester for accessing parts of their health record from the data center. Meanwhile, the user should send a consent μ to the data center to inform the center about the authorization or permission. Afterwards, the requester will request for the patient's medical data from the center. The data center checks the validity of the data request γ by using the consent μ from the user. If the verification outputs valid, the data center re-encrypts the data M' with the re-encryption key provided by the user and sends the ciphertext M'' to the data requester. During this phase, the ciphertext M', under the public key of the user, is transformed into the ciphertext M'' under the public key of the data requester. As a result, the requester is able to access the original data M by decrypting M'' with their private key.

If the re-encryption key was to be generated via traditional algorithms, it would be vulnerable to collusion attacks. Traditional re-encryption keys are only related to the user's private key and the data requester's public key. The user's other health information may be accessed if the data requester re-encrypts it and the data requester decrypts it. As a result, we propose a shared secret k_1 between the user and the data provider for each record. Thus, the health information M is encrypted by the user's public key as well as the the secret k, i.e., $M' = Enc(X_u, M, k_1)$. Meanwhile, the user also introduces a shared secret k_2 with data requesters for each individual consent token. The secret k_2 is included in the re-encryption key such that only the intended data requester is able to access the specific data.

4.3.2 Consent Based Access Control

Unlike traditional permission-based access control mechanisms, where permissions are based on user identity, security label or role, the consent-based access control mechanism guarantees the permission by sending a consent token to the data requester. The consent token is neither an identity nor a security label. It is the product of an agreement between the user (data owner) and the data requester, so it is also not a role as the data requester cannot access the user's other non-consented data.

For notification of consent, the user also sends a consent token to the data center. The data requester is required to send a token to data center in order to access data. The data center is able to check the validity of the token by comparing it with the consent received from the user. Additionally, the user has the privilege to terminate the consent whenever he/ she chooses. Therefore, data centers are able to retroactively revoke access to user information.

The consent-based access control mechanism guarantees data security through

three conditions:

- The data requester is compelled to negotiate with the user in order to receive a consent token. Consequently, the capacity to access records is controlled by the user, which protects their data. Additionally, consent tokens cannot be forged as the data center can verify tokens.
- The consent token is only valid for access of specific data which is granted by the user. Thus, data is still protected when requesters have access tokens.
- Whenever patient medical data is accessed, the patient will be notified to ensure there is no abuse or use without their consent.

4.3.3 Protocol Description

The proposed CBAC protocol is composed of the following steps.

4.3.3.1 System Initialization

Given the security parameter λ, the TA generates the system parameters $params = (q, g, \mathbb{G}_1, \mathbb{G}_2, \hat{e}, H_0, H_1, H_2, H_3, \sigma(\bullet))$. Where $\mathbb{G}_1$ and $\mathbb{G}_2$ are finite cyclic groups with the same prime order q, and g and h are generators of $\mathbb{G}_1$ and $\mathbb{G}_2$, respectively. The bilinear pairing $\hat{e}$ is a mapping: $\mathbb{G}_1 \times \mathbb{G}_1 \rightarrow \mathbb{G}_2$. The secure hash functions are as follows:

$$H_0 : \mathbb{Z}_q^* \rightarrow \mathbb{Z}_q^*$$

$$H_1 : \{0,1\}^* \times \mathbb{Z}_q^* \rightarrow \{0,1\}^*$$

$$H_2 : \{0,1\}^* \times \mathbb{Z}_q^* \times \mathbb{Z}_q^* \times \{0,1\}^* \rightarrow \{0,1\}^*$$

$$H_3 : \{0,1\}^* \rightarrow \{0,1\}^*$$

$$H_4 : \mathbb{G}_1 \rightarrow \mathbb{G}_2$$

The item $\sigma(\bullet)$ is a general signcryption algorithm implemented by the data senders, which completes signing and encryption in one logical step. Signcryption can achieve non-repudiation, data confidentiality, and data integrity [22], [46]. The TA generates a public/private key pair (X_i, x_i) for the entity $i \in \{u, d, s, c\}$, i.e., the user u, data provider d, data requester s, and data center c [1].

4.3.3.2 Data Generation

Data generation is completed through the interaction between the user and data provider. The user with identity ID_u randomly chooses $u_1 \in \mathbb{Z}_q^*$ and computes $T_1 = g^{u_1}$. T_1 is introduced for the generation of the secret key shared between the data provider and the user for the health record m. Thus, each record m has a corresponding random number, which is used to generate a condition c (or the secret value) shared between the patient and data requester. It can be used by the patient to achieve fine-grained access control. The user goes to the clinic (or other data provider) for treatment with the value $\sigma(T_1)$, where $\sigma(T_1)$ is the signcryption of the user on T_1. In subsequent communications, messages are all signcrypted as $\sigma(\bullet)$ of the sender, where $\bullet$ is the message in transmission. After the diagnoses and lab tests, the data provider will record the user's health information M in the format as shown in Figure 4.2, including user identity ID_u, health record m, record time t_r and recorder ID_d.

[1] In order to protect the user's privacy, the TA also generates a pseudo identity PID_u for the user with identity ID_u and publishes it with the user's public key. The pseudo identity is usually generated from the real identity, i.e., $PID_u = H_3(ID_u)$. Then, the user will use his pseudo identity PID_u for interaction with the data center. However, he should use his real identity ID_u for the interaction with data providers or data requesters to achieve identity authentication. As the data provider and the data requester are public institutes, it is unnecessary to protect their identity privacy.

4.3.3.3 Data Storage

This task is performed through the interactions between the data provider and data center. The data provider computes the shared key $k_1 = T_1^{x_d}$ with his private key x_d. The data provider computes a record ID, where $ID'_u = H_1(M, k_1)$, for the health information M. Furthermore, it encrypts the data M with the user's public key X_u and the shared secret key k_1.

The Data Provider encrypts data as follows

- Randomly chooses $r \in \mathbb{Z}_q^*$ and computes $C_a = X_u^r$
- Next it computes $h_0 = H_0(k_1)$ and $C_b = \hat{e}(g, g)^r * H_4(X_u^{h_0}) * M$

Thus, health information M is encrypted as $M' = (C_a, C_b)$. The data provider sends $\sigma(ID'_u||M')$ to the data center for storage. Upon receiving data from the data provider, the data center stores the data in the format shown in Figure 4.2, including data identity ID'_u, encrypted message M', receiving time t_c, and data sender ID_d.

4.3.3.4 Data Consent Sharing

This step is performed by the user. When a data sharing agreement is made between the user and the data requester, the user sends a consent token β to the data requester for retrieving records from the data center. The user sets the expiry time of the consent token as t_e. The consent token β is generated by the user as follows. The user:

- Randomly chooses $u_2 \in \mathbb{Z}_q^*$ and computes $T_2 = g^{u_2}$, $k_2 = X_s^{u_2}$
- Computes $h_2 = H_2(M', k_1, k_2, t_e), h_k = H_0(k_1) \oplus H_0(k_2)$

Table 4.1: **CONSENT LIST IN DATA CENTERS**

Record ID	PID **of the user**	**Expected Receiver**	**Consent**	**Encrypted Message**	**Expiry Time**
ID'_u	PID_u	ID_s	$h_2 \| \| h_3 \| \| x_r$	M'	t_e

Table 4.2: **CONSENT REVOCATION LIST IN DATA CENTERS**

Record ID	**Revocation Token**	PID **of the user**	**Expected Receiver**	**Expiry Time**
ID'_u	h_2	PID_u	ID_s	t_e

A token is constructed in the format of $\beta = \sigma(h_2||h_k||T_2||t_e)$ and sent to the data requester. Simultaneously, the user sends a notification consent token μ to the data center. The consent token μ is generated by the user. The user:

- Computes $h_3 = H_3(h_2, PID_u, ID_s, ID'_u, M', t_e)$, PID_u is the pseudo identity of the user in the data center for privacy preservation, ID_s is the identity of the data requester [2]
- Computes the re-encryption key $x_r = (x_{r_1}, x_{r_2}) = (X_s^{1/x_u}, g^{H_0(k_2)})$

The consent $\mu = \sigma(h_2||h_3||ID'_u||ID_s||x_r||t_e)$ is then sent to the data center. The data center searches its database for M' with record identity ID'_u and checks $h_3 = H_3(h_2, PID_u, ID_s, ID'_u, M', t_e)$.

If the equation holds, the data center stores the consent as shown in Table 4.1. If no encrypted message M' with the record identity ID'_u is found in the data center, the message is ignored. Notably, a user can terminate a consent that he/she authorizes. They send a consent revocation notification $\theta = \sigma(ID'_u||h_2||PID_u)$ to the data center. Additionally, the data center stores the consent revocation into the

[2]The items (ID'_u, M') are calculated by the user since he/she knows the shared secret key k_1 and the health record M.

consent revocation list (CRL), as shown in Table 4.2. Moreover, both the CRL and the consent list will be updated periodically to remove the expired consents.

4.3.3.5 Data Transmission

After receiving the consent token $\beta = \sigma(h_2||h_k||t_e)$ from the user, the data requester formulates a data request $\gamma = \sigma(ID_s||h_2||t_e)$ and sends it to the data center to fulfill data requirements. When a data request $gamma = \sigma(ID_s||h_2||t_e)$ arrives, the data center first checks Table 4.2. If h_2 is in the revocation list, the message is ignored. The data center checks the consent list in Table 4.1 to decide whether requests have expired. It checks the item $consent$ in Table 4.1 to decide if a consent token with the same value as h_2 is in β. If no consent token is found, the request is discarded. Otherwise, the data center checks whether the expected receiver ID_s in the table is the same as the identity of the data requester. If all the above verifications are valid, the data center re-encrypts the message $M' = (C_a, C_b)$ with the re-encryption key $x_r = (x_{r_1}, x_{r_2})$ and the consent as follows:

- Randomly chooses $w \in \mathbb{Z}_q^*$
- Computes $C_1 = C_a^w, C_2 = x_{r_1}^{1/w}$
- Computes $C_3 = C_b * H_4(x_{r_2}^{x_c})$

The data center sends the re-encrypted message $M'' = (C_1, C_2, C_3)$ to the data requester. Additionally, the data center sends an access notification $\nu = \sigma(ID'_u||h_2||h_3||ID_s)$ to the user to inform them that data was transmitted.

4.3.3.6 Data Reception

The data requester decrypts the message $M'' = (C_1, C_2, C_3)$ as follows:

- Compute $k_2 = T_2^{x_s}$ and $C_b = C_3 / H_4(X_c^{H_0(k_2)})$
- Compute $h_0 = h_k \oplus H_0(k_2)$
- Compute $M = C_b / (\hat{e}(C_1, C_2)^{1/x_s} * H_4(X_u^{h_0}))$

If any variable during the re-encryption or decryption is invalid then the decryption will fail.

4.4 Security Analysis

In this section, we analyze how the proposed CBAC protocol achieves the design goals as described in Section 4.2.2.

4.4.1 The proposed protocol can achieve data confidentiality and integrity

Before sending the data to the data center, the data provider encrypts the health information with the user's public key and their shared key k_1. Only the user with the correct private key and secret key can decrypt it. Moreover, after obtaining the consent token from the user, the data requester is able to access the data by decrypting the re-encrypted ciphertext, which is generated by the data center with a re-encryption key. The re-encryption key can only be generated by the user. Thus, data can only be accessed with the user's consent. In this way, data confidentiality is ensured. Additionally, data integrity is achieved by the signcryption of the data provider on the message $ID'_u||M'$.

4.4.2 The proposed protocol can achieve contextual privacy

For each health record M, there is an unique record $ID'_u = H_1(M, k_1)$ which is sent to the data center to label the health record. The encrypted health information M' is provided by the data provider. Therefore, the data center cannot link data to its original user. The user sends its consent to the data center with a pseudo identity. Consequently, the data provider cannot get any information concerning transactions between the user and other data requesters. Furthermore, even though the data provider knows the user identity of certain records, they cannot derive what information is requested from the data center since all transmission are signcrypted.

4.4.3 The proposed protocol can achieve consent revocation

Consent can be revoked in two ways. There is an expiration time for each consent. Thus, the consent automatically becomes invalid when expiration times are met. Users are also able to terminate consent by sending a consent revocation notification to the data center before expiry dates.

4.4.4 The proposed protocol can achieve mutual authentication

Mutual authentication is achieved through the signcryption of messages. Specifically, only intended receivers are able to decrypt messages, which also authenticates receivers. Moreover, the signature algorithm in signcryption can authenticate senders. Furthermore, the shared secret key k_1 between the user and the data provider, and the shared secret key k_2 between the user and the data requester help the user to authenticate data providers and requesters, respectively. This is because only the data provider and the data requester can compute the shared secret k_1 and

k_2 with the value T_1 and T_2 under the computational Diffie-Hellman assumption.

4.4.5 The proposed protocol can achieve collusion resistance

The re-encryption key is not only generated from the user's private key but also from the shared secret k_2, which is unique and different for each consent. As a result, even if the data requester and data center collude, they are unable to access the non-consented data because they have no information on $H_0(k_1)$.

4.5 Performance Evaluation

In this section, we analyze the computational and storage overhead of our scheme. We then compare it with other data exchanging schemes in cloud environments, which share similar security objectives with our system model.

4.5.1 Benchmarks

The essential task of the proposed protocol aims to share data in a cloud environment. We compare it to other data sharing schemes in the cloud [48], which have similar security objectives. Both the proposed CBAC and [48] use conditional proxy re-encryption to achieve the security requirements, we also use the conditional proxy re-encryption algorithms [44] for comparisons. For the unification of our benchmark, we chose the following two settings:

1. We assume the conditions in the conditional proxy re-encryption of the three schemes are the input of algorithms. Thus, we do not consider the computational overhead for condition generation.

2. We mainly compare the computational overhead at the user ("client U_i" in [48]), data center ("cloud storage" in [48]), and data requester (another client U_j in [48]).

4.5.2 Computational Overhead

As the operations of bilinear pairing and exponentiation dominate the computational overhead of the algorithms, we consider the time consumption of these operations. We denote t_p, t_1, t_2 as the computational cost of bilinear pairing, exponentiation in $\mathbb{G}_1$, and exponentiation in $\mathbb{G}_2$, respectively. In the proposed CBAC protocol, the computational cost of the user is caused by the data encryption and re-encryption key generation. It takes $t_p + 2t_1 + t_2$ to encrypt the the data and $2t_1$ to generate the re-encryption key, thus the computational overhead at the data provider is $t_p + 4t_1 + t_2$. In order to re-encrypts the ciphertext M' , the data center performs three exponentiations in $\mathbb{G}_1$ to output the ciphertext M''. At the data requester, decryption of the ciphertext M'' costs $t_p + 2t_1 + t_2$ overhead.

In [48], the client U_i performs the role of data provider in our system. It takes them $2t_p + 3t_e + t_m$ to encrypt the data and $3t_e + 2t_m$ to generate the re-encryption key $RK_{(i \to j)}$. The cloud storage, which works as a data center in our scheme, carries out $2t_p + 2t_e + 2t_m$ to complete the data encryption. Additionally, the cloud has to execute one exponentiation to generate the partial re-encryption key. Consequently, the computational overhead of the cloud is $2t_p + 3t_e + 2t_m$. In order to decrypt the data, the data receiver U_j consumes $t_p + 2t_e + 3t_m$ overhead.

We applied the conditional proxy re-encryption technique [44] in our system and compared it with our proposed scheme. In [44], it takes $t_p + 4t_e + t_m$ to encrypt the data and $t_p + 4t_e + 2t_m$ to generate the re-encryption key. As a result, if the conditional proxy re-encryption of [44] is used in our system, it will cost the data

Table 4.3: **COMPARISON OF COMPUTATIONAL OVERHEAD**

Scheme	**Data Provider**	**Data Center**	**Data Requester**
[44]	$2t_p + 6t_1 + t_2$	$2t_p$	$2t_p + 2t_1 + t_2$
[48]	$2t_p + 4t_1 + 2t_2$	$2t_p + t_1 + t_2$	$t_p + 2t_1 + t_2$
CBAC	$t_p + 4t_1 + t_2$	$3t_1$	$t_p + 2t_1 + t_2$

Table 4.4: **TIME CONSUMPTION OF OPERATIONS**

Operations	**Time**
Preprocessing Pairing	5.9
Exponentiation in $\mathbb{G}_1$	6.4
Exponentiation in $\mathbb{G}_2$	0.6

provider $2tp + 8te + 3tm$ overhead. Similarly, the data center will take $2t_p + 2t_m$ to re-encrypt the data and the data requester will take $2t_p + 3t_e + 2t_m$ to decrypt the ciphertext, as shown in Table 4.3.

To quantify the running time of the operations, we use the results in [45] as the benchmark for comparisons, as shown in Table 4.4. The processor is a 64-bit, 3.2Ghz Pentium 4. The running times are tp = 5.9ms, t1 = 6.4ms, t2 = 0.6ms, respectively. We evaluate the performance with the similar settings. Computational overheads are compared in Figure 4.3.

From Figure 4.3, we discern that our proposed scheme has the lowest computational overhead for clients and users. Consequently, our proposed CBAC is also suitable for users with resource constrained terminals, i.e., mobile phones. The scheme [44] has the highest computational efficiency at the data center. Thus [44] offers a better performance for the cloud than the other schemes but has a much higher cost for clients. Notably, the scheme [48] and our proposed scheme have

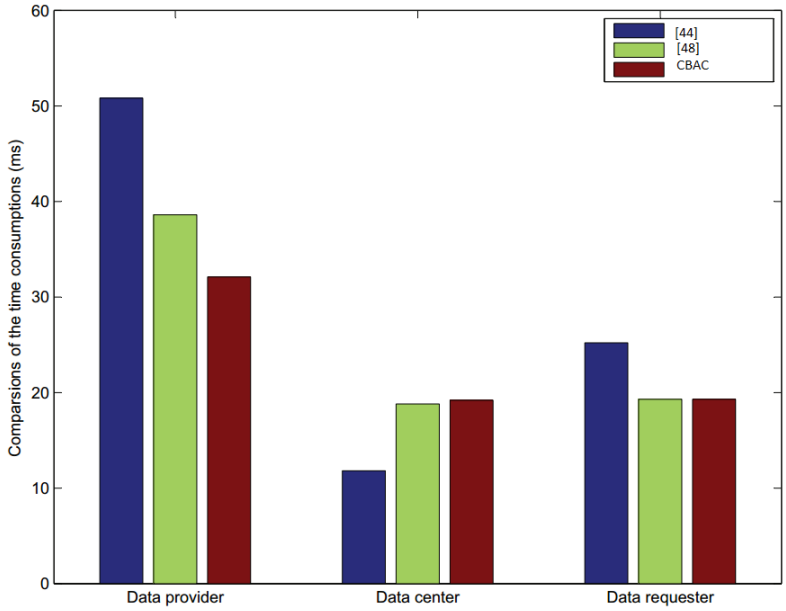

Figure 4.3: Comparison of Time Consumption

Table 4.5: **COMPARISON OF CIPHERTEXT SIZES**

Schemes	**Original Ciphertext Size**	**Re-encrypted Ciphertext Size**
[44]	$3\|G_1\| + \|G_2\| + \|m\|$	$5\|G_1\| + \|G_2\|$
[48]	$2\|G_1\| + 2\|G_2\|$	$2\|G_1\| + 2\|G_2\|$
Proposed Scheme	$\|G1_\| + \|G_2\|$	$2\|G_1\| + \|G_2\|$

almost the same time consumption for data requesters, while the proposed scheme outperforms the [48] for user (data providers / clients).

4.5.3 Ciphertext size

We denote $|G_1|, |G_2|, |m|$ the size of the elements in $\mathbb{G}_1, \mathbb{G}_2$ and the message. In the proposed CBAC, the original ciphertext M' is composed by C_a and C_b, the length of which is $|G_1|$ and $|G_2|$, respectively. Thus, the size of the original ciphertext is $|G_1|+|G_2|$. The re-encrypted ciphertext $M'' = (C_1, C_2, C_3)$, with size $2|G_1|+|G_2|$. The

Table 4.6: **STORAGE OVERHEAD OF THE PROPOSED SCHEME**

Entity	Components	Storage
Data Center	Health Information	$2\|S\| + \|H\| + \|G_1\| + \|G_2\|$
	Consent	$3\|S\| + \|H\| + 3\|G_1\| + \|G_2\| + 2\|h\|$
	Consent Revocation	$4\|S\| + \|H\|$
Data Provider	Health Information	$2\|S\| + \|H\| + \|m\|$

size of the original ciphertext and re-encrypted ciphertext of the three schemes are compared in Table 4.5. As shown in the table, the proposed scheme has the smallest size among the three schemes for both the original ciphertext and the re-encrypted ciphertext [3].

4.5.4 Storage Overhead

We analyze the storage overhead at the data center and data provider as they store all user data, which increases with the amount of users. Note that we only provide the storage needed for one user's health information. The total storage can be calculated through multiplication. We denote $|S|$ and $|H|$ as the size of the identity for all the entities and the size of the receiving time t_c or record time t_r, respectively. For the data provider, it stores the user's original health record M, which is composed by user identity ID_u, health record m, record time t_r and recorder ID ID_d. The storage overhead at the data provider is $2|S| + |H| + |m|$.

At the data center, the user's health information, consent list and the consent revocation list are stored. As displayed in the table of the Figure 4.2, the user's

[3] As all the messages are signcrypted by the sender, the communication overhead of the proposed scheme is determined by the signcryption algorithm. Thus, we do not analyze the communication overhead of the proposed scheme. Instead, we analyze the storage overhead of the proposed scheme.

health information in the data center includes data ID ID'_u, encrypted data M', receiving time tc, and sender ID ID_d. The storage overhead of the health information is $2|S| + |H| + |G_1| + |G_2|$. From Table 4.1, we can calculate the storage overhead of the consent as $3|S| + |H| + 3|G_1| + |G_2| + 2|h|, |h|$ denotes the length of output of the hash function H_2 and H_3. One consent revocation list takes up $4|S| + |H|$ space in the data center. The storage overhead is shown in Table 4.6.

4.6 Concluding Remarks

This chapter proposed a consent-based access control (CBAC) mechanism for secure and privacy-preserving health information exchanges. The proposed scheme achieves data security and privacy preservation by introducing consent-based access control, where consent can only be generated by the authorized user. A proxy re-encryption algorithm is proposed to achieve data sharing between the data center and the intended data requester without disclosing the content to the data center. Additionally, a condition is integrated into the re-encryption key to achieve collusion resistance. Moreover, mutual authentication and contextual privacy are realized by using the public key and a pseudo identity for users.

Chapter 5

Fairness Aware Privacy Preservation

This chapter proposes the fairness-aware and privacy-preserving (FAPP) protocol for online third party non-medical entities. More specifically this protocol is meant to address online health insurance systems. In the FAPP protocol, a user's health condition is encapsulated into a ciphertext with random numbers and sent to the health insurance company. The company will be unable to access the plaintext without prior user permission. However, the company will still be able to verify user integrity based on the ciphertext. In contrast to current health insurance schemes where insurance quotes are calculated by the company, the quote is calculated by the user based on the company's public policy in the proposed FAPP protocol. Additionally, the company is able to determine whether users have cheated when generating quotes. Furthermore, a concept of privacy-preserving quote is proposed. A user's health details cannot be derived from the quote alone. Security analysis demonstrates that the proposed FAPP protocol can achieve privacy-preservation and transparency.

5.1 Introduction

The advance of information and communication technology has revolutionized industries. The insurance industry, for example, offers online quotes to their customers. Such online insurance quoting systems provide convenience and savings for both companies and their consumers. Unfortunately, the handling of sensitive customer health data has introduced several privacy and security challenges [2], [47]. Privacy schemes and mechanisms have been developed in the past to secure e-Health systems and ensure that they conform to local health information legislation [27], [24]. However, health insurance companies require a myriad of sensitive information prior to issuing a quote and providing insurance. Additionally, how a quote is calculated remains a mystery to most insurance clients. Furthermore, an insurance company may deny coverage to a client if they deem them to be a future liability. This creates a challenge to secure online insurance application systems and creating fairness.

Unfortunately, completely obscuring a client's health details poses new challenges to insurance companies, such systems may be susceptible to abuse from clients. For instance, a client may be deceptive. They may attempt to deceive insurers concerning their insurance application and not disclose their previous smoking habits or chronic illness. This is one problem plaguing the insurance industry, also known as insurance fraud [35]. Consequently, designing a privacy scheme for this particular scenario presents several unique challenges:

1. The company should not be allowed to access the user's health information while it is able to generate a quote from the messages provided by the applicant.

2. Users are encouraged to provide true information to the company. Otherwise,

the dishonest behaviors will be discovered by the company.

3. Observers of data transmission between the applicants and company are unable to derive a user's health details through eavesdropping.

In order to address the above challenges, this chapter proposes a fairness-aware and privacy preserving (FAPP) insurance application system. Firstly, we propose the concept of privacy-preserving quotes. Users are encouraged to apply for the health insurance, it is difficult for the company or other entities to derive a user's health status from the quote alone. Based on the quote mechanism, an insurance application protocol is designed with fairness-aware and privacy preserving capabilities. This is done by encapsulating user health details into a ciphertext with random numbers. The company is unable to access the plaintext, but can discern if users are truthful when a claim is made.

5.2 FAPP System Model

This section introduces the system model and the concept of privacy-preserving quotes. Design goals are presented later in the section.

5.2.1 FAPP System Architecture

An online insurance system is composed of three entities, including the trusted authority (TA) such as CISRO (Canadian Insurance Service Regulatory Organization) [14], insurance company, and users, as shown in Figure 5.1. We assume that a user u with identity ID_u wants to apply for an insurance from a health insurance company c with identity ID_c. He/she should provide the company with some necessary information, including his/her name, birth date, and so on, which cannot be

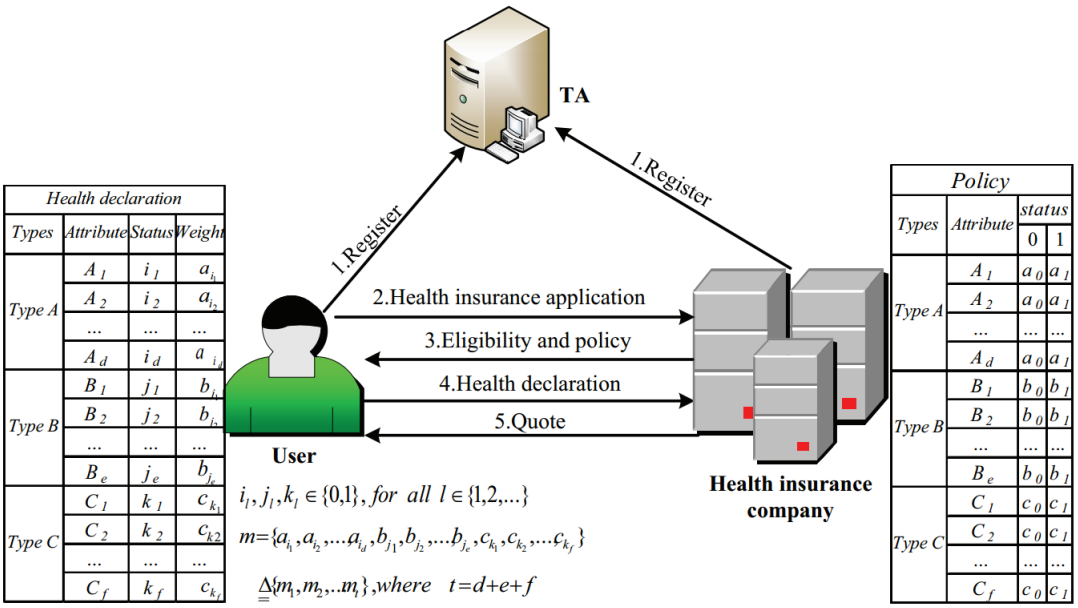

Figure 5.1: Processes of Health Insurance Applications

kept secret to the company. After checking the eligibility of the user, the company sends him/her a policy for fairness, which includes a detailed quote, as shown in Figure 5.1. The quote is determined by the user's attributes, each of which stands for a related health condition (a disease or a habit). In order to avoid deriving a user's health details or habits from the quote, it should be carefully designed so that one quote corresponds to multiple attribute combinations. In doing so, the k-anonymity technique [52], [36] is adopted. Consequently, all related attributes are categorized into three attributes[1], i.e., Type A, Type B, and Type C. Different types of attributes have different weights in affecting the quote, while for all the attributes in one type, they have the same weight, as shown in the *policy* Table of Figure 5.1.

In the Policy Table of Figure 5.1, the notation "0" denotes that an applicant does not have an attribute or is healthy without the disease or bad habit that the attribute

[1]The attributes can be classified into more than or less than three types. In the scheme, three types are chosen as an example.

Table 5.1: **INSURANCE RATES AND QUOTES**

Insurance Rate	$[W-, W_1)$	$[W_1, W_2)$	$[W_2, W+]$
Quotes	Q_1	Q_2	Q_3

indicates, and "1" denotes that an applicant has an attribute [2]. The value a_0 and a_1 denote the weights of attributes in Type A with status "0" and "1", respectively. Notably, the attributes of the same type with the same status should be given the same weight in the quote. For example, the value b_0, b_1, c_0, c_1 for attributes of Type B and Type C, respectively. It is assumed that there are in total t attributes for which the health insurance company are concerned about and that there are d, e, and f attributes in Type A, Type B, and Type C, respectively, where $t = d + e + f$. In order to describe privacy-preserving quote clearly, we will first define an insurance rate.

Definition 1: Insurance Rate - Let $i_1, i_2, ..., i_d \in 0, 1$ denote the user's attributes $A_1, A_2, ..., A_d$. Correspondingly, $j_1, j_2, ..., j_e \in 0, 1$ and $k_1, k_2, ..., k_f \in 0, 1$ denote the user's attributes $B_l, B_2, ..., B_e$ and $C_l, C_2, ..., C_f$, respectively. The user's insurance rate is the summation of the weights of all the attributes, expressed by:

$$M + u = a_{i_1} + a_{i_2} + ... + a_{i_d} + b_{j_1} + b_{j_2} + ... + b_{j_e} + c_{k_1} + c_{k_2} + ... + c_{k_f}$$

We assume that there are three quotes for the insurance rate. The quotes are fixed by setting lines for the insurance rate, as shown in Table 5.1. $W-$ and $W+$ denote the low bound and upper bound of the insurance rate. In the proposed scheme, they are computed by:

[2]For simplicity, we assume that an attribute only has two statuses, i.e., "0" and "1". In reality, some attributes could have more than two statuses, and the proposed scheme can easily be extended, for example, by introducing more symbols.

$$W- = d \times a_0 + e \times b_0 + f \times c_0$$

$$W+ = d \times a_1 + e \times b_1 + f \times c_1$$

The upper bound of Q_1 and Q_2 are W_1 and W_2 respectively, which are set by the company.

5.2.1.1 Anonymity analysis

Given the quote, the possibility of divulging the health status of the a depends on the corresponding insurance rate and the number of combinations having the same rate. Without loss of generality, we assume that there are α values in the insurance rate range of $[W-, W_1)$, denoted by $W-, W-+1, W-+2, ..., W-+(\alpha-1), where W-+\alpha = W_1$. Each insurance rate may be generated by several combinations of health statuses. We assume that $W-, W-+1, W-+2, ..., W-+(\alpha-1)$ are generated by $L_1, L_2, ..., L_\alpha$ combinations of health statuses, respectively, as shown in Figure 5.2. Then, given Q_1, the possibility of discerning the health status of the user is expressed as:

$$p_1 = \frac{1}{L_1 + L_2 ... + L_\alpha}$$

In order to protect the user's privacy as much as possible, the weights of the attributes should be skillfully setup so that there are multiple health status combinations for one insurance rate. This will reduce the possibility of the company to deriving a user's health status from the quote. For example, we may set $a_0 = b_1$ and $b_0 = c_1$. Obviously, the smaller p_1 is, the higher level of anonymity we have.

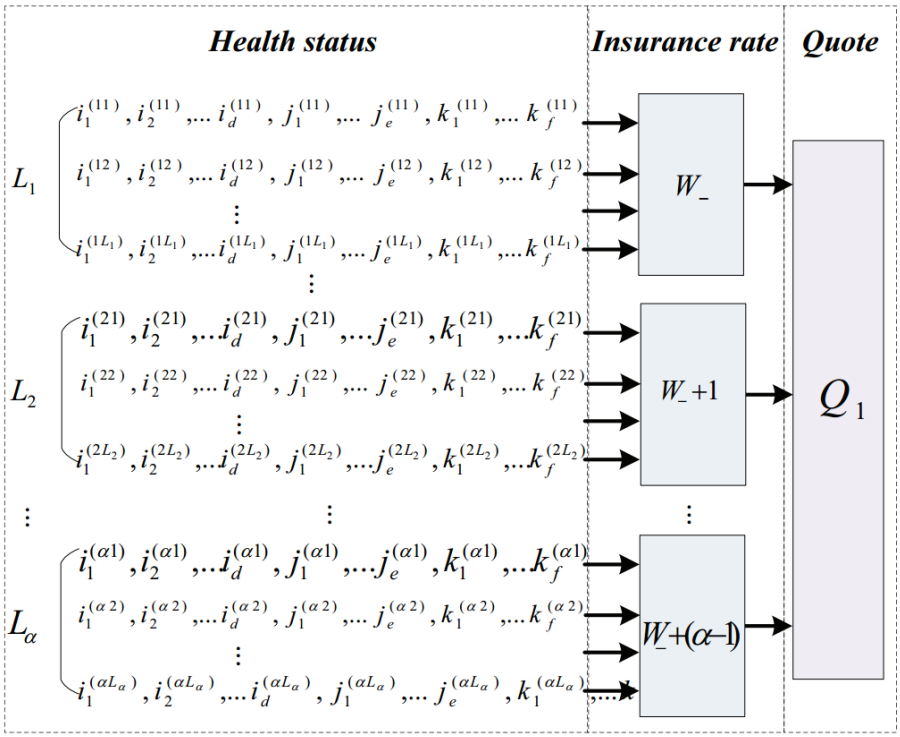

Figure 5.2: Privacy Preserving Quote with k-anonymity

5.2.2 FAPP Design Goals

Security: The user's data should be protected during transmission and storage.

Privacy-preservation: 1) The company is unable to access the user's health status or other sensitive data except information which is necessary for health insurance applications, these include names, addresses and birth date, etc. 2) Any other entities, besides the company and the user, are unable to derive health information or quote by observing transmissions between the user and company.

Fairness: 1) The company should be unable to deny that they have received the user's data provided for a quote. Furthermore, they are not allowed to access sensitive health information. 2) The users should provide truthful information to the company. Otherwise, the dishonest behaviors should be detected by the company.

For example, when a user later makes an insurance claim, the company can ask the user to submit his/her medical history to determine whether the user had lied on his/ her insurance application.

5.3 FAPP Protocol Description

In this section, we propose the FAPP insurance application protocol, which can be used by both the user and the company to calculate quotes according to the policy for transparency and fairness. However, the user's specific health details should remain hidden from the company to achieve privacy-preservation. Consequently, in the health declaration process of Figure 5.1, the health information should be in the form of a ciphertext which cannot be decrypted by the company. The company is able to determine the applicant's quote from it. More specifically, the user's health status information is denoted by

$$\{m_1, m_2, ..., m_t\} \triangleq \{a_{i_1} a_{i_2}, ..., a_{i_d}, b_{j_1}, b_{j_2}, ..., b_{j_e}, c_{k_1}, c_{k_2}, ..., c_{k_f}\}.$$

The composition of the proposed scheme is explained in detail below.

5.3.1 System Initialization

Given the security parameter γ, the TA generates a large prime q. P is a generator of cycle group G, which is on ECC with order q. The TA randomly selects $s \in \mathbb{Z}_q^*$ as the master private key and computes the public key X_P where $X_P = sP$. Moreover, the TA chooses a secure hash function: $H_1 : \mathbb{Z}_q^* \times \mathbb{Z}_q^* \rightarrow \mathbb{Z}_q^*$ and a certificateless signcryption algorithm $Sgn(\bullet)$. The system parameter is published as $params = (q, P, X_P, H_1, Sgn(\bullet))$.

5.3.2 Registration

Both the user and the insurance company register with the TA for key generation. We denote $i \in u, c$ the user u or the company c. These steps are performed by the entity i and the TA interactively.

- The entity i randomly chooses $x_i \in \mathbb{Z}_q^*$ as its secret value and computes $X_i = xiP$ as its public key, and sends it to the TA.
- The TA randomly selects $y_i \in \mathbb{Z}_q^*$ and computes $Yi = y_iP, z_i = y_i + sH_1(ID_i, Y_i, X_i, X_P)$ for the user with partial public key X_i.
- The partial private key z_i is sent to the user through a secure channel and the public key (X_i, Y_i) is stored in the public tree by the TA.

The full private key of entity i is (x_i, z_i). The full public key is (X_i, Y_i). The entity i may judge the validity of the partial private key by checking whether $Y_i + H_1(ID_i, Y_i, X_i, X_P)X_P = z_iP$.

5.3.3 Health Declaration with Privacy Preservation

Upon receiving the user's application for the insurance, the company checks the eligibility of the user and randomly chooses $f_1, f_2, ..., f_t \in \mathbb{Z}_q^*$. It then computes $F_1 = f_1P, F_2 = f_2P, ..., F_t = f_tP, F = f_1 + f_2 + ... + f_t$. The company sends $(F_1, F_2, ..., F_t)$ to the user. The user then randomly selects $r_1, r_2, ..., r_t \in \mathbb{Z}_q^*$ and hides their health information by computing

$$U_1 = m_1Y_c + r_1X_c + x_uF_1, h_1 = H_1(m_1||r_1)$$

$$U_2 = m_2Y_c + r_2X_c + x_uF_2, h_2 = H_1(m_2||r_2)$$

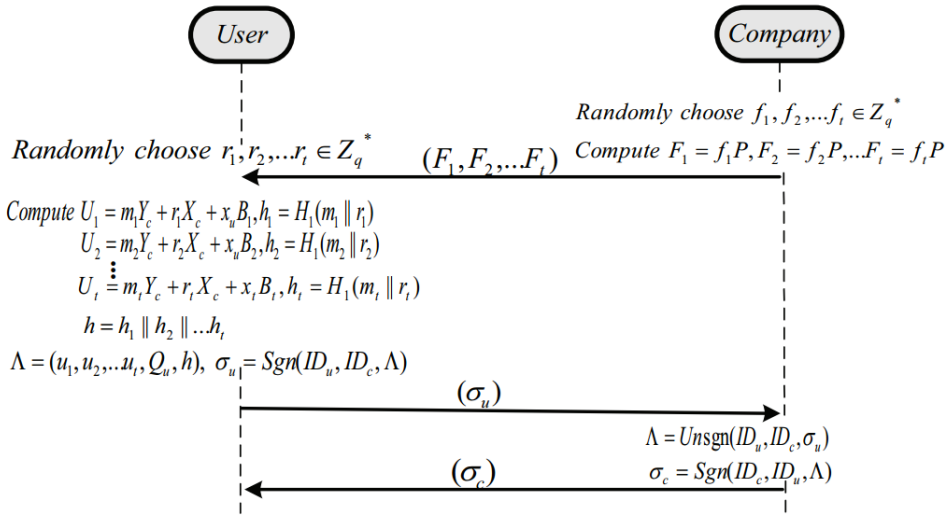

Figure 5.3: Proposed FAPP Protocol

...

$$U_t = m_tY_c + r_tX_c + x_uF_t, h_t = H_1(m_t||r_t)$$

and

$$h = h_1||h_2||...||h_t$$

The user later computes his insurance rate as $M_u = m_1 + m_2 + ... + m_t$.

It is worth noting that we assume insurance companies make their insurance policies public. As a result, the user knows his/her eligibility and how much they have to pay for their policy. This can assist in achieving fairness by preventing the abuse of insurance companies. Instead of sending M_u to the company directly, the user checks Table 5.1 for a corresponding quote Q_u. Thus, the user formulates their data as $\Lambda = (U_1, U_2, ..., U_t, h, Q_u)$ and creates a signcryption on it, obtaining $\sigma_u = Sgn(ID_u, ID_c, \Lambda)$. The user sends σ_u to the company, as shown in Figure 5.3.

After receiving σ_u, the company unsigncrypts it with their private key and the user's public key, getting $\Lambda = Unsgn(ID_u, ID_c, \sigma_u)$. The company stores Λ as the

evidences in the format of Table 5.2. Simultaneously, the company also signcrypts on Λ, formulating $\sigma_c = Sgn(ID_c, ID_u, \Lambda)$, and sends it to the user as an evidence for their policy approval.

Remark 1 *1) The user should sign on Λ because $(U_1, U_2, ..., U_t, h)$ acts as evidence to provide non-repudiation of the user. On the other hand, the information of quote Q_u can only be accessed by the health insurance company and should be kept confidential during the transmission. Therefore, both signature and encryption are required for the health declaration. In order to reduce the computational complexity, we adopt signcryption to achieve the security objectives. 2) The user is also required to collect evidence from the company to show that they provided a true declaration of health to the company should an argument appear. Consequently, the company should also make a signcryption on Λ as an evidence for the user.*

Table 5.2: **EVIDENCE AVAILABLE TO COMPANY**

USER ID	Public Key	Evidence	Data
ID_u	X_u	$\sigma_u, f_1, f_2, ..., f_t$	$U_z, U_2, ..., U_t, h, Q_u$

5.3.4 Argument Disposal

A user may make a health insurance claim. If the company suspects fraud, the user will be asked to provide their medical history, $(m'_1, , m'_2, ..., m'_t)$. Each m'_i corresponds to m_i for all $i \in 1, 2, ..., t$. In this case, the user sends the company

$$(m'_1, m'_2, ..., m'_t, r_1, r_2, ..., r_t).$$

The company then checks whether the following equations holds

$$U_i \triangleq m_i' Y_c + r_i X_c + x_u F_1, h_i \triangleq H_1(m_i' || r_i)$$

for all $i \in 1, 2, ..., t$. If all the equations hold, it has been proven that the user provided true health details[3]. The company then continues checking whether $M_u' = m_1' + m_2' + ... + m_t'$ corresponds to a distinct health rate Q_u. Upon completing all verifications, the company will compensate the user for their insurance claim. However, if the user is proven to have falsified details they will be rejected.

The traditional insurance application procedure violates the user's privacy because the user needs to send all of their health details $m_1, m_2, ..., m_t$ to the company. In order to address this problem, we propose to send only the insurance rate M_u' and the summation of the random number R to the company for verification, where $M_u' = m_1' + m_2' + ... + m_t'$ and $R = r_1 + r_2 + ... + r_t$. The company checks

$$U_1 + U_2 + ... + U_t = M_u' Y_c + R X_c + F X_u, (1)$$

where $F = f_1 + f_2 + ... + f_t$. In the next section, we prove that the proposed protocol is insurance-rate cheat resistant. Thus, if the equation holds, the user is proved to provide the true insurance rate to the company. It is worth noting that the insurance rate M_u may leak some information of the user. However, it has better privacy protection than the traditional method where $(m_1, m_2, ..., m_t)$ has to be provided.

[3]In Section 5.4, we prove that the scheme is health-declaration-cheat resistant

5.4 Security Analysis

In this section, we prove that the proposed protocol can resist a health-declaration-cheat and insurance-rate-cheat. Furthermore, we show that the proposed protocol can achieve privacy preservation and fairness.

5.4.1 Health Declaration Cheat Resistance

The user falsifies health declarations (statuses) to the company to reduce their quote. The following lemma demonstrates that the proposed scheme can achieve health declaration cheat resistance.

Lemma 1 The user is able to cheat on the health status only when he is able to solve the ECDLP.

Proof Without loss of generality, we assume that the user cheats on m_1 by setting $m_1 = m_1' - d$, where $d = a_1 - a_0$ and m_1' is the true health status. The user provides $U_1 = m_1 Y_c + r_1 X_c + x_u F_1$ to the company. When a claim is made, the user should disclose to the company their true health details m_1' with another random number r_1'. If the user is able to select the right r_1' to pass the verification, they will be successful in falsifying the health information. In order to pass the verification, (m_1', r_1') should satisfy the following equation:

$$U_1 = m_1' Y_c + r_1' X_c + x_u F_1,$$

Thus, we have

$$m_1 Y_c + r_1 X_c + x_u F_1 = m_1' Y_c + r_1' X_c + x_u F_1. (2)$$

Replace m_1 with $m_1' - d$, Eq. 2 becomes

$$(r_1 - r_1')X_c = dY_c. (3)$$

In Eq. 3, given d and r_1, the user has to solve ECDLP for obtaining the value of r_1' [4]. Consequently, under ECDLP assumption, the proposed scheme is health-declaration-cheat resistant.

5.4.2 Insurance Rate Cheat Resistance

The user may send a false quote to the company even though they provided the true health information. The following lemma shows that the proposed scheme can achieve insurance-rate-cheat resistance.

Lemma 2 The user is able to cheat on the insurance rate only when he is able to solve the ECDLP.

Proof We assume that the user provides the false insurance rate $M_u^* = M_u - e$, where $M_u = m_1 + m_2 + ... + m_t$ is the true insurance rate and e is an integer selected by the user. In order to pass the verification of Eq. 1, he has to send the company a false $R^* = R + \delta$, where $R = r_1 + r_2 + ... + r_t$ is the true summation of the random numbers. Given e, if the user can find the correct δ to pass Eq. 1 with (M_u^*, R^*), he will succeed in falsifying the insurance rate. In order to pass the verification of the company, the following equation should hold:

$$M_u^* Y_c + R^* X_c + F X_u = U_1 + U_2 + ... + U_t (4).$$

[4] Since d and Y_c are known to the user, the scale multiplication product of $(r_1 - r_1')X_c$ can be obtained. For simplification, we denote $r_1 - r_1' = a, x_c = b$, thus $X_c = bP, (r_1 - r_1')X_c = abP$. To solve Eq. 3 over r_1' is to actually solve the following problem: Given abP and bP, compute a.

Eq. 4 can also be presented as

$$M_u Y_c - eY_c + RX_c + \delta X_c + FX_u = M_u Y_c + RX_c + FX_u. (5)$$

From Eq. 5, we find that δ should satisfy

$$eY_c = \delta X_c. (6)$$

Similar to the proof of Lemma 1, the user should be able to solve ECDLP for obtaining the value of δ. Therefore, the proposed scheme is insurance-rate-cheat resistant under the ECDLP assumption.

5.4.3 Privacy Preservation

In the proposed scheme, the user's health details m_i for all $i \in 1, 2, ..., t$ are encapsulated as $U1, U2, ..., Ut$ with random numbers $r1, r2, ..., rt$. Thus, the company is unable to access the content m_i, because they do not know r_i. Moreover, even though the user provides their quote Q_u to the company, the company only has a small possibility of guessing the user's health statuses because one quote can be generated through different combinations of health statuses. Additionally, the observers, who view all transmission between the user and company, know nothing about the user's health information because the data Λ is signcrypted by the user and the company. Only the intended receiver, the company or the user, can unsigncrypt it. Even if an adversary unsigncrypts the ciphertext σ_u or σ_c, they would only receive encapsulated information $U_1, U_2, ..., U_t$, which provide no information about $m_1, m_2, ..., m_t$ without the knowledge of $r_1, r_2, ..., r_t$.

5.4.4 Fairness

The proposed scheme provides fairness for both the users and the company. In the existing health insurance system, the users passively accept the quotes from the company without prior knowledge on how quotes are calculated. Whereas, in our proposed scheme, quote specifics are included in the public policy such that users can calculate quotes themselves. However, users are compelled to provide honest health information to the company. Otherwise, false information will be detected by the company according to *Lemma 1* and *Lemma 2*. This provides fairness for the company. Additionally, the user and the company's signcryption σ_u and σ_c provide non-repudiation of sending and receiving the data Λ for the user and company, respectively.

5.5 Performance Evaluation

In this section, we analyze the computational and communication overhead of the proposed FAPP protocol.

5.5.1 Computation overhead

We denote tm as the time consumed for one scalar multiplication in G, t_s and t_u as the time consumption for the signcryption and unsigncryption, respecitively. In the proposed FAPP protocol, it takes the user three scalar multiplications to compute each ciphertext U_i for $i \in 1, 2, ..., t$. The total computational overhead for the t ciphertexts is $3t \times t_m$. Additionally, the user has to signcrypt on Λ and unsigncrypt on σ_c. Thus, the computational overhead of the user is $3t \times t_m + t_s + t_u$. For the company, it takes $t \times t_m$ to compute $F_1, F_2, ..., F_t$ and $t_u + t_s$ to complete the signcryption

Table 5.3: **CBAC COMPUTATIONAL AND COMMUNICATION OVERHEAD**

Overhead	User	Company
Computation	$3t \times t_m + t_s + t_u$	$t \times t_m + t_u + t_s$
Communication	$\|\sigma\|$	$t\|G\| + \|\sigma\|$

and unsigncryption algorithms, as shown in Table 5.3.

5.5.2 Communication Overhead

We denote $|G|$ and $|\sigma|$ as the size of an element in G, and the size of the signcryption, respectively. In FAPP, the user sends the signcryption σ_u to the company, generating $|\sigma|$ communication overhead. The company sends $F_1, F_2, ..., F_t$ and σ_c to the user. Thus, the communication overhead is $t|G| + |\sigma|$, as shown in Table 5.3.

5.5.3 Concluding Remarks

In this chapter, we introduced the concept of privacy-preserving quotes for online health insurance systems. Based on the quote mechanism, we proposed a fairness-aware and privacy-preserving (FAPP) insurance application protocol, which can achieve health-declaration-cheat resistance and insurance-rate-cheat resistance under the ECDLP assumption. The proposed FAPP protocol protects user health information privacy by encapsulating data into a ciphertext with random numbers. Moreover, FAPP is fairness-aware because quote details are publicly available and users can compute their own quotes. However, users are still unable to falsify their rates and quotes without being detected.

Chapter 6

Conclusions and Future Work

In this chapter, we conclude the main contributions of this thesis. We also outline potential research directions in security and privacy-preservation for electronic health record systems.

6.1 Conclusions

Electronic health record exchanges are crucial functions for modern healthcare systems. These components are fundamental in providing quality care and enabling a larger spectrum of services. These may include online health insurance tools. Due to health record sensitivity, privacy-preservation is a crucial issue for electronic health record systems. A framework which protects patient information during data exchanges is essential for a reliable healthcare network. Furthermore, modern healthcare access control frameworks should also accommodate patient consent. Therefore, this thesis proposes a new consent-based scheme to manage access control for sensitive health information while including patient consent in data exchanges.

A potential application of our above contributions is the accommodation for health insurers in electronic health networks. Currently, record retrieval may result in an excessive collection of sensitive data by insurers. This is a clear violation of privacy and has major ramification if insurers stockpile previous collected patient profiles. In order to address this issue, this thesis proposes a fairness-aware and privacy-preserving (FAPP) protocol for online health insurance systems. In summary, the main contributions of the thesis are fourfold.

- We proposed a framework for electronic health record systems, in which data providers offer the original health information (encrypted for privacy preservation) to the data center. The user is able to ensure the integrity of the data. Other entities are unable to access the data without prior consent from the user.

- We proposed a consent-based access control mechanism for secure and privacy-preserving health information exchanges. The data requesters must negotiate with users to access their health information. After reaching an agreement, the user sends a consent token to the data requester and a consent notification is sent to the data center. The consent token is carefully designed so that only the intended data requesters are able to access the data.

- We proposed a concept of privacy-preserving health insurance quotes by adopting the k-anonymity technique. The attributes are divided into different types and are endowed with different weights. In this case, attributes denote user health conditions or habits. Attributes of the same type have the same weights. One quote can correspond to multiple attribute combinations. Therefore, it is difficult for entities to guess the health details of a user.

- We designed a fairness-aware and privacy-preserving insurance application

protocol. In FAPP, the health insurance policy is public. Thus, the user is able to calculate their own quotes. The user's health status information is encapsulated into a ciphertext with random numbers to achieve privacy-preservation. Nevertheless, the company is able to check whether the user provided an honest declaration after a claim is made.

6.2 Future Work

There remain many challenges for security and privacy-preservation in electronic health record systems. Based on the work done in this thesis, the following topics are recommended for further work in this area. The first area of research being social network assisted electronic health record exchanges. In these EHR systems, the owners of records are people who form social networks. Consequently, the security and privacy preservation solutions of social network can be explored in EHR systems. However, it is challenging to construct reliable social ties and social trusts to improve the security and privacy in electronic health record systems. Another area of interest is cloud-based security and privacy-preservation in EHRs. In electronic health record systems, health records are stored in data centers, which we can considered as the cloud. Security and privacy preservation in the cloud have drawn considerable interest from research community and industry. Securing and preserving privacy in the cloud presents its own challenges. Considering the nature and sensitivity of electronic health record, successfully and securely merging these two areas is of considerable interest. Enhancing privacy and security in e-Health is a challenging and meticulous task; however, it is a challenge well worth pursuing.

Bibliography

[1] Abbas, A., and Khan, S. U. A review on the state-of-the-art privacy-preserving approaches in the e-health clouds. *IEEE Journal of Biomedical and Health Informatics 18*, 4 (2014), 1431–1441.

[2] Acquisti, A., Brandimarte, L., and Loewenstein, G. Privacy and human behavior in the age of information. *Science 347*, 6221 (2015), 509–514.

[3] Al-Riyami, S. S., and Paterson, K. G. Certificateless public key cryptography. In *Asiacrypt* (2003), vol. 2894, Springer, pp. 452–473.

[4] Anderson, R. J. A security policy model for clinical information systems. In *Security and Privacy, 1996. Proceedings., 1996 IEEE Symposium on* (1996), IEEE, pp. 30–43.

[5] Appari, A., and Johnson, M. E. Information security and privacy in healthcare: current state of research. *International journal of Internet and enterprise management 6*, 4 (2010), 279–314.

[6] Appelbaum, P. S. Privacy in psychiatric treatment: threats and responses. *Focus* (2003).

[7] Barreto, P., Deusajute, A. M., Cruz, E., Pereira, G., and Silva, R. Toward efficient certificateless signcryption from (and without) bilinear pairings. *Preprint* (2008).

[8] Beale, T. Iso 18308 conformance statement. *The openEHR Foundation 2006* (2002).

[9] Becker, M. Y., and Sewell, P. Cassandra: Flexible trust management, applied to electronic health records. In *Computer Security Foundations Workshop, 2004. Proceedings. 17th IEEE* (2004), IEEE, pp. 139–154.

[10] Boneh, D., and Boyen, X. Efficient selective identity-based encryption without random oracles. *Journal of Cryptology 24*, 4 (2011), 659–693.

[11] Brandom, R. Uk hospitals hit with massive ransomware attack, May 2017.

[12] canadian healthcare.org. Canadian health care.

[13] Church, E. Canada's health-care bill to top $219-billion this year. *The Globe and Mail* (2015).

[14] CISRO. Canadian insurance services regulatory organizations (cisro), 2017.

[15] Coiera, E., and Clarke, R. e-consent: The design and implementation of consumer consent mechanisms in an electronic environment. *Journal of the American Medical Informatics Association 11*, 2 (2004), 129–140.

[16] Demiris, G., Afrin, L. B., Speedie, S., Courtney, K. L., Sondhi, M., Vimarlund, V., Lovis, C., Goossen, W., and Lynch, C. Patient-centered applications: use of information technology to promote disease management and wellness. a white paper by the amia knowledge in motion working group. *Journal of the American Medical Informatics Association 15*, 1 (2008), 8–13.

[17] eHealth Ontario. ehealth ontario.

[18] Ferraiolo, D. F., Sandhu, R., Gavrila, S., Kuhn, D. R., and Chandramouli, R. Proposed nist standard for role-based access control. *ACM Transactions on Information and System Security (TISSEC) 4*, 3 (2001), 224–274.

[19] Group, C. Healthcare Challenges and Trends. Tech. rep., CGI Group, 2004.

[20] Hu, J. *Privacy-Preserving Data Integration in Public Health Surveillance*. University of Ottawa (Canada), 2011.

[21] INCITS, A. Incits 359-2012. *Information Technology-Role Based Access Control* (2012).

[22] Jing, X. Provably secure certificateless signcryption scheme without pairing. In *Electronic and Mechanical Engineering and Information Technology (EMEIT), 2011 International Conference on* (2011), vol. 9, IEEE, pp. 4753–4756.

[23] Kalra, D. Cen pren 13606, draft standard for electronic health record communication and its introduction to iso tc/215. *CEN/TC 251* (2004).

[24] Lin, X., Lu, R., Shen, X., Nemoto, Y., and Kato, N. Sage: a strong privacy-preserving scheme against global eavesdropping for ehealth systems. *IEEE Journal on Selected Areas in Communications 27*, 4 (2009), 365–378.

[25] Lin, X., Sun, X., Ho, P.-H., and Shen, X. Gsis: A secure and privacy-preserving protocol for vehicular communications. *IEEE Transactions on vehicular technology 56*, 6 (2007), 3442–3456.

[26] Liu, W.-H., and Xu, C.-x. Certificateless signcryption scheme without bilinear pairing. *Ruanjian Xuebao/Journal of Software 22*, 8 (2011), 1918–1926.

[27] Lu, R., Lin, X., and Shen, X. Spoc: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency. *IEEE Transactions on Parallel and Distributed Systems 24*, 3 (2013), 614–624.

[28] Mercuri, R. T. The hipaa-potamus in health care data security. *Communications of the ACM 47*, 7 (2004), 25–28.

[29] Muzyka, D., Hodgson, G., and Prada, G. The inconvenient truths about canadian health care, 2012.

[30] News, C. EHealth Scandal a $1B waste auditor, 2009.

[31] Ni, Q., Bertino, E., Lobo, J., Brodie, C., Karat, C.-M., Karat, J., and Trombeta, A. Privacy-aware role-based access control. *ACM Transactions on Information and System Security (TISSEC) 13*, 3 (2010), 24.

[32] Oecd. Oecd statistics.

[33] O'Keefe, C. M., Greenfield, P., and Goodchild, A. A decentralised approach to electronic consent and health information access control. *Journal of Research and Practice in Information Technology 37*, 2 (2005), 161–178.

[34] Paar, C., and Pelzl, J. *Understanding cryptography: a textbook for students and practitioners*. Springer Science & Business Media, 2009.

[35] Paper, S. W. Combating insurance claims fraud-how to recognize and reduce opportunistic and organized claims fraud. Tech. rep., SAS Institute, 2012.

[36] Qiu, F., Wu, F., and Chen, G. Privacy and quality preserving multimedia data aggregation for participatory sensing systems. *IEEE Transactions on Mobile Computing 14*, 6 (2015), 1287–1300.

[37] Rindfleisch, T. C. Privacy, information technology, and health care. *Communications of the ACM 40*, 8 (1997), 92–100.

[38] Russello, G., Dong, C., and Dulay, N. Consent-based workflows for healthcare management. In *Policies for Distributed Systems and Networks, 2008. POLICY 2008. IEEE Workshop on* (2008), IEEE, pp. 153–161.

[39] Samy, G. N., Ahmad, R., and Ismail, Z. Threats to health information security. In *Information Assurance and Security, 2009. IAS'09. Fifth International Conference on* (2009), vol. 2, IEEE, pp. 540–543.

[40] Sandhu, R., Ferraiolo, D., and Kuhn, R. The nist model for role-based access control: towards a unified standard. In *ACM workshop on Role-based access control* (2000), vol. 2000, pp. 1–11.

[41] Sandhu, R. S. Role-based access control. *Advances in computers 46* (1998), 237–286.

[42] Schloeffel, P., Beale, T., Hayworth, G., Heard, S., Leslie, H., et al. The relationship between cen 13606, hl7, and openehr. *HIC 2006 and HINZ 2006: Proceedings* (2006), 24.

[43] Senese, S. V. A study of access control for electronic health records. Master's thesis, Governors State University, 2015.

[44] Shao, J., Wei, G., Ling, Y., and Xie, M. Identity-based conditional proxy reencryption. In *Communications (ICC), 2011 IEEE International Conference on* (2011), IEEE, pp. 1–5.

[45] Shi, E., Bethencourt, J., Chan, T. H., Song, D., and Perrig, A. Multi-dimensional range query over encrypted data. In *Security and Privacy, 2007. SP'07. IEEE Symposium on* (2007), IEEE, pp. 350–364.

[46] Shi, W., Kumar, N., Gong, P., and Zhang, Z. Cryptanalysis and improvement of a certificateless signcryption scheme without bilinear pairing. *Frontiers of Computer Science 8*, 4 (2014), 656–666.

[47] Shi, Y., Fan, H., and Xiong, G. Obfuscatable multi-recipient re-encryption for secure privacy-preserving personal health record services. *Technology and Health Care 23*, s1 (2015), S139–S145.

[48] Son, J., Kim, D., Hussain, R., and Oh, H. Conditional proxy re-encryption for secure big data group sharing in cloud environment. In *Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on* (2014), IEEE, pp. 541–546.

[49] Spending, H. Canadian Institute for Health Information, 2016.

[50] Weng, J., Deng, R. H., Ding, X., Chu, C.-K., and Lai, J. Conditional proxy re-encryption secure against chosen-ciphertext attack. In *Proceedings of the 4th International Symposium on Information, Computer, and Communications Security* (2009), ACM, pp. 322–332.

[51] Wu, L., Li, J.-Y., and Fu, C.-Y. The adoption of mobile healthcare by hospital's professionals: An integrative perspective. *Decision Support Systems 51*, 3 (2011), 587–596.

[52] Wu, S., Wang, X., Wang, S., Zhang, Z., and Tung, A. K. K-anonymity for crowdsourcing database. *IEEE Transactions on Knowledge and Data Engineering 26*, 9 (2014), 2207–2221.

[53] Xiong, H., and Qin, Z. Revocable and scalable certificateless remote authentication protocol with anonymity for wireless body area networks. *IEEE transactions on information forensics and security 10*, 7 (2015), 1442–1455.

[54] Zhang, A., Bacchus, A., and Lin, X. Consent-based access control for secure and privacy-preserving health information exchange. *Security and Communication Networks* (Dec. 2016).

[55] Zhang, A., Bacchus, A., and Lin, X. A fairness-aware and privacy-preserving online insurance application system. In *IEEEGC16: Freedom through Communications* (Sep. 2016), Globecom.