EHR SECURITY AND PRIVACY:
ENCOUNTERING HONEST-BUT-CURIOUS ATTACKS THROUGH
SELECTIVE MULTI-LEVEL ACCESS CONTROL POLICY

Approved by:

Dr. Dr. Suku Nair(Advisor)

Dr. Stephen Szygenda

Dr. Khaled Abdelghany

Dr. Frank P. Coyle

Dr. Tyler Moore

EHR SECURITY AND PRIVACY:
ENCOUNTERING HONEST-BUT-CURIOUS ATTACKS THROUGH SELECTIVE MULTI-LEVEL ACCESS CONTROL POLICY

A Ph.D. Dissertation Presented to the Graduate Faculty of the

Lyle School of Engineering

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Computer Science

by

Bilal Ibrahim I. Alqudah

(B.A., Mu'tah University, Mu'tah, Karak, Jordan, 1999)
(M.S., Southern Methodist University, Dallas, TX, USA, 2007)

May 16 , 2015

Ibrahim I. Alqudah, Bilal B.A., Mu'tah University, Mu'tah, Karak, Jordan, 1999
M.S., Southern Methodist University, Dallas, TX, USA, 2007

EHR Security and Privacy:
Encountering Honest-But-Curious Attacks Through
Selective Multi-level Access Control Policy

Advisor: Professor Dr. Suku Nair
Doctor of Philosophy degree conferred May 16 , 2015
Ph.D. Dissertation completed April 22 , 2015

The expansion in automation, digitalization, and network communication in the health care sector provided advantage and developed concerns regarding privacy protection and data security. The knowledge provided by medical and auxiliary data can reveal identity with high accuracy. The value of Electronic Medical Records comes from its content, the amount of personal information it hold, and its impact if disclosed to the public. Being identifiable based on non quasi-identifiers gives an indication of the problem complexity where data cannot be contained in one location.

Consequences of releasing medical information could be job loss, increased insurance rates, identity theft, sexual crimes, and discrimination based on health problems. Although hospitals and medical facilities are trust-worthy parties, attacks on patient's privacy can come from inside those facilities. Almost 50% of privacy violations came from a person who works inside hospital [92]. This type of attack can be classified as honest-but-curious attack (HBC) where the attacker is an honest person and authorized to access system resource but he could abuse his access rights to learn more information.

The main goal of this research is to provide a framework targeting HBC attackers whether they are active or passive in hospitals, as a known context. The framework identifies risk assessment, data segmentation, data sharing, fine granularity access rights, and patient participation in data protection as factors in a bigger formula in privacy protection.

In Risk assessment, the framework provides a process of risk assessment and the compliance with the regulations, standard, and provides a method of exchanging compliance results without disclosing the interior policy details. In the area of data sharing, the framework provides a communication protocol to build trust relationships in healthcare network where data exchanged based on a quantified trust association. The exchange process provides the ability to filter patients private data before sharing according to its sensitivity and according to patients privacy preferences.

In access control policy, the framework provides a novel approach for fine-granularity access control where access is granted in a segment level rather than file level. The access control policy provides a solution for mutual access in the same role, granting access rights selectively, and revoking access rights using compound key structure. We provide various implementations for cryptographic access controlling and multi-level access controlling.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

First, I thank all my teachers , professors, and everyone who participated in teaching me a letter ...

Thank you all!

To my great example in life, my parents.

To the soul of my mother, Zenah.

To my father, Ibrahim.

To my beloved wife, Rima and my children, Furat, Aynaa, Awn , Ellen, Lemar.

To my brothers and sisters who support me all the time.

I dedicate this work

Chapter 1

INTRODUCTION

Due to the accelerated development of facilities providing health care service and the wide acceptance of computer and network technologies, regulations specifying privacy protection as requirement for heath care systems [47]. Concerns like information misuse, privacy violation, identity theft, data loss, and cyber security are evolving rapidly.However, depending on traditional access control, encryption, and physical security may not be sufficient in an environment where attacks from inside and outside are equally likely to occur [92].

Honest-but-curious attack (HBC) is a well-known adversary model with a slight difference in definition, [15, 36, 74] but confirming a common factor of its nature as an internal attack. The attacker in HBC is a person who is not required to break the rules or hack the system to get access to information. The domain of the attack is information centric systems and its goal is to collect information.

Based on U.S. Department of Health and Human Services report [50], in 2012 about 71.8% of clinics (office based) are using some sort of electronic medical records. However, only 39.6% only meet the requirements for a basic system and 23.5% have fully functioning system. The selection to Electronic Medical Records (EMR)came from the fact that EMR systems is a data centric system known to have HBC model [92,102]. However, the environment is assumed to be trusted and closed where information should be used for a known purpose, yet attacks under the category of HBC do exist with serious consequences.

As defined in [76]: "Electronic medical records (EMRs) are a digital version of the paper charts in the clinician's office. An EMR contains the medical and treatment history of the patients in one practice". Using EMRs provide more advantages than conventional paper medical documents such as:

1. Increasing the level of health care provided.
2. Reducing medical errors and health-care cost.
3. Reducing health-care disparities.
4. Reducing diagnostic test duplication.
5. Reducing administrative expenses.
6. Tracking data over time.
7. Engaging patients and their families.
8. Easing patients tracking in terms of follow up or future checkups.
9. Monitoring and improving overall quality of care within the practice.

The expansion of collecting information about individuals has increased over time. For example, in the State of Illinois in 1983, a birth certificate contained four fields of 280 bytes of data. However, by 1996 it became 1864 bytes, and the amount of data saved per health care visit jumped from "zero"bytes in 1983 to 633 bytes. In 1999 a typical electronic birth certificate contained 226 fields [93]. The data collected consists of information like parent's information and addresses, race, tobacco usage, alcohol usage, medical risk factors, obstetric procedures, abnormal conditions of newborn and other details. The amount and type of information uniquely identifies the subject of the file can also specify other family members. It is a point where much private information exists in one place.

An estimated saving for long-term cost of a national health care information system established is around $77.8 billion annually [89]. This large savings will contribute to enhancing the health-care sector, service provided and improving its quality by expanding the umbrella of health-care coverage and affordable benefits. Identifying individuals based on medical information provided by EMR is not difficult. According to El Emam *et al.* [29], 63% to 87% of the US population is identifiable based on some distinctive characteristics.

Being identifiable based on non quasi-identifiers gives an indication of the problem complexity where data cannot be contained in one location. In this instance, it is not only personal information that identifies a person. The knowledge provided by medical and auxiliary data can reveal identity with high accuracy. The value of Electronic Medical Records comes from its content, the amount of personal information it hold, and its impact if disclosed to the public. Consequences of releasing medical information could be job loss, increased insurance rates, identity theft, sexual crimes, and discrimination based on health problems.

Although hospitals and medical facilities are trust-worthy parties, attacks on patient's privacy can come from inside those facilities. Almost 50% of privacy violations came from a person who works inside hospital [92]. According to Los Angeles Times, around 150 staff members have access to parts of a patient's information while receiving health care services [35]. The size of a health care industry based on The American Hospital Association is massive [103] covering 5,686 registered hospitals with total expenses of $859,419,233,000.

Since EMRs contains many details about patient's private life, this could drive legitimate honest-but-curious users to seek some information without patient's consent. Currently, neither hospitals nor patients are able to control what a legitimate user can or connot see. It is estimated that about 12 million users will connect to a national

healthcare network which will expose medical records to misuse and attacks [4].

The agreement on EMR systems importance, as well as the shortages in conventional security suffers from inability to protect patient privacy from insiders attacks requires farther research in privacy protection. Information need to be classified and segmented based on established criteria in order able to selectively disclose or conceal data on a need-to-know basis. The standards on which data should be segmented upon remains a point of research and study, a work group has been established to discuss segmentation and its implementation [37].

Assuming segmentation is taken care of by using existing standards such as HL7 [52]. This research studies data segment security and how segmentation serve the purpose of privacy protection using selective access control policies. Currently, the practice to guarantee the privacy and confidentiality of EMR is to encrypt records and implement some access control policy such as RBAC.

An access control system initiates the proper access control policies where each entity in the health care system must present certain credentials to gain access to resources they need. Encrypting information guarantees its confidentiality, while access control policies (i.e. RBAC, DAC, and MAC) offer system resources protection from unauthorized access [51]. However, neither of the solutions solve the problem of HBC adversary model as authorized users abusing given or granted rights.

The main goal of this research is to provide a framework targeting HBC attackers whether they are active or passive in hospitals, as a known context. The framework identifies risk assessment, data segmentation, data sharing, fine granularity access rights, and patient participation in data protection as factors in a bigger formula in privacy protection. The problem is defined in chapter 3, chapters 4 through 7 cover the framework components. The framework is a product that enables patients to

participate in protecting their own information. Chapter 8 provides an overview of the software developed for the framework. However, a complete solution implementation requires more time than the life of this research.

1.1. Problem Overview

The adversary model of honest-but-curious presumes the existence of a legitimate user who is trustworthy, but acts against patient's privacy driven by curiosity to collect more information. As mentioned before, this category performs 50% of attacks through accessing medical records in both electronic and paper formats. This model, HBC, is the point of research conducted in this work and how this threat can be mitigated.

It is considered a privacy violation if someone gleans information from medical record, such as HIV status or habits, without medical need or data owner consent. It cannot be considered an attack in terms of causing immediate harm or denial of service because the attacker is a legitimate user. Yet, from privacy point of view, the act of accessing information without a need is considered an act of attack and a violation of privacy roles.

Despite having access control policies, authorized users who have access to the system could potentially violate patients' privacy by accessing files for different purposes other than providing health care service. Users classified as HBC attackers will not be affected by how strict access control policies are , or by encryption strength since they are authorized system users. Figure 1.1 authorized participants in a secure EMR system [63] with a potential to violate patients' privacy.

Different access control policies, such as attributed RBAC, can limit privacy breaches if an adversary attempted to get access to protected health information that is not assigned to their role. The complexity of managing roles and attributes for all users if access granted in a user level requires a significant efforts as well as continuous maintenance if done by organizations. Information sharing and data transfer are additional problems that access control policies will not solve, especially if EMR is transferred outside its host system. The proposals of data segmentation [12, 73]

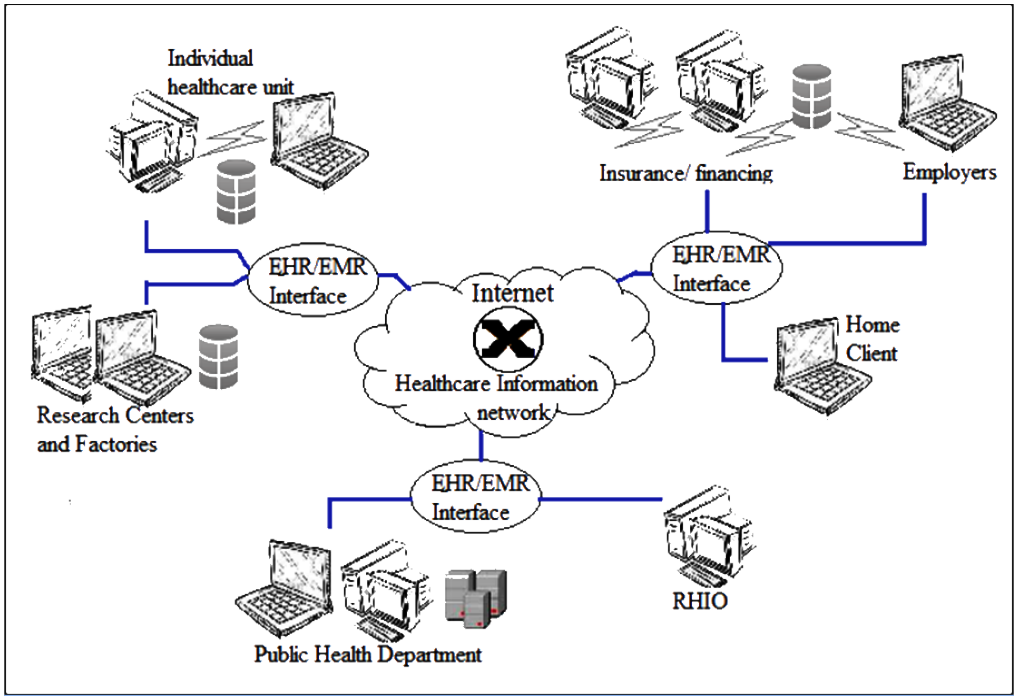


Figure 1.1. EHR accessed by authorized parties

provide sufficient evidence on the importance of data classification for privacy protection. However, the official proposal did not provide a mechanism to protect segmented information.

The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy and Security Rules provided a national standards for security and privacy rules to protect health information (PHI). However, HIPAA standards and rules suffer from shortcomings ,shown in Table 1.1, that effects privacy protection [21].

Table 1.1: HIPPA shortcomings and areas need to consider

No.	Area HIPAA need to cover
1	Patients receive notice about privacy practices, however data security operates behind the scenes are out of patient's control.
2	Patient's consent for using his/her medical information is not required if it is used or disclosed for treatment, payment, or health care operations.
3	Patients entire medical information may become available, even if it is irrelevant to current treatment and thought to be protected.
4	Patient's private health information can be used for marketing and may be disclosed without authorization.
5	Business associates of a covered entity can receive protected health information (PHI) without a patient knowledge or consent.
6	Some disclosures could be made to law enforcement without a warrant or court order.

1.2. Information Sharing

Information availability is essential to provide an appropriate level of health care to patients as well as financial savings [89]. However, sharing information introduces patient's privacy to more threats, and new security safeguards will be required to protect it. In the health care network, information can be shared between different entities and facilities for many reasons [4] including reducing the cost of examinations, labs, and providing faster service. However, communicating parties may not be applying the same access control policies and security safeguards, which can lead to privacy breaches. Health care systems described as open loop systems, where no feedback is provided to patients about privacy breaches, have resulted from information sharing.

The growing need to share health information, even with the public, raises concerns about privacy protection and information security. However, the harm caused by disclosing a medical record can result in irreversible consequences on businesses and the public. Regardless of their importance, yet health care records are considered a "personal" and "private" issue for individuals and different from other types of information.

privacy and security for financial records, for example, are governmental interest, investment interest and worth paying for from risk point of view. As an example of governmental interest, the Swaziland government safeguards those individuals who have accounts in Swaziland banks [70]. In the U.S., banks are required to keep the records private in the mutual interest with the customers. Better privacy and protection make a positive reflection on business volume. On the other hand, the cost-benefit in addition to who-owns and who-pays relations is different for medical records.

In banking systems, the costs to implement privacy protection policies are justified by the benefits banking systems are getting. In health sector, the customer benefits from information sharing and data transfer more than the providers do. So far, the benefits for health care providers do not justify the cost. As being discussed, this is one major obstacle in the way of adopting EMR solutions and applying advanced privacy protection methods.

Information sharing between trustworthy parties does not guarantee patient's privacy is protected against honest-but-curious (HBC) attacks. When two or more parties share information, pre-sharing steps should be taken, such as trust negotiations, policy exchange, and authentication to guarantee information security in transit and after delivery [3, 28].

Users of health care network should be granted access to resources filtered based on some quantified trust association (relationship), the type of resources requested, and exchanged attributes (i.e. security credentials, certificates). Building a trust relationship through negotiations facilitates exchanging electronic medical records, policy disclosure, exchanging private policy information [28].In a dynamic trust negotiation [3], parties exchange certificates and other attributes (such as security policies) to build a trust relationship that allows secure resources exchange . However, adding attributes to control trust relations on top of existing access control policies, RBAC for example, is a very complicated process [61].

The framework provided in this research supports risk assessment through surveying internal policies. Evaluating policies depends on metrics provided by HIPAA and NIST. Trust associations established between hospitals are based on hospitals' compliance with standards in the provided survey and after exchanging policies..

1.3. Anonymity

The benefits of information sharing have been described in the section on Information Sharing 1.2. The goal of information anonymity is to protect patient identity protection against insider and outsider attacks. Information anonymity can protect patients' privacy preserved during data transfer or process. However, information cannot be shared, with some parties, without being anonymized through removing the basic identifiable information from the shared record.

Studies shows that it is possible to re-identify individuals using anonymized records. El [30], El Emam and Dankar show that an expert witness was able to re-identify 18 out of 20 individuals through anonymized records. Other cases reported like re-identifying of individuals from Chicago homicide records and the social security death index. By matching several datasets, it becomes more possible to get more detailed information about individuals.

Many study cases show that data will not be fully anonymous by stripping the set of identifiable information from protected health information (PHI) will make it safe to share. The set of data fields specified by HIPAA does not cover each and every variable that might identify an individual. However, fine-granularity access control policy can help in limiting access to the surrounding auxiliary information that would accommodate the process of re-identification. A thorough discussion of data anonymity and fine-granularity will be provided in Section 1.7.

1.4. Context Awareness

"The laws that cover privacy of medical information vary by situation. And, confidentiality is likely to be lost in return for insurance coverage, an employment opportunity, your application for a government benefit, or an investigation of health and safety at your work site. In short, you may have a false sense of security " [22]. The term of *context-awareness* is defined by the conjunction of one or more of the

following elements: content or type of information accessed, access time, patient's location, patient's condition, and authorized user's location.

On one hand, the elements of context-aware access control highlight privacy concerns for users and patients. For instance, specifying the location of patient or user at a certain time and storing them in a log file can violate their privacy. Another issue is the security of the logging information, where it can be saved and what security should be provided. Another concern is information content management, where data has to be classified based on its sensitivity. On the other hand, applying a context-aware access control policy based on data content, patients' practitioner auxiliary information dynamically forms a high overhead regarding maintenance cost.

It is found in the literature a study trying to solve the problem of context-awareness while considering contextual parameters. An attempt to build RBAC on top of DAC is proposed by [58] that addresses identification, authentication, authorization, and access control. Another example found in [8] that provides a version of RBAC under time constraints introducing the TRBAC. The GTRBAC [53] provides a method to grant or revoke access of a role based on temporal parameters. Where [9] proposed, a context aware access control will limit access to a specific geographic location as a service of GEO-RBAC model. However, some proposals that merges RBAC and other access control policies, in part, are not considering the healthcare sector in particular.

1.5. Identity Theft and Identity Protection

A Booz Allen Hamilton report [43] defined the medical identity theft as "the misuse of an individual's personally identifiable information (PII) such as name, date of birth, social security number (SSN), or insurance policy number to obtain or bill for medical services or medical goods. Medical identity theft may occur with or without

the identified individual's consent or knowledge."

According to [96], a report in 2003 showed the ability to use de-identified prescription records to uniquely identify 2.3% of individuals. This ration increase to 6.1% where two individuals can be identified by the same set of information. The information used in the study in nine states are:

1. Drug, dose, and refill information.
2. Patient diagnosis
3. Patient ZIP code (derived from pharmacy ZIP code).
4. Prescription fill date.

The level of accuracy in re-identifying patients increased when data is not segmented and more than one record found for identified patients. Unsegmented information with no proper access control policy increased the threat of re-identifying patients. According to [43] , medical identity theft could result in:

1. Decreased accuracy in medical records.
2. Risking patient health and compromising care because of a possible inaccurate health records.
3. Increase in th cost of healthcare on patient and healthcare system.

Multiple sources [1, 31] reported incidents where patient identities had been misused or stolen by insiders. For example, an admissions employee in New York Presbyterian Medical Center stole patients' names, phone numbers, and some social security numbers and sold them to identity theft groups. The process of limiting or avoiding medical identity theft goes through three stages; prevention , detection, and recovery.

Among the recommendations of [43] to prevent and detect identity theft is to develop a method which allows patients to access and control their own medical records. In our research, we focus in the prevention stage through enhanced access control policies.

1.6. Policy Exchanging

Issues of medical records privacy and protection seemed to be closely tied and related. For example, compliance with HIPAA or NIST standard and conducting a proper risk assessment assures, to a certain limit, the implementation of the mandatory security techniques. However, the level of compliance is not necessarily the same across the health care network. Since privacy and medical records should be preserved in the network, there should be a level of trust between communicating parties prior to PHI exchange.

Goldstein and Rein [41] discussed the importance of establishing a common platform for medical information exchange. The report describe the effect of having different *internal policies* on areas like exchange method (e.g information pull or push) , what types of information can or cannot be exchanged, and other legal concerns.

It is important to differentiate between *internal policies* and *patients' policy*. Internal policy and its compliance with standards such as HIPAA, applying its requirements guarantees the existence of the minimum conditions for practicing and operating. Internal policy can affect how data is used and its security, however, there are other aspects not covered by that policy such as:

1. The information shared.
2. The purpose of sharing.
3. Who can access the shared information.

4. The number of parties involved in exchanging medical or personal information.

5. Hospital, facility, or agency maturity in data exchanging and internal management.

The policy exchange problem cannot be solved by implementing internal policy, or by the compliance with a standard such as HIPAA. The focus of standardization is to establish a common infrastructure to evaluate and operate healthcare providing facilities. Alternatively, applying patients' customized policy aside from hospital policy can contribute in solving such problems. It is important to address a wide spectrum of concerns in the provided solution for privacy protection and medical information security.

1.7. Fine-Granularity Access Controlling

An essential requirement for providing a flexible and manageable privacy-centric framework for a healthcare system is information granularity and data presentation. Data presentation, normalization, or categorization can be performed based on data logical value or content. However, classifying medical information is a complex problem. For example, medicine name can disclose the nature of a sickness or a health condition. Health information can be revealed by the disclosure of other auxiliary information such as the name of treating doctor or hospital name. The knowledge delivered by *auxiliary information* can be used to derive or infer more critical information about an individual. However, the principle of fine-granularity access control introduced new challenges such as :

1. Mutual access. At a certain time t Several parties , n, can acquire access to the same data segment mutually. However, at $t + 1$, only $n - 2$ parties should be granted access to that specific segment. This introduces the problem of

managing access control tables for many segments and many parties.

2. Key managements. The amount of keys generated to manage access rights will grow rapidly. Each user should maintain a large number of keys to be able to access a set of files.

3. Key distribution. In the case of revoking the right of access from a user, a new set of keys need to be generated and redistributed to all authorized users.Also, the old set of keys has to be invalidated to prevent unauthorized accesses.

4. Encryption. Segmented data has to be re-encrypted whenever access is granted or revoked from users. This process significantly increases the overhead on the EMR system in data encryption.

5. Scalability vs. complexity. Fine-granularity access control methods can work efficiently if the number of users and the number of files are relatively small. However, when the problem size is scaled up. The increase in number of keys will become logarithmic. Considering the size of medical information files which can include images and videos, the increase in number of encryption and decryption operations will become high as well as the number of I/O operations.

The idea of fine granularity access control is not new in the medical field. Some research [10] suggested that the granularity at the levels of medical and administrative data, the category of the medical record, and who is requesting access. However, other researches [112] shows the difficulty of implementing fine-granularity access control policy while maintaining system efficiency and data availability .

1.8. Challenges and Research Objectives

Development in health IT is driven by the need to lower the cost of provided health care services, preventing medical errors, improving service quality, and expanding access to affordable care. Also, health IT participates in the early detection of infectious outbreaks around the country, improving the ability to track chronic diseases. The de-identified collection of information from the medical field for price verses quality provides a value-enabled measurement of development in that field. However, those goals face obstacles and challenges technically, financially, and encounter resistance from providers. This section is covering some of those challenges and explains our research objectives.

1.8.1. Challenges

Based on the American Hospital Association annual survey in 2015 [103], the total number of registered hospitals is 5,686 with total expenses for all U.S. registered hospitals $859,419,233,000. The total number of admissions is 35,416,020 with total capacity of all registered hospitals 914,513 beds. The provided numbers shows a very large industry with a high level of heterogeneity revealing many challenges. However, we will be narrowing the problem size to what our area of focused interest it.

1.8.1.1. Environment

Electronic medical records security is a concern because of the context surrounding them and the way in which they are used. The characteristics of medical field environment include, but are not limited to the following:

1. Access reason. Electronic medical or health records can be accessed by different authorized parties in the system based on least need-to-know principle . However, there are no guarantees that the rule will be respected, and information

may be reached without any need.

2. Distributed databases: Data can be found in different locations and formats, such as pharmacies, laboratories, clinics, hospitals, and insurance companies.

3. Heterogeneous infrastructure. The problem of heterogeneity on implementation can be found within the same facility or hospital. One hospital can run many systems(e.g. pharmacy, X-Ray, filing) from different vendors and each system treats data differently. However, this problem becomes more complex when scaled to a network of hospitals with different vendors, policies and implementations.

4. Open-Loop system. In a healthcare network, there is no sufficient feedback provided to the patients if data abused or mistakenly disclosed.

5. Dynamic trust model. The trust level among entities in a health care network varies based on time, standards compliance level, and events. For example, if data compromised in a trusted hospital, its trust association with other entities should change based one the loss severity.

1.8.1.2. Technical

In the light of the subsection 1.8.1.1 describing the environment, other technical problems affects the security of health records such as:

1. Legacy systems. The way old systems built versus the new systems that rely on web services make it difficult to implement standards and security procedure without too many changes in both sides. This present a very high cost on providers, preventing them from adopting new technologies for higher costs.

2. Access control:

(a) Content awareness: In the health care system, access to data is granted based on roles and functionality on object (report or file) regardless data type or category. For example, RBAC grants access on reports to some roles (e.g. discharge report to accountant) concern for the security level of the data.

(b) Context awareness: Some data need to be accessed from a specific location and in a specific time interval. For example, some data should not appear during a regular checkup but it is necessary when the same patient is in an emergency room. Similarly, data should not be viewed after discharged from hospital.

(c) Applying patient's security preferences. Another challenge is to apply the patient's preferences on who will access information and how without interfering with hospital policies. Applying data owner's security roles should not prevent healthcare providers from performing their patient care duties.

(d) Information presentation. To be able to apply fine-granularity and selective access controlling policies, data should be presented in a way that allows it. Data must segmentation and classified to establish the foundation of applying multiple policies simultaneously for further control.

3. Encryption and emergency access. The problems of securing electronic health records, encryption, emergency access, and key distribution highly effects the proposed solutions and implementation.

4. Information sharing. Information sharing between medical facilities helps to lower the cost of services provided, however, several factors must be considered. (a) What data can or cannot be exchanged? (b) On what criteria can the decision be made? (c) What possible risks will information sharing introduce

and how can they be minimized?

5. Document presentation: In addition to the information content, and presentation, it is important to determine what criteria will be used to format data prior to the exchange. A unified formatting (e.g. HL7 messaging system) should be negotiated and agreed upon to guarantee accurate interpretation.

1.8.2. Research Objectives

The previous section presented to the problems facing electronic health and privacy protection, showing the wide guidelines for solution development. Relying on the conclusions driven by the latter discussion, the objective of this research is to provide a framework for privacy protection satisfying the following areas:

1.8.2.1. Information Presentation and Categorization

Classifying each single file content into categories, tagged segments, and related contents lays down the foundation for fine-granularity access control and protecting information privacy. However, enforcing new standards will require additional time and resources that participating entities and hospitals are hesitant to invest. Utilizing what is already implemented as a global standard, such as $HL7$ categorization and tagging, , does not require many changes on existing systems. What remains, then, is to provide a ranking system to classify information based on its sensitivity and importance. Our goal is to match HL7 headers with HIPAA standards as a minimum level of information ranking. Moreover, patients can further improve the level of accuracy by specifying their own preferences in addition to standards implemented previously.

1.8.2.2. Policy Exchange and Trust Negotiation

To be able to conduct business, each covered entity (hospital) in a federated health care network should comply with standards, such as HIPAA, to verify it *readiness*. However, hospital compliance level varies based on many factors, including: entity size, type of services provided, and the methods of implementing standards and requirements. This variation labels hospitals with security levels and business priorities. This means that disclosing the way policies and standards are implemented and the

amount of investment in that area can harm the business itself.

To engage in a trust relationship with another party, it is important to exchange standards compliance level, security provided, and internal policy implementation. A reasonable information sharing and policy exchange protocol should respect the following points to prevent leaking internal business values:

1. Not revealing the level of implementing standard and its value to the entity.
2. Keep entity's internal policy secrecy.
3. Provide a truthful information about that entity.
4. Match different policies and areas of coverage.

After the necessary information is exchanged, such as policy and standards, the new knowledge should be evaluated to determine how trustworthy the other party is. Quantifying exchanged policies to create a numeric representation of trust association is not a one-to-one matching process, as well be described in the incoming sections. In a complex structure of trust relationships, immediate association can be affected by transitive trust relations with other entities too and it will play a role in the newly established trust relations and modifying old trust levels. Figure 1.2 illustrates for a directed trust association between hospitals where it is common that $v_i \neq v_j$ for any two hospitals where $i \neq j$.

The related components of the proposed solution consists of a special data structures to exchange policy analysis, negotiate trust levels, and transitive trust negotiations. A negotiation protocol is needed to state the rules of building trust association between entities.

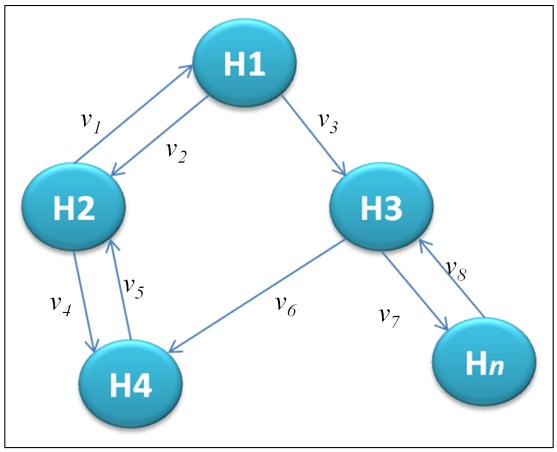


Figure 1.2. Directed trust associations example between hospitals

1.8.2.3. EHR Exchange

Exchanging private anonymized or complete electronic health records is a big challenge for health care providers. There are many obstacles to information sharing such as:

1. Information ownership and copy rights. The problem of who owns the information and who pays for it, who created the information and has the right to reuse it, information accuracy, and legal consequences of acting upon that information can prevent some entities from sharing patients informations. For example, can a doctor claim a specific diagnosis or case for himself as a copyrighted material? Why should a hospital pay the cost of keeping medical records secure if not allowed to use them or exchange them?

2. Lack of standardization. It is hard to find a universal recognized standard for formatting information, exchange protocols, and messaging systems that are compatible with both old and new systems. However, one major factor of preventing covered entities from updating their systems are time needed, high

cost, and backward compatibility.

3. Fear of privacy leak. The willingness to exchange information decreases when considering security factors and information privacy. There is no guarantee of how a securely transmitted file will be used by the receiving entity.

Such factors are holding back the progress of information sharing and deprive patients from getting improved level of health care service.

1.8.2.4. Selective Access Control

A common access control policy used in the health care sector is role-based access control (RBAC), where users are granted access based on their roles in the system. One advantage of RBAC is the ease of assigning privileges and revoking them based on business needs. However, it does not go to the level of controlling access based on data content or based on users as individual or security clearances. The cost of applying a multi-level access control policy on top of RBAC will add the additional responsibility of administration and user management if done by hospitals. Also, RBAC does not allow a direct relation between users and objects. It is more to *operations*, *role* assignment.

Having a mixture of security levels within the same object or file makes it difficult to control what to reveal and what to protect. Many services can be provided if there is a system allowing some sort of *selective access control policy* that can run along with RBAC, providing the ability to link users to object, and to be *content-aware*. However, the solution provided should reduce the need for administration, key (re)distribution, and be able to overcome the problem of user-object separation in RBAC without interfering with RBAC work.

1.9. Research Overview and Thesis Organization

The field selected for implementing the framework is the health sector where 50% of attacks come from insiders. The high complexity of the health care system requires more sophisticated solutions than simple or direct access control policies. The solution provided decomposes the privacy protection problem into sub-areas, including the current security metrics and procedures implemented, and finding common factors for an optimized solution.

The proposed framework provides a comprehensive set of tools to encounter attacks such as honest-but-curious. The framework covers the scope of environment preparation through risk assessment and checking for compliance level with standards. The next stage is building a trust network between hospitals by defining trust associations. Hospitals can use the provided policy exchange protocol, also provided by the framework, to build trust association. Data exchanged in the trusted network can be filtered based on patient privacy settings and by an organization's policy of exchanging protected data.

In the level of local privacy protection effort, the framework utilizes data segmentation and categorization provided by standards like HL7 to enable multi-level security access control. Each segment or segment category can be assigned different roles of access rights and it operates at a user level rather than role level. The provided fine-granularity access control policy enables data owners to embed their own preferences without interfering with the existing system. This structure lowers the cost of policy implementation since it is running independent from RBAC.

The research is divided into nine chapters, the first chapter introducing the work and the area of research. The second chapter surveys the work and the previous research on the field, where the third chapter provides the problem formulation and areas of interest. Chapters four, five, six, seven, and eight describe the solution frame-

work. Chapter four describes the risk assessment process and compliance policies. Chapter five presents the policy exchange process and how trust tokens are produced and evaluated where chapter six shows segmentation policies and the preparation for fine-granularity access controlling policy. It describes the concept of decoupling privacy and security and the concept of segment gateways. Chapter seven discusses compound key structure and minimizing key distribution process when access rights are modified dynamically. Finally, chapter nine shows the implementation for the framework components and demonstrates how a fine-granularity access control policy can run with RBAC without major changes.

Chapter 2

LITERATURE REVIEW

In 2001-2002 , the former Chief Executive of the NatWest financial group, Sir Derek Wanless, undertook an independent review of the UK National Health Service (NHS). Sir Derek said "the NHS is much more complex than any business I have involved in, not just banking. The reason is the multiplicity of outcome the NHS can have. In most businesses there is the bottom line, most things translated into cash. But in health there are a whole series of outcomes - there are almost as many as there are patients." He also commented "I started off being critical of the IT systems in the NHS but as I went on I gained an appreciation of how difficult it is to get it right. This was difficult enough in banking but nothing like as hard as in the NHS. Even areas such as patient records, which appear to be a basic requirement for the health service, raise complex issues for IT designers".

Piero Bonatti [66] argues that the new open scenarios require scalable authentication techniques to replace the old methods, and suitable for the dynamic nature of virtual organizations. Such organizations combine hospitals, clinics, and outpatient services in a single entity. All previous needs are the same needs for other systems but are different in approach. Solutions should be revolutionary rather than evolutionary. The special nature of the Electronic Medical Records justifies the increased attention and the need for a real time fraud detection system, affordable that has more than role based access controlling policy, is scalable, supports authentication, allows secure data sharing, and resolve the data integrity problem.

Amalia R. and Catherine E. [67] reported that in 2006, John D. Macaulay , the vice president of Health Care and Life Science, talked in his speech before the American Health Information Community about the future of the online health systems, EMRs, and how they are different from the financial systems. According to Macaulay, there are more than 500 phishing and other attacks *per day*. The fact that banks are seeking to adopt a second factor of authentication shows how password and username are not enough to protect users (The Federal Financial Institutions Examination Council, FFIEC Guidance). The financial systems allowed difficult risk assessment plans to tolerate a certain amount of fraud, and balance the cost of security measurement against the risk of attack. In contrast, the level of risk is very different for patients and practitioners .

2.1. The Health Insurance Portability and Accountability Act (HIPAA)

The goal of the Health Insurance Portability and Accountability Act of 1996 is to protect protected sensitive health information, how it is used and to whom it is disclosed. However, the level of compliance and the technologies used to apply standards cause a form of incompatibility among EMR systems in healthcare network. Each document issued by HIPAA specifies required security standards in a specific area, and also specifies the possible risks and risk management strategies. For example, in HIPAA security guidance for remote use of the electronic protected health information (ePHI), the use of dual-factor authentication is proposed to mitigate log-on/password losses. This second factor could be using a security question (e.g. "Favorite fruit?"). Hence implementing HIPAA standards and risk management strategies varies from one hospital to another, such that the level of protection provided to transmitted data will not be the same in the hospitals network.

While HIPAA standards attempt to cover many areas in ePHI security and privacy topics, areas like integrity and availability standards, data auditing standards did not cover information recovery if needed. If a piece of information intentionally destroyed, the ePHI owner cannot determine the value of the destroyed information if the proper mechanisms are not provided to regenerate the lost information. Summarizing HIPAA requirements into security matrices will simplify policy bridging between health care providers. Security Standards Technical safeguards [48] specifies a matrix of general security standards a covered entity must comply with according to HIPAA regulations. Table 1.1 shows that HIPAA standards suffer from some shortcomings that affect information security and patient privacy. It deserves mentioning that security rule does not specify a certain technological solutions to give more flexibility in security measure selection.

By narrowing down the set of common parameters in the healthcare environment to match HIPAA required and addressable standards, it becomes applicable to find common platform for security negotiations. This goal can be achieved by even partial implementation of HIPAA standards by creating a schema of (technology, goal) and (practice, goal) for compliance check. In a single hospital, it is possible to identify actors (users),operations, roles, information (EHR), security safeguards (SSO, certificates, user name/password, etc...), physical access control (security guards and cameras, badge reader door locks, and so on), policies and rules (e.g. privacy agreement), and other system parameters to arrange and govern the facility operation. In its effort for standardization, HIPAA classified security needs for covered entities into the following categories [47]:

1. Security for remote use

2. Security standards administrative safeguards

(a) Security Management Process

(b) Assigned Security Responsibility

(c) Workforce Security

(d) Information Access Management

(e) Security Awareness and Training

(f) Security Incident Procedures

(g) Contingency Plan Evaluation

(h) Business Associate Contracts and Other Arrangements

3. Security standards organizational policy

4. Security standards physical safeguards

   (a) Facility Access Controls

   (b) Workstation Use

   (c) Workstation Security

   (d) Device and Media Controls

5. Security standards technical safeguards

   (a) Access Control

   (b) Audit Controls

   (c) Integrity

   (d) Person or Entity Authentication

   (e) Transmission Security

2.2. Anonymity

There is no real definition of how much information, from the non-identifiable set, can specify a person. According to El Emam at el. [29], 63% to 87% of the U.S. population are identifiable depending on some distinctive characteristics. However, a subset of auxiliary information, such as job, race, color, and sex narrows down the set of possibilities considerably. By adding one more detail like car type, it becomes more likely to identify a person. Having medical information kept secure by hospitals does not mean it is private [92]. Regardless, of whether a record is active or archived, disclosing anonymized records or granting access to them without restrictions can lead to serious consequences.

Amalia and Tucker [7] claimed that only 41% of the U.S. hospitals implemented a basic EMR system in 2005. According to the study, the adoption of EMR could save the U.S. about $34 billion through higher efficiency and safety. However, the state privacy regulations restricting information release reduced the EMR adoption by 24%. The violation of privacy law in California fined Kaiser Permanente $200K when one of its employees disclosed patients' medical information on a blog.

In a risk assessment study [97], Alexander Berler et al. categorized data in the health system into major categories such as medical data, personal data, patient management data, indexing data, administrative data, clinical protocols, financial/l-ogistics, system data, and MIS. The study classified the critical nature of information based on its importance and the impact of loss into catastrophic for medical data, where the rest scaled from critical to marginal and minor. In medical records, the immediate intervention, recovery process, and the consequences have a high impact on the implementation cost, required technologies, and all other quantitative requirements of the security system.

Privacy attacks on electronic medical records vary based on data context. An active attack can cause denial, termination, or prevention of service, whereas, a passive attack on off line private information can add to the previous harm the probability of identity theft, and disturbs the natural flow of life. However, active attacks can be initiated by authorized users from inside hospitals. Christian Stingl [92] mentioned that 50% of the attacks targeting the information systems are conducted by insiders.

In [29]. El Emam et al. provides a comparison between common heuristics to evaluate data importance of HIPAA-specified ID fragments shown in Table 2.1 and non-HIPAA specified IDs such as providers' information. Latanya Sweeney [95] provided an example for identifying William Weld, who was a governor of Massachusetts, after GIC Company released his medical records. By direct matching with the voters list, the set of data reduced to six people with the same birth date, half of them were men, and he was "the only one in his ZIP code."

Table 2.1: Initial quasi-identifiers

Number	HIPAA regulated Patient identifiers
1	Account numbers
2	Name(s) of relative(s)
3	Biometric identifiers
4	Names
5	Certificate/License numbers
6	Medical record number
7	Dates
8	Photographs and comparable images
9	Device identifiers

Continues on next page

Table 2.1 – Continued

10	Postal address
11	Email addresses
12	Social security number
13	Fax numbers
14	Telephone numbers
15	Health plan numbers
16	Vehicle identifiers including license plate numbers
17	IP address numbers
18	Web URL's
19	Any other unique identifying number, characteristic, or code

Several researchers tried to contribute to the field of patient anonymity through adding noise, extracting identifiable information, developing anonymizing software, duplication, and altering information [2, 85, 93–95]. however, no solution was able to accomplish real data anonymization.

2.3. Identity Protection

To insure confidentiality and authenticity of patients' electronic medical records, Hui-Mei et al [19] propose a patient-identity security mechanism, an identity cipher/decipher, and user-authentication protocol during the transit and at final destination. They related the cipher text to the patients' identifying data to the patient EMR. The use of public key infrastructure (PKI), certificates, and the dynamic cookies came in verification and identification of patients and grant access to the user.

Hui-Mei et al. used a symmetric-key algorithm like AES to generate a cipher text, hash function (e.g. SHA-1) to produce a message digest, and public key algorithm (e.g. RSA) for digital signature creation. The architecture is as follows: Healthcare Certificate Authority (HCA) manages the certificates and the public-private key pairs, the EMR database that stores the encrypted EMRs, the EMR authentication server, which is the connection point between the HCA, the Encrypted EMRs server, and the user terminal machine via the authentication protocol. The users utilize an IC token or keyboard to access their EMR.

The mechanism depends on using the patient medical information (PMI) , the patient SSN, and the patient identification data (PID), all processed by two functions, a logical-based function ,F, to encrypt/decrypt the information mentioned, a data hiding function, D, to hide the embedded password within the PID. A ciphertext name V, works as the identifier of the encrypted EMR, the I_k used as a key to decrypt the PID. Both V, I_k generated using D taking PID and the SSN as input. The PMD and a random number R ciphered using F, to prevent R from being disclosed, R encrypted using an administrator password.

They defined their own functions of data hiding and encryption depending on shifting and bitwise XOR logic operations. The test is done on HL7 EMR file. The authors claim that using EMR-related ciphertext as the identifier will prevent the

unauthorized users from using the information from the EMR to disclose the patient identity. However, the use of agent-EMR-related ciphertext allows healthcare providers, where the permission is based on their privilege assigned, to control the clinics access to the EMR. In some cases, out-of-network clinics request access to a specific EMR of a patient. The data custodian either rejects the request, to protect the patient privacy, or accepts the request based on some credentials the requester provides to establish a legitimate order.

Alfred C. Weaver et al. [108] propose a framework solution using Microsoft .Net to implement the system. In this proposal, an authentication web service manages trust levels, issues authorization tickets, and uses biometric devices to establish user identity. An authorization web service determines what data may be accessed, in what way and by whom. All records and images are encrypted using AES with 256 bit keys. The federated trust-sharing arrangements take place when off-network entities request access to the data. None of XML, Simple Object Access Protocol (SOAP), Web Service Description Language (WSDL), or the Universal Description, Discovery, and integration (UDDI) implements security.

The access level is classified based on the requested service. Hospital administrators define the trust level required for each service available. For example "access to patient records requires a trust level equal to or greater than that provided by fingerprints."Had the access device supported iris scans, that modality would also have been an acceptable option because the trust level of iris scans exceeds that of fingerprints. The protocol and the description define many points of authentication, authorization, encryption and decryption.

David K. Vawdrey et al. Vawdrey et al. [107] described an authentication and access control service based on trust negotiation, which enables two parties with no pre-existing knowledge to establish sensitive transactions via mutual disclosure of

attributes contained within digital credentials. Trust negotiation targets individuals from outside a local security domain granting them safe access to sensitive data and services. Sensitive resources such as services, data, credentials, and other policies are protected by access control policies from unauthorized access.

Attribute credentials, which are protected by public key cryptography, depends on access control policies as well. [107] describe the trust negotiations by passing a request from one party to another asking for information. In its turn, the provider will send his disclosure policy to the requisite. Then the primary care provider discloses EMR after encrypting it using a shared secret key. Identity is verified by comparing received credentials with those issued by a trusted third party like the national licensing association.

2.4. Access Control Policies MAC, DAC, and RBAC

In 1992, David Ferraiolo et al. [33] introduced a non-discretionary role-based access control policy (RBAC), which is more appropriate and centralized than the Discretionary Access Control (DAC) [27, 75]. RBAC provides a new access control policy that satisfies non-military systems by assigning rights to the roles rather than users. RBAC does not support data ownership in user level since user actions and rights are represented by the privileges granted to his or her role in the system. RBAC used as a base for other access control policies because of the benefits it provides such as easy implementation, low maintenance cost, and its flexibility in managing access rights based on users role in the system.

Bertino et al. [8] introduced the TRBAC, or Temporal RBAC, as an extension of RBAC to support periodic role enabling and disabling, which provides more control and compliance with HIPAA constraints. Junzhe Hu et al. [91] introduced further steps to authenticate and grant privileges depending on the context to comply with

the HIPPA regulations [46]. This approach, context-aware security, went further than (user, roles, permission, session) in RBAC to refine the authentication criteria. It includes time and place of access, data objects, and operations in addition to other criteria such as context constrains.

Schwartmann [87] discuses the issue emerging of role-base access control policies and attribute-based access policies, such as ‶certified Dr. A″treats ‶Patients X″rather than ‶the certified Dr. treats patients″. Where in the first case Dr. A is restricted to Patient X which will fulfill the least-rights principle. To overcome the shortcoming of RBAC system, An extension to the standard with attributable permission and roles is introduced. Classes of permissions are defined by attribute-definition associated to permission objects, thus, constraining role activation. An UML presentation of the Attribute-Definition and Attribute-Values show that Attribute-Definition holds a unique name and type within the authorization subsystem.

The concept of activation-context presented as the activation of a role is constrained by all attribute-definitions that is implicitly attach to it through permission assignment. To permit a doctor to be an attending physician of patients, say A, B and C, an activation-context is created for each patient. Activation of a role requires that the activation-context holds attribute-values for all attribute-definitions of that role.

Nishiki, K. and Tanaka, E. [71], propose a distributed system based on authentication and access control agents characterized by network federation, autonomic policy decision, and dynamic access control for context-aware services. The motivation for the work is the ubiquitous use of the computing environment and expanding service market abilities by establishing an identity based secure service platform. Thus, they proposed and implemented an authentication and access control agent framework for context-aware services.

A context-aware policy engine has a a security policy database where the access and authentication roles are kept. The policy engine executes actions according to the received changes in context. An authentication ticket, presented using SAML, can include context data, where the access ticket contains user attributes such as sex and age. Authentication and access agents are used to manage the system configuration. At runtime the client-side authentication module and the device control module are loaded simultaneously. Once the agent receives the access ticket, it will customize access accordingly, and issue the service ticket. Network domains are contacted through the authentication agents to create a federation association. Some issues still unsolved include network mobility support, support for wireless only LAN, lack of heterogeneous networks such as ad-hoc, trust relationship among agents assumed to be available, and where we should develop it. Such trust domains have been reported such as X-GTRBAC framework, and identified device interconnection where network resource virtualization should be developed.

Qingfeng He and Annie I. Anton [44] tackle the issue of access control policies and not considering them at the time of requirement analysis activities. The method integrates policy specification into the software development process, ensures consistency across software artifacts, and provides prescriptive guidance for how to specify the Access Control Policies (ACPs) and the Security and Privacy Requirement Analysis Tool, SPRAT, to support the method. As the authors claim, the two major issues in the access control system are (a) defining correct and complete policies to control users access to the system and its resources, and (b) ensuring the resulting policies comply with the system requirements and high-level security/privacy policies. The DAC access controlling policy allows the Trojan horse attack, and MAC prevents the information leak allowed by DAC. However, MAC is still vulnerable to covert channel attack. RBAC became an ANSI standard and used in commercial products like

Oracle 9i.

ReCAPS is designed to derive (ACPs) from security and privacy policies since the ACPs come from requirements as well as high-level security and privacy policies. ReCAPS focuses on ensuring compliance between ACPs, requirements, and design. ReCAPS integrates policy specification with requirements analysis and software design. There are two main kinds of activities in ReCAPS: analysis and refinement. High-level security and privacy policies should be specified as system requirements. Mid-level policies are instances of high-level policies within a specific systems context. To specify these policies, one must examine system requirements to identify users and their interactions with the system, and system designs (e.g. database design) to identify the data to be protected. Focusing on mid-level policies offers two major advantages.

In [69], Oveeyen Mooniana et al. compare the Role-Based Access Con- trol techniques like Role Based Access Control (RBAC), Dynamic RBAC (DRBAC), Context Based Access Control (CBAC), Proximity-Based Access Control (PBAC), Team-Based Access Control (TBAC), and Task-Based Access Control (TMAC) in context awareness, dynamicity, flexibility, scalability, Centralized vs. distributed access, and user mobility, reliability and performance. The study proposed the HCRBAC as an access control system for collaborative context-aware healthcare services in Mauritius Republic to provide accessibility based on context and role, where personnel from police departments or hospitals are allowed to access the records without compromising data integrity or confidentiality.

However, HCRBAC should gather rigorous authentication, and ease of access based on context. Biometrics and active sensors are used to determine the level of accessibility, where the information is passed to an authentication system. If authenticated, the user identification information is filtered by the access-control engine

based on this context. To get the context information regarding time and role, a database of work schedules and roles is used. Information sharing is based on trust relations between service providers, where each entity passes the required authentication information regarding its users who need for access. The work is a combination of RBAC, CBAC, and PBAC. Table 2.2 shows a comparison between RBAC, MAC and DAC in the area of focus.

Table 2.2: MAC, DAC and RBAC focus areas

Standard	Description
MAC	restricting access to objects based on the sensitivity (label)
	Cares about the formal authorization of the individual
	Each user has to have a rank (Privacy clearance) $(User \leftarrow Privacy\,level \rightarrow Resource)$
DAC	Fits Ownership: Data owner can grant access without system admin
	Single level system (no hierarchy systems/subject) organization
	Management{Cost, Risk}: owners (patients) has to knows each and every detail, increase risk of privacy loss
RBAC	Users "subjects" cannot pass access permission like DAC
	Not based on multilevel security requirements
	Does not support individual control over data
	Concerns about $\{Function, Role\}$ of subjects/users rather than who the subject/user is

Other factors, such as scalability, ability to share information,individuals control over their information, fine-granularity access control, and context awareness are very important in deciding the appropriate access control policy to implement. Table 2.3 provides a comparison between DAC and non-DAC access control policies in those areas. Discretionary Access Control (DAC) has major draw backs mentioned in [51]

such as:

1. No flow control on data allowing it to be copied from one object to another.

2. The owner is the one who decides access privileges, not the overall system policy. The organization policies not applicable in such cases.

3. After transmission, there are no limitations on how the receiver will use the received data.

Table 2.3: MAC, DAC and RBAC properties

Name	Scalable	Share	Individual Control	Fine-Granularity	Context-Aware
1- DAC	No	Yes	Yes	No	No
2- NDAC					
2.1- MAC	Yes	No*	No	No	No
2.2-RBAC	Yes	Yes*	No	Yes **	No
2.3- Temporal					
2.3.1 *WFMS*	No*	No	No	N/A	No
2.3.1 *Chinese wall*	Yes	Yes	No	N/A	N/A

Some models or implementations developed for DAC are Take-Grant Model Lampson in year 1974, Graham-Denning in 1972, Harrison-Rizzo-Ullman in 1976, Griffiths-Wade 1976, and Originator Control 1989. In general, DAC suffers from transitive trust (e.g. read operations) and being vulnerable to Trojan horses. The non-DAC policies such as MAC are in some cases vulnerable to covert channels attacks. Another drawback is the fact that is designed to implement enterprise policies, application specific, and not flexible enough to implement users or patients preferences. The Mandatory Access Control (MAC) model which introduces an indirection between the users and

the resources through security labels. A user can see the information if he/she has the proper security clearance assigned to the information, where the subject or patient does not have any control over it. An example models of multi-level security is Bell-La Padula confidentiality which demonstrates how data ownership is implemented.

2.5. Context-Awareness

Jakob E. Bardram et al. [5] describe the concept of Proximity-Based User Authentication protocol, as a usability-wise ideal for UbiComp systems. The process depends on using the smart cards with Java capability (JavaCard) for identification and cryptographic calculations, using a context-awareness system to confirm the user location, and implement a security fall-back strategy. . The authors consider the problem of logging into many computers during the day and if all users are logging out properly. If not, they leave the system open, which is a security hall through usability since working in more than one location with the same patient is a common practice in hospitals.

The idea of proximity-based login depends on authenticating the user to login the computer while physically approaching it. If the healthcare providers adopt the Active- Based Computing or (ABC), the task a nurse or a doctor is involved in can follow them wherever they move. The Proximity-Based authentication can provide an easy login-logout mechanism for them. For the context-awareness and verifying the user location and physical presence close to the system and carryout the authentication process, smart cards and RFID technology can provide that information. However, the need for a fall-back mechanism still high. The proposed authentication process depends on public key-private key, shared key, and Nonce to verify the user identity.

Dekker [23] and Ferraiolo et al. [34] asserted that the distributed nature of the e-health systems, the increase of access policies, and roles lead to provide a distributed administration model. As distributed databases, and the web-based application became a common trend for development, the need to build trust [110] and enhance the password authentication [111] became a field of interest to provide security management for health systems. Another important aspect in privacy is the accidental or intentional information disclosure where a single click can reveal a large number of records..

It is worthy to mention that privacy rules set the standard for who is allowed to have access to PHI, which may be in electronic, oral, or paper form. Whereas, the security rules focus on EPHI, the electronic Protected Health Information, and sets the standards to insure that only the authorized people ill get access to the EPHI [78]. By not dictating the technology used in protecting the EPHI, HIPAA rules provide the required flexibility for each entity to select the technology within its abilities that comply with the standards.

The context of accessing the PHR dictates the nature of the safeguard implementations. While the records are in use, these may include username-password, security questions, auditing, logging, role-based access control, temporal access control, context-aware access control, and temporal context-aware access control. Another technique is used when records need to be disclosed for research or by a third party to maintain anonymity.

2.6. The Obstacles to Health IT Adoptions

As with other industries, the health sector faces many obstacles in the process of adopting an EHR system. We are summarizing the major difficulties as follows [7, 18, 70].

1. Asymmetry of costs and benefits associated with EHRs, a survey in 2006 found that 94% of hospitals consider the initial cost of the EHR as a significant barrier. The total EHR system for nationwide cost has been estimated to exceed $100 billion [99] .

2. Backward compatibility and old MR, the associated cost with modifying the old system or moving to the electronic systems, and the fear of losing information forms another obstacle for EHR.

3. Lack of national interoperability standards and information sharing. In 2004, President Bush issued an executive order for applying a national interoperable health IT network, EHR for all Americans in a 10 year period. Steps in that direction are still being made but are not sufficient to achieve the goal.

4. Public fear of errors, misuse, lack of control of their personal information are the result of lack of education and poor judgment by data custodians and users.

5. Resistance to new technology is a common problem in computerization and digitizing systems.

6. State privacy regulations, on one hand, state privacy regulations impose additional costs on hospitals to provide more protection for the EHR. While on the other hand, the service provider might obligate to financial loss for accidental discloser of information.

7. State privacy regulations, on one hand, impose additional costs on hospitals to provide more protection for the EHR. On the other hand, the service provider

might incur significant financial loss for accidental disclosure of information. Approximately 25% of U.S. adults believe the Privacy and Security of Personal Health Information is significantly diminished with the move to EHR. Many valid concerns such as data availability in electronic system, information integrity, confidentiality, freedom of data controlling (access controlling) by data owners and hospitals, and reporting auditing services also need to be addressed.

Narenda [70] provided some surveys results from 2005 indicating that 79% of Americans still consider privacy and security as their major concern. Most e-health projects only consider privacy and confidentiality at the end of the project when it is extremely difficult and costly to address [57, 70]. In 2006, [40] pointed that 73% in New Zealand concerned about the privacy and security of their medical information.

2.7. Privacy Violation Results

The results of a privacy breach are not limited to violation of the law. It can go further than that if the information are misused. Some of the major areas where both patient and provider could be affected if their privacy breached are listed here.

1. Discrimination based on health information
2. Job loss, or decreases the ability to obtain a job
3. Increase in insurance rates
4. Subornation[1], by forcing patients or providers to do things against law.
5. Crimes, since an attacker can choose his victim based on facts from the medical record he has with confident to win.

[1]To induce (a person, especially a witness) to give false testimony. To bribe or induce (someone) unlawfully or secretly to perform some misdeed or to commit a crime.

6. Identity theft, both patients and providers, such as doctors, are subject to this threat. The Federal Trade Commission reported that identity theft is the fastest growing crime in the U.S [26] .

7. According to U.S Federal Trade Commission, privacy breached can cause devastating and irreversible effects on patients subject to the attack, financially, socially, or health wise.

The impact of the attack may not be bounded by specific time. It could last from the time of the initial breach and lead to more breaches, creating a sequence of violations. For instance, the subornation could lead to further crimes or force people to provide a cover for more violations, as well as identity theft. The latter discussion shows the special nature of the Electronic Health Records, the difficulties facing the adoption of EHR, and what the privacy violation effects are on both patients and providers.

In the next section we will discuss, in general, from where should we start, and what are the available solutions and the best practices. Table 2.4 shows who are included in the HIPAA rules of privacy. The rule applies only to those who are classified as covered entities players or actors, described as a health care provider, a health plan, and a health clearinghouse.

Table 2.4: Covered entities in HIPPA

Covered Entities	Description
Health Care Provider	Doctors, Clinics, Psychologists, Dentists, Chiropractors, Nursing Homes, Pharmacies: only if they transmit any information in an electronic form in connection with a transaction for which HHS has adopted a standard.
Health Plan	Health insurance companies, HMOs company health plans Government programs that pay for health care, such as Medicare, Medicaid, and the military and veterans health care programs
Health Care Clearinghouse	This includes entities that process nonstandard health information they receive from another entity into a standard (i.e., standard electronic format or data content), or vice versa.

2.8. Health Level Seven International Overview

Health Level Seven International (HL7) was founded in 1987 as nonprofit standard development organization. It provides standards and frameworks for the health sector for information exchange, sharing, and integration. Typically hospitals and healthcare providing facilities have different computer systems for health records, billing, administration. When communicating, heterogeneous systems use an interface or dialog protocol to exchange information. HL7 attempts to provide the methods and standard required for hospitals to interact. Some of what HL7 provides are the document standards (e.g. HL7 CDA) application standards(e.g HL7 CCOW), conceptual standards(e.g. RIM) [68] , and messaging standards(e.g. HL7 v2.x, v3.0) [52]. Among all HL7 standards, message standards hold a special importance because they specify how information is ordered and exchanged between participating parties.

Health Level 7 standard gains wide attention because of the way it define health environment parameters and objects, the thing that attracted many researchers to studying and evaluating it [3, 32, 62, 90]. Subsequently, Bhatti et al [10]proposed a context-aware policy-based approach for federal healthcare databases (RHIOs), elaborating RBAC in policy-based systems, and HL7 standards. Bhatti attempts to address the requirements for security management in RHIOs from the perspective of database system principles.

In our work, we focus on how data is stored, categorized, segmented, and the relation between information spanning multiple segments. Using HL7 information classification, placement policy, and segmentation, it is highly possible to get preclassified information based on its type which is reflected by its security level. Table 2.5 shows an example of data classification and segmentation in HL7. The area of data segmen- tation and information ranking still requires more specific study for the healthcare systems. However, HL7 still provides adequate level of classification to

establish EMR privacy.

Table 2.5: Example of HL7 data segments headers

Tag	Description	Tag	Description
ACC	Accident	ADD	Addendum
AFF	Professional Affiliation	AIG	Appointment Information - General Resource
AIP	Appointment Information - Personnel Resource	AIS	Appointment Information
APR	Appointment Preferences	ARQ	Appointment Request
AUT	Authorization Information	BHS	Batch Header
BLG	Billing	BPO	Blood product order
BPX	Blood product dispense status	BTS	Blood Product Transfusion/Disposition
CDM	Charge Description Master	CER	Certificate Detail
CM1	Clinical Study Phase Master	CM2	Clinical Study Schedule Master
CON	Consent Segment	CSP	Clinical Study Phase
CTD	Contact Data	CTI	Clinical Trial Identification
DB1	Disability	DG1	Diagnosis
PD1	Patient Additional Demographic	PDA	Patient Death and Autopsy

Chapter 3

PROBLEM FORMULATION

3.1. Introduction

The Honest-But-Curious (HBC) adversary model represents a high risk to confidential information and organizations' internal security. HBC attacks are initiated by trusted and legitimate users trying to obtain information out of curiosity, not to provide a legitimate service or complete an authorized task. However, HBC attacks do not trigger any type of alarm or action since the attackers are legitimate users with proper access rights.

The cost and effort of modifying or installing new security systems to face HBC attacks are incomprehensible based on the pervasiveness of the changes needed. Encountering internal security threats and attacks against confidential information requires more than traditional access control policies such as MAC, DAC, RBAC, or MLS. Moreover, developing new policies to accommodate more settings and actors requires developing new adaptable techniques to replace old techniques such as ACL, capability lists, access control matrix, and grouping.

This research provides a framework to solve the problems of policy exchanging, ranking remote covered entity compliance level with standards, creating a trust association between negotiating parties, applying patient's privacy roles, and providing a selective access control policy one-time key distribution. The approach can be implemented in environments other than the medical field and healthcare information.

3.2. Problem Definition

In the medical field, several types of information with a various levels of security may be present in different formats and locations. Electronic medical records might contain health information, financial records, and demographic data. However, all data categories are reduced to the general concept of Electronic Health Records (EHR). This collection of information can be anonymized through extracting the identification data, so that a specific person cannot be pinpointed. However, claims that anonymized information will not reveal the identity of a subject in them have been invalidated by several studies [2, 85, 93–95].

3.2.1. Problem Statement

Assuming a network of N hospitals, where any hospital $H_x \in N$ in the set is a covered entity that is required to comply with HIPAA standards. A member hospital, such as H_x, has the ability to interact and exchange patient information with any other hospital H_i as well as other hospitals within the network. The covered entity H_i is assumed to host an electronic health record F for a patient such as P. The set Z represents all employees working at hospital H_i such that $U \subset Z$ is a subset of the employees assigned to provide healthcare services to patient P. However, it is possible that there is an attacker U_w who is a member of Z but not U.

The risk factor that an HBC attacker, such as U_w, presents on privacy is given by the function $W(S_i)$ where S_i is the part or the segment of F attracting U_w the most, and has a high value to the patient or public. The HBC attacker U_w is a legitimate user with the proper access rights who might belong to U, $U_w \in U$. Each user, including U_w, has been assigned access rights based on RBAC access control policy. In RBAC all users within the same category have the same access rights and can perform the same functionalities.

Within the data structure of electronic health records in the EMR system, each file follows the HL7 standard with a proper XML structure. The EMR file consists of segments. Each segment has a header, attributes and content. Segments with matching headers contain the same *type* of information. For example $< PID >$ denotes a segments with patient identification data. HL7 structure makes it easy to exchange medical records in the healthcare network since it follows a well-defined XML standard.

Access control policy RBAC does not limit data availability based on file content or based on levels of security presented in the file. Moreover, implementing multi-level access control policies over RBAC requires developing another layer of security to deal with individuals rather than roles as in RBAC. The HBC adversary model represents a high risk when it is not possible to exclude or grant access to individuals selectively. Another problem is protecting EHR/EMR information when transmitted to a remote system. Not all EMR systems apply the same roles and settings for access control. Currently, it is not possible to control files and who can see what based on patient's preferences and privacy roles.

This research aims to provide the ability to negotiate trust relationships and create security associations based on compliance levels with standards like HIPAA. Moreover, it will provide the ability to filter information transferred on the network between entities depending on the security level each association has. Another service is the ability to implement patient's privacy and security roles a long with hospital roles. Services to protect privacy are provided through developing a selective access control policy that allows data owners and hospitals from applying different policies cryptographically with a minimum or no keys redistribution.

3.2.1.1. Honest But Curious Adversary Model

The problem statement presented HBC attacks as an insider attack started as a passive attack. The goal of that attack initially is to learn information out of curiosity without premeditation to disclose it. The information learned by the attacker is outside the scope of that needed information to perform a job or provide healthcare service to the victim. The attacker is a legitimate user and allowed to access resources, however this state creates two challenges. The first challenge is how to locate the attacker before performing the attack. Secondly, determine how to locate the targeted information or the hot region or feature attracting HBC users to perform the attack.

To formally define who the attacker is, a logical sequence of events has been driven towards attack completion. The HBC attacker can be viewed like a *hibernating threat*, which means that they will not start the attack unless triggered by a motivation. In a set of legitimate users Z there is a subset U such that a member of that subset like U_w creates a risk factor W when he/she accesses a piece of information such as S_w.

$$W(S_i) \propto U_w \colon \forall S_w \neq \phi \tag{3.1}$$

$$U_w \subset Z, U_w \cap U\, can\, = \phi \colon U\, assigned\, users\, to\, patient\, P\, from\, set\, Z \tag{3.2}$$

As Equation 3.1 explains, the risk factor W can be minimized by minimizing the presence of the motivating factor S_w. However, it is not possible to limit the risk by reducing the factor of U_w because it is a human factor and unknown to us as explained in Equation 3.2.

3.2.1.2. Policy Exchange

In a heterogeneous network consisting of different implementations, multiple standards (e.g. HIPAA , NIST SP 800, COBIT), variety of policies and security metrics , with a diverse business goals, it is very hard to establish one coherent system that

fits all. However, this diversity can be distilled to a common factor and relationship based on the following factors:

1. Target: Who is the standard or policy is targeting within the health care system or network.

2. Goal: What is the desired achievements and result of implementing policy and standard.

3. Coverage: What areas of interest or aspects policies should cover (i.e. privacy protection, compliance check, education and other areas of interest).

4. Criticality or Importance: Each policy has critical points that must be fulfilled for the implementation to be adequate.

As it is known, disclosing management policies, internal business values, and compliance with standards can put business existence and goals at risk. However, that hidden information is critical to know whether or not sensitive information can be exchanged between two parties. By relying on the previous points of categorization, it is possible to define circles of policy coverage and overlapping areas without revealing business secrets. Moreover, it becomes possible to exchange compliance results and level of implementation based on {Target, Goal, Coverage} and create a policy ranking system. This method of evaluating entities or hospitals compliance with standards can lay a foundation for exchanging private information while retaining a considerable level of individuality.

3.2.1.3. Trust Negotiation

Proper protocols and data structures need to be developed to enable two or more communicating entities to establish a trust relationship, or a trust association, in

a health care network. A protocol is required to build trust relationships, exchange information, and control communication between two entities in a health care network with no previous knowledge of one another.

A trust association is a term introduced by the framework as a two-way relationship between hospitals. The Trust Association is quantified by evaluating the opposite side's degree of compliance with privacy and security standards assumed by law. The degree of compliance in this research is called a **trust level** and is expressed as a percentage. Trust level is a negotiable value and can be adjusted. Two data structures, Trust token (Tt) and Trust association token (Ta) in Section 5.2, have been developed to summarize, carry, and exchange policy information and compliance data as well as negotiating the trust level of a trust association.

Another requirement is to find a way to evaluate the information delivered and match it with the local policy implemented like *Ranking system*. However, developing a ranking system in a traditional way of matching one-to-one entries, or evaluating weights can give false results. One challenge is to find the correct balance between local policy and standard compliance level with the one received in analytical way without getting enough information from the other side.

The issue of transitive trust presents another problem in building a trust relationship with unknown party. By transition trust, we mean depending on the circle of trusted parties of the requester as a qualifier to build a new relation. Transitive trust plays a very important role in deciding whether or not to establish a trust association with a new party. Assuming the level of trust based on Tt exchange was less than the acceptable threshold to exchange information, the other party can exchange a new type of information called Tl or trust list. The receiver can use Tl delivered to correct the trust level previously computed to higher values. However, this also can cause a threat of forgery and passing fake Tl with high values from previous relations

with untrusted parties.

3.2.1.4. Information Sharing

Information sharing has many advantages that promotes its importance and make it an essential component of health care systems. These include lowering health care cost, speeding up health care service delivery, and providing more accurate information about medical history and current health issues of the patient. A subject patient in medical records F might receive healthcare services in a different facility not related to the first hospital treated in such as H_j. In such case, to provide the best health care service to patient P, H_i and H_j may need to exchange the EHR of P.

In addition to the common factors between H_i and H_j, it is difficult for both facilities to apply the same level of security and access control policies. This will define a function of trust level Q for each of them different from the other such that $Q(H_i) \neq Q(H_j)$. The possible risk level W will be defined as a directed association between H_i and H_j such that $H_i \xrightarrow{W} H_j \neq H_j \xrightarrow{W} H_i$ for any $i, j \in N$ where N is a set of healthcare providing facilities. This concludes that one of the two hospitals might perform a high risk in patient's privacy if the corresponding EHR/EMR shared.

3.2.1.5. Information Segmentation

As mentioned in Equation 3.1 , information presentation is among the main factors of triggering HBC attacks against private records. The difficulty of protecting information mainly comes from the data at all security levels being accessed based on a user role. In RBAC, which is the common access control policy used in the healthcare sector, users are assigned permissions based on their function within the system rather than security clearance or individually. However, if the data owner's desire is to grant access to some portions of his/her data, but not others, it is not pos-

sible under RBAC. Since information in any healthcare system is classified according to the use of it and to the content type, why not store it in accordance with some existing standard such as HL7 2.8?

Segmenting data based on similarity of content, level of security, its need, or type makes it difficult to accurately define the proper category. It is very hard to find an automated system that has the full capability to classify any given set of information based on a certain criteria. The lack of automated categorization systems makes the task of data segmentation one of the hardest in this area. Even under the most sophisticated access control policies, misplaced information is a high threat to privacy. Determining how information is segmented or classified is outside our study scope. However, it is assumed that information is classified and stored in XML files using HL7 standards. In HL7, each segment has a tag which describes the content of that segment. This classification makes it easier to associate tags with security levels based on its content. The level of accuracy in this approach is acceptable and serves the purpose of this study.

3.2.1.6. Selective Access Controlling and Segment Gateway

Granting access based on the content type or category of data segments rather than the file containing those segments establishes a fine granularity access controlling policy. Neither RBAC nor multi-level access control policies get to that level of distinction between object and its content (if possible) without dramatically increasing the overhead selectively. In the case of RBAC it is not applicable to grant access based on security level. Access control policies and the mechanisms used to implement them do not consider the content of files, but are limited to what systems physically see. However, we are assuming a logical structure for electronic health/medical records and we are trying to control access based on that structure and not its physical

presence.

Another issue is the mixed level of security found in the same file, where a single file can contain sensitive and non-sensitive information at the same time. This mix of security levels requires a new approach to solving the problem of access control, and should take into consideration interfacing with RBAC and providing a sort of granting and preventing access based on individuals and roles simultaneously.

Our approach depends on controlling access on the conceptual level of accessing data using cryptography. Each segment can be encrypted using a specific key based on its content's security level. The key can then be shared with a subset of users who are allowed access to that segment. This means that one file can be encrypted using several keys. The problem with this approach is key management, key distribution, and performance. Solving the problem of key management and performance can be achieved by adopting a technique of segment gateways. Segment gateways can be viewed as another layer of keys encrypting the first layer of data encryption keys where the problems of key distribution/redistribution, and granting/revoking access rights are resolved by the technique of selecting the gateways values, key generation process, and a second layer of secret keys.

Figure 3.1-A shows a high level perspective data access policy under RBAC and regular session registration. Normally in RBAC users get access to a file if the role assigned to them has access to that file category. By file category, we point to file type , not the object or the instance of that category. However, in Figure 3.1-B, the framework introduces levels of clearance for HL7 tags within the system as high, medium, low. Since data is categorized under tags, it becomes possible using that new record meta data to decide if the user requesting access has the proper clearance. The same concept applies when different hospitals communicates to share information. Patients can specify those levels of security to lock certain categories of information.

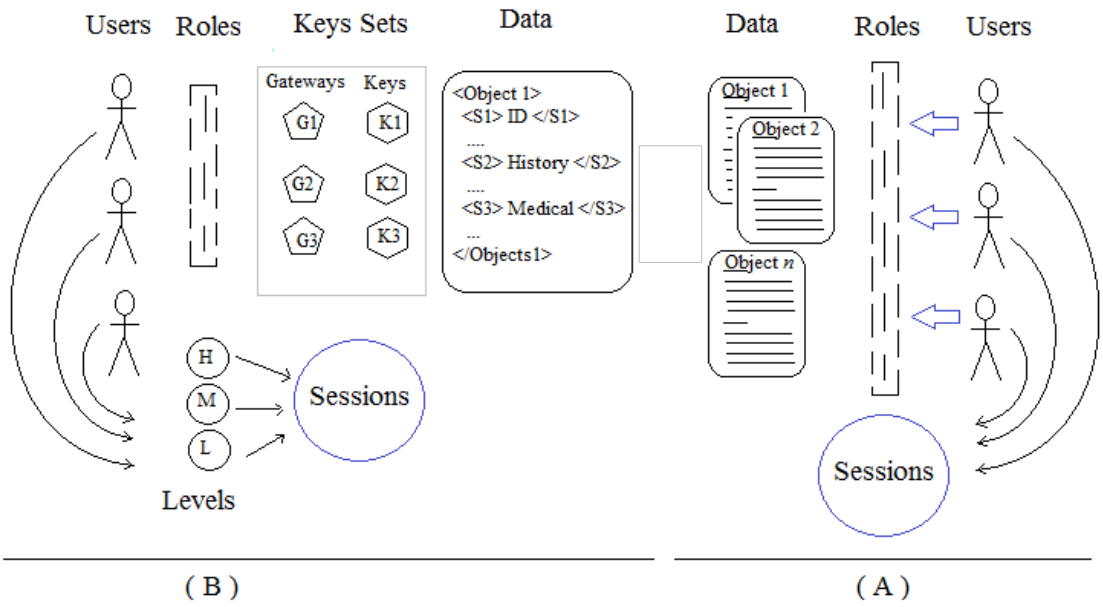


Figure 3.1. Transforming environment from traditional RBAC to gated segments

3.3. Difficulties in Role-Based Access Control Policy

The previous Section 6.2 and its subsections described, on one hand, segmentation benefits. On the other hand, it showed the complexity level involved in understanding and categorizing information. The following example is used as a case study illustrating some of the problems in identifying the infrastructure and environment where the EMR will be accessed and resides.

Assume patient P in hospital H has information stored in several files, where each file has different types on information. Let P has N files, and each file has M segments. The total amount of relevant information $D = N \times M$. The scope of total system users Z including nurses, doctors, and other hospital personnel having access to all Z. However, only X number of users, where $X \subset Z$ assigned to take care of P and have a reason to check some information such as medication and the time to provide it. Another set of users shares access to P's medical records such

as R, where X, R is the team who takes care of P. Hospital H implements RBAC to control its electronic system and to manage operations and users presence in the system. However, RBAC makes no distinction between users as individuals. RBAC grants privileges to roles rather than users. This case can be formulated as follows:

$$X, R \subset Z \tag{3.3}$$

$$Z \quad \underrightarrow{access} \quad N \times M \tag{3.4}$$

$$X, R \quad \underrightarrow{Assigned} \quad P \tag{3.5}$$

$$\forall\ Role\ \exists\ Role_{X_i} = Role_{X_j},\ Role_{R_i} = Role_{R_j} \quad : i \neq j \tag{3.6}$$

$$\bigcap_{i=1}^{|X|} Privilege\,(X_i) \quad \neq \quad \{\Phi\},$$
$$|x|\ users\, can\, have\, same\, duties\, and\, operations \tag{3.7}$$

Equation 3.4 shows that not only the assigned users can access the information, all users under RBAC with a proper role have access. However, narrowing down the list of users depending on RBAC is not possible as shown in Equation 3.6. It is possible to exclude a specific user from accessing information if a special role is created for each user excluded from a certain privilege to avoid sharing rights as stated by Equation 3.7. The problem of isolation between role members prevents the implementation of selective access controlling.

From a different perspective, Figure 3.2 shows that RBAC does not allow direct mapping between (objects-permission) and users. This fine line of separation makes it difficult to exclude certain users from accessing information if the role allows it. The structure of RBAC makes it almost impossible to maintain access control list or prevention list for a specific segment and tie it to a certain user.

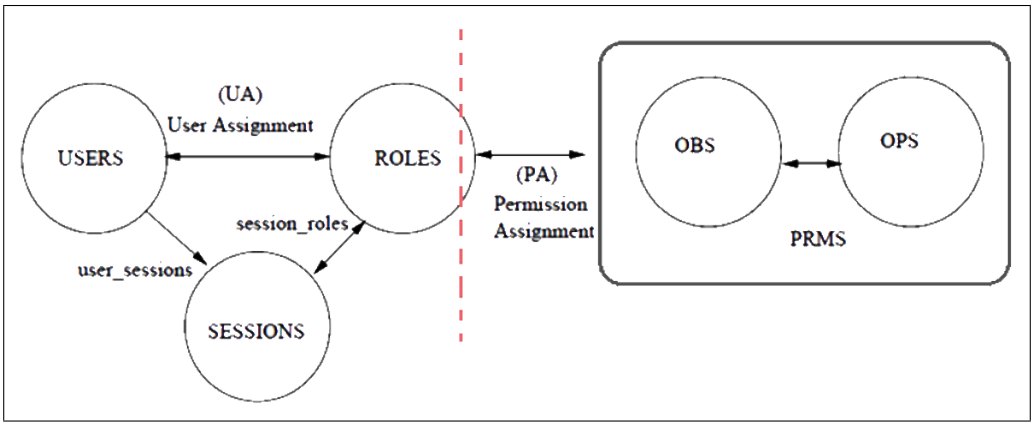


Figure 3.2. RBAC system and the fine line between objects and users

To tackle the problem of access control without disturbing the RBAC system, we designed the following procedure of data classification and tagging shown in Figure 3.3. Since the EMR data is stored in XML format using HL7 standard, each segment of the data should have a tag that describes the data presented under it. This tagging process will classify data based on its content or category. Whenever data is repeated in the same EMR or in a different file, it will be classified the same way the first occurrence is classified with the same tag value. Therefore, assigning security level to the tag will allow similar segments of data to be grouped into one security level. The resulting tagging-level is stored in a relation called TL where each tag has one security level.

Each role in the system can be assigned one or more security clearance levels depending on the type of information required to accomplish their specific function. However, this exposes the problem of RBAC flat access based on role. Not only will the assigned personnel will have access to the information, other users under the same role may also gain access. On one hand, if the patient's desire is to limit access of a certain segment to one user out of the four assigned users, RBAC will not provide

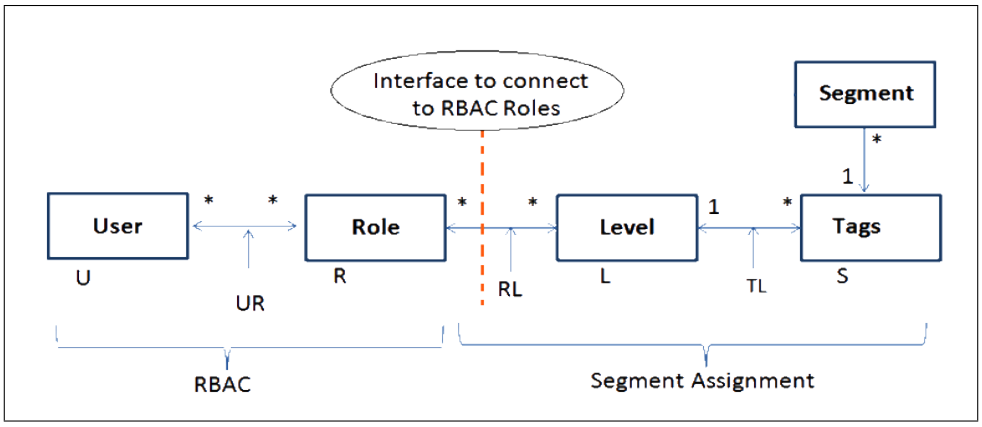


Figure 3.3. Tagging segments and connection to RBAC

this feature. On the other hand, RBAC does not provide a direct connection between users and object-privilege since it is provided through user-role and role-privilege. This problem shows that normal access control policies will not provide an effective, scalable solution in such cases. This drives us to seek a new solution through what we call segment gateways, providing selection and prevention on the user level with RBAC.

3.4. Summary

This chapter describes the problem of honest-but-curious attacks and security of medical records. It defines what the attack is, its source, and what factors can affect information sharing such as policy exchange, trust levels, and information presentation. It also explains the need to find new methods to implement selective access control policies and development of new techniques to implement them. The chapter provides a brief description about the venues of the solution and the areas it covers.

Chapter 4

RISK ASSESSMENT

4.1. Introduction

This chapter describes the process of risk assessment and the compliance with the regulations, standard, and provide a method of exchanging compliance results without disclosing the interior policy details. However, the process of surveying security policy and standards should produce an exchangeable and quantifiable policy. The result of surveying can be exchanged and evaluated by other parties without revealing its worthiness in the source. Also, this chapter describe policy quantifying to provide a numeric representation of compliance and risk assessment. Another area discussed in this chapter is the problem of heterogeneous standards and surveys, this chapter will introduce to the idea of evaluating surveys by considering multi-factor evaluation method based on (target, goal, coverage) metrics rather than one-to-one comparison.

4.2. Compliance

Complying with a certain policy and regulations, such as HIPAA, helps the health care sector improving overall performance and enhance its credibility regarding privacy protection. Each healthcare provider can apply certain privacy policy, standards, and has its own internal policy to manage relation and legal liability with other parties like insurance companies and patients. However, covered entity in a health care network has a different response for policy violations and scale the risk of policy violation in a different way from other entities on the field. Thus, exchanging hospital internal

policy, business values, compliance with standard, and risk management plans is not practical for building a trust association with another covered entity or hospital in the network.

Figure 4.1 shows the compliance process components that can be used to govern facility work; those components are dynamic and can expanded to include other policies. In this stage of the framework component design, a covered entity security level will be identified, quantified, and weighted according to its compliance with the used policies standards in a "Risk assessment" phase. However, the security framework takes into consideration the risk analysis made by security and compliance personnel in the facility to bridge the gap between business priorities and security needs. However, the overall security will not be compromised, the framework will repeat the risk evaluation step after each policy modification or each reported incident inside the facility as well as other communication hospitals.

4.2.1. Defining Compliance Policy and Standards

A covered entity, or a hospital, can define one or more standard to comply with, however, HIPAA standards are required by law. In this research HIPAA and NIST SP 800 standards are used as a showcase of evaluating security and compliance levels. The HSR toolkit (HIPAA Security toolkit) is a surveying tool provided by NIST [72] that integrates both HIPAA standards and NIST SP 800 standards in one survey based on the similarity between them. Since the survey provided is not considered as a risk assessment tool or a compliance tool, we have built our own tool that can adopt the survey and use it as a risk assessment tool. The NIST survey has the advantage of being in the middle of two standards, the NIST and HIPAA standards, the thing that allows us to negotiate more than one policy.

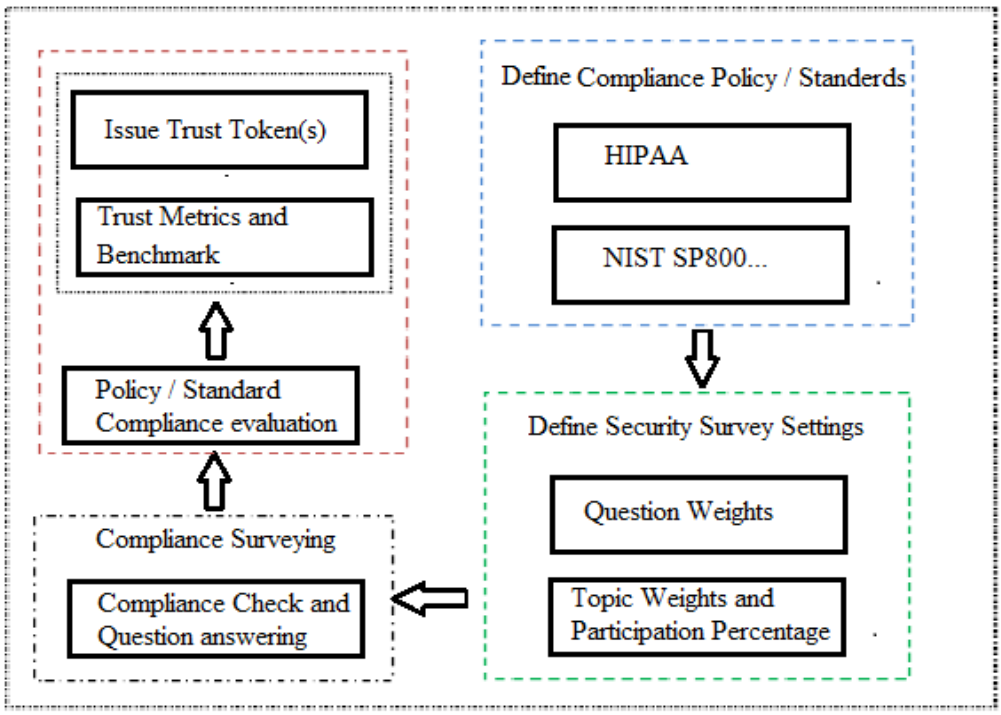


Figure 4.1. Policies prototypes for compliance and evaluation

The survey consists of several sections covering HIPAA standards, each subsection on the survey has a set of question to determine the compliance for that section. The adopted questioner from HSR toolkit combines standards and information from the different resources shown in Table 4.1. Combining different resources will create a circle of intersected policies and standards of which we can use as a reference point of trust negotiations with wider spectrum of coverage.

The design adopted for the survey can be used of any other standard other than the surveys mentioned before. For instance, an entity can use pure HIPAA standard to follow without conjunction with any other standard. However, the only thing required in our model is the survey structure which consists of the main topic (mt), sub topics(st) , questions(qs), and a set of possible answers or responses (rs) for a question. Equations 4.1 through 4.2 are used to generate a quantifying method for

compliance level with the standard surveyed:

Table 4.1: NIST HSR toolkit survey resources

Resource	Title
NIST Special Publication 800-66	An introductory resource Guide for Implementing the HIPAA security Rule.
NIST Special Publication 800-53	Recommended Security Controls for Federal Information Systems and Organizations.
NIST Special Publication 800-53A	Guide for Assessing the Security Controls in Federal Information Systems and Organizations for Building Effective Security Assessment Plans
HIPAA Security Rule	The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy and Security Rules
HITECH Act	Health Information Technology for Economic and Clinical Health Act

$$s = \{mt\} : s \neq \Phi \tag{4.1}$$
$$mt = \{st\} : mt \neq \Phi \tag{4.2}$$
$$st = \{qs\} : 1 < |st| \tag{4.3}$$
$$rs \in \{Yes, No, Not\ Applicable, Not\ Answered\} \tag{4.4}$$
$$response(qs) \in rs \tag{4.5}$$

The adopted survey consists of five main topics covers the following areas; Administrative safeguards, physical safeguards, technical safeguards, organizational requirements, and policies and procedures and documentation requirements. Each main

topic consists of a set of subtopics, where each subtopic covers a standard or a requirement under that main topic. For example, the main topic *Administrative Safeguards* has subtopics such as *Security Management Process* , which is a standard, and *Risk Management* which is a requirement. On the third level of the survey there is the question level where a subtopic can have more than one question, figure 4.6 shows an example of the structure.

4.2.2. Dependency Analysis and Survey Truthfulness

NIST assessment survey does not show clearly the relation between questions and categories. However, there is a goal behind the question itself, for example asking if there is any integrity verification applied, is there any backups and log files found. It is noticeable that logging can be used for integrity check in a database. In another case, a question goal can be checking for access control policy implementation, proper access assignment and separation of duties. Another question can say: is the data classified based on its sensitivity and who can access what. Both questions are focusing in access controlling and information confidentiality. Thus, it is the goal or the aim behind the question is what specifying the relations and dependencies. The focus should be on "why the question has been asked from the beginning?" and the area of coverage. The following example shows the concept using the NIST HSR toolkit survey;

$$if X \longrightarrow B\,,\ B \longrightarrow C \Longrightarrow X \ \longrightarrow B \wedge C,\ both\ has\ to\ be\ true\ at\ the\ same\ time \quad (4.6)$$

$$if Y \longrightarrow X \longrightarrow B \oplus C$$

$$this\ does\ NOT\ imply\ that\ y \longrightarrow B \oplus C,$$

$$however\ it\ implies\ that\ y \longrightarrow ((B \oplus C) \wedge X) \quad (4.7)$$

The following tables

Table 4.2: Dependency sequence

164.308(a)(1)(i): SP 800-66 4.1.1 Security Management Process: Identify Relevant Information Systems	
Symbol	Description
X	Has your organization defined the frequency of your Risk Assessment policy and procedures reviews and updates?
B	Has your organization reviewed and updated your Risk Assessment policy and procedures in accordance with your defined frequency?
C	Has your organization developed, disseminated, reviewed/updated, and trained on your Risk Assessment policies and procedures?
164.308(A)(3)(i): SP 800-66 4.3 Workforce Security: Administrative safeguards	
P	Has your organization implemented policies and procedures to ensure that any and all staff, employees, and workforce members have appropriate and only appropriate, access to ePHI; and to prevent the staff, employees, and workforce members who do not have access to ePHI from obtaining access to ePHI?
164.308(a)(1)(i): SP 800-66 4.1.1: Security Management Process: Identify Relevant Information Systems	
Q	Has your organization identified the types of information and uses of that information and the sensitivity of each type of information been evaluated (also link to FIPS 199 and SP 800-60 for more on categorization of sensitivity levels)?
164.308(a)(1)(ii)(B): Risk management (Required).	
Z	Do your organization's current safeguards ensure the confidentiality, integrity, and availability of all ePHI?

The figures 4.2 through 4.4 show the relations between categories found in NIST HSR toolkit. The figures summarize NIST special publications and HIPAA standards in a directed graph of dependencies.

4.2.3. Surveying

The following assumptions has been made to guarantee the survey accuracy and its authenticity: The survey should be completed by a third party that has no interest with the health care facility. A third party, such as HIPAA inspector, can perform surveying task to avoid any possible conflict of interest. The personnel who is required to fill the survey should be regionally recognized by the authorities to inspect and manipulate the designated survey. The responsibilities of the security personnel are limited to the lowest level of the questioner only, the questions level.

However, survey quantifying and evaluation is not among the responsibilities of the surveyor. Weights (importance level) of a question within its category is specified ahead of survey completion based on business priorities and by the entity itself. The second level is the category participation percentage that specifies the category shares on its level. Those two categories, weights and percentages, can be affected by the security goals and the business goals. Hence they can be specified in conjunction between the entity's security personnel and any other party from the circle of decision making inside the facility. For further discussion, scoring and ranking methodology will be discussed in depth in the scoring section 4.3 later in this chapter.

Upon the completion of the survey, an inspector will sign the survey and submit it to the security framework which will issue a "security token"which will be used for trust negotiations between any two communicating parties that applies the same security layer. The security token will be signed by security personnel form the entity subject of the survey using the entity's private key to prove its identity. Up

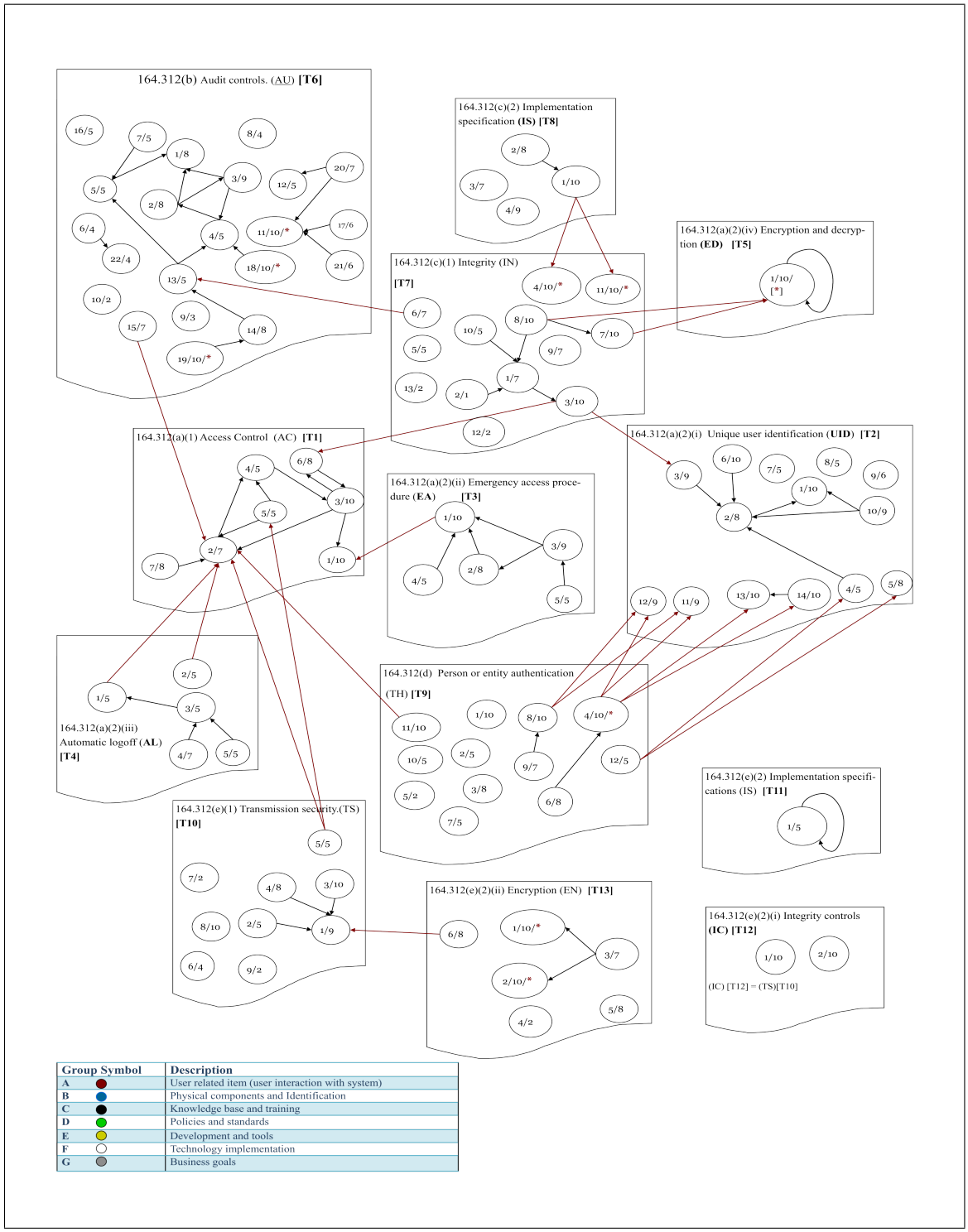
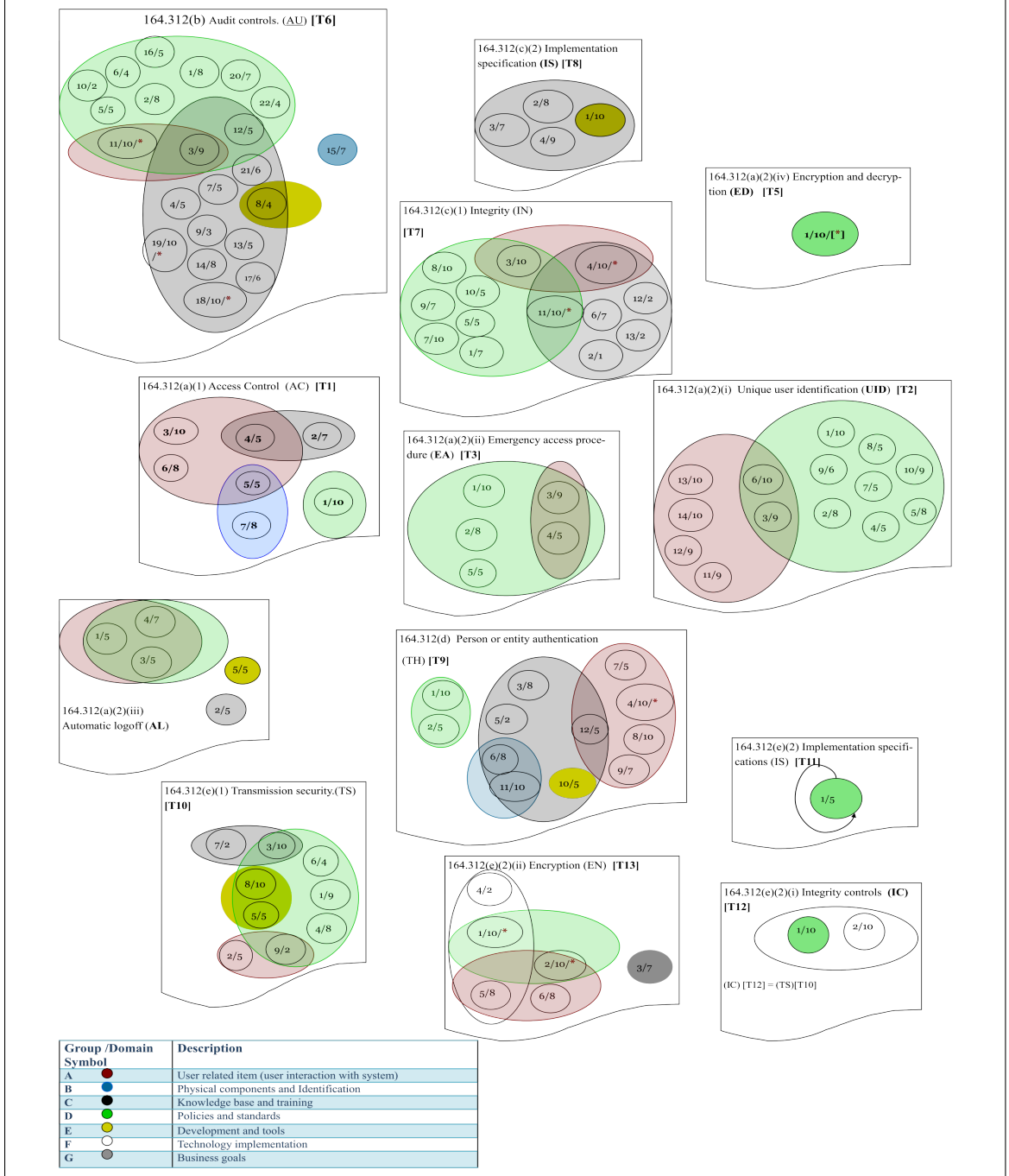
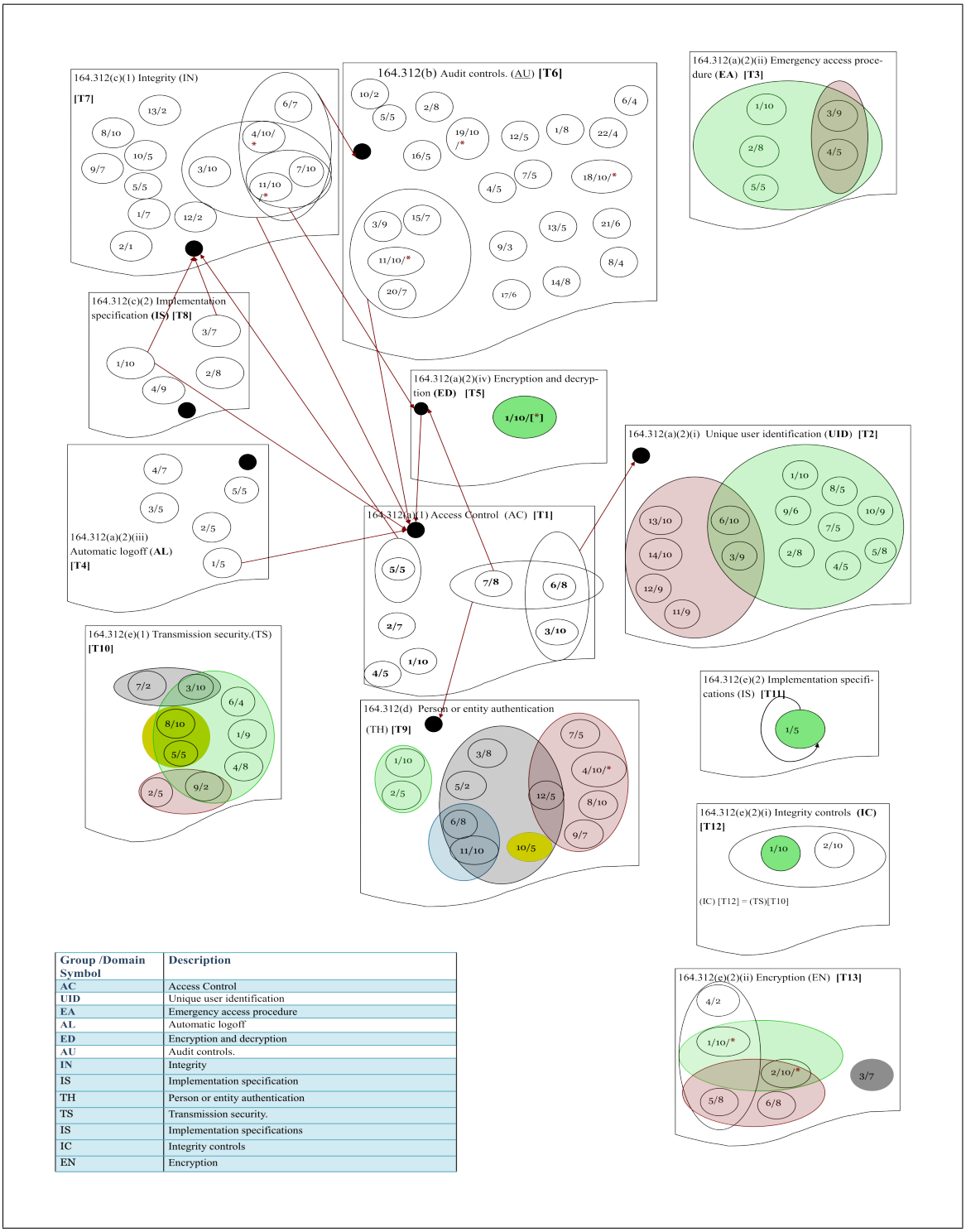


Figure 4.2. Questions dependencies graph



Group /Domain Symbol	Description
A	User related item (user interaction with system)
B	Physical components and Identification
C	Knowledge base and training
D	Policies and standards
E	Development and tools
F	Technology implementation
G	Business goals

Figure 4.3. Areas of target coverage



Group /Domain Symbol	Description
AC	Access Control
UID	Unique user identification
EA	Emergency access procedure
AL	Automatic logoff
ED	Encryption and decryption
AU	Audit controls.
IN	Integrity
IS	Implementation specification
TH	Person or entity authentication
TS	Transmission security.
IS	Implementation specifications
IC	Integrity controls
EN	Encryption

Figure 4.4. Classification based on questioner goals

to this point, the survey is completed, however, if any new techniques implemented to enhance security, the same process should be repeated to issue a new certificate. Therefore, a new survey and inspection is required for any change in the security policies, metrics, or policies and any old security token issued before that date will be invalidated.

4.3. Scoring

Despite the fact that NIST survey provided with HSR toolkit does not provide a quantifying methodology to evaluate risk or share security information, we have developed scoring and ranking methodologies that helps us to transfer the same survey to its next level. Each survey question has four possible answers from a set such as x= yes, no, not applicable, not answered. If the area covered by that question is applied then the answer will be "yes"with 100% of the weight specified for that question or area. However, if the answer was "no"; the result will be 0% of the assigned weight for that question. The other two values can be specified base on the risk level for that area.

If a certain category does not apply for an entity, then the answers will be "Not Applicable". If the percentage calculated is based on 0%, the facility might lose points for a "not applicable"category response. As a result, the "not applicable"category can be given a neutral value of 50% , or we can change the participation percentage of the overall category on the top level assignment ratio. Another solution for "Not applicable"is to reduce the weight to 0. In another case; if an entity does not provide an answer to certain questions in the survey, this means that the category can be (yes, no, or Not Applicable). Based on this, we can assign it to get the 20% of the category weight or change the participation percentage on the top level to 0% to avoid faulty evaluation or scoring. The upcoming Section 4.3.1 provides a complete description of assigning weights and ratios to survey questions and sections.

4.3.1. Weighted Tree

Providing more than one level of ranking the survey result as shown in 4.1 helps to avoid the misleading results of risk assessment and information sharing. As Figure 4.5 shows, the physical safeguards branch participates in a percentage of 20% of the total

policy importance. Out of the 20% assigned to the *physical safeguards* branch, the "implementation specifications"gets 18.8% participation ratio *in its level*. However, "implementation Specifications"category has a total of seven questions. Each question has its own "Maximum Possible to achieve". A question weighted by the importance to its category. A small set of questions can get as much representation into its category as all other questions.

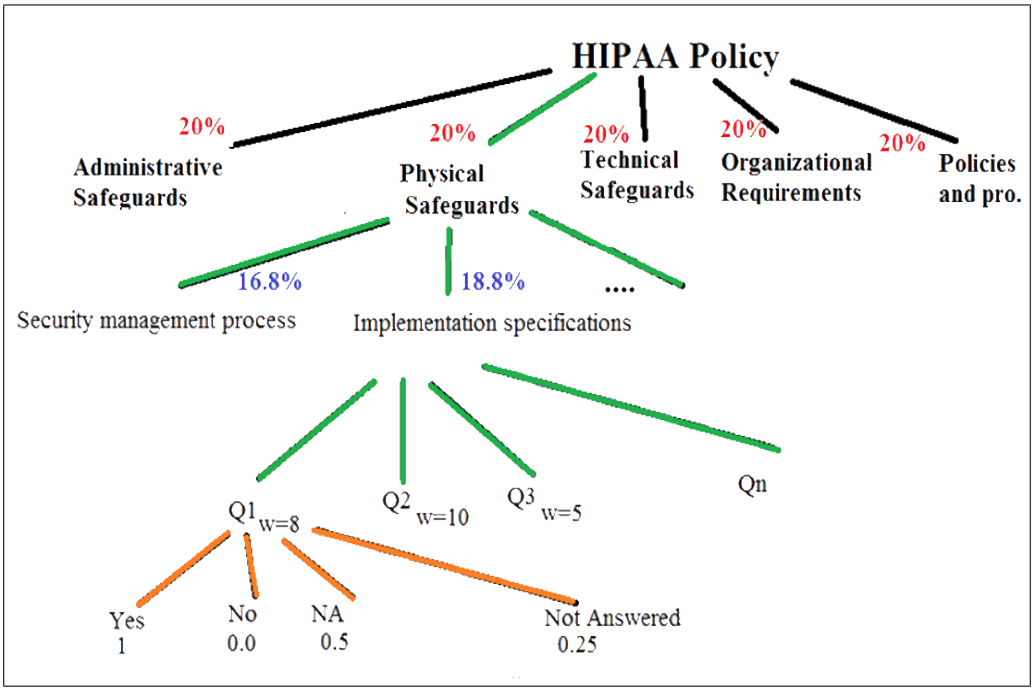


Figure 4.5. Survey structure and scoring method

For instance, assume that (S) a subtopic like "implementation specifications"has a set of questions (Q) where each question assigned a weight (W) in a scale from 0 to 10. The possible score a question can get out of its weight (Qs) is determined by the answer participation ratio Qr ="yes"=1.0, "no"= 0.0, "not applicable"= 0.5, "not answered"= 0.25. The topic has a participation percentage (Pp) specified based

on its importance in its category. Thus, the scores (Ps) that a subcategory achieves based on the answers provided as given by Equation 4.8.

Equation 4.8 gives a *sub topic* participation score P_s based on its weight:

$$Ps_j = \frac{\sum_{i=1}^{n} Qs_i}{\sum_{i=1}^{n} W_i} \times 100 \tag{4.8}$$

Question score is given by Equation4.9 as follows :

$$Qs_l = Qr_j \times W_l : \ 1 \leq j \leq |Qr| \tag{4.9}$$

Equation 4.10 represents the *participation ratio* of a category or subcategory on its level:

$$Pr_i = Ps_i \times Pp_i \tag{4.10}$$

Equation 4.11 represents the *Participation Ratio constraint* where the total survey value should not exceed 100%.

$$\sum_{i=1}^{n} Pp_i = 100 \tag{4.11}$$

The participation ratio (Pr) of a certain main topic like "Physical Safeguards" on its level is calculated based on the participation percentage assigned on its category as:

Equation 4.12: computes the *main category* participation ratio

$$Pr_x \sum_{i=1}^{n} Pr_i \tag{4.12}$$

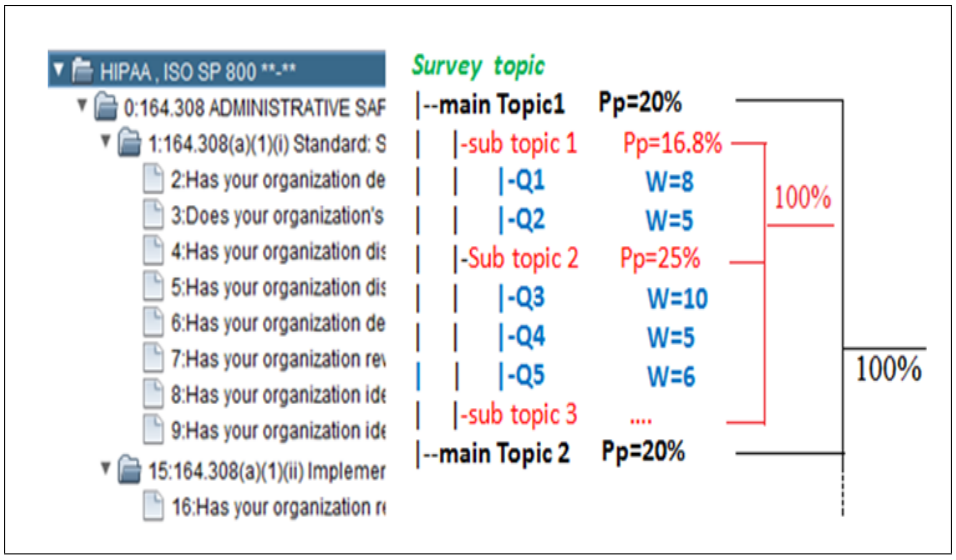


Figure 4.6. An example of scoring and evaluating the security survey

The set of Pr and the set of Ps form the basic high-level negotiation scoring system in building a trust and security security association as described in section 4.3.2.

Figure 4.6 shows an example of evaluating and quantifying the survey. In this example we are showing the process of computing the weights and the participation ratios in the policy subject of the survey. As shown in Figure 4.6, main topic 1 Pp= 20% subtopic 1 Pp=16.8% , where $Q1$ and $Q2$ assigned weights of 8, and 5 respectively subtopic 2 $Pp = 25\%$, where $Q3$,$Q4$, and $Q5$ are assigned weights of 10,5, and 6 respectively. The answers for $Q1$ to $Q5$ in order are yes, yes, yes, yes, not applicable. The ratios for the answers is yes=1, no =0, not applicable = 0.5, not answered = 0.25. Based on the values provided, $Q1$ score is calculated using Equation 4.12. Question Score is $Qr1 = W \times Qr = 8 \times 1.0$. The complete questions scores will be $A, B = 8, 5, 10, 5, 3$.

To compute the scores for subtopic 1 using Equation 4.8

$$Ps_1 = \frac{(8+5)}{13} \times 100$$

$$Ps_1 = 100$$

For the second subcategory

$$Ps_2 = \frac{(10+5+3)}{21} \times 100$$

$$Ps_2 = 85.71$$

								Qs based on possible answers			
			Pp(H1)	Pp(H2)	Pp(H3)	Pp(H4)	W	Yes	No	Not App	No Answer
								1.00	0.00	0.50	0.25
Mt			**0.20**	**0.25**	**0.30**	**0.25**					
	St1		**33.33**	**50.00**	**40.00**	**25.00**					
		Q1					**9**	9	0	4.5	2.25
		Q2					**5**	5	0	2.5	1.25
		Q3					**2**	2	0	1	0.5
		Q4					**3**	3	0	1.5	0.75
	St2		**33.33**	**15.00**	**10.00**	**30.00**					
		Q5					**10**	10	0	5	2.5
		Q6					**2**	2	0	1	0.5
		Q7					**5**	5	0	2.5	1.25
	St3		**33.33**	**35.00**	**50.00**	**45.00**					
		Q8					**7**	7	0	3.5	1.75
		Q9					**2**	2	0	1	0.5
		Q10					**8**	8	0	4	2
		Q11					**6**	6	0	3	1.5

Figure 4.7. Survey example for four hospitals

The value of (Ps) gives an indication of the compliance level for that subcategory without revealing the question importance level for the facility. In the example, the first category scored 100% of all possible scores, where the second subcategory scored 85.71% of all possible scores. The participation ratio for a subcategory among other subcategories in its level is specified by the participation percentage (Pp).

$$
\begin{aligned}
Pr_1 &= Ps_1 \times Pp_1 \\
&= 100 \times 0.168 \\
&= 16.8 \\
Pr_2 &= Ps_2 \times Pp_2 \\
&= 85.71 \times 0.25 \\
&= 21.4275
\end{aligned}
$$

The summation of all (Pr) values in a category or a subcategory of its child nodes will form the total participation score (Ps) for that node. Knowing the participation percentage of a specific category node and its score will give us the (Pr) of it in the next level of main category. In one hand, Figure 4.7 shows a survey representing four hospitals $H1...H4$. It also shows the weights for the questions and the participation percentages for each category within the survey. Figure 4.8 shows the survey results for each hospital and the scoring values. The values we are focusing on are the Ps values and Pr values which represent the scores and the ratios respectively.

In the previous example, questions' weights have been fixed for all hospitals, however, the subsection Pp varies from one hospital to another since the risk analysis is different from one section to another. As described previously, the question weight importance ratio will not affect the category importance among other categories, but it makes the category biased in a certain direction based on the risk analysis. The Pp value is the value that represents the importance of the category in the survey.

4.3.2. Scaling Schema

Quantifying the survey categories, subcategories and questions weights for each subcategory provided levels of comparison among policies without disclosing an entity's specific policy considerations. To make this goal more feasible, and to prevent

	H1	H2	H3	H4
Answers(St1)	NYYY	Y,N,A,S	YYYN	YYAS
Score(St1)	10.00	10.75	16.00	15.75
Ps(St1)	52.632	56.579	84.211	82.89
Pr(St1)	17.544	28.289	33.684	20.72
Answers(St2)	YYY	YYN	YYA	YYS
Score(St2)	17.00	12.00	14.50	13.25
Ps(St2)	100	70.588	85.294	77.94
Pr(St2)	33.333	10.588	8.5294	23.38
Answers(St3)	YYYY	Y,N,A,S	YYYN	YYAS
Score(St3)	23.00	12.50	17.00	14.50
Ps(St3)	100	54.348	73.913	63.04
Pr(St3)	33.333	19.022	36.957	28.37
Score(Mt1)	**84.21**	**57.899**	**79.17**	**72.48**
Pr(Mt1)/20	16.842	14.475	23.751	18.12

Figure 4.8. Surveys results and scores for four hospitals

entities from policy recreation for other participants, the scores will be shared in terms of {High, medium, low} concept rather than numbers as shown in Figure 4.9. Considering $H3$ as the entity performing the evaluation, the survey results from the previous example for the four hospitals, $H1$ trough $H4$, can be represented as {Trusted, Low, Trusted, Medium} respectively. However, the ranking does not provide a numeric values that can be used to reproduce or deduce the original survey for sure, but it can only be estimated through a large set of probabilities. The impact of this ranking process is shown in the data shared between the participating entities and the general level of trust regardless the type of information communicated.

As shown in the previous example, hospitals $H1$ and $H3$ fall in the same trust category (Trusted) which is a preliminary agreement for information exchange. However, this classification can give a false indication when the difference between the Mt score values is high but they fall in the same category. The same happens when the

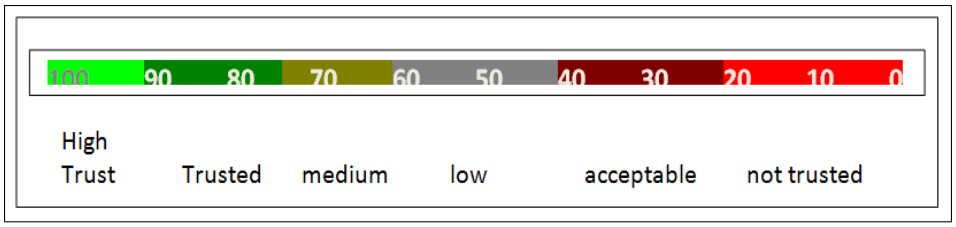


Figure 4.9. Ranking survey topics based on scores values

difference of scores is small and they fall in two different categories. The following example explains such case:

In the latter example, Mt_1 (H_1)= 84.21 where Mt_1 (H_3)=79.17, and Mt_1 (H_4)= 72.48.

$$
\begin{aligned}
|\triangle Mt_1(H_1, H_3)| &= 5.04 \\
|\triangle Mt_1(H_3, H_4)| &= 6.69 \\
However: & \\
\lceil HighTrust \rceil - \lfloor HighTrust \rfloor &= 10 \\
\lceil Trusted \rceil - \lfloor Trusted \rfloor &= 15 > |\triangle Mt_1(H_3, H_4)|
\end{aligned}
$$

Since $\triangle Mt_1$ within the same category, such as "High Trust", can reach up to 10 points, and in "Trusted"category up to 15 points, we can sense the need for new metrics to measure trust between entities. As the example shows, H_3 and H_4 have a difference of 6.69 score points, however they are not falling in the same security category. This difference in classification could result in limitations and restrictions on information sharing. Although the difference between H_1 and H_3 in scored points is close to the difference between H_3 and H_4, those in the same category can have better information exchange. This leads us to a new perspective of ranking that depends on "Rank Threshold"rather than "Categorization"based on a range of values.

4.3.3. Circle of Trust

To solve the problem of categorization, each participating entity can specify its acceptable threshold for each main topic in the policy. This method depends on considering the current score of a topic, such as Mt_1, the center of a "trust circle". The ranking is based on $\triangle Mt_1(H_x, H_y)$ rather than the position of the scores on the ranking benchmark. The following example illustrates the concept and the difference between the "trust circle"method and the previous method.

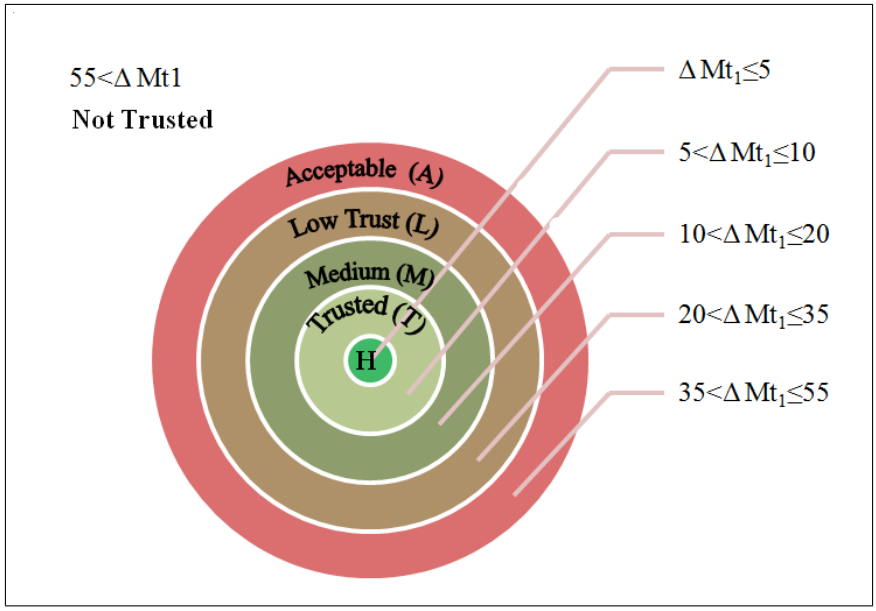


Figure 4.10. Trust Circles (TC) based on score threshold

To explain trust circles by example, assume that H_3 is the entity that is measuring trust or security levels for the other three hospitals ($H_1, H_2, and\, H_4$). Using the same numbers in the previous example, we observe that both $|\triangle Mt1(H_3, H_1)|$ and $|\triangle Mt_1(H_3, H_4)|$ falls in the same circle of trust regarding $H3$. Using TC approach, both H_1 and $H4$ classify as "trusted"and can communicate with $H3$ in the same secu-

rity clearance level. However, the assumption of using the trust threshold metric only will lead to a false sense of security when a transitive trust relationship is established with a new entity like H_5. The reason behind the false trust stems from the fact that a low-level security entity can establish a trust relationship with another low-level security entity to get a security token of high trust then use it as a referral token with a trust worthy entity.

To mitigate the problem of depending in the trust circles (TC), we used both metrics, the TC based on threshold and the ranking method shown in Figure 4.9. The ranking method will classify the security level for an entity based on its compliance with a policy. The TC method will classify the other entity's similarity of policy implementation to the local entity. Thus, the result of applying both metrics will give a comprehensive overview of policy implementation and similarity aspects between participating entities. Applying both ranking method on our latter example of the four hospitals, we found the following:

Table 4.3: Definition of variables

Hospital	Score	$\triangle Mt_1(H3, Hx)$	Ranking	TC
H1	84.21	5.04	Trusted	Trusted
H2	57.89	21.28	Low	Low
H3	79.17	0	Trusted	High
H4	72.48	6.69	Medium	Trusted

As Table 4.3 shows, the ranking method specifies the compliance of the entity with the standards used, such as HIPAA, whereas the TC shows the difference between both entities of applying that standard. Using both metrics help in the next step of data filtering based on the patient's consent and his/her classification of personal

information security. The two proposed methods have different goals and reasons, one for building a trust relationship, which is the TC metric. The other metric, which is the ranking metric, is used for information classification and risk assessment.

This chapter showed an example of evaluating both policy implementation and the closeness of implementation between entities. Also, it provides the infrastructure of the roles of surveying, quantifying, and evaluating the risk of communication between covered entities. However, there is a need to show how to communicating sensitive information and survey results. Also, there is a need to describe the negotiation process to agree on a certain level of trust for communicated information based on the output of quantifying policy implementation and trustworthiness. Chapter 5 describes the process and the data structure needed to build a link of trust, such as trust tokens and trust negotiation algorithms and protocol, based on the metrics provided in this chapter

4.4. Summary

This chapter describes the process of risk assessment and the compliance with the regulations, standards, and provides a method of exchanging compliance results without disclosing the interior policy details. However, the process of surveying security policy and standards should result an exchangeable and quantifiable outcome. Thant result can be exchanged and evaluated by other parties without revealing its worthiness in the source. Also, we describe policy quantifying to provide a numeric representation of compliance and risk assessment. Another issue is heterogeneous standards and surveys. This chapter will introduce to the idea of evaluating surveys by their goals rather than one-to-one comparison.

Chapter 5

POLICY EXCHANGE

5.1. Introduction

Building a trust association between two covered entities (hospitals) involves exchanging some parameters that describe security practices and policies. The disclosure of internal policy, the level of compliance with standards, and implementation is not an acceptable practice because of its consequences on business values and interest. However, the level of trust available in the network or a trust association determines what data can or cannot be exchanged. This chapter describes the policy exchange mechanism, the data structure used to carry the required information, and methods to enhance trust association rank.

5.2. Trust Token

A fundamental part in a communication protocol is specifying a data carrier that defines information presentation , order, type, and how to translate transmitted data. The first component is to define is trust token (Tt) and its role in trust negotiations between healthcare providers or covered entities. The specified data structure of Tt provides the ability to communicate and negotiate information preserving policy secrecy, authenticity, and enabling future negotiations. Figures 5.1 and 5.2 show the data structures for trust tokens developed to conduct trust negotiations.

Figure 5.1 illustrates the design of a trust token (Tt) used to exchange policy and entity information. The Tt classifies information into two categories based on

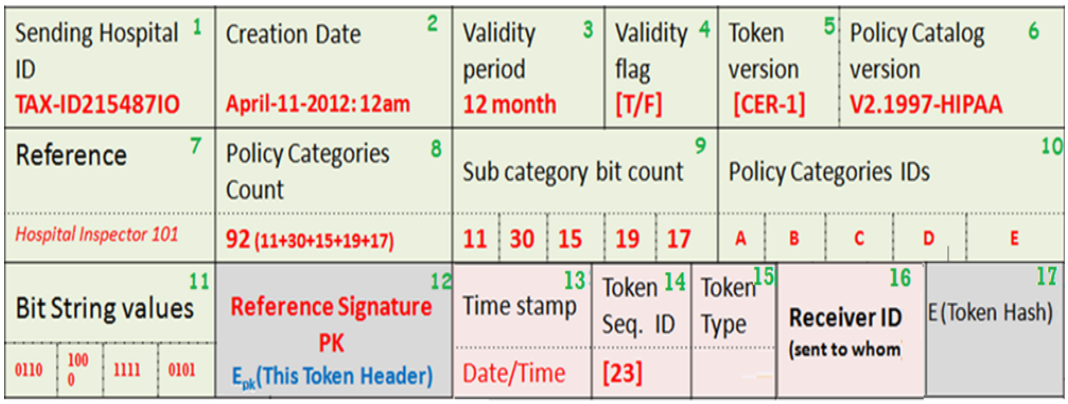


Figure 5.1. Policy exchange trust token (Tt)

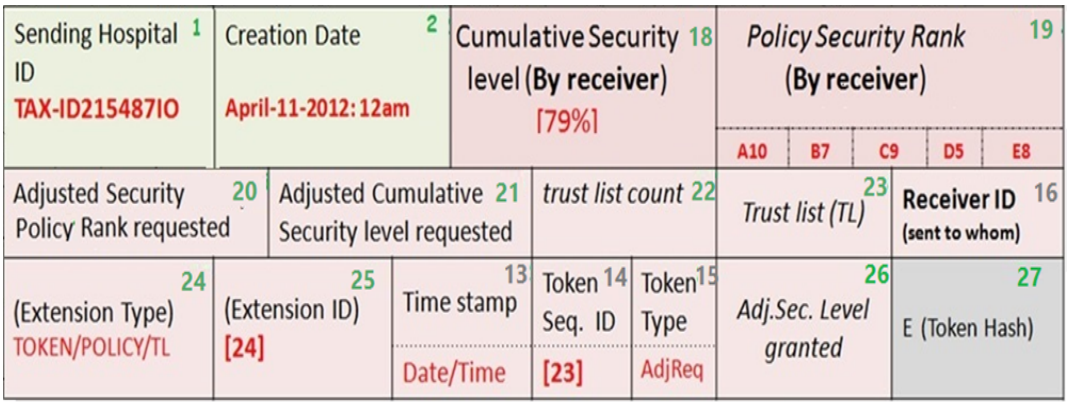


Figure 5.2. Trust association negotiation token (Ta)

how frequently the content changes. Entries 1 through 12 on trust token diagram describes the static information section where data forms the Tt header. The second part of Tt structure is the dynamic information section that covers fields 13 through 17. The first section contains relatively static information since it will not change frequently. This section remains intact during the cycle of re-certification and Tt re-evaluation unless changed by the assigned inspector (*i.e.*, HIPAA inspector). The second part holds the dynamic information exchanged and manipulated by communicating covered entities or hospitals while building a trust association. The trust

token provides the necessary information to exchange policy, negotiate trust level, and support patient's privacy setting at the origin. However, Tt authenticity cannot be proven unless signed using the corresponding holder's key and attaching it to the trust token as a "Reference Signature." Figures 5.1 shows the fields of the trust token and provides brief description for each entry. Some fields holds single or multiple values. The following list provides a description for trust token content.

Hospital ID: Each hospital or covered entity is required to have its own unique ID (*i.e.*, Tax identification number or any other nationally recognized information) that can identify the entity with no ambiguity. This field is provided by the entity who owns the Tt and it should remain the same during the lifetime of Tt. If this value modified or changed, all consequent trust relationships and security bound must be rebuilt.

Creation Date: This field specifies the time of creating a Tt for a certain entity. The time stamp specifies the starting date of when this Tt activated.

Validity Period: A numeric value specifies how long Tt is valid after the creation date. After this period has expired, the trust token needs to be re-certified. However, this is an advisory field more than a constraint. An entity can continue using Tt as long as it is not revoked or rejected by other parties. The reason for this is due to the context of how the trust token is used and the need to get the job done without risking patient's life or health.

Validity Flag. A logical value, true/false, specifies whether the current Tt is valid and can be used for further conversations. Note that this field can be updated to $false$ even without Tt being expired.

Token version. This field is used to control versioning, compatibility, and future development of Tt. Each trust token version has its own structure, by specifying version. The core of the security layer implemented dictates which structure to use

to be able to read the content of any received token. Versioning supports system robustness and different types of trust tokens when used.

Policy Catalog Version. Each Tt is issued after surveying entity compliance against certain standards. The policy catalog version provides information about which version of the standards have been used in the survey to create the trust token. This field can support backward compatibility in case of standards developments and allow covered entities to continue using their valid old Tt until the new Tt prepared based on the new version of standard.

Reference. This field holds the ID and the required information about the security personnel preparing the Tt for the covered entity. The reference person should have the authority and the qualifications to issue Tt and sign it digitally using the covered entity's private key.

Policy Category Count. As explained in Section 4.2.3 of Chapter 4, each policy consists of main categories and sub categories. Policy category count represents the total number of all subcategories in the survey. This field helps in verifying the policy and the number fields expected in the transmitted policy for integrity check purposes and validating information accuracy for the current Tt version.

Policy Categories IDs. Each main category is assigned a unique identification value that represents it in the trust token. This field contains a list of all main category identification values or IDs. For example, if the survey has five main categories, this field will hold five different IDs.

Sub-Category Bit Count. This field contains the number of bits expected to represent each main category in the survey. It holds a list of values where each value index matches its Policy Category ID index in the previous field.

Bit String Values. This field represents a list of blocks, and each block has bit string. The number of bit-string blocks should match the number of policy category

IDs and relate them to using their location and index. Each two bits represents a question answer in the received policy. The number of bits in each block is specified by "subcategory bit count"field.

Reference Signature. A security personnel must sign the previous fields to guarantee their authenticity using the entity private key. The "static fields"or Tt header is the only portion signed by the hospital private key. The receiving hospital, or covered entity, will use the sender's public key to decrypt this entry to verify that the attached information has not been changed.

Time Stamp: Date/time field. This field shows when the token has been manipulated and sent in either direction. Both the sender and receiver must change it when communicated.

Token Sequence ID. Integer value, set to zero when a new conversation or negotiation is started. The value is auto-incremented by one each time the trust token is communicated. It helps both parties to decide how long they can negotiate a trust level in a single session.

Token Type. Each token has a specific type which reveals its purpose. Types can be selected from Table 5.1. Each step of negotiation has to have a token type passed to the other entity. In conjunction with "Token Sequence ID"field, this field shows the negotiations progress.

Table 5.1: Trust token tags

Tag	Description
new	New conversation
reply	New reply
TrustACK	Trust acknowledgment
Continued on next page	

Table 5.1 – continued from previous page

TrustRej	Trust Rejection
AdjReq	Adjustment Request
AdjRej	Adjustment Rejection
fin	End of conversation

Encrypted Token Hash. This field holds the value of the Tt after being hashed and encrypted using the sender's private key and/or receiver public key, depending on the level of security and authenticity required.

Cumulative Security Level. This value provides an indication of the overall compliance level with standards based on the recipient security metrics and values. It is used to help the parties engaged in negotiations to decide whether to continue or cease negotiations based on the level of security scored.

Policy Security Rank. A list of values in which each entry represents a rank for a main category in the standard used. This value is manipulated by the recipient. The rank provided represents the format of {High trust, trusted, medium trust, low, not trusted}. The data inquiring hospital, which is the owner of the initial Tt, will get the response of his policy implementation ranking stored under this category. More description regarding the method of evaluation and negotiations will be provided in section 5.4 on trust negotiation and protocols.

Adjusted Cumulative Security Level Requested. Provides Tt owner with the capability to request a certain level of trust and security if the calculated level of security by the recipient does not meet the minimum required level for data exchanging process. The requested level has to be justified by other values and entries in Tt.

Adjusted Cumulative Security Level Requested. This optional field allows the Tt owner to ask for a new cumulative security level that allows it to move from initiating trust negotiations to exchange protected health information.

Trust List Count. In a health care network, a relatively new covered entity can use a list of references who trust that entity as evidence of good reputation when asking for adjusted security levels. This field helps when the facility is relatively small and there is not much standard compliance is required from it, or when the facility is still in the evaluation process. This field holds the number of entities that have a trust relationship with Tt holder.

Trust List (TL). The trust list is a data structure holding information about the entities establishing a trust relationship with Tt holder. The number of entries is specified by the "trust level count "field. This entry helps the recipient decide whether or not to override the estimated security level with the "adjusted security level "requested by the Tt holder. The trust list will be described in more detail in Section 5.3.

Receiver ID. The Receiver ID field holds the unique ID of the recipient covered entity. Each covered entity has its own unique ID (e.g. Tax identification number).

Extension Type. Optional, this field , if manipulated, defines the type of the extension trust token attached or following this Tt. Possible extension types can be a token for another token with same policy; a policy: another token with new policy implemented; or the new extension has more trust lists.

Extension ID. An integer value in this field shows how many extensions have been used. The field is an auto-incremented value when extensions exist.

Adjusted Security Level Granted. This field specifies the level of security granted to the sender, if requested, that overrides the real evaluation of the policy. It is not necessary for the recipient to respond with the same security level requested

by the sender, however, a new value can be negotiated.

Table 5.2: Dependency Sequence: Initial Trust Token (Tt) fields

ID	Field Name	Description
1	Hospital ID	Trust Token (Tt) owners ID
2	Creation Date	The date when the Tt created
3	Validity Period	Indicates how long this Tt is valid starting from the creation date
4	Validity Flag	Shows if this Tt is valid or revoked during its validity period
5	Token Version	Keeps track of the token version, for compatibility check
6	Policy Catalog Version	Specifies which version of the standards were used to create this Tt upon, for compatibility check
7	Reference	The ID for the security personnel responsible for creating this Tt
8	Policy Categories Count	Shows how many bits in all categories covered by policy/standards contained and included in this Tt
9	Sub Category Bit Count	An array-like field contains, for each category, the number of bits in that category's bit string presentation. Multi-values
10	Policy Categories IDs	The policy category IDs within the survey conducted and represented in this Tt. Multi-values
11	Bit String Values	Bit string representation for the answers of the survey. Multi-values

Continued on next page

Table 5.2 – continued

12	Reference Signature	Tt digital signature, prepared by the Reference security personnel. The entity private key used to sign it
13	Time Stamp	The time of the communication initiation
14	Token Sequence ID	a sequential number for this Trust token (automatic numbering)
15	Token Type	The type of the current Trust token communicated
16	Receiver ID	To whom this Tt was sent (the receiver ID)
17	Encrypted Token Hash	Integrity check hash code for the Trust token. Encrypted using senders private key

Table 5.3: Trust association negotiation token (Ta)

ID	Field Name	Description
18	Cumulative Security Level	The overall security level of this Tt evaluated and manipulated by the receiver entity based on the receiver entity policy settings.
19	Policy Security Rank	Numerical ranking for each policy category implemented by Tt owner using the receiver entity security settings. Multi-values
20	Adjusted Security Policy Rank Request	Manipulated by sender. Represents a request for different security rank other than what the receiver estimated. Multi-values

Continued on next page

Table 5.3 – continued

21	Adjusted Cumulative Security Level Requested	Manipulated by sender. Represents a request for different security rank than the receiver estimated
22	Trust List Count	The number of entities who trust the sender or the Tt holder
23	Trust List	The entities IDs who trust this Tt holder. Multi-Values
24	Extension Type	Shows the type of the extended token, if found
25	Extension ID	A unique ID for the extension, if found, for the current transaction
26	Adjusted Security level granted	The new security level granted to this token to override the computed one. Manipulated by receiver
27	Encrypted Token Hash	Integrity check hash code for the Trust token. Encrypted using senders private key

5.3. Trust List

The function of Trust List (TL) is to provide more information about how many different entities trust the Tt holder. This information improves the level of trust with any new party negotiating to build a trust association. In this design, trust association between two entities can change based on what other associations the communicating entities have had in the past. This methodology provides a new technique of adding credibility to the healthcare dynamic network, and provides a method of motivating healthcare entities to apply improved security metrics. Entities participating in negotiations can request adjusted security levels based on business

needs. However, the request to adjust security level granted based on policy evaluation can be rejected by the other party if there are not enough reasons to justify it. By providing a list of trust associations built with other entities, a requester can prove its trustworthiness regardless of its compliance level with some areas of standards. The list of trust can be manipulated based on the trust tokens negotiated and approved between Tt holder and other entities. Regarding its design, Trust Lists or (TL) consists of the following fields:

Sender ID. The unique ID for the covered entity communicated by Tt holder. This ID is taken from the Tt used to negotiate trust association with the sender.

Date. The date when the trust association established between the two negotiating entities is recorded here.

Exp. Lists the expiration date for this trust list.

Trusted Destination. This is the unique ID of the covered entity that holds the Tt where the TL resides and initiated the negotiations.

Categories Counter. A numeric value indicates how many fields of the security policy the other party or covered entity assumed trust association with Tt holder.

Security Rank(s). A multi-value field shows the security rank given or estimated by the other parties for each category included in the policy subject of negotiation. These fields are manipulated based on the values extracted from the Tt used to build trust association with the entity mentioned in the "Trusted Destination "field.

Categories. A multi-value field shows the policy category IDs mentioned in the current list. The fields are ordered respectively according to their appearance in the "Security Ranks "fields.

Sender's Signature. This is a self-descriptive field which indicates the digital signature for the attached TL to guarantee its authenticity and integrity. The sender

signs the TL using his private key. As it is shown in the example given, a Tt holder cannot sign or modify the values mentioned in this list.

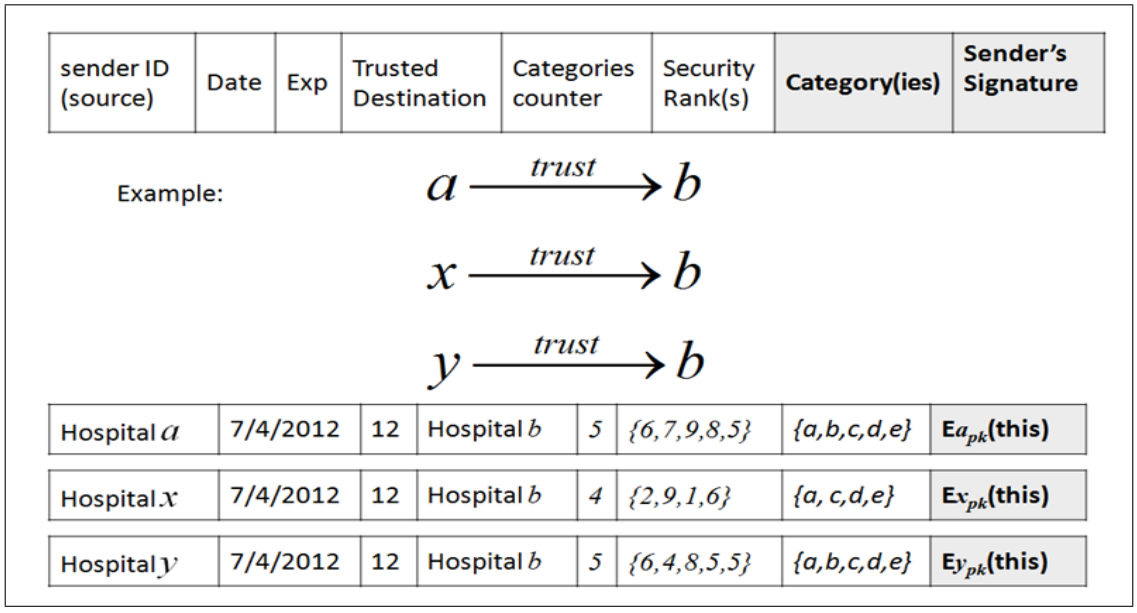


Figure 5.3. Trust list design and relations

Fields 22 and 23 in Figure 5.2 indicate whether or not a TL is associated with the communicated Ta. Field 22 specifies how many TL are expected to appear in Field 23. If no list is attached, those fields will show zero value for the number of trust lists and a *null* value for attached lists. Both parties can keep track of trust associations created to use if required in future negotiation as a reference or as recommendation materials toward new associations.

5.4. Trust Negotiation and Protocols

We have explained in previous sections surveying policies, trust token structure, trust lists and how the information required for establishing trust relationships is inferred and utilized. The goal of the trust association is to protect the privacy of

a certain type of information from being disclosed if receiver is not complying with security standards and policies at a given level. The first step in deciding what information can or cannot be shared is building a trust association between the participating covered entities. Trust association is formed based on many parameters exchanged regarding policy implementation and standards compliance survey. Such associations play the role of information filter for transferring portions or segments of EHR fulfilling trust association parameters.

Table 5.4: Trust token tags required for negotiating trust association

Token Type Tag	Description
new	Specify a new session tag, the first token tag exchanged for new conversation
reply	First reply message header from receiver to a message tagged with $< New >$
TrustACK	First reply to a successful/ accepted association request, responded to a message tagged with $< Reply >$
AdjReq	A reply tag to a successful accepted association request, responded to a message tagged with $< TrustACK >$ requesting a different level of security other than what the data guardian/holder evaluated and granted.
TrustRej	First reply to a unsuccessful / rejected association request, responded to a message tagged with $< Reply >$
Continued on next page	

Table 5.4 – continued

AdjRej	A reply tag to a unsuccessful / rejected adjustment request message , responded to a message tagged with $< ReqAdj >$
Abort	To terminate trust association (negotiations) in any stage
fin	Last tag exchanged to finish successful negotiations to establish a trust association. Replied to a message tagged with $< TrsutACK, TrustRej, AdjRej >$

Trust associations allow different parties to exchange electronic medical records and important information while preserving its privacy. However, trust association will not create, and are not intended to create encryption schema. The target is information privacy in a domain where attacks can happen from insiders abusing access rights. The Honest-But-Curious model affects the data in the environment where security and compliance assumed all the time and all users are authorized to access the information. Thus, it is important to provide a security protocol to control the process of electronic medical records transfer and sharing. It is necessary for trust associations to be aware of the EMR contents and categories. In this section we will describe the necessary steps to negotiate and establish the trust association between two entities.

The flowchart in Figure 5.4 illustrates the negotiation process to build a trust association between two parties. It also describes the process of further negotiations to elevate trust association rank. In the following section, a detailed negotiation algorithm will be provided for initiation of a trust association and how its used to exchange information. Table 5.4 shows the possible request types Tt and Ta can

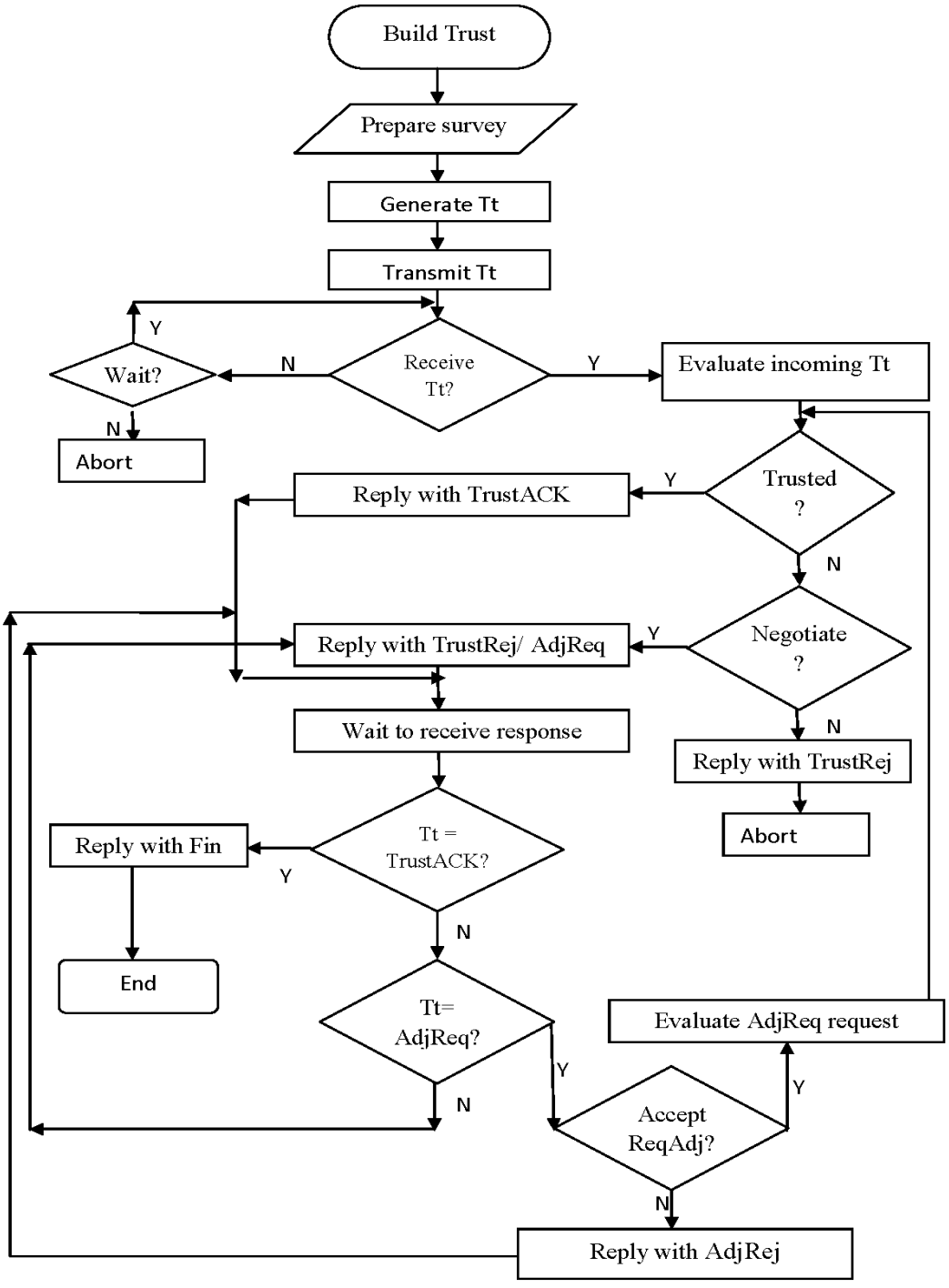


Figure 5.4. Trust negotiation flowchart

handle.

5.4.1. Starting New Association Process

When two parties start negotiating for the first time, they have to refer to a central location or to a local database to get the initial communication link established. In our protocol we assume all *secure links* are established one step ahead. The goal of this protocol is to create a level of trust, not a secure communication link.

Table 5.5: Creating new trust association process

step	Description
1	Start
2	If(Tt_{local}==null)
3	Tt_{local} = CreateTrustToken(localPolicy)
4	guestpolicy = receiveIncommingPolicy()
5	guestFork= new ForkLocalPolicyTree(Tt_{guest}).implement()
6	Tt_{guest}= CreateTrustToken(guestFork)
7	Trust_Algorithm= Select_Evaluation_Policy().getPolicy(n): *returns how policy will be evaluated*
8	Trust_Algoritm.Evaluate (Tt_{local}, Tt_{guest})
9	If (Trust_Algorithm.isTrusted())
10	Ta= new Trust_Association(Tt_{guest})
11	Reply(Ta).status(TrustACK)
12	else
13	Reply(Ta).status(trsutRej)
14	End if : END

Table 5.5 shows the basic negotiation process with a simple response of $< TrustACK >$ for accepted negotiations, and $< TrustRej >$ if rejected. However, the initiator or the target still have the option to go further in negotiations by replying in line 13 with $Ta.status = AdjReq$. Trust association token or Ta used after both parties exchanges their Tt information.

5.4.2. Negotiate Rejected Trust Association Request

In case of rejecting a request for creating a trust association as in section 5.4.1 , the initiator can reply with a trust association Ta token asking for modification. In this step, there is no policy exchange process since it occurred in the previous step. The following algorithm describes the steps of negotiating existing Ta or a rejected one. Table 5.6 shows the algorithm for adjusting the result of negotiated trust association.

Table 5.6: Ta negotiation process

step	Description
1	Start
2	Ta_{guest}= ReceiveIncomingTrustAssociation()
3	If(Tt_{guests}.has($Ta_{guest}.getID()$)==false)
4	Exit(-1)
5	If($Ta_{guest}.tokenType$== "adjustReq")
6	If(Ta_{guest}.TrustListCount!=0)
7	TL_{guest}=get($Ta_{guest}.TrustList$)
8	$Eval_1 += Tt\_Repository\ _C ALL(evaluate(TL_{guest}\ ))$
9	If ($Ta.MoreTrustList$) Go to 6
	Continued on next page

Table 5.6 – continued

10	$Eval_2$= Evaluate($Ta_{guest}.AdjectedSecurityRankRequested$, $Ta_{guest}.PolicySecurityRank$)
11	Eval= Aggregate ($Eval_1$,$Eval_2$)
12	Create $Ta_{reply} = new(Ta(\text{Eval}))$
13	Reply(Ta_{reply})
14	END

5.5. Summary

This chapter describes the process of building a trust association between two covered entities (hospitals). It also provides the logical structure for the tokens/-packets used to transmit policy and standards implementation for security practices and policies. This chapter describes how to negotiate the level of trust available in the network or the trust association between hospitals. It also covers and determines what data can or cannot be exchanged in the negotiations stage. Moreover, this chapter delivers the necessary trust negotiation protocols required for any two parties in the network to build a trust relation that governs and filters protected and personal information to be exchanged.

Chapter 6

SEGMENTS GATEWAYS: LOOSELY COUPLED DATA-KEY STRUCTURE

6.1. Introduction

A major challenge in access controlling in the health-care sector is providing a fine-granularity access control policy that does not require major changes in existing systems. The new policy should be able to provide the ability to patients to participate in securing their information. It should also not increase the amount of maintenance cost or the time needed to access information without degrading system performance. This chapter shows how access rights can be controlled dynamically using a second layer of encryption keys called *segments gateways*. This chapter also introduces the problem of information segmentation and how it is a necessity for privacy protection.

6.2. Motivations for Data Segmentation

There are various motivations for data segmentation in medical field divided between complying with federal laws, enhancing individuals' participation in privacy protection , to functional and organizational improvements. Some federal laws restrain access to mental health information or HIV records or genetic information [78]. However, other states like Illinois have restrictions on accessing information related to mental health, spinal cord injuries, head injuries, alcoholism, cancer and genetic tests [42]. In more strict cases , it is prohibited to disclose some medical information even in emergencies. For example, in Massachusetts it is prohibited to disclose

HIV results even in emergencies. Another example, CORHIO, the Colorado Regional Health Information Organization decided not to exchange data instigated from mental health clinics. M. Goldstein [41] describes information and data segmentation in the medical field and details many areas of discussions and interests.

Regardless HIPAA regulations and state laws, individuals may want to keep some portions of their information confidential or secret for various reasons. Reasons include the desire to restrict access to certain data for fear of job loss, and preventing discrimination, physical harm, or social stigma. Aside from improving privacy protection, patients' participation in reviewing medical records can enhance the accuracy of diagnosis and reduce medical errors. A survey in [104] shows about 68% of patients have concerns about their medical records, about 52% trusted what doctors told them, and 57% trusted organizations.

Regarding information sharing, 42% of the patients in the sample said they do not feel comfortable about sharing their information and only 31% are comfortable with it. However, about 15% of patients on the study will hide some information from their doctors if they had system sharing for information with others. This shows the need to provide some method to apply patients' preferences on how the data will be used and accessed other than the current access control policy dictated by the facilities providing healthcare services.

Forms completed by patients at the first stage of admission at a hospital cover medical history, health survey, insurance forms, and other information . Types of information provided by patients include identity and other quasi identifiers such as race, gender, occupation, medical history, allergies, address, emergency contacts, and so forth. Following a known standard when collecting information such as XML or messaging standard like HL7 provides initial segmentation in all stages of data acquisition process. Such standardization provides the ability to embed a data sensitivity

level, or privacy level meta data to a record. However, traditional access controlling policies does not look into data types or data sensitivity since access to data usually granted based on objects or based on functionality.

6.2.1. Segments and Segmentation

Before introducing the segments gateways, it is useful to define segments, and what we mean by gateway. A basic definition of data segmentation as it appeared [37] is "the process of sequestering from capture, access or view certain data elements that are perceived by a legal entity, institution, organization, or individual as being undesirable to share."However, this definition does not cover segmentation policy, scope, or granularity level as stated by [37].

Another definition for data segmentation can be found in [45] "Data segmentation is the term often used to describe the electronic labeling or tagging of a patient's health information in a way that allows patients or providers to electronically share parts, but not all, of a patient record."It also states that "Data segmentation helps providers comply with specific state and federal laws, helping to keep the 'sensitive' portions of a patient's electronic record private."

In our proposal and from an implementation point of view, regardless the process, a segment can be defined as a piece of related and meaningful information that can be classified, tagged, and stored as one atomic unit. A single segment can be used to perform a job or take an action based on its content. For example, information about medicine name and dose can be stored in one segment and accessed daily when medicine is provided to a patient. Another example is allergies segment and so on.

There is no limitation on the size of a segment or its content. Hence, a certain segment should contain one type or class of information such as personal identification information, medical information, history, or any other type of information. A certain

object or a file may contain many segments where each segment has its own tag or header. Segment security level can be specified based on the segment on file or for tag level. Each segment is assigned a tag and can be accessed by a role with proper security clearance matching the tag security level if the policy of multi-level security is enforced.

In the data segmentation approach, all segments categorized under a certain tag should contain the same type of information. An example of existing standards of segmentation is HL7 standards. The HL7 formatted message has a tag for each segment of information [6]. The tag value gives an indication about the data itself, for example the <PID> tag indicates a segment for Patient Identification information in HL7 message. More about the adopted HL7 messaging system can be found in the HL7 background study in Section 2.8.

Segment importance can be assigned in two ways; automatically by the EMR system as initial classification, or manually by the data owner. Some information categories are known to be highly secure and specified by HIPAA regulations [48], where other information can be classified by the patient as protected information. For example, a patient can classify a certain category, say medical history, as medium, allergies as low security, and habits as highly secure. This classification will apply to all data segments which fall under that category even if resides in multiple files.

In our proposal, segment privacy and segment security translate into two layers of implementation; segment gateways and encryption keys. Furthermore, securing data segments for each patient can be done through encrypting segments as follows.

1. Single key use for all segments. A single key can be used to encrypt all segments. However, there are two issues with using a single key for all segments. First, if the key used to encrypt all categories compromised, all segments can be decrypted. The second issue is performance. Whenever the key is updated or

changed, all information has to be decrypted and re-encrypted using the new key.

2. Segment level assignment. Each segment can be assigned a unique key. This approach provides a higher level of security than single key approach. Nevertheless, key management/generation for small sizes is not a trivial task for large databases.

3. Category level assignment. Each category has a unique key. This approach depends on segmentation process accuracy , however it saves a considerable cost in key management. The problem with this approach is being unable to assign different level of privacy for a certain segment in a category since all segments within a category has same security level.

We will be describing access control more in the upcoming sections and how it can help controlling accessibility without access control policy implemented. This will help in protecting private information outside its environment when shared with other parties.

6.2.2. Technical difficulties

Despite the advantages data segmentation provides, there are many technical difficulties associated with the process. Those difficulties come from multiple sources such as information, followed procedures, regulations, implementation, and many others. The following list identifies some difficulties with the process of data segmentation [42].

1. Segment selection. How will the data be classified and segmented and what criteria will be used?

2. Legacy systems. Some systems are old and need to be updated before segmenting information .

3. Locating data in the system. Same information can be found in many places in the medical record. HIV can be an example where the red cell count, support group information in the free text area of the file, medication, other diagnoses, *e.g. Kaposi's sarcoma*, all are indications of HIV.

4. Lack of terminologies and codes. Each structured data requires a set of identifiers and codes to organize information. Sharing and classifying information is a big hassle when there is no single standard to follow.

5. Data accuracy and dependability. As segmentation provides the infrastructure for more privacy protection, it creates another problem in decision making. To what limit data shown provide the necessary knowledge for fast medical action without being liable for a wrong procedure occurred because of missing or hidden information.

6. Dumb output. Providing a closed intelligent system that is used to retrieve segments related to the treatment only and based on the context. Such *rule engines* still need to be developed.

7. Keeping track of patients' consents. How will the system apply patient preferences when data retrieved?

The previously mentioned obstacles in data segmentation show the difficulty of achieving the correct formula for segmentation. Clearly, those problems increase the difficulty of protecting segmented information and complicate privacy protection. However, new problems evolve when viewed from security perspectives,

1. Encryption and key management. Encrypting segments carry different levels of

complexity since it can be done in different levels. Each approach has its own level of complexity as described in 6.2.1.

2. Identical-role mutual-access for a single segment. Depending on how keys are managed and the key distribution method, what is the most efficient approach if access is granted based on keys ownership? This problem becomes more visible if selective access control policy takes place within the same role.

3. Replay attack and dynamic user-patient assignment. There is no guarantee that some party will not try to access the data by reusing the same credentials used in previous sessions by another user in the same role.

4. Emergency access. Since data will be encrypted, not providing a method for accessing records based on patients' preferences will disclose information in unintended way.

5. Key management and key storage. What type of key management and storage will be utilized; public-private key sets, secret keys, or session keys? What is the best way to ensure privacy protection?

6. Applying patient's preferences. Does providing patients with the ability to apply their own preferences interfere with legacy systems implementation and internal policies? Any proposed solution should take in consideration cost effectiveness.

Such implementation difficulties play an important role in how privacy protection solutions should be delivered. For example, providing an access control table for each EMR per patient before segmentation requires intensive management. This problem of access control will increase dramatically if EMRs are segmented because of the amount of information generated and number of policies to maintain.

Other factors should be taken into consideration such as auxiliary data related to medical records. Table 6.1 shows general properties for medical environments to

participate in solution formation and technologies used to secure segmented EMRs. Other factors like system implementation, infrastructure heterogeneity, and dynamic trust modelling affecting privacy protections solutions must be applied to EMR system. As data is transferred from one site to another, segmentation techniques, access control policies at the recipient system may not provide the same level of protection and support provided in the sender system.

Table 6.1: EMR environment characteristics

Nature	Description
Distributed access	Parties from different places requiring access to the EMR repository.
Distributed information (database)	Personal information scattered among different databases or within the database itself.
Heterogeneous infrastructure	Not all information custodians apply the same security metrics or use same equipments or supports similar services
Open-Loop system	Insufficient feedback provided in case of abuse or unplanned disclosure.
Dynamic trust model	Trust level associated with one entity varies based on time, location (temporal/special), and/or subject condition

Since the segmentation problem itself is not the main focus of our research, this chapter provides an overview of segmentation and why it is needed. It also clarifies the need for common standards in information segmentation and classification. It

will be beneficial for future research and for proposed framework completion to cover the segmentation problem in general.

6.3. Segments Gateways : Decoupling Privacy and Security

Our framework provides an encryption scheme to comprehend mutual access, implementing patients' security preferences, is scalable for large systems, and enhances privacy protection based on fine-granularity selective access controlling. Another concept introduced in segments gateways is layering. Providing two layers of keys (gateways and encryption keys) reduces the cost of encrypting segmented EMR by avoiding data re-encryption. In this approach, EMRs and its logical structure will be transparent the user lowering the risk of HBC attacks.

6.3.1. Segmentation Gateway Overview

As described previously, each EMR in the system is an XML file that contains classified information based on tagging. Each data segment classified by a tag will be assigned two keys. The first key,K_R, is the closest to the data and used to encrypt the data itself providing data security. The second key, S_k, is for controlling access. All gateway keys are members of a set called SG such that $K_r \in SG : K_1$ is the first key in SG. As shown in Figure 6.1, segment gateway is another layer of keys used to grant access to the first layer of keys or the segment key (K_R). Each segment has its own key (K_R) to encrypt the physical data, and they are not to be shared with system users.

As shown in Figure 6.1, to get access to segment F_1 for example, the user should have access to SK_1 then K_1 to be able to decrypt F_1. However, RBAC does not support or maintain access lists to control data within files. A case study is used to demonstrate how segmentation gateways will provide access to multiple users with a selectivity option without interfering with an existing access control policy such as RBAC.

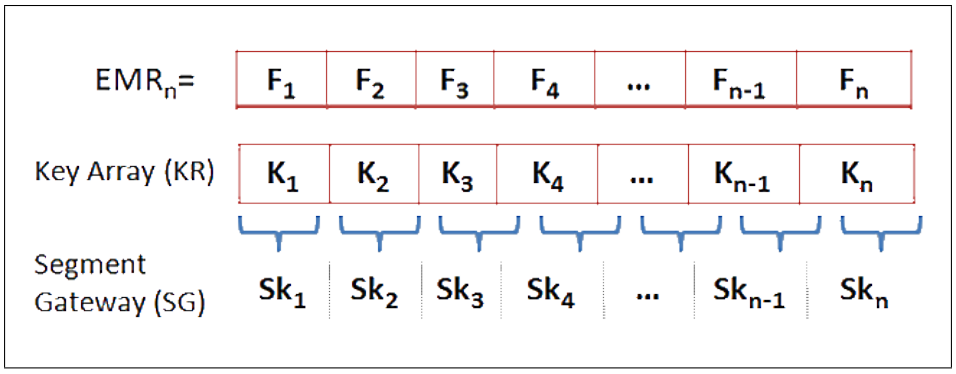


Figure 6.1. Segments gateways overview

Assuming three health care personnel (e.g. nurses) with the same role are assigned to provide the proper care to patient P. Based on RBAC, all three users will have the same accessibility to patient P information. P wants to exclude one nurse from accessing some type of information but allow the others. To make it easy, it is assumed that P has only four segments on his EMR data file. Each segment has its own encryption key and segment gateway. 7.1 shows the segments and keys and the following relations showing Ps desire to access controlling list.

1. $U = U_1, U_2, U_3$, where U is the set of users.
2. $U_1 = S_2, S_3$
3. $U_2 = S_1, S_2, S_3$
4. $U_3 = S_2, S_4$

The idea of accessing the segment is based on the ability of the user to present evidence that he has the key for the segment or has been assigned one. The evidence can be the key itself, by comparing the presented key with the one on the corresponding gateway. The system will either grant access if they are equal or prevent access

otherwise. However, the users in such scenarios need to keep so many keys for each segment in the system, which is not efficient or secure for the following reasons:

1. The number of encryption keys is tied to the number of records if access granted based on encryption key ownership. Since the database size is very big, it will not be efficient to grant access depending on encryption keys.

2. Each time the keys are updated, a new key distribution for all users should be generated.

3. If one or more user needed to be excluded from accessing any segment, an update for gateway keys is required, in addition to key redistribution.

4. No efficient method for storing and managing keys has been found for very large scale problems. Each file can present too many segments and/or too many users.

Proving segment gateways will solve the previous problems by extracting access rights from a security aspect. The second layer of keys will be dedicated to access controlling only without looking into data security.

6.3.2. Effect on Access Rights

As described in the example in Section 6.3, all three users share access to segment S_2. Granting access to users should not relate to carrying the encryption key for S_2, however, it will be allowed to cross the gateway of S_2. All three users will be given a value as proof that they are allowed to access records such as SG_2. The same value SG_2 will be the gateway value in the database level for segment S_2.

In case we need to re-encrypt segment S_2 with a different key, we are not required to grant access to all users again since there is no relation between the two keys. Other problems will appear when we revoke access rights from one user and leave the other users. This problem will be solved in the upcoming chapter on key management.

6.4. Summary

This chapter discusses information segmentation, why it is needed and its advantages. Difficulties in information segmentation are reflected in the nature of solutions provided for fine-granularity access controlling. However, problem size, implemented policies, and procedures followed in performing duties in facilities like hospitals are added complexity.

In this chapter we described the fine granularity access controlling using segments gateways dynamically separate from internal security. Mutual access was introduced in Section 6.3 and Section 6.3.2 where the advantage of keeping security apart from access rights helps in more effective management.

Chapter 7

FINE-GRANULARITY ACCESS CONTROL POLICY USING COMPOUND KEYS

7.1. Introduction

Chapter 6 introduced the problem of data segmentation,fragmentation as well as problems of mutual access. In this chapter we will be providing a solution for mutual access in the same role, granting access rights selectively, and revoking access rights using compound key structure. As described before, when more than one user is assigned to the same role, it is difficult to selectively grant or revoke access using the proposed segment gateway. To solve this problem, we introduced the concept of complementary sets and compound keys.

The concept of complementary sets relies on providing prepared common factors in all keys supplied to users for future use. Whereas, the concept of compound keys depends on creating a key with more than one identifier, this chapter provides more than one solution for the problem of mutual access. However, we provided several methods of implementation at different costs, performance and space requirements. The basic assumptions of lowering key redistribution, allowing data owners to apply personal preferences, and minimizing the modification requirements to existing systems and implementations.

The first approach is αCL, where α refers to the complementary set and aims to replace ACL. The second approach is βCL, or the *Bit* control list and it depends on the concept of place holders in a bit vector. Each approach has more than one way

of implementation. However, the main difference comes from how segment gateways manipulated and the method of constructing the compound keys. There will be a set of assumptions for each approach before going into details in each section.

7.2. Assumptions

Before proceeding with αCL or βCL , the following assumptions hold for both approaches:

1. In both approaches, segment gateways are used for organizing access rights and not intended for information encryption or data protection.

2. The concept of segment gateway assumes the presence of structured files following some standard. In this framework it is assumed that $HL7$ standard implemented.

3. Keys are distributed to users in encrypted certificates where a user cannot decrypt it or has no access to certificates' content.

4. For each active patient in the hospital there is a conceptual presentation in the system that ties the patient to users.

5. Each user is issued a certificate for each patient to whom he or she provides service.

7.3. αCL Using Complementary Sets

Segment gateways are introduced to provide more flexibility on access controlling and encounter the problem of honest-but-curious attacks on private information. The idea of selectively granting or revoking access to a certain user or users can be implemented through access control policies. However, these access control policies can affect the functionality of the health care system. So we came up with the idea of providing for each user an access key "C."The key "C"is calculated based on need-to-know or least-to-know role. Starting with the a simple case of three users and five segments, each user will be assigned a key as follows:

$$C_i = \prod_{x=1}^{n} SG_{i,x},$$

$$\textit{where } C_i \textit{ is the key for user number } i, \textit{ and } SG_{i,x} \textit{ is the corresponding segment gateways for user } i \tag{7.1}$$

In Equation 7.1, any user i in the set will be granted access to a selected set of segments from n segments related to his assignment. For the following assignment for three users :

1. $U = U_1, U_2, U_3$, where U is the set of users.
2. $U_1 = S_2, S_3$
3. $U_2 = S_1, S_2, S_3$
4. $U_3 = S_2, S_4$

Certificates should contain keys selected from table 7.1 for users U_1, U_2, and U_3 as :

$$\begin{aligned} C_1 &= \{3, 5\} \\ C_2 &= \{2, 3, 5\} \\ C_3 &= \{3, 7\} \end{aligned}$$

Table 7.1: Patient P's EMR and keys

Data Segment	S_1	S_2	S_3	S_4
Encryption Key	K_1	K_2	K_3	K_4
Segment Gateway	2	3	5	7

Any of the three users can present his key (C_i) as evidence to the system of his being assigned the key. The access control system will check the key against the gateway key for a match, and, if found, the user will gain access to that specific segment as in Equation 7.2 where x is the segment index and i is the user index.

$$Boolean\, \beta_{i,x} = ((select\, count(C_{i,x})\, from\, SG) > 0)\, ?\, true\, ;\, false \tag{7.2}$$

The value of β is true if and only if SG has a match with any of certificate C_i values and access will be granted. Access will be rejected otherwise. For example, if user U_1 attempts to access segment S_1 he will submit his certificate which is $3, 5$. However, by computing $\beta_{5,1}$, the result will be $\beta_{5,1} = false$ since the key C_1 does not contain the appropriate key value 2 for S_1. The same scenario will occur if U_2 tries to access S_4 or U_3 tries to access S_1 or S_3. Access will not be granted unless the key submitted has the unique gate value included. Still, the problem of mutual access is not solved for users with same degree of clearance or in the same role.

7.3.1. Generating the Complementary Set α

The idea of complementary set α came from the fact that access rights will not be tied to the segment itself. It is possible to use pre-compiled certificates for a certain number of users. The number of users covered by a complementary set should be the maximum expected number of users to take care of a patient. We found that the average number of persons taking care of a patient at a time is 4 and the maximum might reach 10 in its worse cases.

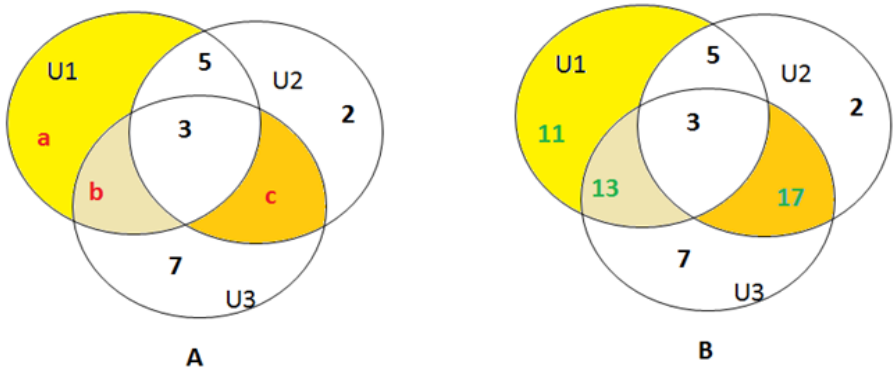


Figure 7.1. Locating the complementary set α elements

Creating a complementary set is an easy process. It depends on creating a common factor between any n number of users accessing the same data. In Figure 7.1-A, the diagram represents each user by a circle with all possible intersections. For each region assigned a unique key identifying that area as a key holder. Figure 7.1-A shows three colored areas that have no value assigned because there was no segments needed to be assigned to those regions. This caused the problem of mutual access within the same role.

The complementary set proposal solves this problem by assigning values to unallocated areas. Using this feature, we can manipulate the colored regions $\{a, b, c\}$ with any unique value that is not used for any existing regions. The new selected

values are $\{a = 11, b = 13, \textit{and}\, c = 17\}$. This set will be called the complementary set $\alpha = \{11, 13, 17\}$. The selected complementary set α should fulfill all conditions that SG complies with in addition to the following condition in Equation 7.3.

$$SG \cap \alpha \;=\; \phi \tag{7.3}$$

This condition provides the ability to selectively deny access to some segment for a certain use, or grant access exclusively to a user or a subset of users in this initial distribution. The complementary value in region a provides the feature of exclusive access to U_1 on any segment assigned that key. The other two values in b and c provide access to the users sharing them, but not the other, on any segment or set of segments they are assigned to. If this stage is done *before* the keys are distributed, then the new keys will be different and each key will carry all the desired features.

Common factors between any set of users as well as the unique factors are provided prior to keys creation. So it is possible to replace any gate value with a factor that covers a certain set of users and exclude others. Note that there is no need to re-generate keys, redistribute certificates, or change anything other than the gate value to control access. The following values for keys are the ones to be distributed at the initial state where initial $\alpha = \{11, 13, 17\}$.

$$\begin{aligned}
\alpha &= \{\mathbf{11,\ 13,\ 17}\} \\
C_1 &= 3, 5, \mathbf{11\ ,\ 13} \\
C_2 &= 2, 3, 5, \mathbf{17} \\
C_3 &= 3, 7, \mathbf{13,\ 17}
\end{aligned}$$

7.3.2. Granting Access Rights

Based on the previous key assignment with complementary set integrated, (see Table 7.2), it is possible to grant access to U_2 on segment S_4 without affecting the access rights of U_2 . This can be done by selecting a common factor between U_3 and U_2, however the selected value should be unique and both users have it in their certificates.

Table 7.2: EMR segments, keys, and segments gateways

data segment	S_1	S_2	S_3	S_4
encryption key	K_1	K_2	K_3	K_4
initial assignment	2	3	5	7
adding S_4 to U_2	2	3	5	**17**
all users access S_4	2	3	5	**3**

As key assignment shows, only $C_2 = 2, 3, 5, \mathbf{17}$ *and* $C_3 = 3, 7, \mathbf{13, 17}$ carry the value of 17 where C_1 does not. The same will happen if we need to grant access to all users. The third row of key assignment shows changing the segment value to 3 will allow all users to get access to S_4 .

7.3.3. Revoking Access Rights

Revoking access rights from a user should not affect other users who have access to the same segment. Using the same example we used before, if we intend to deny user U_1 the access right to segment S_3, all we need to do is to replace the gate on S_3 with another value. The new gate value should allow other users who have access to the same gate to get access without changing their keys. In our case U_1, U_2 are the

only users who has access to S_3.

$$SG_3 = 5 : (SG_3 \notin C_1) \wedge (SG_3 \notin C_2)$$

$$SG_3 = 2 \rightarrow SG_3 \in C_2$$

Table 7.3: Exclusive accecss to a segment

data segment	S_1	S_2	S_3	S_4
initial assignment	2	3	5	7
U_2 only access S_3	2	3	**2**	3

The new gate value will deny U_1 the right to access segment S_3 but not U_2. The new key used for S_3 will be 2 which is only held by U_1 as shown in figure 7.3.

7.3.4. General Complementary Set creation for n Users

Testing system tolerance for errors and unexpected new users can be done by adding a new user to the system that is not considered when α set created. Assuming a new user joined the system and needs access to some segments such as U_4 with assignment of $\{S1, S2, S3\}$. The initial solution is to compute C_4 such as:

$$C_4 = \{SG_1, SG_2, SG_3\} \, from \, the \, initial \, vector \, SG$$
$$= \{2, 3, 5\}$$

The problem becomes U_2 and the new user, U_4, has the same access rights. If we intend to prevent U_2 from accessing segment S_1 with $SG_1 = 2$, access will be revoked automatically from U_4 also because:

$U_4 \subseteq U_2$

The solution for such cases is to expand our initial complementary set α to include other values that can be used to distinguish a certain user. Assuming our initial $\alpha = 11, 13, 17, 19$, we can compute $C4 = \{2, 3, 5, 19\}$, now we can use α_4 which is 19

to replace the gate on S_1 and allow C_4 exclusively access S_1.However, there are no other common factors or intersection areas between the new user and the old ones to cover mutual access problem.

$$\forall U_x \, \exists C_x \, : \left| C_x - \bigcup_{y=1, y \neq x}^{n} C_y \right| \geq 1, \, \forall x \leq n \tag{7.4}$$

$$\forall U_x \, U_y \exists C_x, C_y \, : C_x \cap C_y \neq \phi, \forall x \leq n, \forall y \leq n, \, x \neq y \tag{7.5}$$

Relation 7.4 specifies that, for any key in the assigned user set, it is a must to have special factors distinguishing that user from other users. Relation 7.5 guarantees the existence of overlapping between any user and a set of $n-1$ of the assigned users. Figure 7.2 shows how those relations are translated when factors are assigned. As it appears in the figure, in region "A" user U_1 has a unique factor from SG not shared with other users. However, in region "M" both users 1 and 2 have at least one unique common factor, where in region "L" users from 1...3 share a new common factor. The common region represents the intersection of all sets together which will contain the factors providing access to all users for any selected segment.

The number of the independent sets required for this approach is given by Equation 7.6, which specifies the value of items in the complementary set α. In our example, the size of the complementary set $\alpha = 15$ different unique values.

$$\begin{aligned} |\alpha| \quad &\geq \quad \sum_{i=1}^{n} {}^{n}C_i \\ & \quad n = Maximum\, number\, of\, assigned \\ & \quad users\, with\, different\, permissions \end{aligned} \tag{7.6}$$

By implementing Equation 7.6 on a case of four users we found that size should not be less than 15 at any given time if we want to use only unique numbers.

1	A
2	B
3	C
4	D
43	E
42	F
41	G
432	H
431	I
421	J
32	K
31	L
321	M
21	N
4321	O

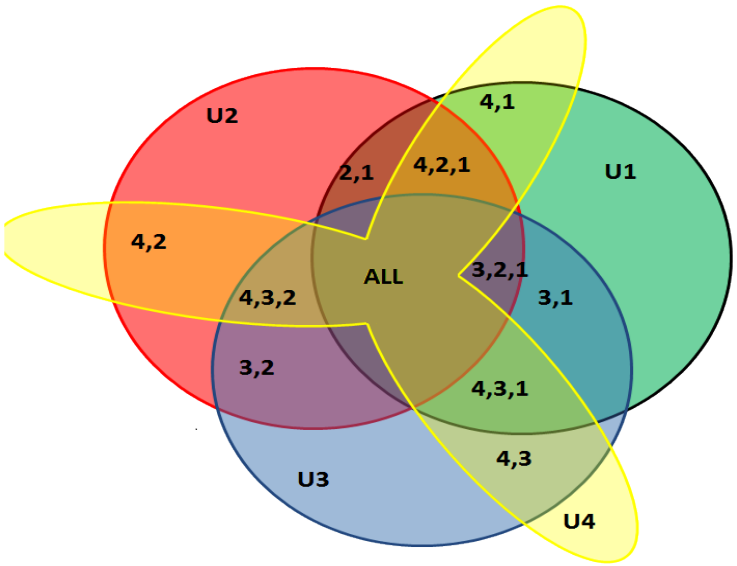


Figure 7.2. Complementary set selection for users α for 4 users

$$
\begin{aligned}
|\alpha| &\geq \tfrac{4!}{1!(4-1)!} + \tfrac{4!}{2!(4-2)!} + \tfrac{4!}{3!(4-3)!} + \tfrac{4!}{4!(4-4)!} \\
&\geq 4 + 6 + 4 + 1 \\
&\geq 15 \; or \; 2^u - 1
\end{aligned}
$$

Consequently, adding a new user to the system requires a pre-existing large complementary set to be able to accommodate that addition with the same flexibility of access control. This means that all keys should be calculated based on a problem size of n before the initial distribution to avoid key redistribution in future. We need to keep in mind that the relations we are providing is preparation for the worst scenarios. However, it is possible in some cases to find a common divisor within the numbers used for other gateways by factoring the keys to find a common divisor. The following example shows the case where we add a fourth user to our previous example without using α of 15 values assuming the new user accessing segments with gates $\{2, 5, 7\}$, then $C4 = \{2, 5, 7\} \cup \alpha_a$; $\alpha_4 = \{d, e, f, g, h, i, j, o\}$ is a unique value for C_4. Recalling

the previous example, the new keys for all users will be as follows after considering the complementary set values:

$$C_1 = \{3, 5, \textbf{ a, g, i, j, l, m, n, o}\}$$
$$C_2 = \{2, 3, 5, \textbf{ b, f, h, j, k, m, n, o}\}$$
$$C_3 = \{3, 7, \textbf{ c, e, h, i, k, l, m, o}\}$$

To prevent user U_4 from accessing S_4 with the value 7 on its gate, we replace the gate value with another one that allows user U_3 only from accessing it such as c. In such case user U_3 is the only user who can access segment S_4. If we intend to prevent user 1 and 2 from accessing segment S_3 with gate value 5 and allowing only user 4 to access it, we can select the replacement gateway value to be d. It could be possible to find a replacement value for the gateway from within the used numbers to replace any other gateway for access controlling. The future work will show the algorithm used to select the replacement gateway values.

7.4. Implementation of Complementary Sets

There are several way to implement α set principle depends on the environment and resources. The main component affected by implementation is user certificate.

- indexed αCL. This approach depends on having an index for each entry on the source CL. The index is used as a reference to the key in the original CL. The size of the index is 2 bytes used as a segment gateway in a reverse check approach.

- Hashed αCL. The hashed approach depends on storing the key hash in front of the data segment for comparison.

- Singleton αCL. Access control is based on the mathematical relationship between keys, where keys are prime numbers and unique in the set α.

- Challenged αCL. This approach assumes a cryptographic access control policy where keys used in α are used to encrypt the first layer of keys , the data encryption keys.

7.4.1. Challenged αCL

The challenged approach is a cryptographic method where the values used for the keys in αCL can be used for encryption. This method does not require strong keys or a strong encryption algorithm since it is used for access controlling in a closed trusted environment. Algorithms like $skip32$ or Tiny Encryption Algorithm (TEA) can serve the purpose here.

In this approach, α set values are small sized encryption keys, depend on how the system is configured where each key is a unique value. Granting and revoking access rights follow the same method shown in Section 7.3.3. Segment gateways are used to **encrypt** the first layer of keys. However, segment gateways are stored in only two locations. One is the user certificate and the second is the α set for each user. There is no physical presence for SG in the database containing medical records.

To get access , a user should submit an encrypted certificate with set of key, only the EMR system will be able to decrypt it. The EMR system will create a session for the user and retrieve the list of keys from the certificate. In a brute force version , the system will try to decrypt as many level one keys as possible (the data encryption keys) with those keys retrieved from the certificate. The result will be a set of decrypted keys from layer one (the first layer) that can be used to decrypt segmented data related to it.

The primary advantage of this approach comes from the fact that no segment gateway key is stored in the database. Another advantage is that data encryption keys are somehow provided with extra security from the segment gateways. It also

provides record security when transferred to another hospital since it depends on who holds the keys to decrypt records.

However, the disadvantage comes from computation where the storage is limited to the number of patients in the system only. This means if a hospital has a capacity of 1000 patients , then we need to keep a 1000 α set each with 2^P entries for each patient as long as they are active patients.

7.4.2. Singleton αCL

In the previous approach in Section 7.4.1 , segment gateways values (SGs) are unique and stored as a list in the certificate. The difference in this approach is that all SG values are **prime numbers**. Certificates distributed to users will not contain a list of values anymore, but it will contain one value which is the product of SG values assigned to him or her.

To get access, a user must submit a properly encrypted certificate to the system. The EMR system will decrypt it to retrieve the single-value key. The user will be granted access to the record encryption key (which is not secured by the primary key as an encryption key) if :

$$key_{cert_x} \ (mod) \ SG_i = 0 \tag{7.7}$$

For example , if a user with initial assignment granted a key from α set as $c_i = 3, 5, 7$ then the single value key will be $c_i' = 3 * 5 * 7$ or 105. Only segments with segment gateways participated in c_i will allow access to the next level of encryption keys, otherwise access will be denied.

This approach is very appropriate for systems that does not support list comparison and have the ability to do simple math. It also beneficial for systems with a

small number of users. Space complexity in this approach and number of keys in the database is manageable since the gateway value is one digit. However, the downside will be the product value stored in the certificate and in finding many prime numbers for large systems.

7.4.3. Indexed αCL

The approach of indexed α set was developed to minimize the size of storage in the database consumed by segment gateway values when specified as encryption keys. In this method the complementary set and access rights are the same as the previous two approaches where user certificates still have the values from α. However, the list of keys in the certificate implemented in a hash map where each segment gateway has a specific index.

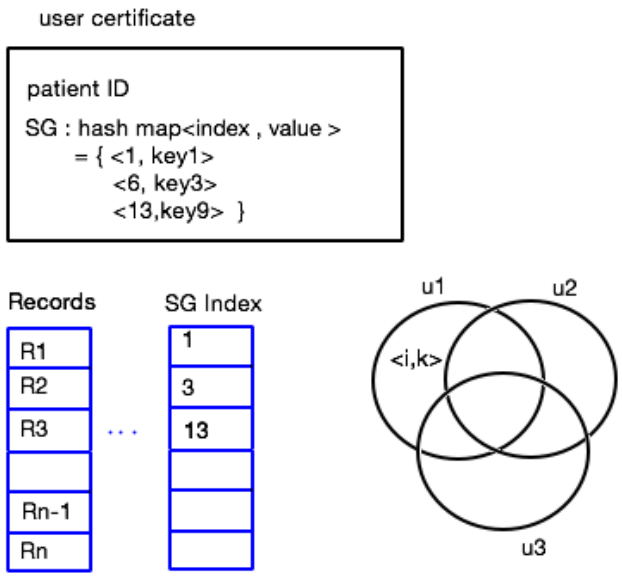


Figure 7.3. Using inverted indexing for large keys

Figure 7.3 shows user certificate, segment gateways, and the initial creation of the complementary set and α set. Granting access should follow the following algorithm:

1	for each record (i=0 to n)
2	key=certificate.get(SG[i])
3	try key_2=decrypt(K,key)
4	if key_2!=null then
5	return decrypt(R_i, key_2)
6	else
7	continue
8	end if
9	end

Table 7.4. Inverted index approach for indexed αCL

This approach will distribute the cost of large key values into certificates. The number of certificates is tied to the number of active patients in the hospital. To avoid having accidental disclosure, it is still possible to come up with a number of integers from 1 to $t * 2^p$ and use them as indexes for SG level.

Table 7.4 shows the algorithm used to implement the inverted index approach. In step 2 of the algorithm, the value of the segment gateway will be used as an index to get the corresponding key from the hash table in the certificate. If the hash table returns a key, the key will be used to decrypt data encryption key for the segment guarded by that gateway value as shown in step 3. In step 4, there will be another round of decrypt if the previous step , step 3, returns a valid key, which is step 5. Otherwise, the algorithm will continue to the next segment.

7.4.4. Hashed αCL

This approach mixes features from indexed and challenged approaches in one. In the hashed approach , it is possible to save the hash value for the key as a segment gateway value, whereas the complementary list consists of the key values used to encrypt level 1 keys. When a user submits a certificate, the EMR system will de-

crypt the certificate, then try to hash each key and match it with hashes from the segment gateway. If a match found, the corresponding key will be used to decrypt that segment.

This approach has advantages where it does not keep a certain index or order for keys. With a good hashing function, the size of the hash can be small enough not to be an overhead in the database. Still the trade off in this approach is the hashing and comparison time.

7.4.5. Quantifying Complementary Set Implementation

The following factors define the complexity of the proposed complementary set solution. These factors will be compared with the closest known solution for the same problem which is a traditional access control list structure that ties record with user ID. Factors include:

1. total number of patients (T)
2. active patients (t)
3. number of low level segments in a medical records/patient (R)
4. number of service providers / patient (P)
5. size of α set = 2^p
6. total number of ACTIVE records in the system = t*R

For ACL, the total number of entries to control access to the data is :

$$t * R * (P + 2) + [(T - t) * R * 2] \tag{7.8}$$

where $t * R * (P + 2)$: total number of active records in the system , $P + 2$: the two users are the administrator account and the data owner or patient account. Simplifying the first equation 7.8 :

$t * R * (P + 2) + [(T - t) * R * 2]$

$= t * R * (P + 2) + [(T - t) * R * 2]$

$= tRP + 2tR + 2TR - 2tR$

$tRP + 2TR = R(tP + 2T)$

The total number of entries for access control list implementation will be given by $R(tP + 2T)$. The formula shows that the number of entries increases at the same rate the number of records increased. It is also tied closely to the over all number of patients in the system.

For αCL, the complexity depends on the implementation and how will it be used. The total number of entries used in complementary set approach is given by :

$$(T * R) + (t * 2^p) + (T - t) \tag{7.9}$$

Where $T * R$: total number of entries in the whole system.

$(t * 2^p)$: active entries kept in system as a key repository.

$(T - t)$:keys for inactive records , one key used by the administrator and data owner as alpha set keys.

Simplifying the previous relation :

$$(T * R) + (t * 2^p) + (T - t) = TR + t * 2^p + T - t \tag{7.10}$$

7.4.6. αCL and ACL: Complexity Based on Number of Keys Comparison

It is noticeable that ACL in Relation 7.8 depends on the data size in the system, number of users accessing the data, and number of active record. The space complexity to provide access controlling will be affected by all factors t, T , P , and R. However, in the proposed solution the growth will be tied to the number of keys kept in the system given by $2^p * t$. However, space complexity relate to key index size , depending on implementation technique.

The increase in size looks exponential, however, the factor p has been described as the upper limit of possible health care providers to a patient at a certain time, which is limited to 10, where (t)is the number of active patients in the system. By active patients we mean the patients currently in the facility or the hospital receiving service.

Both ACL number of entries given by Relation 7.8 and αCL number of entries given by Relation 7.10 has a common factor of $T * R$. By subtracting that factor from both relations , the final comparison will be between :

$$ACL = R(tP + T) \tag{7.11}$$

$$\alpha CL = t * 2^p + T - t \tag{7.12}$$

By comparing Equation 7.11 and Equation 7.12 , to validate growth speed , two factors are considered: number of keys complexity and storage complexity. The proposed framework will be better than the ACL approach if the following condition holds for a number of keys or access control entries : $2^p * t + T - t < R(t * P + T)$. We notice that $2^p * t$ is a value with predictable limits given by $4 \leq P \leq 10$, $1 \leq t \leq Hospital\ capacity$ and $t \ll T$, where R and T are not limited with upper bound. That upper bound can be expressed by a constant value z.

By this , the comparison factors are limited to the unbounded variables, by substituting the constant values and the bounded values by factors, where $c = 2^p * t$, $m = t * P$ the inequality will be:

$$c+T-z < R*(m+T) \longrightarrow T < R*m+T*R : c, z, m \ll T \; and \; c, z, m \ll R \quad (7.13)$$

Figures 7.4 and 7.5 show the comparison between using the alpha set and using traditional access control list entries. However, space consumption for entries still depends on how each approach is implemented.

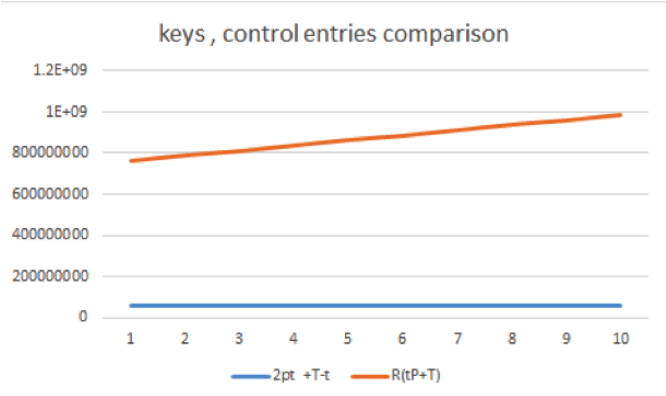


Figure 7.4. Comparing αCL entries with ACL entries

By presenting the ratio between the number of keys in αCL and ACL , we can see that αCL improves the solution when the data increase. The reason behind that is αCL keys are not tied heavily with the data or user identifier size .

T	R	$2^{P}t$ +T-t	R(tP+T)	($2^{P}t$ +T-t) / (R(tP+T))
10000	200000	55713900	710000000	0.078470282
11000	201000	55714900	733650000	0.07594207
12000	202000	55715900	757500000	0.073552343
13000	203000	55716900	781550000	0.071290257
14000	204000	55717900	805800000	0.069146066
15000	205000	55718900	830250000	0.067110991
16000	206000	55719900	854900000	0.065177097
17000	207000	55720900	879750000	0.063337198
18000	208000	55721900	904800000	0.06158477
19000	209000	55722900	930050000	0.059913876

Figure 7.5. Growth rate in number of key comparison in growing records

7.5. βCL

7.5.1. Compound Key Structure

The compound key can be defined as a key holding more than one identifier. A compound key can be assigned to any entity to be uniquely identified within its environment. Our approach defines two types of compound keys, the user identification key (UIK) and the group identification key (GIK) as Bit vectors where:

GIK uniquely identifying a group in the system where:

$$GIK_y = \{g_y\} : \quad g_y \oplus g_z \neq 0 \quad | \forall\ x, zx \qquad \neq z \tag{7.14}$$

UIK uniquely identifying a user such that:

$$UIK_x = \{p_x, g_z\} : \quad p_x \wedge p_y = 0 \quad | \, x \neq y, \\ p_x, p_y \in g_z \quad \forall \, x, y \tag{7.15}$$

Figure 7.6 explains the relation between groups and user place holder key or (UIK). Group identification key cannot match as stated by Equation 7.14. However, the user identification key is unique within its group as a place holder, but not across groups, as shown in Equation 7.15.

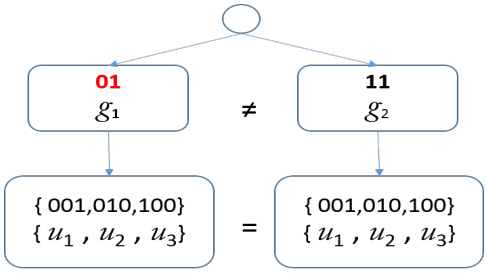


Figure 7.6. group key (GIK) and user keys (UIK)

As specified in Equations 7.6 and 7.14, keys are not tied to a known user in the system and do not hold user identifying information such as name, number, or ID. The keys are generic and grant access to their holders, if they will be granted within a security certificate.

Each active patient in the electronic medical system will be represented by a group with a unique group identifier. Each user in that group will be identified by a place holder bit in a bit vector as shown in Figure 7.6. By this assignment, even though two users have same location in their groups, we still can differentiate between them using the group IDs.

7.5.1.1. Gateway Manipulation and Granting Access Rights

In the previous two sections,we discussed segment gateways structure 6.3.1 and compound key structure 7.5.1. In this section we will be describing gateway-key assignment based on groups, key assignment based on users, and a combined approach to solve group-user-segment assignment.

7.5.1.2. Assigning Access Based on Group ID

Granting exclusive access on a record to a certain group can be done by assigning the corresponding group GIK as a segment gateway for that segment. To access data, a user should submit a security certificate issued from the hospital with a proper UIK described in Equation 7.15. The user in that group is will be simply verified by:

$$boolean\ granted = UIK_n \rightarrow g_z \oplus SG_y \rightarrow g \tag{7.16}$$

Assigning a group identification key as a gateway value serves the same goal $RBAC$ is serving without differentiating between users.

7.5.1.3. Assigning Access Based on Users Bit Vector

As described by Figure 7.6, each group has a set of fixed number of possible users. Assuming a group of three users , the bit vector $|v| = 3$ where each user has a place holder such that $v_{u_n} = 1$ if the user have access. otherwise $v_{u_n} = 0$ as shown in Equation 7.17:

$$access_{u_n} = \begin{cases} true \mid v_n = 1 \\ false \mid v_n = 0 \end{cases} \tag{7.17}$$

If a bit vector v grants access to users u_1, u_3 then $v = 101$ will be used as a gateway value for the desired segments to protect. However, granting access based on bit

vectors as gateways will allow users from other groups to access data without being assigned to them. This is because v is a place holder bit vector unrelated to user IDs as stated in Equations 7.14 and 7.15.

7.5.1.4. Group Lookup Table

A solution for the problems mentioned in the previous section is to create a lookup table where entries represent $[Object, Group]$, $[Patient, Group]$, or $set\ v = [group\ vector, users\ vector]$. However, assigning access to groups based on patients will be much less costly that objects assignment since objects are system resources.

- The number of records are large and grow in size each time a new file or record added.
- Objects require continuous maintenance after each operation.

Assignment based on group will be done one time only at patients' admission time.

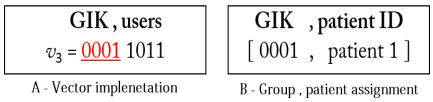


Figure 7.7. Use group ID vs lookup table

Figure 7.7 shows the implementation of single vector (A) and group-patient (B) assignment. By comparing both approaches, we found that a single vector serving as a gateway will:

- Increase space consumed by segment gateway since it will replicate the same information with its maximum size in front of each segment.

- limit the number of groups to the number of bits designated for group IDs. For instance, 4 bits will be 2^4 groups. In the case of having 1000 patients, groups will require 10 bits vector in addition to users vector.

7.5.1.5. Implementation

Figure 7.8 shows a high level design for how to implement $[group, patient]$ assignment and its relation with existing systems such as RBAC. Users in the system still controlled and managed by RBAC; whereas the proposed policy operates as an internal solution to filter what information disclosed based on previous assignment.

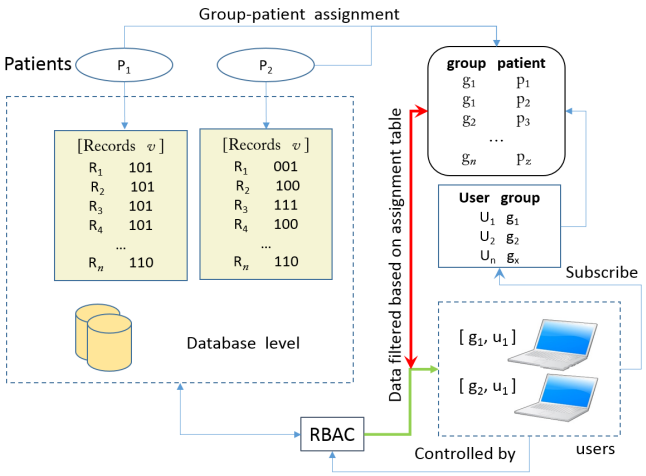


Figure 7.8. High level design for <patinet-group-user-segment> relation

The algorithm in Table 7.5 explains how data will be returned to authorized users. In addition to describing the process of the user-patient assignment process, it also specifies user identification bit, and certificate generation.

7.5.1.6. Revoking Access Rights

1	define *bit vector* $\|v\|$ = max number of users/group
2	$\forall patients : p_i$ define group g_i with unique ID gik_i
3	define user certificate as $u_{cert} = \{gik, uik\}$
4	user-patient assignment $u_x, p_i \ \forall x, i$: 1- set group: $u_x \xrightarrow{joins} g_i$ 2- set ID: a- set new $v'_i = v$ b- from 0 to $\|v\|$: $find(v[j]) = 0$ c- set $v'_i[j] = 1$ d- set $uik_x = v'_i$ e- set $u_{x\ cert^{P_i}} = \{gik_i, uik_x\}$
5	demanding access: a- $u_x \xrightarrow{submit} u_{x\ cert^{P_i}}$ b- $(matchx\ cert^{P_i} \rightarrow g, g_i) \stackrel{?}{=} true$ go to 6 else ignore
6	segment level access: return $S_k \&\& (v'_i \& sg_k)$

Table 7.5. Users' registration and granting access process

Revoking access rights from a user can be performed in three different ways. However, the cost will vary depending on which method is used.

1. Selective access rights revoking (segment level). The first method of revoking access rights is from the database or segment level by flipping the value for the bit representing the user in the SG bit vector from 1 to 0. For example, flipping bit number 1 in any segment gateway will prevent any user in position 1 in any group accessing that segment from accessing the specified record. For selective access within a group, this is the only way to prevent a certain user from accessing a certain record.

2. Revoking all access rights. Revoking all access rights from a user in all records can be done in one of the following ways or both at the same time.

   (a) Invalidate users' certificate used to access patients' data. This solution does not require database update. However, if the user requests access after the certificate is revoked, a new certificate should be issued to the user.

   (b) Delete user entry from $[user - group]$ lookup table. Since it is the only tie between patient and user, deleting user entry will deassign the user to the patient. However, the user will still hold the old certificate and can get access again by reassigning him to the patient.

Providing different revocation policies allows more flexibility in managing access rights in different levels.

7.5.2. Scalability and Complexity

Depending on its implementation, fine-granularity access controlling can introduce space consumption and might increase access time.

7.5.2.1. Scalability

The implementation of the proposed access control policy allows it to serve large data sets without interfering with previous security implementations. Scalability can be studied in terms of:

1. The number of users in a group. Depending on system settings, the size of the bit vector v is tied to the maximum number of users expected to request access to a record concurrently. Each user will be represented by 1 *bit* place holder in a bit vector v. For a group of 8, $|v| = 8$, and 2 *bytes* for 16 users.

2. The number of groups in the system. There is no upper limit for how many groups the proposed policy can support. Each patient is represented by one entry in the [*groups_patients*] lookup table.

3. Scaling for large databases. The overhead introduced by the proposed schema affects active patients only and does not reflect on inactive patients or history files that are not accessible.

7.5.2.2. Space Complexity

A main concern in fine-granularity access control policy is space usage. The proposed policy introduced $[group - user]$ lookup table , $[group - patient]$ lookup table, user certificates , and bit vectors v for each patient record. For the new data structures introduced, the amount of space used for each category is constant, has a known upper bound, or increases in a linear fashion within boundaries. For a facility with R members providing healthcare services, maximum assignment limit of m, an average number of records per patient in the system S, and number of active patients t, where T is the commutative number of patients in the system.

- $[group - user]$ lookup table: number of entries for this table is $\leq t * R$ since each patient is represented by a group ID.

- $[group - patient]$ $entries = t$ for active patients only.

- User certificates: for each patient-user association, there should be a certificate issued to that user. Total space consumption is given by Relation 7.18:

$$c \leq R * m \tag{7.18}$$

- Bit vector entries: number of entries $e_v = S * t$. In its worse case, entries can be persistent for each record $e_v = S * T$.

7.5.2.3. Possible Improvements

The policy design of key layering, can be improved by focusing on the key presentation methodology for both v and GIK. To elaborate more on compound keys, lookup tables can be avoided if the bit vector v is developed to represent both users and groups uniquely. However, the process should be collision resistant.

Another area of development is providing a database wrapper to eliminate physical storage for the bit vector v in the database. The wrapper works like an interface between the database conceptual layer and fine-granularity filter.

To date, the proposed policy has not been implemented practically in areas such as electronic medical records, financial system, mobile device security , or for hardware devices to moderate and manage access rights.

7.6. Summary

The approach gives data owner, which can be different from the data custodian, from applying his or her own privacy protection layer on top of the institute access controlling policy. Another advantage the presented policy provide is data interoperability in a secure fashion. Data can be exchanged and used outside its environment and be accessed only if the proper set of keys is exchanged. A partial key delivery will provide partial access rights. The policy minimized the need of key redistribution through using bit vector v independent from user identity, however, it provided the ability to modify access rights dynamically without informing the key holders or the users.

To be able to use the policy efficiently, data should be categorized logically through refining data, tagging, and structural presentation. In some areas, such as the medical field, providers who use the HL7 standard can easily use the fine-granularity policy. Other areas use XML or KML files also will be able to implement the policy as well.

Chapter 8

IMPLEMENTATION

This chapter demonstrates the implementation of fine-granularity access controlling policy in a simulated small system. The implementation covered a simulated RBAC control access system where users are assigned to roles. A set of operations-objects defined in the system and rights granted to the roles in the system.

The implementation provides a full control panel for the RBAC system to manage roles, files, operations, and roles. The fine-granularity access control policy implemented with RBAC and both policies are running in the same time. Two different views provided to demonstrate how the developed fine-granularity policy can operate and implement patient's policy without conflicting with RBAC policy.

Another software has been developed to survey security policy for healthcare providers. The provided software uses NIST survey in the HSR toolkit to check for hospitals compliance level with the standards. Survey implementation provides the ability to start new survey, specify the importance and priorities in standards implementation in the hospital, store a survey, extract security token, import security token from a different hospital, compare forging policy with local policy and show the differences in a graphical and analytical view.

8.1. Survey Implementation

Chapter 4 covers the theoretical aspect of risk assessment and its importance in building a trust relationship between entities before exchanging critical data. We have provided a software similar to the HSR toolkit, NIST HIPAA Security Toolkit, with improvements. These improvements enable hospitals to create internal policy, create a security certificate, and evaluate incoming certificates based on local metrics.

The tool developed can establish a weighted tree within the survey to help in quantifying compliance level. Figure 8.1shows the software developed for preparing the survey and assigning weights to each branch of the survey. It also shows a graphic representation of the progress on completing the survey.

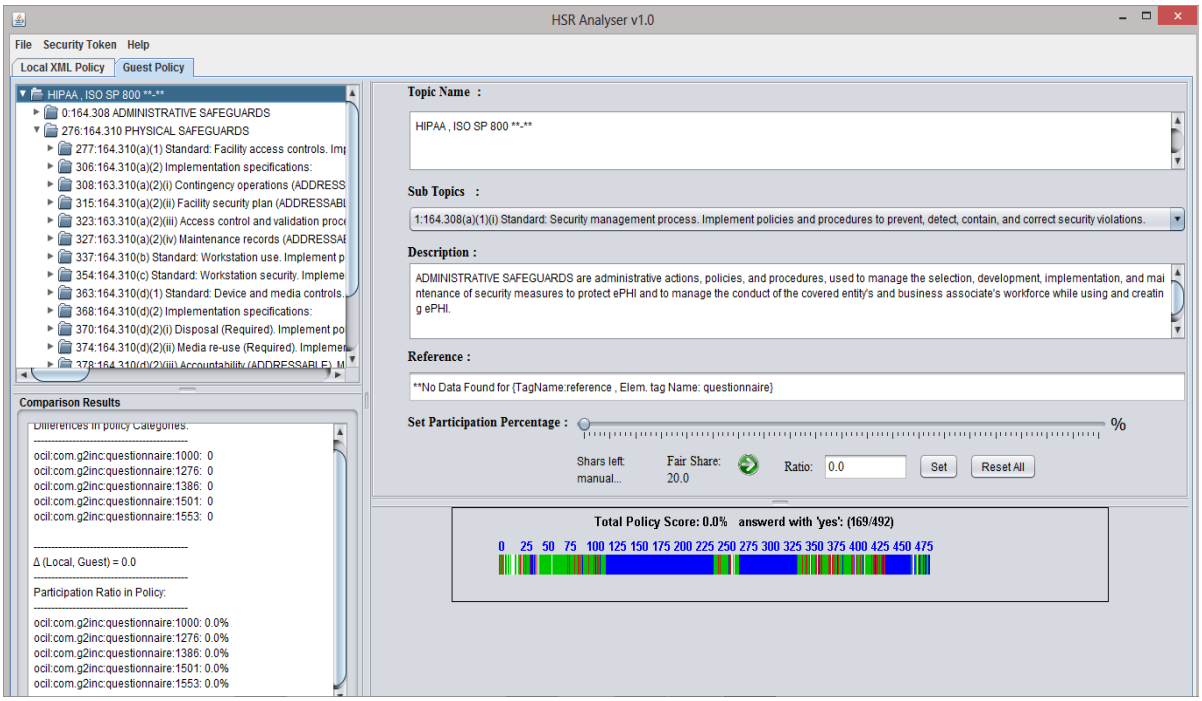


Figure 8.1. Main category setup in a survey

Figure 8.2 shows another step in setting up business priorities for the local entity which is importance level. This is an internal policy and not transferable with security

token. However, any policy exchange token received will be evaluated based on local settings. Evaluating remote policies under local business goals provides a better understanding of common factors that can be found between communicating entities.

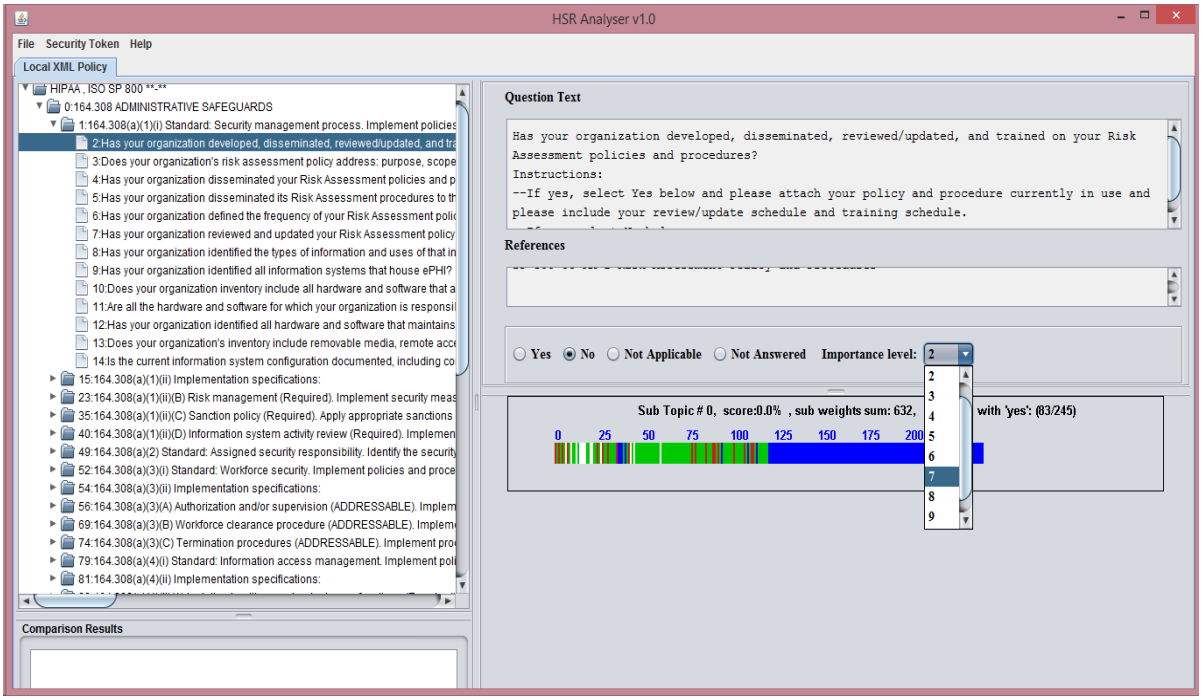
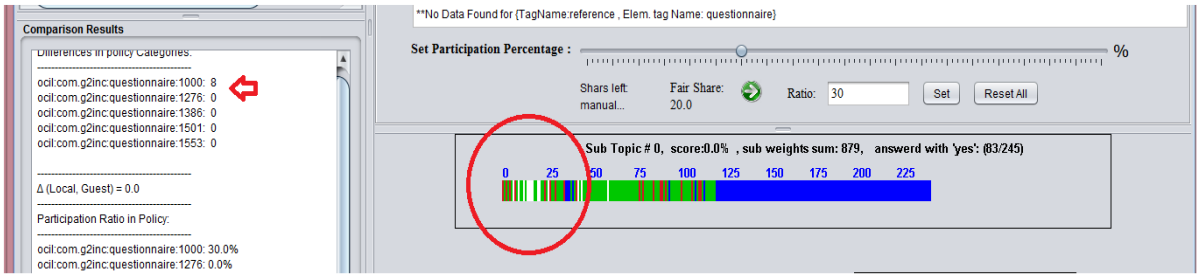


Figure 8.2. Questionnaire importance level setup

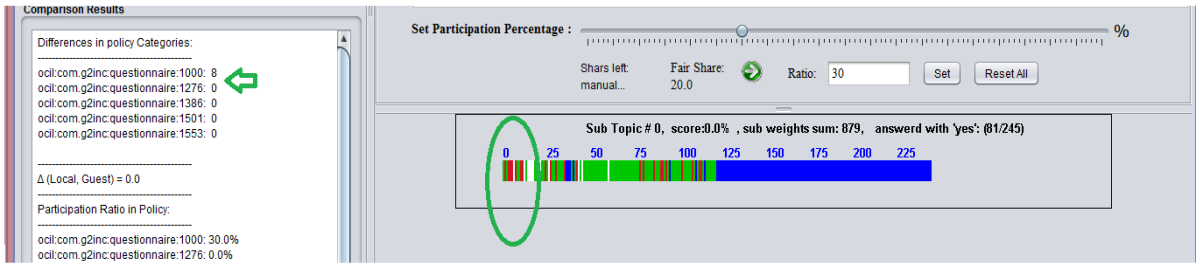
For incoming security tokens, the evaluation is done side-by-side in the local system showing a graphic representation of differences and a summary for those differences. Using the evaluation results, it becomes possible for the security framework to determine the level of trust in one direction. The other side will perform the same operation and get its own level of trust value.

Figure 8.4 shows the incoming security token and the local security token used to build the graphic presentation and derive a local evaluation for a foreign policy

A security token can be stored and forwarded. Each token has a version, issue data, serial ID, and a validity period. However, it does not carry any information



A- incoming/imported policy from outside entity evaluation



B- local policy evaluation

Figure 8.3. Evaluating local and incoming policies and showing differences

about category weight or the importance level of any question because it represents internal policy. We assume that this token will be encrypted when exchanged and can be extracted by the receiving party using some known key to both entities.

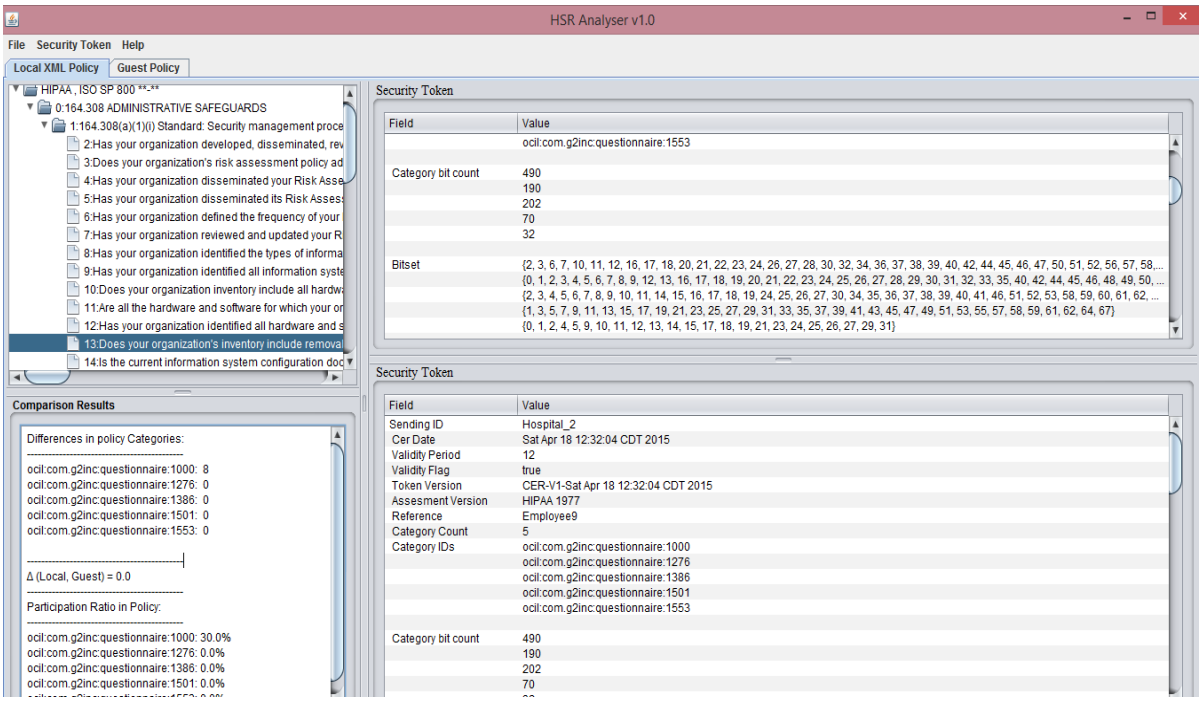


Figure 8.4. Extracted local security token and the received token

8.2. Fine-Granularity Access Control Implementation

Chapter 7.1 discusses fine-granularity access controlling and its implementation in several techniques. It also described the ability to selectively grant access rights based on patients preferences. This section provides an implementation for fine-granularity access control policy on top of RBAC. Both implementations are provided with the control panels to demonstrate how it works.

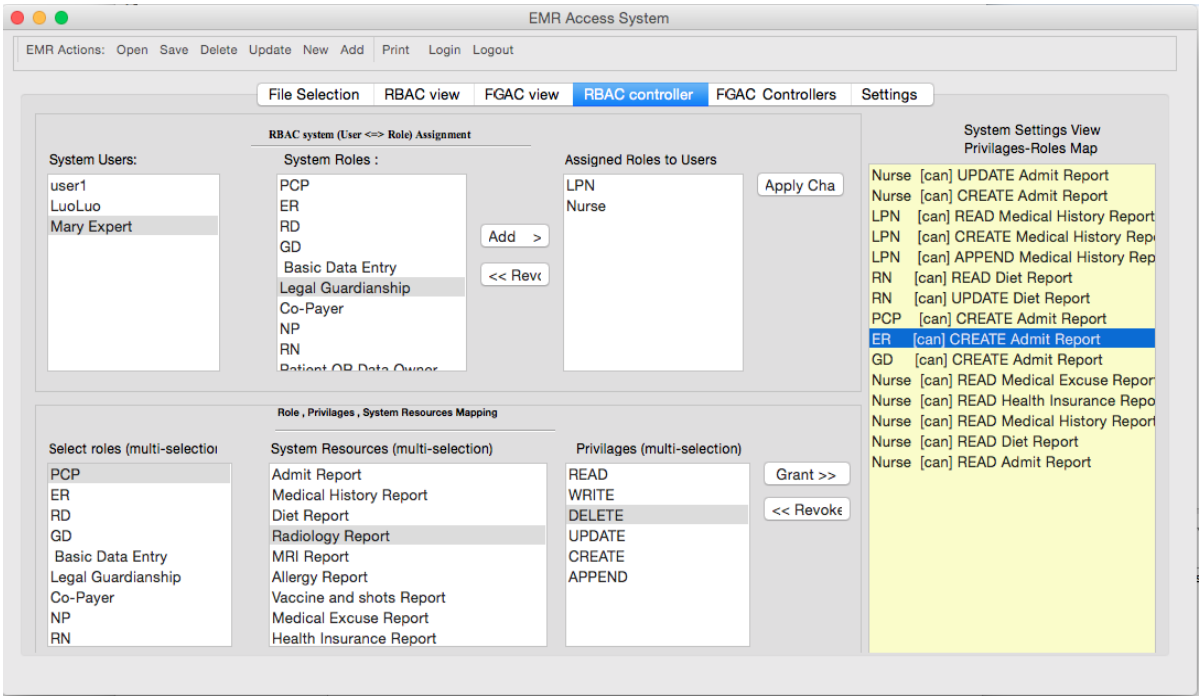


Figure 8.5. RBAC control panel

Figure 8.5 shows the implementation of RBAC system. The implementation provides the ability to assign roles to users, and assign privileges to roles. It also provides the functionalities for checking whether a certain role has the proper permissions to perform any operation on an object, such as updating or creating an admit report or request.

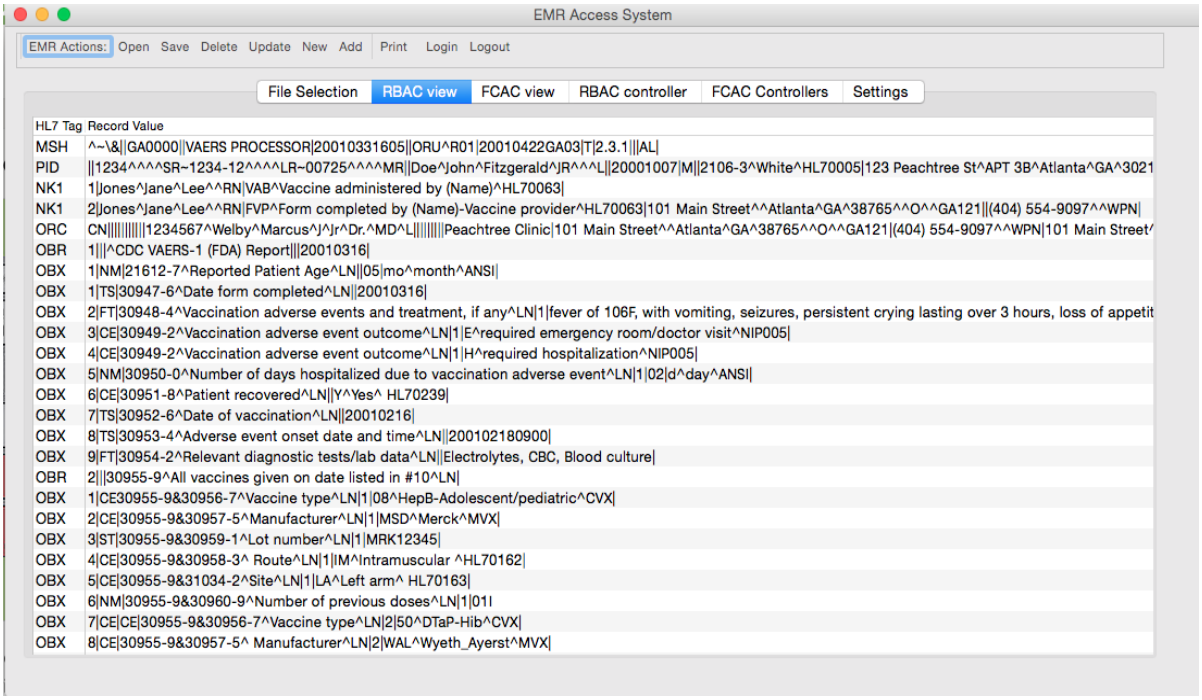


Figure 8.6. Viewing a report under RBAC control only

When an authorized user tries to access a record, RBAC checks for the role of that user if it has been assigned the permission for that operation. RBAC in this stage does not look into other settings like user settings or fine-granularity access rights. Regardless of users preferences, which is already set in this test case, Figure 8.6 shows all records displayed to the user.

However, under the proposed fine-granularity access control policy operating within RBAC, Figure 8.7 shows some records where the user is not authorized to view. The decision to protect those records has been taken by the fine-granularity policy based on user preferences.

Figure 8.8 shows the control panel of the fine granularity access control policy. The user can select who can see what. The system provides the structure of α set based on the number of users taking care of the patient.

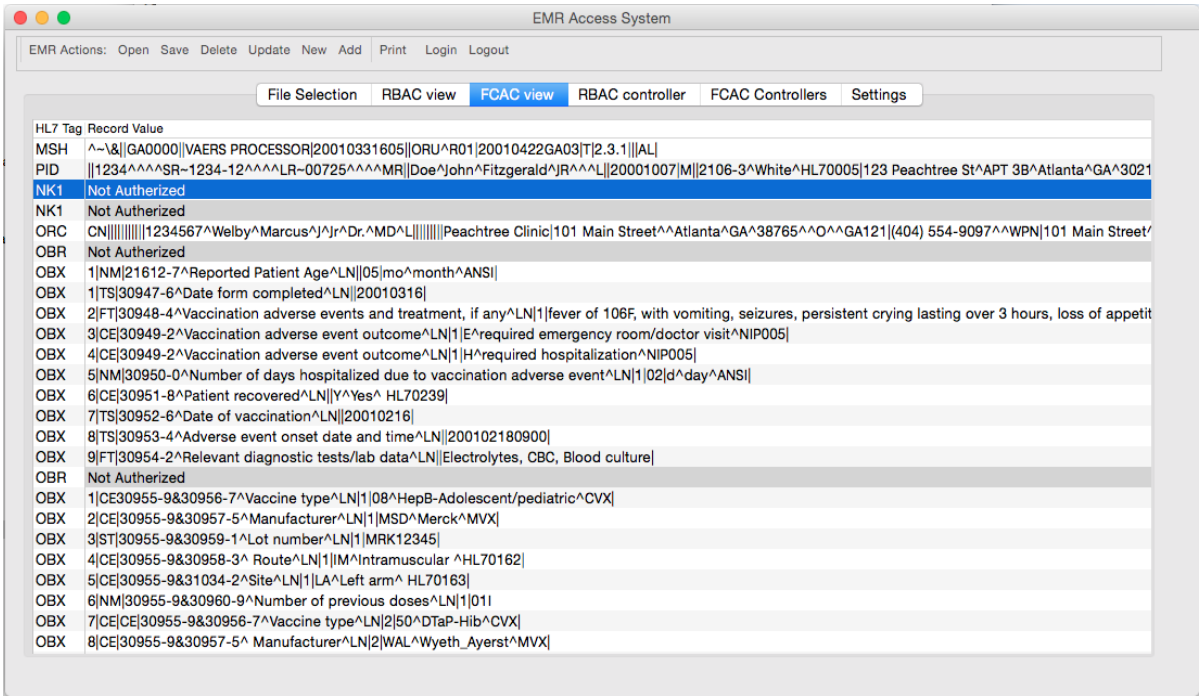


Figure 8.7. Viewing a report under RBAC control with fine granularity policy implemented

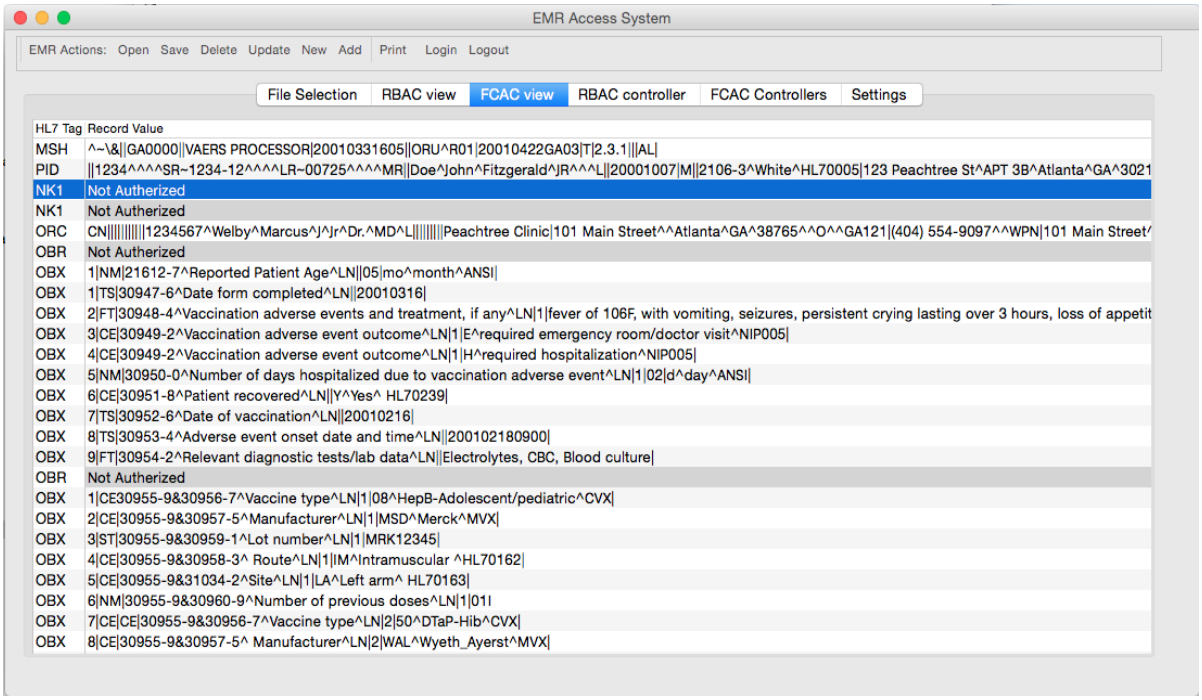


Figure 8.8. Fine-granularity access control policy, user selectivity

8.3. Summary

Implementing a fine-granularity access control policy on top of existing RBAC provides practical proof that the theory does offer extra services for more privacy. However, the implementation cycle has shown the need to develop more implementation standards and publications with proper guidance. A well-documented resource for segmentation can be found in HL7 standards, work groups like S&I and their framework, and through tools like the Security Content Automation protocol, provided by NIST.

Chapter 9

SUMMARY AND CONCLUSIONS

9.0.1. Summary

The proposed framework explains the importance of personal information and the need to implement sophisticated and novel access control policies to protect privacy. This need came from the fact that data owners or data subjects should have the ability to perform a role in their own privacy protection.

The case study of protected health information (PHI) and electronic medical records (EMR) proves the need for such policies even in a closed, trusted environment such as hospitals. The nature of attack from the honest-but-curious adversary model showed how a well-known policy, such as RBAC, will not be able to provide enough privacy protection for private medical information.

This framework introduces a solution for information sharing problem between covered entities who might exchange information and specified by HIPAA standards and regulations. It discusses facility evaluation and its compliance with standard and how that can be used to build a trust relationship in the network. Also, the framework describes the method of policy exchanging without revealing internal business goals and priorities.

The communication protocols and data structure provided allows information exchange under the governance of trust association. That association enables entities to filter private data before transmission to comply with trust level between entities. Taking into consideration the previous achievements, our future work will investigate the following areas:

1. Internal policy evaluation
2. Policy exchange and communication protocol
3. Filtering data based on trust association
4. Selective fine-granularity access control policy over RBAC

Finally, we provided an implementation for the cryptographic fine granularity policy and a tool for surveying security metrics in hospitals. The policy provided has many iterations and applications, depending on the host environment.

9.0.2. Future Work and Possible Improvements

The framework developed in this dissertation covers the major difficulties in privacy protection and provides a practical solution for those areas. However, there are some aspects of improvement and development the framework developed in this dissertation covers the major difficulties in privacy protection and provides a practical solution for those areas. However, there are some aspects of improvement and development.

In the area of policy matching and communication, more surveys need to be collected and analyzed from different sources. The result of the analysis can be used in the dependency network to be able to match compliance level with different standards as described in Risk Assessment chapter, section 4.2.3. Based on the result of surveys, multi-level security filters for data exchange can be tested and optimized to match user preferences. Building privacy filters requires more investigation and research in auxiliary information and data mining to avoid disclosing protected information.

The framework implementation can be improved and developed in a form of libraries, plugins, and information wrappers for easy integration with running systems. For example, the segment gateways can be implemented as a wrapper hiding the internal structure of the medical record. This improvement will allow EMR systems to use the wrapper regardless the internal structure of the file or the standards used other than HL7.

The proposed framework is not limited to the area of medical records privacy protection, it can be implemented in many other areas where a- data has different levels of security b- many users can access the same information c- low level of certificate exchange needed. Some of those areas are financial records, mobile devices, operating hardware with different privileges and different resources.

Another area of research is HBC adversary detection model. The proposed framework does not provide a method to detect and measure HBC attacks in complex environments like medical records. It is possible to integrate the prevention and detection to provide better dynamic adjustment to the prevention system. The calibration process should enhance privacy protection when a learning system of the two components works together.

9.0.3. Conclusion

The proposed framework integrates many areas of privacy protection and draws a roadmap for future development in the field of privacy protection. It has been proven that users privacy can be enhanced even in complex environments, such as electronic medical records, with the presences of adversary model hard to detect like HBC attacks. It is clear that more research and work need to be done in data classification and segmentation for better security and privacy protection. Areas of improvement and development in many areas still open as mentioned in the previous section 9.0.2.

REFERENCES

[1] PHIprivacy.net Medical Privacy: Who's losing your data? Who's selling it? . http://www.phiprivacy.net/, March 2013.

[2] Aggarwal, G., and Feder, T. Approximation algorithms for k-anonymity. *Journal of Privacy ...* (2005), 1–18.

[3] Ajayi, O., Sinnott, R. O., and Stell, A. Dynamic trust negotiation for flexible e-health collaborations. In *Mardi Gras Conference* (Baton Rouge, Louisiana, USA, Ajayi2008), p. 8.

[4] Appel, J. M. Why shared medical database is wrong prescription, December 2008. Website: http://articles.orlandosentinel.com/2008-12-30/news/OPappel30_1_medical-records-medical-system-electronic-medical. Access date: February 2013.

[5] Bardram, J. E., Kjr, R. E., and Pedersen, M. . Context-aware user authentication supporting proximity-based login in pervasive computing. In *In Proceedings of the 5th International Conference on Ubiquitous Computing (UbiComp 2003)* (Seattle,WA, USA, October 2003), S. Berlin, Ed., vol. 2864/2003 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 107–123.

[6] Beeler, G., Huff, S., Rishel, W., Shakir, A.-m., and Walker, M. Message development framework. *HL7, Inc*, December (1999).

[7] Berler, A., Spyrou, S., Monochristou, E., Tolias, Y. A., Konnis, G., Magglaveras, N., and Koutsouris, D. Risk assessment in integrated regional healthcare networks. In *Interoperability & Security in Medical Information Systems* (May 2007), F. Makedon and J. Ford, Eds., vol. 2, The Electronic Journal for E-Commerce Tools & Applications (eJETA), eJETA.org.

[8] Bertino, E., Bonatti, P. A., and Ferrari, E. Trbac: A temporal role-based access control model. *ACM Trans. Inf. Syst. Secur. 4*, 3 (2001), 191–233.

[9] Bertino, E., and Catania, B. GEO-RBAC: a spatially aware RBAC. *Proceedings of the tenth ...* (2005), 29–37.

[10] Bhatti, R., Moidu, K., and Ghafoor, A. Policy-based security management for federated healthcare databases (or rhios). In *HIKM* (Sheraton Crystal City Hotel, Arlington, VA, November 2006), International Workshop on Health Information and Knowledge Management (HIKM 2006), pp. 41–48.

[11] Biham, E., and Shamir, A. Differential fault analysis of secret key cryptosystems. In *CRYPTO '97* (1997), pp. 513–525.

[12] Bilal Alqudah, S. N. *Biomedical Engineering: Health Care Systems, Technology and Techniques*, 1 ed. Springer, August 2011.

[13] Bonatti, P. A. Rule languages for security and privacy in cooperative systems. In *COMPSAC (1)* (2005), IEEE Computer Society, pp. 268–269.

[14] Boneh, D., DeMillo, R. A., and Lipton, R. J. On the importance of checking cryptographic protocols for faults. In *Advances in Cryptology - Eurocrypt 97* (1997), Springer-Verlag, pp. 37–51.

[15] Božović, V., Socek, D., Steinwandt, R., and Villányi, V. I. Multi-authority attribute-based encryption with honest-but-curious central authority. *International Journal of Computer Mathematics 89*, 3 (2012), 268–283.

[16] Brandner, R., van der Haak, M., Hartmann, M., Haux, R., and Schmcker, P. Electronic signature for medical documents integration and evaluation of a public key infrastructure in hospitals. In *Methods of Information in Medicine* (2002), vol. 41, pp. 321 – 330.

[17] Capgemini. http://www.capgemini.com/, March 2010.

[18] Castro, D. Improving health care: Why a dose of it maybe just what the doctors ordered. Tech. rep., ITIF, October 2007.

[19] CHAO, H.-M., TWU, S.-H., and HSU, C.-M. A patient-identity security mechanism for electronic medical records during transit and at rest. *Medical Information and The Internet in Medicine 30*, 3 (September 2005), 227–240.

[20] Claredi, I. C. X12 de-identification utility. Software, Ingenix Claredi Classic, 2003.

[21] Clearinghouse, P. R. Medical privacy in the electronic age, April 2013. Website: https://www.privacyrights.org/fs/fs8a-hipaa.htm. Access date: February 2013 , Revised February 2013.

[22] Clearinghouse, P. R. Medical privacy in the electronic age, April 2013. Website: https://www.privacyrights.org/fs/fs8-med.htm. Access date: October 2013 , Revised April 2013.

[23] Dekker, M. A. C., Crampton, J., and Etalle, S. Rbac administration in distributed systems. In *SACMAT* (2008), pp. 93–102.

[24] Department of Defense. *Trusted Computer System Evaluation Criteria*. Dec. 1985.

[25] Dhem, J. F., Koeune, F., Leroux, P. A., Mestr, P., Quisquater, J. J., and L.Willems, J. A practical implementation of the timing attack. *CARDIS* (1998).

[26] Dimitropoulos, L. Privacy and security solutions for interoperable health information exchange: Phase ii. In *The AHRQ Health IT web conference* (2008).

[27] DoD. Trusted computer system evaluation criteria, DECEMBER 1985.

[28] Elisa Bertino, Elena Ferari, A. C. S. Trust-x: A peer-to-peer framework for trust establishment. *IEEE Transactions in Knowledge and Data engineering* (October 2003).

[29] Emam, K. E. Heuristics for de-identifying health data. *IEEE Security & Privacy 6*, 4 (2008), 58–61.

[30] Emam, K. E., and Kamal, F. Protecting privacy using k-anonymity. *Journal of the American Medical Informatics Association 15*, 5 (August 2008), 627–637.

[31] FairWarning, Inc. Privacy Surveillance in Healthcare. Tech. rep., FairWarning, Inc., 2008.

[32] Fernandez, E. B., and Sorgente, T. An analysis of modeling flaws in hl7 and jahis. In *SAC* (2005), pp. 216–223.

[33] Ferraiolo, D. F., and Kuhn, D. R. Role-based access controls. In *15th National Computer Security Conference* (National Institute of Standards and Technology, Technology Administration, U.S. Department of Commerce , Gaithersburg, Md. 20899 USA, 1992), pp. 554 – 563.

[34] Ferraiolo, D. F., Kuhn, D. R., and Chandramouli, R. *Role-based access control*. Computer Security Series. Artech House, 2003.

[35] Foreman, J. At risk of exposure, June 2006. Website: http://articles.latimes.com/2006/jun/26/health/he-privacy26. Los Angeles Times Access date: February 2013.

[36] Fotiou, N., Arianfar, S., Särelä, M., and Polyzos, G. C. A framework for privacy analysis of icn architectures. In *Privacy Technologies and Policy*. Springer, 2014, pp. 117–132.

[37] Fridsma, D. The standards & interoperability (s&i) framework, April 2015. Website: http://wiki.siframework.org/. Access date: April 2015.

[38] Friedlin, F. J., and McDonald, C. J. A software tool for removing patient identifying information from clinical documents. *Journal of the American Medical Informatics Association 15*, 5 (oct 2008), 601–610.

[39] Gafurov, D., Helkala, K., and Svendsen, N. K. Security models for electronic medical record. *Telektronikk* (2005).

[40] Galpottage, P. A. B., and Norris, A. C. Patient consent principles and guidelines for e-consent: a new zealand perspective. *Health Informatics Journal 11*, 1 (2005), 5–18.

[41] Goldstein, M., and Rein, A. Consumer consent options for electronic health information exchange: policy considerations and analysis. *Office of the National Coordinator for Health IT* (2010).

[42] Goldstein, M., and Rein, A. Data segmentation in electronic health information exchange: Policy considerations and analysis. *Washington DC: Office of the National Coordinator for Health IT* (2010).

[43] Hamilton Allen Booz. Medical Identity Theft Environmental Scan. Tech. rep., Booz Allen Hamilton for the US Department of Health and Human Services, 2008.

[44] He, Q., and Antón, A. I. Requirements-based access control analysis and policy specification (recaps). *Information & Software Technology 51*, 6 (2009), 993–1009.

[45] HealthIT. Data segmentation, April 2015. Website: http://www.healthit.gov. Access date: April 2015.

[46] HHS. US Department of Health and Human Services. Web site, March 2014.

[47] HIPAA. Security 101 for covered entities, 11 2005.

[48] HIPAA. Security standards: Technical safeguards, 11 2005.

[49] Hopper, N. J., and Blum, M. Secure human identification protocols. In *ASIACRYPT 2001, LNCS 2248* (2001), Springer-Verlag, pp. 52–66.

[50] Hsiao, C.-J., Hing, E., and Ashman, J. Trends in electronic health record system use among office-based physicians: United states, 2007-2012. *Nat Health Stat Rep 75* (2014), 1–17.

[51] Hu, V. C., Ferraiolo, D. F., and Kuhn, D. R. Assessment of access control systems. *National Institute of Standard and Technology* (2006), 60.

[52] International, H. L. S. Health level seven international, October 2008. Website: http://www.hl7.org. Access date: February 2013.

[53] Joshi, J., Bertino, E., Latif, U., and Ghafoor, a. A generalized temporal role-based access control model. *IEEE Transactions on Knowledge and Data Engineering 17*, 1 (Jan. 2005), 4–23.

[54] Junzhe, and Weaver, A. C. A dynamic, context-aware security infrastructure for distributed healthcare applications. In *Pervasive Security, Privacy and Trust (PSPT 2004)* (University of Virginia , Charlottesville, VA 22904, 2004).

[55] Kandasamy, V., and Papitha, E. Flexible access control for outsourcing personal health services in cloud computing using hierarchical attribute set based encryption. *2013 International Conference on Information Communication and Embedded Systems (ICICES)* (Feb. 2013), 569–571.

[56] Kelsey, J., Schneier, B., Wagner, D., and Hall, C. Side channel cryptanalysis of product ciphers. In *ESORICS 98* (1998), September, Springer-Verlag, pp. 97–110.

[57] Kerr, K., and Norris, T. Telehealth in new zealand: Current practice and future prospects. *Journal of Telemedicine and Telecare 10*, 1 (2004), 60–63.

[58] Khan, M. F. F., and Sakamura, K. Context-aware access control for clinical information systems. *2012 International Conference on Innovations in Information Technology (IIT)* (Mar. 2012), 123–128.

[59] Kocher, P., Jaffe, J., and Jun, B. Introduction to differential power analysis and related attacks. *Cryptography Research 1666* (1999).

[60] Kocher, P. C. Timing attacks on implementations of diffie-hellman, rsa, dss, and other systems. In *Proceedings of Advances in CryptologyCRYPTO 96* (1996), Springer-Verlag, pp. 104–13.

[61] Kuhn, D., Coyne, E., and Weil, T. Adding attributes to role-based access control. *Computer*, June (2010), 79–81.

[62] Lebak, J. W., Yao, J., and Warren, S. Hl7-compliant healthcare information system for home monitoring. In *Proceedings of the 26th Annual International Conference of the IEEE EMBS* (Department of Electrical & Computer Engineering, Kansas State University, Manhattan, KS, USA, 2006).

[63] McLean, V. Electronic Health Records Overview.". *... Mitre. National Institutes of Health National Center for ...*, April (2006).

[64] Messerges, T. S., Dabbish, E. A., and Sloan, R. H. Investigations of power analysis attacks on smartcards. In *USENIX Workshop on Smartcard Technology* (May 1999), pp. 151–61.

[65] Messerges, T. S., Dabbish, E. A., and Sloan, R. H. Power analysis attacks of modular exponentiation in smartcards. *CHES'99, LNCS 1717* (1999), 144–157.

[66] Microsoft. Connected health framework architecture and design blueprint, March 2009.

[67] Miller, A. R., and Tucker, C. E. Privacy protection and technology diffusion: The case of electronic medical records. In *MANAGEMENT SCIENCE* (July 2009), vol. 55, Informs, p. 10771093.

[68] Model, H. R. I. Hl7 reference information model (rim), October 2009. Website: http://www.hl7.org/implement/standards/rim.cfm. Access date: February 2013.

[69] Mooniana, O., Cheerkoot-Jalima, S., Nagowaha, S. D., Khedoa, K. K., Doomuna, R., and Cadersaiba, Z. Hcrbac an access control system for collaborative context-aware healthcare services in mauritius. *Journal of Health Informatics in Developing Countries 2* (2008), 10–21.

[70] Narendra, C. P. Ehrs: Fear of breach? the new zealand public's opinion. Master's thesis, University of Otago (Information Science), 2006.

[71] Nishiki, K., and Tanaka, E. Authentication and access control agent framework for context-aware service. In *Applications and the Internet Workshops, 2005. Saint Workshops 2005. The 2005 Symposium on* (2005), pp. 200 – 203.

[72] of Science, N. I., and Technology. Hipaa security rule tookit, 11 2011.

[73] of the National Coordinator for Health IT, O. Data segmentation in electronic health information exchange: Policy considerations and analysis.

[74] Olson, L. E., Rosulek, M. J., and Winslett, M. A generalized honest-but-curious trust negotiation strategy for harvesting credentials.

[75] PATRICK R GALLAGHER, J. Computer security subsystem interpretation of the trusted computer system evaluation criteria, September 1988.

[76] Peter Garrett, J. S. Emr vs ehr what is the difference?, January 2011. Website: http://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/emr-vs-ehr-difference/. Access date: February 2013.

[77] Power, D. J., Slaymaker, M., and Simpson, A. C. On the construction and verification of self-modifying access control policies. In *Secure Data Management* (2009), pp. 107–121.

[78] Pritts, J. The state of health privacy: an uneven terrain: a comprehensive survey of state health privacy statutes. *Institute for healthcare research and policy , Georgetown University 2*, 2 (1999).

[79] Rahn, R. W., and de Rugy, V. Threats to financial privacy and tax competition. Tech. Rep. 491, Cato Institute, October 2003.

[80] Ratha, N. K., Connell, J. H., and Bolle, R. M. An analysis of minutiae matching strength. *AVBPA '01 Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication* (2001), 223–228.

[81] Reed-Fourquet, L., Lynch, J. T., Martin, M. K., Cascio, M., Leung, W.-Y., and Ruenhorst, P. P. Managing information privacy & security in healthcare the chime-trust healthcare public key infrastructure and trusted third party services: A case-study. Case study, Healthcare Information and Management Systems Society (HIMSS), Jan 2007. CHIME Inc., Wallingford Connecticut.

[82] Rindfleisch, T. C. Privacy, information technology, and healthcare. In *COMMUNICATIONS OF THE ACM* (1997), vol. 40, pp. 92–100.

[83] Roy, S., and Chuah, M. Secure data retrieval based on ciphertext policy attribute-based encryption (CP-ABE) system for the DTNs.

[84] Rubik, E. Spatial logical toy, 1983.

[85] Ryan, A. M., and Williams, R. General k-anonymization is hard. Tech. rep., In Proc. of PODS04, 2003.

[86] Schneider, J. K. Positive outcomes implementing biometrics in multiple healthcare application. In *TEPR 2001 Conference and Exhibition* (2001).

[87] Schwartmann, D. An attributable role-based access control for healthcare. In *International Conference on Computational Science* (2004), vol. 3039, pp. 1148–1155.

[88] Shamir, A. Identity-based cryptosystems and signature schemes.

[89] Sharona Hoffman, A. P. Case western reserve university professors call for regulation of electronic health records, October 2008. Website: http://blog.case.edu/case-news/2008/10/30/ehrregulation. Access date: February 2013.

[90] Smith, B., and Ceusters, W. Hl7 rim: An incoherent standard. In *Studies in Health Technology and Informatics* (University of Buffalo , NY. USA, 2006), vol. 124, p. 133138.

[91] Stell, A., Sinnott, R., and Ajayi, O. Secure, reliable and dynamic access to distributed clinical data. In *Proceedings of the LSGRID2006: Yokohama* (University of Glasgow, National e-Science Centre, Glasgow, G12 8QQ, UK, 2006).

[92] Stingl, C., and Slamanig, D. Privacy-enhancing methods for e-health applications: how to prevent statistical analyses and attacks. *IJBIDM 3*, 3 (2008), 236–254.

[93] Sweeney, L. *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, 1 ed. Elsevier Science, Washington, DC, November 2001.

[94] Sweeney, L. ACHIEVING k-ANONYMITY PRIVACY PROTECTION USING GENERALIZATION AND SUPPRESSION. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10*, 05 (Oct. 2002), 571–588.

[95] Sweeney, L. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10*, 5 (may 2002), 557–570.

[96] Sweeney, L. Patient Identifiability in Pharmaceutical Marketing Data. *Harvard University, Cambridge, MA, WP-1015* (2011), 1–22.

[97] Privacy Choices for Your Personal Financial Information. http://www.ftc.gov/bcp/edu/pubs/consumer/credit/cre29.shtm. *Federal Trade Commission* (March 2010).

[98] Smart Card Alliance . http://www.smartcardalliance.org/pages/smart-cards-applications-healthcare. *Healthcare Applications* (Nov 2010).

[99] Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure and Mathematics, and Applications Commission on Physical Sciences

and National Research Council. *For the Record: Protecting Electronic Health Information*. National Academies Press, 1997.

[100] DE-ID Data Corp. http://www.de-idata.com/.

[101] http://www.safety-and-security.de/. Safety and security systems in europe. *Authentication System based on Network ID Cards for Critical Environments* (2010).

[102] Office of Civil Rights. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf.

[103] The American Hospital Association Annual Survey. http://www.aha.org/research/rc/stat-studies/fast-facts.shtml.

[104] Undem, T. Consumers and health information technology: A national survey.

[105] van Eck, W. Electromagnetic radiation from video display units: An eavesdropping risk. *Computers and Security 4* (1985), 269–286.

[106] Varshney, U. *Pervasive Healthcare Computing: EMR/EHR, Wireless and Health Monitoring*. Springer, 2009.

[107] Vawdrey, K., D., Sundelin, L., T., Seamons, E., K., Knutson, and D., C. Trust negotiation for authentication and authorization in healthcare information systems. In *Engineering in Medicine and Biology Society* (2003).

[108] Weaver, A. D., J., S., Snyder, I., Dyke, A. M. V., Hu, J., Chen, J., Mulholland, X., and T. Marshall, A. Federated, secure trust networks for distributed healthcare it services. In *Industrial Informatics, 2003. INDIN 2003. Proceedings. IEEE International Conference on* (Aug 2003), IEEE, pp. 162 – 169.

[109] Win, K. T., Phung, H., Young, L., Tran, M., Alcock, C., and Hillman, K. Electronic health record system risk assessment: a case study from the minet. *Health Information Management 32* (2004), 43–48.

[110] Wu, Z., and Weaver, A. C. Dynamic trust establishment with privacy protection for web services. In *ICWS* (2005), pp. 811–812.

[111] Yang, Y., Deng, R. H., and Bao, F. Fortifying password authentication in integrated healthcare delivery systems. In *ASIACCS* (2006), pp. 255–265.

[112] Zhang, R., and Liu, L. Security Models and Requirements for Healthcare Application Clouds. *2010 IEEE 3rd International Conference on Cloud Computing* (July 2010), 268–275.