Scientific Research Publishing

E-Health Telecommunication Systems and Networks, 2018, 7, 1-42
http://www.scirp.org/journal/etsn
ISSN Online: 2167-9525
ISSN Print: 2167-9517

Privacy Preserving Risk Mitigation Approach for Healthcare Domain

Shaden S. Al Aqeeli, Mznah A. Al-Rodhaan, Yuan Tian, Abdullah M. Al-Dhelaan

Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia
Email: rodhaan@ksu.edu.sa

How to cite this paper: Al Aqeeli, S.S., Al-Rodhaan, M.A., Tian, Y. and Al-Dhelaan, A.M. (2018) Privacy Preserving Risk Mitigation Approach for Healthcare Domain. *E-Health Telecommunication Systems and Networks*, **7**, 1-42.
https://doi.org/10.4236/etsn.2018.71001

Received: February 14, 2018
Accepted: March 26, 2018
Published: March 29, 2018

Abstract

In the healthcare domain, protecting the electronic health record (EHR) is crucial for preserving the privacy of the patient. To help protect the sensitive data, access control mechanisms can be utilized to restrict access to only legitimate users. However, an issue arises when the authorized users abuse their access privileges and violate privacy preferences of the patients. While traditional access control schemes fall short of defending against the misbehavior of authorized users, risk-aware access control models can provide adaptable access to the system resources based on assessing the risk of an access request. When an access request is deemed risky, but within acceptable thresholds, risk mitigation strategies can be exploited to minimize the risk calculated. This paper proposes a risk-aware, privacy-preserving risk mitigation approach that can be utilized in the healthcare domain. The risk mitigation approach controls the patient's medical data that can be exposed to healthcare professionals, according to their trust level as well as the risk incurred of such data exposure, by developing a novel Risk Measure formula. The developed Risk Measure is proven to manage the risk effectively. Furthermore, Risk Mitigation Data Disclosure algorithms, $RIMIDI_0$ and $RIMIDI_1$, which utilize the developed risk measures, are proposed. Experimental results show the feasibility and effectiveness of the proposed method in preserving the privacy preferences of the patient. Since the proposed approach exposes the patient's data that are relevant to the undergoing medical procedure while preserving the privacy preferences, positive outcomes can be realized, which will ultimately bring forth quality healthcare services.

Keywords

Access Control, Healthcare, HIPAA, Risk-Aware, Risk Mitigation

1. Introduction

In the healthcare domain, Patients' electronic health records (EHR) [1] contain

sensitive and detailed information regarding their health issues and diagnosis, such as their psychological and mental disorders, abortions, substance abuse and much more. The release or access of such private data by unauthorized entities, whether intentionally or accidentally, can pose serious consequences for those individuals; they could face social judgment and embarrassment, difficulties in getting employed as well as obtaining and maintaining insurance policies [2]. To help overcome such consequences, several legislations and regulation rules have been issued in efforts to maintain privacy and bring patients more control over their data such as the Health Insurance and Accountability Act, HIPAA, legislation [3]. Therefore, the privacy of such medical records has been an important issue and, thus, must be preserved and has been under intensive research [2].

To protect the privacy of the patients, several approaches can be employed. Privacy can be preserved using statistics, cryptography or by policy [4]. In privacy by statistics, anonymization techniques [5] can be utilized to hide the identity of the patients before the data is released to third parties. Privacy by cryptography [6] allows a patient's data to be protected using security principles such as encryption mechanisms. Finally, privacy by policy [7] encompasses employing authentication and authorization rules and constraints that need to be enforced upon access to the health records to preserve the privacy. The techniques can be combined to deliver more robust outcomes as required by the application and system administrators [4].

While containing sensitive information, an EHR of an individual patient must be accessed by staff or healthcare professionals, such as doctors, to deliver an accurate diagnosis of the patient's current condition based on her previously stored information such as her previously diagnosed record [2]. Evidently, delivering such patient a viable diagnosis while preserving her privacy cannot be realized using anonymization techniques because anonymization involves concealing a patient's identity among various other patients and the data is meant to be released for analytical and scientific purposes [5]. In contrast, policy and cryptography mechanisms can be utilized for protecting the patient's privacy since the access to the EHR is protected by authentication, authorization and encryption processes [6]. In effect, only authorized healthcare providers are allowed access to carry out medical procedures.

Several mechanisms can be utilized to protect the patient's sensitive data. Access control (AC) is a major and well-known security technique utilized to limit or restrict access to specific data sets by controlling access rights and privileges to resources [6]. There are several types of access control: discretionary access control (DAC) [8], mandatory access control (MAC) [6], role-based access control (RBAC) [9] and attribute-based access control (ABAC) [10].

Although access control acts as the first line of defense against illegitimate access, an issue arises when the authorized healthcare professionals abuse their access rights to a patient's health records [11]. Such situation puts the sensitive data under increased risk of leakage or exposure as well as goes against the pri-

vacy preferences of the patient. Traditional access control mechanisms fall short of defending against such types of misuse. Such scenario sets the stage for searching for a method to assess the risk associated with the access request to a particular resource.

In efforts to assess the risk associated with a legitimate access request, several metrics can be employed. One of the legitimate measures is calculating the trustworthiness of a user requesting access. Trust can be used as a means to forecast a user's behavior towards a resource by analyzing past behavior [12]. Consequently, the more a user behaves as predicted, the higher his trust degree and, thus, the lower the risk associated with his access request. Such metric can guard against malicious actions of authorized users as well as encourage good behavior towards the system's resources [11].

Since trust values increase and decrease based on users' past behavior, adding a risk assessment element to an access control scheme allows dynamicity and adaptability [13] as opposed to other traditional models. After an access request is assessed and considered risky but within a tolerable interval, risk mitigation strategies can be employed to lower the impact of the associated risk [14]. Several works proposed the incorporation of risk awareness into existing access control models, such as RBAC [15] and ABAC [13], while others have proposed risk-based access control as a new breed of access control techniques [16].

Due to their inherent features, Risk-aware access control models have been under research. Existing works in the literature can be divided broadly into two types of efforts: risk assessment and risk mitigation. In the former, risk assessment investigated the ways to assess the riskiness of an access request before granting access to the resources [17]-[21]; namely, quantifying the risk. In the latter, efforts focused on approaches and strategies that bring down the risk that has already been assessed [23] [24] [25].

To mitigate the risk, several research efforts considered mitigation by imposing a set of obligations [22] that the user needs to fulfill either before or after access is allowed [21]. Obligations are usually a set of rules that the user is expected to behave according to, and can be monitored by the system to evaluate the user's actions towards the resource for future access requests. Alternatively, other studies focused on mitigating the risk by employing security methods such as increasing encryption measures or activating automatic alerts [23].

This research tackles the issue of preserving the privacy preferences of the patients by restricting access to their records using means of risk assessment methods. More specifically, the research aims at proposing a novel risk mitigation approach by searching for a suitable set of patient's relevant data objects that are safe to be exposed to the healthcare professional while maintaining privacy preferences when access in considered risky; of which ultimately can provide helpful insights in delivering quality healthcare services. **Figure 1** shows a medical scenario for the proposed risk mitigation approach.

As illustrated in **Figure 1**, when a doctor requests access to a patient's health record, the risk mitigation approach needs to assess the riskiness of the access

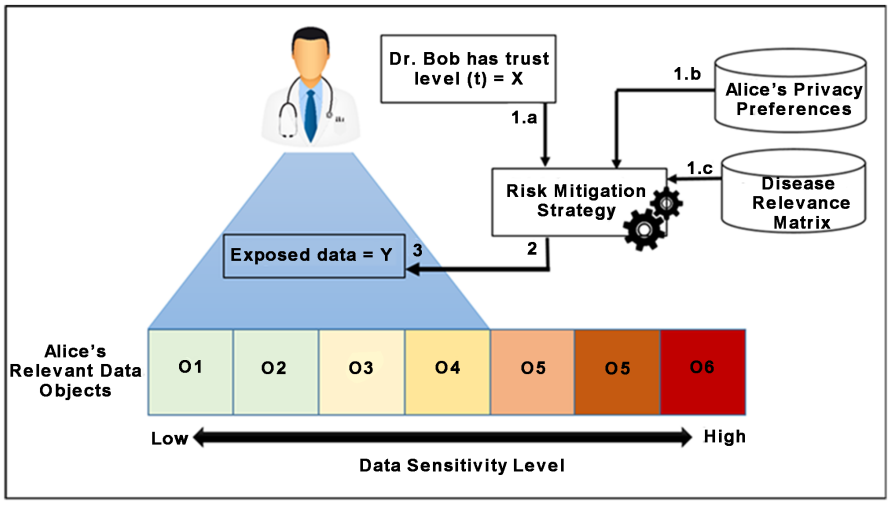

Figure 1. A medical scenario for the proposed privacy preserving risk mitigation method.

request. To assess the risk, the following steps are performed. In the beginning, the trust level of the doctor (1.a), the privacy preferences of the patient (1.b) and the relevance information from the Disease Relevance Matrix (1.c) are all obtained. Such data is considered as an input for the risk mitigation approach (2), which operates and then decides the suitable set of the patient's relevant data that can be exposed to the doctor (3) and which does not undermine the privacy preferences of the patient. The patient's data are associated with sensitivity weights, which represent the privacy preferences regarding each data object. The sensitivity weights are mapped to numerical values scaled from [0, 1] such that for two data objects o_i and o_j, with the corresponding weights w_i and $w_i > w_j$, if $w_i > w_j$ then o_i is considered as more sensitive than o_j and vice versa. These sensitivity weights are assumed to be supplied by the corresponding patients when they fill out the medical forms for their healthcare procedure and are entered into the system by the medical staff.

The health problems can be assumed to be classified according to the ICD-11, which is the eleventh revision of medical classification for diseases and health problems devised by the World Health Organization (WHO) [26]. In this regard, the various health issues can be arranged into groups, which contain further classification into sub-groups. Each health issue is given a code that makes it distinguishable among other health problems.

This research proposes and develops a novel privacy preserving risk mitigation approach for the healthcare domain. The risk mitigation method, based on access risk and data exposure, is highly valuable in situations where the revelation of a subset of the patient's relevant data can provide valuable insights in pursuing medical diagnoses while protecting the privacy preferences. As a result, the proposed approach is HIPAA compliant and dynamic, with risk-aware and privacy preserving properties. The main contributions of the research are as follows:

- A novel risk mitigation approach for the healthcare domain that links data

exposure to the risk incurred of an access request based on privacy preferences. The approach can be utilized to augment and extend current access control schemes that already incorporate trust evaluation.

- A Risk Measure formula that can calculate the risk associated with the exposure of the patient's data.
- A Risk Mitigating Data Disclosure algorithm, RIMIDI, which realizes the Risk Measure formula. Accordingly, the set of data objects that can be safely exposed should have some relevance and correlation to the ongoing medical diagnosis.

The remainder of this paper is organized as follows: Section 2 presents the research background. Section 3 presents the related work regarding access control schemes that incorporate risk assessment and the mitigation approaches that can be applied to lower the risk incurred. Section 4 presents the proposed Risk Mitigation approach, which incorporates the derived and developed Risk Measure formula. The mathematical proof for the Risk Measures is presented as well as the Risk Mitigation Data Disclosure algorithm, RIMIDI that realizes the Risk Measure formula. Section 5 presents the implementation, experimental results and discussion. Finally, Section 6 concludes the research and suggests possible future directions.

2. Background

This part presents an overview of the background information, where some preliminaries regarding privacy preferences are presented. Moreover, because the risk mitigation approach assumes the utilization of the Disease Relevance Matrix and trust evaluation, some background is presented for each assumption.

Before the beginning of the derivation process in Section 4, several factors and issues need to be stated and handled.

Issue 1: Privacy Preferences

The main goal of this research is to propose a risk mitigation approach that can protect the patient's private data while providing the health care provider the required access to their related health records to bring them quality healthcare services. More specifically, the research intends to provide a risk mitigation method that brings tailored results to each patient according to their privacy preferences. That is, for two patients, patient A and patient B, who have the same set of previously diagnosed and stored diseases, disease $o_1, o_2, \cdots, o_n$ and who are being treated by the same doctor of trust level t, the risk mitigation approach, utilizing the derived risk measure formula, will provide and reveal patient's data based on the patients' individual preferences. The reason is that each patient has assigned different privacy values to their diagnosed diseases. Therefore, with the exception for the distinguished privacy preferences, the set of revealed diseases will be different when all other variables are of equal values. **Figure 2** illustrates such concept.

To realize the previous outcome, the diagnosed diseases need to be associated with the patient's privacy preferences, which are expressed as privacy weights. In

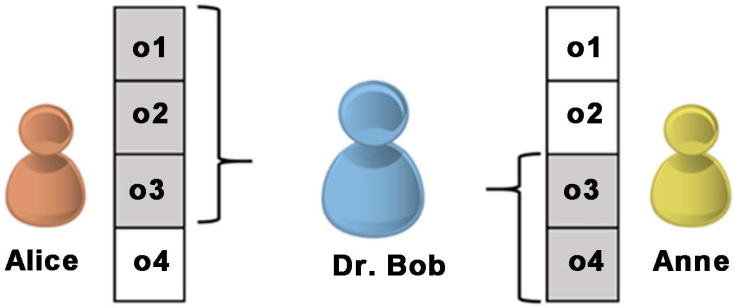

Figure 2. Two patients who have the same set of diagnosed disease objects, o1, o2, o3 and o4, and who are treated by the same doctor will have different privacy preferences and, hence, different data exposure.

practice, the medical data for the patients does not include the privacy weights. Therefore, the privacy preferences must be simulated. The assumption is that when the patients come for a medical purpose, they supply their privacy preference when they fill out the medical forms.

The privacy preferences can be divided into categories, according to their severity or impact, as advised by the NIST 800-60 publication [27] to *low*, *moderate* and *high*. That is, the higher the privacy preference value, the more sensitive the data, and hence, the more severe the impact of exposure.

Each category is mapped to a numerical value indicating its sensitivity. That is, for two data objects o_i and o_j having an associated privacy weights of w_i and w_j, respectively, if $w_i > w_j$ then the conclusion is that o_i is more sensitive than o_j and vice versa. The privacy weights can be realized using the formula:

$$f_{priv} = P \times O \to R \tag{1}$$

where f_{priv} is a function that generates a discrete real number, between $[0,1]$ such as $\{0.1\},\{0.2\},\cdots$ and so forth, representing the privacy weight value of object O_i by patient *p*. In this research, the assumption is that privacy weights are supplied as discrete real numbers ranging from $[0,1]$ where 0 indicates no privacy preference and 1 indicates maximum preference.

Issue 2: Compliance with the Privacy Rule of HIPAA

HIPAA privacy rule contains regulations and rules that must be followed to preserve the privacy of the patient's health records. To comply with HIPAA rules, the medical data of the patient, that can be revealed to medical professionals without obtaining a patient's consent, must be relevant and related to the current medical procedure such that it is beneficial for advancing the treatment efforts. For example, it is well established in the literature of medicine that there exists some correlation, or relevance, between heart disease and high blood pressure [28]. On the contrary, there is no established link between heart disease and some other disease. To realize this objective, relevance information between the different diseases needs to be acquired; which is the primary objective of the assumption of the Disease Relevance Matrix (DRM). The Disease Relevance

Matrix is utilized to serve two main purposes. First, since it includes relevance information regarding the different diseases, the DRM helps in achieving HIPAA compliance. That is because the risk mitigation approach will consult the DRM before further evaluating the set of returned patient relevant diseases for the risk associated with data combinations. Second, the DRM greatly reduces the searching space for our proposed algorithm the Risk Mitigating Data Disclosure (*RIMIDI*) algorithm.

This research assumes that the relevrform the correlational analysis. As proposed in [29], the relevance of the meant information has been calculated and stored in the DRM. Nonetheless, one effective method of calculating relevance information between the diseases is to pedical data is found by analyzing the correlation incurred of the different records. That is, the probability for a disease to have some significant relevance to another disease is computed by analyzing the requesting behavior of the doctors in the system concerning some medical record. For instance, for a medical record of type t belonging to a patient u, the doctor issues an access request to that record of which contains some access purpose p. Therefore, if there are multiple requests of multiple doctors to access medical records of type t_i to serve some purpose p_j, then it is concluded that there exists some correlational degree between the two. In this regard, relevance information can be captured and stored in the Disease Relevance Matrix and updated periodically.

$$DRM\left(o_i, d_j\right) = \begin{cases} 1, & \text{if } o_i, d_j \text{ correlated} \\ 0, & \text{if } o_i, d_j \text{ not correlated} \end{cases} \quad (2)$$

For simplification purposes, the DRM will assume that, for two diseases, a relevance value of 1 indicates an existing link between the diseases; otherwise, if it is equal to zero, then it indicates that there is no relevance or link between the diseases. **Table 1** illustrates the DRM.

Issue 3: Trust Evaluation

In efforts to assess the riskiness of an access request, trust calculation can be employed. Trust of an entity requesting access can be defined as analyzing past behavioral patterns with regards to system resources for evaluating future access requests [12]. Nevertheless, several methods can be used to assess the trustworthiness of an entity. One intuitive, but static, trust evaluation is by issuing security clearances for the users by the system administrators [17]. Furthermore, when an entity issues request access to system resources, several parameters can be evaluated to calculate the trust level. For example, the trust level of a user requesting access from outside the hospital can be evaluated against the security level of the network connection, security clearances of the user, the time of which the access has been requested and so forth [13]. Moreover, recommender systems can be used for recommending trust levels for users; especially when the users are not known to the hospital's system a priori but the organization to whom they belong has some collaborations with the hospital that uses the recommender system [30].

Table 1. Disease Relevance Matrix: every disease (D) has correlation information with the set of other diseases. If correlation value = 1 then there is some relevance to the two intersecting diseases; otherwise, they are not correlated.

	D_1	D_2	D_3	D_4	D_5	D_6	i
D_1	0	1	1	0	1	0	…
D_2	1	0	1	0	0	1	…
D_3	1	1	0	1	1	1	…
D_4	0	0	1	0	0	0	…
…	…	…	…	…	…	…	…

3. Literature Review

Several efforts have been conducted to extend access control models with risk measures to achieve the related objectives. After risky access is deemed tolerable, some works propose applying risk mitigation approaches to lower the impact of the calculated risk further. Furthermore, some works propose applying risk-aware access control in specific domains such as healthcare.

3.1. Risk-Aware Access Control Models

Risk-aware access control models (RAAC) [31] are considered as new types of access control which incorporate risk assessment functions into making the access decision. RAAC facilitate the data sharing securely in highly dynamic environments. In a risk-aware access control paradigm, two steps are required to make an access decision. First, risk assessment is carried out to estimate the risk incurred if access is granted. Second, based on the outcome of risk assessment, when a risky request is considered within acceptable thresholds, risk mitigation strategies can be employed to lower the risk. Risk mitigation approaches are obligatory actions [22] that need to be fulfilled and automatically monitored to assess risk and establish the notion of trust.

The United States National Institute of Standards and Technology (NIST) has defined risk-aware access control based on the proposal in [32]. It includes the system model for assessing risk by incorporating operational need, situational factors and risk measures. Based on the previous work, a risk-adaptive attribute-based access control model (RadAC) has been proposed in [13]. The main contribution is a definition of a conceptual risk model that incorporates risk calculation based on different attributes of subjects, object and environmental elements such as attributes of connections, session and so on. The risk assessment function was not specified and was left up to administrators.

A model has been proposed in [18]. It measures the risk associated with an access request using several parameters; trustworthiness of the subject requesting access, the cost incurred of granting such access to the object and the security policies defined within the access control model. While providing risk analysis to the various consequences, the model is deemed static since trust scores have no means to be updated to reflect the risk adaptability of RAAC.

Moreover, a benefit and risk-based access control model (BARAC) is proposed where vectors of benefit and risk are constructed for each system transaction [33]. Furthermore, a special graph is constructed based on the configuration. When a transaction is issued, the system computes the risk and benefit incurred by such action. If the overall system benefit outweighs the risk, the transaction is carried out. However, the system is also deemed static and updating leads to various problems, as explained by [34].

In another work, a fuzzy multiple level security model that incorporates quantified risk assessment in enforcing access control has been proposed in [17]. In their work, the subjects are associated with clearance levels and the objects are associated with sensitivity levels. An access request is assessed using the difference of the clearance level of the subject requesting access and the sensitivity level of the object. If the difference is large, the risk is high and vice versa. The authors utilized the widely accepted definition of risk assessment as defined by NIST [29] of which calculates risk as the likelihood of an event multiplied by the possible impact. Similarly, [19] have utilized fuzzy inference techniques to calculate risk values for enforcing access control. The risk is calculated primarily based on the sensitivity of the object and secondarily by the security clearances of the subjects. In their work, they show that fuzzy inference can be a good approach to estimating risk. However, both works do not consider the past behavior of subjects in calculating the risk.

To overcome the limitations associated with the previous multiple level security models [17], a trust and risk access control model (TRAAC) [21] has been devised. The access control model calculates the trust and maps the incurred risk to dynamic risk mitigating intervals, which are also mapped to corresponding obligations, of which a user is expected to comply with.

In efforts to quantify risk in the medical field, a risk assessment model has been presented in [11] where the trust of the users requiring access is considered and updated. The work is based on the principle of "Need to Know" such that a doctor requesting access to a patient's information is considered as a low-risk request if the data needed is relevant to the doctor's job. The model requires a relevance function of which the authors explicitly stated that its true shape is unknown or defined yet. While promising, determination of relevance of an access request is a difficult task and machine learning techniques can be employed to mine for relevance patterns as shown in their extended work in [27]. However, the main contribution to the later work is to assist patients in making consent decisions to comply with one of the HIPAA privacy rules and no mitigation strategies have been devised.

To augment the existing access control schemes with risk awareness, dynamic Risk-based decision methods which add trust value and risk value to access control decision point such that it becomes dynamic with the 'history of use' recorded for each user is proposed in [20]. When access is granted, the formula is consulted to add reward or penalty points to the user which will further assess in making future access decisions. To make the system more rational and quickly

adjust to access evaluations, the Exponential Weighted Moving Average EWMA [35] has been utilized. However, the work does not specify what the system does after granting or denying access and no risk mitigation strategies are proposed. Similarly, the work in [36] has considered amplifying the current access control scheme, the RBAC model in the cloud, with trust calculation which helps the data owners in deciding the stored data of which will ultimately reduce the risk incurred of legitimate access. In another study, authors in [37] assessed the risk associated with outsourcing the data to be stored in the cloud using two risk factors: the sensitivity of the data to be outsourced and the degree of how transparent is the privacy control of the cloud service provider. In both works, risk mitigation strategies have not been employed.

Moreover, risk management principals to security have been utilized in [38] to develop risk assessment framework. The developed formulae to quantify risk based on the equation by NIST, which defines Risk as the likelihood multiplied by the Impact, is devised in three different ways. In the first, the primary weight is defined by the subject trustworthiness. In the second, it is defined by the object sensitivity. In the third, it is defined by computing the difference between the previous two. Later, based on NIST definitions, actions were mapped to impact descriptions. While their work is intuitive, they demonstrated how important it is to make observations and derive equations from them. Their work has some limitations; it does not address the insider attack. Also, beyond calculating impact, nothing is proposed to lower or mitigate the risk calculated.

Moreover, auto-delegation in access control schemes has been proposed by [39] and further investigated and extended by [40]. In an auto-delegation system, the features of delegation approaches and the "break-the-glass" emergency based approaches are combined such that the limitations of each approach are minimized and the flexibility of the access control system is improved. An auto-delegation framework based on the risk associated with a user's availability is proposed and presented in [40]. In the framework, each user is associated with a probability of their availability such that in critical situations, the system can delegate access rights to the most qualified available user. A use case in the healthcare domain is presented, however, no implementation has been conducted.

3.2. Risk Mitigation for Access Control Models

Risk mitigation [29] approaches can be employed to lower the risk indicated by an access request. Several works have proposed different approaches to mitigate the risk. Obligations [21] [22], which are user actions that need to be fulfilled and monitored by the system, are considered as a type of risk mitigation strategies. An example of an obligation is by allowing access to a system resource provided that the user will be obligated to use the resource in a fair manner, such as complying with the average download rate [16]. The system monitors and logs the user's actions to be evaluated in future access requests. Others considered lowering the risk by employing dynamic countermeasures such as increasing the

security measures to bring back risk to acceptable levels [23]. An evolution-based genetic algorithm is devised to help lower the risk incurred from the access request by manipulating the security metrics that the organization employs. Effectively, when a request is deemed as risky, mitigation involves devising a security countermeasure such as increasing the length of the encryption key or notifying the system's administrator of potentially risky behavior.

Moreover, anonymization techniques have also been utilized as a risk mitigation strategy to lower the risk associated with data release. In [25], anonymization using k-anonymity measure is used to mitigate the risk associated with data disclosure. In their model, the access control is enhanced with risk assessment such that the risk that can be incurred from querying a dataset is calculated and, then, lowered via returning anonymized data; which ultimately protects privacy. Similarly, the risk of re-identifying data owners is estimated and mitigated using several anonymization techniques; namely, k-anonymity, l-diversity, t-closeness and δ-presence, to find the best risk-mitigating solution for preserving the privacy of the patients [24]. However, while anonymization methods can be beneficial in preserving patients' privacy by hiding their identities among many other patients' data, they cannot be used to help assist in providing tailored medical service for a specific patient as already argued.

It is worth mentioning that some works have considered finding the minimum set of user data, encompassed in user credentials to be exchanged with a communicating server [41] prior to be granted access or as the required personal data usually collected in exchange for providing a service such as bank loan, for the ultimate goal of complying with the Limited Data Collection principle [42]. However, a good inspection of the research studies shows remarkable and distinguishable differences from the one conducted in this research. Mainly, the works above try to find the minimum set of data that can be used to provide the sufficiently enough information for a potential communicating party by utilizing graph theory and constraint satisfaction problem, respectively. In contrast, the proposed approach in this research tries to find the set of a patient's relevant data, which can be accessible by a doctor, based on developing a novel risk measure formula that assesses the riskiness of candidate combinations of data. The problem definitions are different, this research considers the patient's data objects as the variables that are already assigned with privacy values, unlike the other efforts that exhaustively search and explore combinations of variables, which can be assigned, to any value in a pre-defined domain. Moreover, in [41], the problem is to find the minimum set of credentials before allowing access to the system's resource. On the contrary, the proposed approach in this research is concerned with finding the set of exposable data after a user is authorized into the main system. In both mentioned works, the objective is not proposing a risk mitigation approach for access control systems.

On the contrary, the proposed approach in this research is concerned with finding the set of exposable data after a user is authorized into the main system. In both mentioned works, the objective is not proposing a risk mitigation ap-

proach for access control systems. In distinction, this research is a risk mitigation method for access control systems, of which is adaptable to the varying trust level of the doctors and is HIPAA compliant, making it a viable solution for preserving patient's privacy in the healthcare domain.

4. The Risk-Aware Privacy Preserving Risk Mitigation Approach

"Risk" is a terminology that can be defined in different ways depending on the context of which it is used in. Nevertheless, the term is used when there is some degree of uncertainty regarding an outcome [43]. In the field of Probability and Statistics, the concept of risk is used to indicate the result surrounding an expected value. In this section, the Risk Measure formula, which computes the risk incurred upon revealing the patient's private data, is derived. Furthermore, the proposed model is mathematically proven for preserving the privacy of the patient by restricting the exposure of the private data. Finally, RIMIDI algorithm that utilizes the developed Risk Measure formula is presented.

4.1. Risk Measure Formula Derivation

In this section, the objective is to derive a Risk Measure formula that will achieve the intended goal of preserving the privacy of the patients' medical records, which is HIPAA compliant, and while providing them with quality healthcare services. As a result, two formulas will be derived. The generated formulas are then tested with hypothetical examples to assess them. Later, one formula will be nominated and recommended based on the assessment results of the formulas. As previously mentioned, the objective of this research is to find the suitable set of patient's data that can be safely exposed to the healthcare professional, according to HIPAA privacy rule, without undermining their privacy preferences.

To choose the relevant data, the utilization of the Disease Relevance Matrix is proposed. Furthermore, the patient privacy preferences are weights supplied upon filling out the medical forms. As a result, for a set of patient's diseases that have already been diagnosed and stored in the system, the assumption is that there is a corresponding privacy weight.

The direct approach is to analyze the dataset to facilitate the assumption of the suitable type of distribution. That is because the distribution helps in recognizing patterns in the dataset that can be, later, utilized for understanding the central tendency and dispersion of the data [44] as well as projecting or predicting the future data trends. However, in the situation at hand, the patient's data are disease objects that are associated with sensitivity weights identifying their risk of exposure. Unlike the data objects of diagnosed diseases, the associated privacy preferences are not available in a real-world scenario. Therefore, we are entitled to assume the distribution of the patient's data. Nonetheless, to facilitate the formula derivation process, let us consider the following hypothetical example. Patient p has been diagnosed with four different health issues, o_1, o_2, o_3 and

o_4. For each health problem, the patient declares his privacy preference weights, which are shown in Table 2.

The Distribution Assumption:

First, the data objects are assumed to be independent and have discrete sensitivity values. That is, the set of relevant diseases that have been returned from the DRM have discrete sensitivity weights. Therefore, the distribution that needs to be assumed should be of the discrete variable distributions [44]. Furthermore, for individual diseases, the probability is assumed as equally likely. That is, the likelihood of occurrence for any disease is the same among all other diseases. For example, the likelihood of o_1 or o_3 is the same; they each have a probability of 1/4. That is because the total number of relevant disease objects, which denotes the range of the space of the random variable, is equal to 4.

Based on the previous assumptions about the data, the distribution that fits the description is, therefore, the discrete variable empirical distribution [44]. Under this type, for a random variable X that has a space of n independent values, such that its range is $\{x_1, x_2, \cdots, x_n\}$, the probability for each instance is $\frac{1}{n}$; which means that the probability for each instance is equally likely. Therefore, the probability mass function of the discrete empirical distribution is

$$f(x) = \frac{1}{n}, \text{ for all } x\text{'s} \tag{3}$$

Furthermore, since the order of the data is not relevant in a combination instance, the number of combinations generated for n data objects where r objects are taken at a time can be obtained using the following equation [44]:

$$C_r^n = \binom{n}{r} = \frac{n!}{(n-r)!r!} \tag{4}$$

Assuming the distribution of the data facilitates finding and calculating descriptive statistics, of which can help in understanding features about the data and, thus, building the model.

Going back to the hypothetical example, and after assuming the empirical distribution, the statistics of the data can now be calculated to be utilized in the Risk Measure formula. One way to assist in calculating the Risk Measure is to find the mean value of the sensitivity weights. That is, for the n data objects, the average is

$$\mu = \frac{\sum_{i=1}^{n} w_i}{n} \tag{5}$$

where w_i denotes the weight of disease o_i. $\sum_{i=1}^{n} w_i$ denotes the sum of the weights for the n diseases. Table 3 calculates the mean value for the set of sensitivity weights. However, for a combination of m data objects, the average is $\mu = \left(\sum_{i=1}^{m} w_i\right) \Big/ m$,where w_i denotes the weight of disease o_i. And $\sum_{i=1}^{m} w_i$ denotes the sum of the weights for the m diseases. Table 4 demonstrates the mean value for the data in a combination.

Table 2. The set of diseases and the associated privacy weights for patient p.

Disease	Privacy Weight
o_1	0.3
o_2	0.1
o_3	0.9
o_4	0.5

Table 3. The set of diseases, the associated privacy weights and the mean for patient p.

Diseases	Weight
o_1	0.3
o_2	0.1
o_3	0.9
o_4	0.5
Mean	0.45

Table 4. The set of diseases, the associated privacy weights, and the mean for all data combinations.

Diseases	Weight	Mean
o_1	0.3	0.3(0.3/1)
o_2	0.1	0.1
o_3	0.9	0.9
o_4	0.5	0.5
o_1, o_2	0.3, 0.1	0.2((0.1 + 0.3)/2)
o_1, o_3	0.3, 0.9	0.6
o_1, o_4	0.3, 0.5	0.4
o_2, o_3	0.1, 0.9	0.5
o_2, o_4	0.1, 0.5	0.3
o_3, o_4	0.9, 0.5	0.7
o_1, o_2, o_3	0.3, 0.1, 0.9	0.43((0.3 + 0.1 + 0.9)/3)
o_1, o_2, o_4	0.3, 0.1, 0.5	0.3
o_1, o_3, o_4	0.3, 0.9, 0.5	0.57
o_2, o_3, o_4	0.1, 0.9, 0.5	0.5
o_1, o_2, o_3, o_4	0.3, 0.2, 0.9, 0.5	0.45((0.3 + 0.2 + 0.9 + 0.5)/4)

By analyzing Table 4, a serious flaw can be noticed for combination $\{o_1, o_2, o_3\}$. For this particular combination, the patient has placed a very high privacy value for o_3. However, the averaging method has decreased its severity, thus, undermining its risk. Clearly, the average method does not capture the true distance between the different data weights. That is, for that same combination,

there is a significant distance between disease weights 0.9 and 0.1 for example, yet, the averaging method lowered the sensitivity for the disease o_3 of weight 0.9.

To avoid such underestimation, the maximum weight of a combination can be nominated to represent the sensitivity threshold of the corresponding combination. Applying this to the example earlier yields the results shown in Table 5.

As illustrated in Table 5, the most sensitive disease objects have not been underestimated as with the averaging method. However, while choosing the maximum weight prevented underestimation of a data combination, there is still no intuition about how the data is dispersed and spread. That is where the variance and standard deviation can be utilized to provide information regarding the variability, or riskiness, of the data [43] [44]. Since the data is assumed to follow the empirical distribution, the variance (σ^2), the standard deviation (σ) and the mean (μ), for the n data objects, can be obtained as follows,

$$\sigma^2 = \frac{1}{n}\sum_{x\in D} w^2 - \mu^2 \tag{6}$$

$$\sigma = \sqrt{\sigma^2} \tag{7}$$

$$\mu = \frac{\sum_{i=0}^{n} w_i}{n} \tag{8}$$

where w_i denotes the weight of disease o_i. $\sum_{i=1}^{n} w_i$ denotes the sum of the weights for the n diseases, σ^2 denotes the variance, σ denotes the standard deviation and μ denotes the mean. Consequently, the calculated variance and standard deviation for the weights of n diseases would be as in Table 6. Also, the

Table 5. Choosing the maximum weight inside a combination to represent the weight of that combination.

Diseases	Weight	Mean	Max
o_1	0.3	0.30	0.3
o_2	0.1	0.10	0.1
o_3	0.9	0.90	0.9
o_4	0.5	0.50	0.5
o_1, o_2	0.3, 0.1	0.20	0.3
o_1, o_3	0.3, 0.9	0.60	0.9
o_1, o_4	0.3, 0.5	0.40	0.5
o_2, o_3	0.1, 0.9	0.50	0.9
o_2, o_4	0.1, 0.5	0.30	0.5
o_3, o_4	0.9, 0.5	0.70	0.9
o_1, o_2, o_3	0.3, 0.1, 0.9	0.43	0.9
o_1, o_2, o_4	0.3, 0.1, 0.5	0.30	0.5
o_1, o_3, o_4	0.3, 0.9, 0.5	0.57	0.9
o_2, o_3, o_4	0.1, 0.9, 0.5	0.50	0.9
o_1, o_2, o_3, o_4	0.3, 0.2, 0.9, 0.5	0.45	0.9

Table 6. Calculations for the mean, the variance, and the standard deviation.

Diseases	Weight
o_1	0.3
o_2	0.1
o_3	0.9
o_4	0.5
μ	0.45
σ^2	0.09
σ	0.3

standard deviation can be utilized to capture the average distance, or the average deviation, from the mean value of the *maximum* values. As stated earlier, and according to (4), we assume that the data objects will have a total of 15 combinations. Based on the number of objects in each combination, the combinations are divided into *bands* containing the combination instances of the same length.

Therefore, the band contains m instances of the same number of objects. **Table 7** illustrates the different bands generated.

As stated earlier, the distance between the objects' weights needs to be addressed. That is where the standard deviation can be utilized. For the patient's data, the calculation of the standard deviation can provide some intuition about the distance, or variability, between the objects. Let us consider calculating the standard deviation of the whole data objects.

That is $\sigma = 0.3$. Therefore, Risk of combination $X = MAX\ w_i +$ the standard deviation.

When applying the formula to the dataset, the data as in **Table 8** is obtained. However, under this approach, the calculated standard deviation is fixed for all the different bands. In effect, the riskiness of diseases within each band is not addressed since several objects having high weights can be combined in one combination instance but charged the same standard deviation. Therefore, a more reliable approach is to calculate a separate standard deviation for each band for the maximum weights for the instances therein. In this way, the standard deviation is directly affected with the data inside this band.

Calculating a separate standard deviation for each band requires assuming a separate distribution for that band. In other words, since preferences of data combinations cannot be specified by the patient, because it is unrealistic and impractical, we assume that each instance in a band is equally likely to occur. That is, if the number of instances, *i.e.* same length combinations, in a band is equal to m, then each instance has a probability of $\frac{1}{m}$. Therefore, the empirical distribution is, once again, applied in order to calculate the standard deviation. In effect, the calculation of a separate standard deviation for each band captures a more reliable deviation between data objects in that band.

Table 7. The resulting combinations for the data objects can be arranged into bands based on the number of objects in that combination.

Diseases	Band
o_1	1
o_2	
o_3	
o_4	
o_1, o_2	2
o_1, o_3	
o_1, o_4	
o_2, o_3	
o_2, o_4	
o_3, o_4	
o_1, o_2, o_3	3
o_1, o_2, o_4	
o_1, o_3, o_4	
o_2, o_3, o_4	
o_1, o_2, o_3, o_4	4

Table 8. The Risk of a data combination is denoted by calculating the MAX weight + the Standard deviation of the whole data objects.

Diseases	Weight	Mean	Max	Max + SD (0.3)
o_1	0.3	0.3	0.3	0.60
o_2	0.1	0.1	0.1	0.40
o_3	0.9	0.9	0.9	1.20
o_4	0.5	0.5	0.5	0.80
o_1, o_2	0.3, 0.1	0.2	0.3	0.60
o_1, o_3	0.3, 0.9	0.6	0.9	1.20
o_1, o_4	0.3, 0.5	0.4	0.5	0.80
o_2, o_3	0.1, 0.9	0.5	0.9	1.20
o_2, o_4	0.1, 0.5	0.3	0.5	0.80
o_3, o_4	0.9, 0.5	0.7	0.9	1.20
o_1, o_2, o_3	0.3, 0.1, 0.9	0.433333	0.9	1.20
o_1, o_2, o_4	0.3, 0.1, 0.5	0.3	0.5	0.80
o_1, o_3, o_4	0.3, 0.9, 0.5	0.566667	0.9	1.20
o_2, o_3, o_4	0.1, 0.9, 0.5	0.5	0.9	1.20
o_1, o_2, o_3, o_4	0.3,0.2,0.9,0.5	0.45	0.9	1.20

Considering the above discussion, the following section starts the derivation process for the Risk Measure formula, which is utilized in the proposed model.

The Risk Measure 1 Formula:

Based on the discussion of the previous section, the Risk Measure formula is now developed.

Let $o_1, o_2, \cdots, o_n$ denote the set of relevant disease objects for patient *p*. w_i denotes the privacy preference weight of the corresponding o_i. Find all disease combinations. An *instance* is a data combination. For example, the combinations $\{o_1, o_2, o_3\}$ and $\{o_3\}$ are two different instances.

A *band* is a collection of all instances where the number of objects in each instance is equal. That is, the size of the instances in one band is equal. For example, band$_3$ denotes the set of all instances containing 3 disease objects such as $\{o_1, o_2, o_3\}$, $\{o_1, o_2, o_4\}$, $\{o_2, o_3, o_4\}$, $\cdots$ and so forth.

Calculate the standard deviation, *SD*, for all the data $o_1, o_2, o_3, \cdots, o_n$, using (7) as well as the maximum weight, w_i, for each instance.

For every instance, the selected w_i is the highest w_i of any o_i in that instance, and the *Unadjusted Risk Measure* formula is defined as follows:

$$\rho_{ui_0} = MAX\ w_i + SD \tag{9}$$

where ρ_{ui_0} is the Unadjusted Risk Measure. $MAX\ w_i$ is the maximum weight for each instance *i*. *SD* is the standard deviation for all individual objects.

However, the calculated Risk Measure will produce a value that is higher than one. Since the trust level of the doctor is assumed to be in the range $[0,1]$, the Risk Measure value cannot be checked against the trust level value. To overcome the issue, a loading factor, α, is introduced to ensure that any computed Risk Measure value is never higher than 1. The loading factor is added to the standard deviation as follows

$$\rho_{ai_0} = MAX\ w_i + \alpha * SD \tag{10}$$

We need to find α that ensures the risk measure calculated does not exceed 1 That is,

$$\rho_{ui_0} \le 1$$

However, the Unadjusted Risk Measure is not always less than one. Therefore, the highest value of Unadjusted Risk Measure must be selected if it is less than or equal to one. Otherwise, the selected value is 1:

$$Min\left(MAX\left(\rho_{ui_0}\right),1\right)$$

$$Min\left(MAX\left(\rho_{ui_0}\right),1\right) = MAX\ w_i + \alpha * SD$$

$Max\ w_i$ will be the corresponding $Max\ w_i$ for $MAX\left(\rho_{ui_0}\right)$, therefore,

$$Min\left(MAX\left(\rho_{ui_0}\right),1\right) = Corrsponding\ Max\ w_i + \alpha * SD$$

Solving for α, the loading factor is as follows

$$\alpha = \frac{Min\left(MAX\left(\rho_{ui_0}\right),1\right) - Corrsponding\ Max\ w_i}{SD},\ SD \neq 0 \tag{11}$$

Therefore, for every combination instance, the *Adjusted Risk Measure* formula is defined as follows; the incurred Risk is the highest privacy weight of that instance plus the multiplication of the standard deviation, of all individual weights, by the loading factor α, and can be expressed as follows:

$$\rho_{ai_0} = MAX\ w_i + \alpha * SD \tag{10}$$

where ρ_{ai_0} is the Adjusted Risk Measure. $MAX\ w_i$ is the maximum weight for each instance i. SD is the standard deviation for all individual objects. α is the loading factor and is defined as in (11).

As mentioned previously, $Min\left(MAX\left(\rho_{ui_0}\right),1\right)$ ensures that the output of the formula will be less than or equal to 1. That is, if $MAX\left(\rho_{ui_0}\right)>1$, then select $MAX\left(\rho_{ui_0}\right)=1$. Furthermore, in the case where $MAX\left(\rho_{ui0}\right)>1$, then $Corresponding\ Max\ w_i$ would be the highest $max\ w_i$ among all the instances. In other words, the formula first looks up the computed Unadjusted Risk Measure values, ρ_{ui_0}, for all combination instances. If the maximum value found is less than 1, then take that value as well as the maximum weight that represents that combination. Otherwise, if the Unadjusted Risk Measure found is greater than 1, then set ρ_{ui_0} to 1 and select the maximum weight value among all generated instances.

Finally, two minor issues must be resolved. First, in the situation where the Adjusted Risk Measure is equal for more than one instance, then the instance with the highest number of diseases is selected. The reason for this is that, since the DRM is assumed to return the data objects that have some relevance to the ongoing medical treatment, regardless of the degree of relevance, the objective, therefore, is to provide the doctor with the maximum number of data to that can assist in making a better medical decision, which ultimately benefits the patient. Second, in case there are several instances, of the same length, whose Risk Measures are equal, then the selected data would be the one with the lower mean for its diseases within that instance. The intuition behind this is that, for some combinations, having the same length and the same Risk Measure, one could arbitrarily select any combination as the solution. However, the privacy of the patient could be violated. That is, even if two combinations have been evaluated to have the same Risk value, one combination could be riskier than the other. That is, since each combination is represented by its highest sensitivity weight, two combinations could evaluate to the same Risk value. In effect, one would make a closer inspection of the data comprising that combination; the data inside could have very high weights. Therefore, calculating the mean value of the data inside the combination can reveal how risky this combination can be. When evaluating two combinations, selecting the one with the smaller mean value helps in protecting the privacy preferences of the patient.

Proposition 1*:* the patient preferences are protected using Risk Measure 1.

Proof:

Doctor trust value, t, ranges from zero to one

$$0 \le t \le 1$$

And Risk Measure 1, ρ_{ai_0}, ranges from zero to one

$$0 \le \rho_{ai_0} \le 1$$

The risk measure will reveal only the data that has a risk value that is equal to or below the doctor trust, *t*, therefore,

$$0 \le \rho_{ai_0} \le t \le 1$$

According to (10):

$$0 \le MAX\ w_i + \alpha * SD \le t \le 1$$

Since α and SD are positive integers, the last equation can be written as

$$0 \le MAX\ w_i \le t \le 1$$

$$MAX\ w_i \le t$$

$Max\ w_i$ represents the highest preference weight that is selected by the patient, $Max\left(w_1, w_2, w_3, \cdots, w_i, \cdots, w_n\right)$. Therefore, Risk Measure 1 ensures that the exposed data is as per preferred by the patient. □

A closer inspection of Risk Measure 1 formula can reveal some issues. The loading factor, α, denoted by (11), yields a fixed risk load. For example, if $\alpha = 0.5$, then this value is added to all of the calculated risk values. That is, the loading factor α is static and not dynamic, which leads to the next formula improvements. Furthermore, under this approach, the calculated standard deviation is fixed for all different bands of instances. That is, the riskiness of diseases within each band is not addressed since several high weights can be combined in one instance but charged the same standard deviation. Therefore, a more reliable approach is to calculate a separate standard deviation for every band for the max weights for the instances. In this way, the standard deviation is directly affected with the data inside this band.

The Risk Measure 2 Formula:

Let $o_1, o_2, \cdots, o_n$ denote the set of relevant disease objects for patient *p*. w_i denotes the privacy preference weight of the corresponding o_i. Find all disease combinations. Also, as stated earlier, an *instance* is a data combination and a *band* is a collection of all instances where the number of objects in each instance is equal. That is, the size of the instances in one band is equal. In this formula, the assumption is that, in each band, the maximum weights representing each instance follow one distribution which is the empirical distribution. Therefore, a separate standard deviation, SD_i, will be calculated for each $band_i$.

The number of possible combinations generated from *n* values, taken *r* values at a time can be found by (4). For every instance, the selected w_i is the highest w_i of any o_i in that Instance. The *Unadjusted Risk Measure* formula is defined as follows:

$$\rho_{ui_1} = MAX\ w_i + SD_i \tag{12}$$

where ρ_{ui_1} is the Unadjusted Risk Measure. $MAX\ w_i$ is the maximum weight for each instance. SD_i is the standard deviation of the maximum weights representing the instances in $band_i$.

One would notice that, for $band_n$, which includes one instance comprised of all n diseases, the standard deviation is equal to zero. To solve this issue, such band will assume the standard deviation of the immediately lower band, *i.e.* $band_{n-1}$. Generally, if $band_m$ has a zero standard deviation, it will assume the standard deviation of the following lower band, *i.e.* the standard deviation of $band_{m-1}$.

However, as in Risk Measure 1, the calculated Risk Measure will produce a value that is higher than one. Since the trust level of the doctor is assumed to be in the range [0, 1], the Risk Measure value cannot be checked against the trust level value. To overcome the issue, a loading factor is introduced to ensure that any computed Risk Measure value is never higher than 1.

Therefore, to ensure that the risk measure does not exceed 1, a loading factor, π_i, for each band i is introduced and is added to the standard deviation as follows

$$\rho_{ai_1} = MAX\ w_i + \pi_i * SD_i \tag{13}$$

We need to find π_i that ensures the risk measure does not exceed 1. Therefore,

$$\rho_{ui_1} \le 1$$

However, when computed, the Unadjusted Risk Measure does not always yield a value less than or equal to 1. Therefore, the highest value of Unadjusted Risk Measure must be selected if it is less than or equal to one. Otherwise, the selected value is 1:

$$Min\left(MAX\left(\rho_{ui_1}\right),1\right)$$

$$Min\left(MAX\left(\rho_{ui_1}\right),1\right) = MAX\ w_i + \pi_i * SD_i$$

$Max\ w_i$ will the corresponding $Max\ w_i$ for $MAX\left(\rho_{ui_1}\right)$, therefore,

$$Min\left(MAX\left(\rho_{ui_1}\right),1\right) = Corrsponding\ Max\ w_i + \pi_i * SD_i$$

Solving for π_i, the loading factor is as follows

$$\pi_i = \frac{Min\left(MAX\left(\rho_{ui_1}\right),1\right) - Corrsponding\ Max\ w_i}{SD_i},\ SD_i \neq 0 \tag{14}$$

Therefore, for every combination instance, the *Adjusted Risk Measure* formula is defined as follows; the incurred Risk is the highest privacy weight of that instance plus the multiplication of the standard deviation, of all instances in that band, by the loading factor π_i, and can be expressed as follows:

$$\rho_{ai_1} = MAX\ w_i + \pi_i * SD_i \tag{13}$$

where ρ_{ai_1} is the Adjusted Risk Measure. $MAX\ w_i$ is the maximum weight for each instance. SD_i is the standard deviation of the maximum weights representing the instances in $band_i$. π_i is the loading factor for each band and is defined as in (14). Hence, the Adjusted Risk Measure, ρ_{ai_1}, is utilized to cal-

culate the risk measures that, later, will be compared to the doctor's trust level.

If the Adjusted Risk Measure value is equal for more than one instance, then the instance having the larger number of diseases is selected. Furthermore, if there are several instances, of the same length and $Max\ w_i$, whereas the risk measures are equal, the selected data would be the one with the lower mean for its diseases within the instance.

Proposition 2: the patient preferences are protected using Risk Measure 2.

Proof:

Doctor trust value, *t*, ranges from zero to one $0 \le t \le 1$

And Risk Measure ρ_{ai_1} ranges from zero 0 to one $0 \le \rho_{ai_1} \le 1$

The risk measure will reveal only the data that has a risk measure that is equal or below the doctor trust, therefore,

$$0 \le \rho_{ai_1} \le t \le 1$$

According to (13):

$$0 \le MAX\ w_i + \pi_i * SD_i \le t \le 1$$

Since π_i and SD_i are positive integers, the last equation can be written as follows

$$0 \le MAX\ w_i \le t \le 1$$

$$MAX\ w_i \le t$$

$Max\ w_i$ represents the highest preference weight that is selected by the patient, $Max\left(w_1, w_2, w_3, \cdots, w_i, \cdots, w_n\right)$. Therefore, Risk Measure 2 ensures that the exposed data is as per preferred by the patient.

4.2. Coherent Risk Measure

To effectively manage and tolerate an incurred risk, a risk measure formula must satisfy four axioms of translation invariance, subadditivity, positive homogeneity and monotonicity. The risk measure that meets these four axioms is, therefore, a *coherent* risk measure [45].

- *Axiom* 1*:* Translation Invariance—For all instances X and constant β.

$$RM\left(X + \beta\right) = RM\left(X\right) + \beta$$

This axiom indicates that the addition of a sure amount of patient weight increases the risk by the same amount.

- *Axiom* 2*: Subadditivity*—For all instances X and Y,

$$RM\left(X + Y\right) \le RM\left(X\right) + RM\left(Y\right)$$

Subadditivity signifies that merging data does not create extra risk.

- *Axiom* 3*:* Positive Homogeneity—For all $r \ge 0$ and instance X,

$$RM\left(r * X\right) = r * RM\left(X\right)$$

Positive homogeneity describes what happens if there is no data merger benefit. That is, it states that the computed risk of merged data is equal to multiplying the calculated risk by a factor before merging the data.

- *Axiom* 4*:* Monotonicity—For all X and Y with $X \le Y$, then,

$$RM(X) \le RM(Y)$$

Monotonicity indicates that the risk measure will be higher for higher patient preference weight.

Now that the four axioms are stated, we prove that the Risk Measure, ρ_{ui}, satisfies the four axioms and, therefore, is a coherent risk measure. But first, we define the following: *Instance* $X = (w_1, w_2, \cdots, w_n)$, *Instance* $Y = (v_1, v_2, \cdots, v_n)$, *Max Instance* $X = w$ and *Max Instance* $Y = v$.

The Risk Measure is: $\rho_{ui} = MAX\ w_i + SD$, Where,

$$SD = \sqrt{\sum \frac{w_i^2}{n} - \left(\sum \frac{w_i}{n}\right)^2}$$

And instances X and Y are independent.

Axiom 1: Translation Invariance—For all instances X and constant β.

$$RM(X + \beta) = RM(X) + \beta$$

Proof:

$$\rho_{ui}(w_i + \beta) = MAX(w_i + \beta) + SD_1$$

$$SD_1 = \sqrt{\sum \frac{(w_i + \beta)^2}{n} - \left(\sum \frac{w_i + \beta}{n}\right)^2}$$

$$SD_1 = \sqrt{\sum \frac{w_i^2 + 2\beta w_i + \beta^2}{n} - \left(\sum \frac{w_i}{n} + \beta\right)^2}$$

$$SD_1 = \sqrt{\sum \frac{w_i^2}{n} + 2\beta ? + \beta^2 - \left(\sum \frac{w_i}{n}\right)^2 - 2\beta \sum \frac{w_i}{n} - \beta^2}$$

$$SD_1 = \sqrt{\sum \frac{w_i^2}{n} + 2\beta\mu + \beta^2 - \left(\sum \frac{w_i}{n}\right)^2 - 2\beta\mu - \beta^2}$$

$$SD_1 = \sqrt{\sum \frac{w_i^2}{n} - \left(\sum \frac{w_i}{n}\right)^2} = SD$$

$$\rho_{ui}(w_i + \beta) = MAX(w_i + \beta) + SD$$

$$\rho_{ui}(w_i + \beta) = MAX(w_1 + \beta, w_2 + \beta, \cdots, w_n + \beta) + SD$$

$$\rho_{ui}(w_i + \beta) = MAX(w_i) + \beta + SD$$

$$\rho_{ui}(w_i + \beta) = MAX(w_i) + SD + \beta$$

$$\rho_{ui}(w_i + \beta) = \rho_{ui}(w_i) + \beta \quad \square$$

Axiom 2: Subadditivity—For all instances X and Y,

$$RM(X + Y) \le RM(X) + RM(Y)$$

Proof:

$$\rho_{ui}(w_i + v_i) = MAX(w_i + v_i) + SD_1$$

$$\rho_{ui}\left(w_i + v_i\right) = MAX\left(w_1 + v_1, w_2 + v_2, \cdots, w_n + v_n\right) + SD_1$$

Since instances X and Y are independent

$$SD_1 = \sqrt{Var\left(X\right) + Var\left(Y\right)} \le \sqrt{Var\left(X\right)} + \sqrt{Var\left(Y\right)} = SD_x + SD_y$$

$$\rho_{ui}\left(w_i + v_i\right) = w + v + SD_1 \le w + v + SD_x + SD_y = \rho_{ui}\left(w_i\right) + \rho_{ui}\left(v_i\right) \quad \square$$

Axiom 3: Positive Homogeneity—For all $r \ge 0$ and instance X,

$$RM\left(rX\right) = rRM\left(X\right)$$

Proof:

$$\rho_{ui}\left(rw_i\right) = MAX\left(rw_i\right) + SD_1$$

$$\rho_{ui}\left(rw_i\right) = MAX\left(rw_1, rw_2, \cdots, rw_n\right) + SD_1$$

$$SD_1 = \sqrt{\sum \frac{\left(rw_i\right)^2}{n} - \left(\sum \frac{rw_i}{n}\right)^2}$$

$$SD_1 = \sqrt{\sum \frac{w_i^2 r^2}{n} - r^2\left(\sum \frac{w_i}{n}\right)^2}$$

$$SD_1 = \sqrt{r^2 \sum \frac{w_i^2}{n} - r^2\mu^2}$$

$$SD_1 = r\sqrt{\sum \frac{w_i^2}{n} - \mu^2} = rSD$$

$$\rho_{ui}\left(rw_i\right) = MAX\left(rw_1, rw_2, \cdots, rw_n\right) + rSD$$

$$\rho_{ui}\left(rw_i\right) = rMAX\left(w_1, w_2, \cdots, w_n\right) + rSD = r\rho_{ui}\left(w_i\right) \quad \square$$

Axiom 4: Monotonicity—For all X and Y with $X \le Y$, then,

$$RM\left(X\right) \le RM\left(Y\right)$$

Proof:

Assume that $X \le Y$

$$\rho_{ui}\left(w_i\right) = MAX\left(w_i\right) + SD = w + SD_x$$

$$\rho_{ui}\left(v_i\right) = MAX\left(v_i\right) + SD = v + SD_y$$

Since $X \le Y$ for all X and Y, therefore, $w \le v$

$$w + SD_x \le v + SD_y$$

$$\rho_{ui}\left(w_i\right) \le \rho_{ui}\left(v_i\right) \quad \square$$

Proposition 3: the proposed Risk Measure, $w \le v$, is coherent.

Proof:

Since the four axioms for the coherent risk measure are satisfied, this proves that the proposed Risk Measure, ρ_{ui}, is coherent. $\square$

4.3. Risk Mitigation Data Disclosure Algorithm

In this subsection, the Risk Mitigation Data Disclosure (*RIMIDI*) algorithm is

developed. According to the discussion in the earlier parts of Section 4, two versions of *RIMIDI*: $RIMIDI_0$ and $RIMIDI_1$ are presented. This part begins with $RIMIDI_0$ of which is briefly explained, and then is followed by the presentation and more elaboration on $RIMIDI_1$.

The assumption is that trust levels have been calculated and ready, and a mapping function, that maps trust level values to combination risk values, is established and applied. Furthermore, the DRM has been queried and the set of the patient's relevant data has been returned as well as the privacy weights. For every possible combination, the algorithms work by applying several heuristics in order to calculate the risk of exposure using the developed Risk Measure formulas. Afterwards, the data combination that does not violate the patient's privacy preference and has Risk of Exposure ≤ trust is selected. In this sense, the pseudocode for the two algorithms follows the discussion of deriving and developing the Risk Measure formulas in part A of Section 4.

Risk Mitigation Data Disclosure Algorithm ($RIMIDI_0$)

The problem that this research tries to solve is to find the suitable combination that conforms to the constraints of preserving the patient's privacy. That is, exhaustively searching for the best possible combination is impractical and infeasible. Therefore, heuristics must be applied to simplify further and guide the search for a suitable set of data that can be safely exposed.

Let us revisit the ultimate objective of the research: the objective is to find the suitable set of patient data that can be safely exposed to the doctor and without undermining the privacy preferences of the patient. After consulting the DRM for the set of patient data that are relevant to the currently ongoing medical diagnosis, the set of relevant disease information and their corresponding sensitivity weights are returned. As depicted in **Figure 3**, $RIMIDI_0$ starts off by sorting the relevant data in a descending order using MergeSort [46] as in line (3). This step is crucial because, after generating data combinations, it alleviates the expense of checking for $Max\ w_i$ in each instance since it will always be at index 0.

As discussed earlier, the distribution is assumed over the complete set of relevant data and, hence, the standard deviation, *STDV*, is computed to reflect this assumption as in line (18). Later, $Max\left(\rho_{ui_0}\right)$ is calculated and evaluated for $Min\left(Max\left(\rho_{ui}\right),1\right)$ as in lines (19-21). When the Adjusted Risk Measure, ρ_{ai_0}, is computed, and since the objective is to find the maximum number of data objects that can be revealed, with accordance to HIPPA rule of privacy, lines (26-33) show the set of data that can be exposed.

Risk Mitigation Data Disclosure Algorithm ($RIMIDI_1$)

As noted in **Figure 3** and **Figure 4**, both versions of the Risk Mitigation Data Disclosure algorithm, *RIMIDI*, start off by sorting the sensitivity weights in a descending order using MergeSort [46]. As already stated, this step is crucial because it serves multiple purposes. First, since a data combination is represented by the maximum weight ($Max\ w_i$) within that combination, sorting can alleviate the expense of checking every element in a combination to find the Max value. That is, the $Max\ w_i$ of a combination is always at index 0. When the data

```
   Algorithm RIMIDI0
   INPUT : Patien_Relevant_data[ ] , 1 > trust > 0
   OUTPUT: DataCombination OR NULL
1  n ← Size of Patient_Relevant_data[ ]
2  counter ← 0
3  temp_data[ ] = MergeSort(Patient_Relevant_data[ ])
4  for i in 1 to n
5        {if (temp_data[i] == Patient_Relevant_data[0])
6              increment counter
7         else
8              break loop
9  if (counter == n)
10       Request Patient Consent
11       Return NULL
12 else
13       { Max w_i ← temp_data[0]
14         STDEV ← 0
15         MEAN ← 0
16         for j in 1 to n
17               { MEAN = Σ_{j=1}^{n} temp_data[j] / n
18                 STDEV = sqrt( Σ_{j=1}^{n} temp_data[j]^2 / n − MEAN^2 ) }
19       ρ_{ui 0} = temp_data[0] + STDEV
20       if (ρ_{ui 0} ≥ 1)
21             ρ_{ui 0} = 1
22       α = (ρ_{ui 0} − Max w_i) / STDEV
23       for k in 1 to n
24             ρ_{ai 0}[k] = temp_data[k] + α * STDEV
25       for l in 1 to n
26             { if (ρ_{ai 0}[l] ≤ trust)
27                   { StartAt ← l
28                     X ← 0
29                     for m in StartAt to n
30                           { dataCombination[x]
                                   = temp_data[m]
31                             increment x }
32                     Return dataCombination
33                     break loop }}}
34 END Algorithm RIMIDI0
```

Figure 3. Pseudocode for $RIMIDI_0$ algorithm.

combinations are generated by the function *Generate Combinations*, the inclusion and exclusion of data items always yield a combination that has the $Max\ w_i$ at index 0.

For example, consider the following set of relevant diseases with their corresponding privacy weights: $\{\langle o_1, 0.9\rangle, \langle o_2, 0.5\rangle, \langle o_3, 0.4\rangle, \langle o_4, 0.3\rangle, \langle o_5, 0.1\rangle\}$, function *Generate Combinations*, works by generating the corresponding combinations resulting from the submitted patient data based on n and r; which represent the number of relevant diseases and the size of the band, respectively. Table 9 depicts the generated combinations out of the above set of disease weights.

To illustrate, the previous example, in Figure 6, can be exploited. Consider $trust = 0.4$ has been submitted to the system. According to the discussion in part A of Section 4, the computed Risk Measure of a combination exceeds the

```
Algorithm RIMIDI_1
INPUT : Patien_Relevant_data[ ], 1 > trust > 0
OUTPUT: dataCombination OR NULL
n ← Size of Patient_Relevant_data[ ]
k ← 0
exFlag ← false
temp_data[ ] = MergeSort(Patient_Relevant_data[ ])
for i in 1 to n
        if (temp_data[i] ≤ trust)
                increment k
if (k == n)
        for j in 1 to n
                { if (temp_data[j]
                                ≠ Patient_Relevant_data[0])
                        { flag = 0
                          exFlag = true
                          decrement k
                          break loop }
                else
                          flag = −1 }
if (flag == −1)
        { Request Patient Consent
          Return NULL }
binomial_coeff = binomial_coeffecient(n, k)
sdVals
= SdValues(temp_data[ ], n, k, binomial_coeff, trust, exFlag)
if (sdVals == −1)
        { trust is too low
          Return NULL }
else if (sdVals == 1)
        Return dataCombination
else if (sdVals == 0)
        while (k ≥ 2 AND sdVals == 0) do
                { decrement k
                  binomial_coeff = binomial_coefficient(n, k)
  sdVals
= sdValues(temp_data[ ], n, k, binomial_coeff, trust, exFlag) }
END Algorithm RIMIDI_1
```

Figure 4. Pseudocode for $RIMIDI_1$ algorithm.

Table 9. (a) Shows the set of relevant disease objects and their corresponding weights. (b) For instances with the same $Max\ w_i$ and length, the first generated has the highest sensitivity weights while the last has the lowest ones.

Disease Objects	Privacy Weights	Generated Combinations C(5, 4)	
		Disease Objects	Privacy Weights
o_1	0.9	$o_1\ o_2\ o_3\ o_4$	0.9 0.5 0.4 0.3
o_2	0.5	$o_1\ o_2\ o_3\ o_5$	0.9 0.5 0.4 0.1
o_3	0.4	$o_1\ o_2\ o_4\ o_5$	0.9 0.5 0.3 0.1
o_4	0.3	$o_1\ o_3\ o_4\ o_5$	0.9 0.4 0.3 0.1
o_5	0.1	$o_2\ o_3\ o_4\ o_5$	0.5 0.4 0.30 .1
(A)		(B)	

$Max\ w_i$ representing that combination. That is, for trust level $trust = 0.4$, one can be positive that the solution will not include combination instances having $Max\ w_i \geq 0.4$. Therefore, the solution must be in a band which contains in-

stances of $w_i < 0.4$. For this reason, the second heuristic, as shown in lines (5-7) in **Figure 4**, is proposed which estimates the size of the band having the solution by means of calculating the number of *k* objects in the band.

Now that $RIMIDI_1$ estimates the band size having the possible solution, the standard deviation must be calculated prior to further computing the Risk Measure, ρ_{ai_1}, of each combination. Recall from Section 4 that the standard deviation is computed based on $Max\ w_i$ of all instances in the band. Obviously, computing the standard deviation, $STDEV_i$, of each band *i*, requires generating all the instances within that band. For example, if the total number of relevant objects $n = 20$ and the solution is estimated to be in band of size 5, then the total number of of instances that need to be generated is

$C(20,5) = \begin{pmatrix} 20 \\ 5 \end{pmatrix} = 15504$ instances!

Such an approach is, yet again, exhaustive and infeasible. To overcome this issue, the third heuristic is proposed to facilitate the calculation of $STDEV_i$ of any $band_i$ by means of leveraging the binomial coefficient [44]. As shown in **Figure 6**, the $Max\ w_i$ of the generated combinations in $band_4$ are observed to follow the pattern that can be captured by answering the following question: *How frequent does an object sensitivity weight appear in a combination of a band?*

The answer is found in lines (20-21) of **Figure 5**. Specifically, with the aid of *sdValues* function, $RIMIDI_1$ leverages Dynamic Programming [46], by calling *binomial_coeffecient* function in **Figure 6**, to recursively calculate and store the frequencies of every $Max\ w_i$ that can represent a combination, as in lines (6-14) of *sdValues* function in **Figure 5**.

That is, for a combination of *n* data objects and estimated band size of *r*, where the total number of instances $= C(n,r) = \begin{pmatrix} n \\ r \end{pmatrix}$, the number of instances having $Max\ w_i = x_0$ are obtained by

$$C(n-1, r-1) = \begin{pmatrix} n-1 \\ r-1 \end{pmatrix}$$

And the number of instances having $Max\ w_i = x_1$ are

$$C(n-2, r-1) = \begin{pmatrix} n-2 \\ r-1 \end{pmatrix}$$

where $x_0 \geq x_1 \geq \cdots$ and so forth.

Computing the frequencies of $Max\ w_i$ facilitates calculating $STDEV_i$ of $band_i$.

After calculating $STDV_i$, the loading factor, π, is calculated. However, π depends on finding $Min(Max(\rho_{ui}),1)$ as explained in Section 4 Nonetheless, since the data are sorted in a descending order a priori, $RIMIDI_1$ needs only to check $\rho_{ui_1}[0]$ for evaluating the term because it always holds the maximum value, as depicted in lines (22-24) in **Figure 5**. Computing the loading factor

```
   Function sdValues
   INPUT:
        Patient_data[ ], Patient_data_Size, r, nCrTotal, trust, exFlag
   OUTPUT: −1 OR 0 OR 1
1  i ← 0
2  n ← Size of Patient_data_Size
3  Max w_i ← patient_data[0]
4  nCrTotal_ ← nCrTotal
5  ρ_ai 1 ← 0
6        while (nCrTotal > 0) do
7              { if (n == 0 OR r == 0)
8                    break loop
9               else
10                   { decrement n
11                    nCr_i = binomial_coeffecient(n, r − 1)
12                    Frequencies[i] = nCr_i
13                    nCrTotal = nCrTotal − nCr_i
14                    increment i } }
15 for j in 1 to Size of Frequencies[ ]
16       { MEAN = MEAN + (Frequencies[j] × Patient_data[j])
17         Variance = Variance
                      + Paient_data[j]^2 × Frequencies[j] }
18 MEAN = MEAN/nCrTotal_
19 STDEV_i = sqrt( Variance/n − MEAN^2 )
20 if (STDEV_i == 0)
21       Return −1
22 ρ_ui 1 ← Max w_i + STDEV_i
23 if (ρ_ui 1 ≥ 1)
24       ρ_ui 1 = 1
25 π = (ρ_ui 1 − Max w_i)/ STDEV_i
26 for p in 1 to Size of Frequencies[ ]
27       ρ_ai 1[p] = Patient_data[p] + π * STDEV_i
28 for m in 1 to Size of ρ_ai 1
29             { if (ρ_ai 1[m] ≤ trust)
30                   { ρ_ai 1 = ρ_ai 1[m]
31                    bandStartAtElement ← m
32                    break loop } }
33 if (ρ_ai 1 == 0)
34       Return 0
35 else
36       { dataCombination[ ] =
             GenerateDataCombination(bandStartAtElement ,
                   Patient_data[], n, r, exFlag)
37         Return 1 }
38 END Function sdValues
```

Figure 5. Pseudocode for sdValues function.

```
   FUNCTION binomial_coeffecient
   INPUT : n, k
   OUTPUT: C[n][k]
1  for i in 0 to n
2        for j in 0 to MIN(i, k)
3              { if (j == 0 OR j == i)
4                    C[i][j] = 1
5              else
6                    C[i][j] = C[i − 1][j − 1] + C[i − 1][j] }
7  Return C[n][k]
```

Figure 6. Pseudocode for binomial_coeffecient function.

facilitates computing the Adjusted Risk Measure, ρ_{ai}, which is compared against the submitted trust level.

As shown in lines (28-32, 35-37) of *sdValues* function, in **Figure 5**, if $\rho_{ai} \le trust$ evaluates to true, $RIMIDI_1$ generates the data combination that has the corresponding $Max\ w_i$ but with the lowest *MEAN* value, which, due to function *Generate Combinations*, is found as the lastly generated instance with $Max\ w_i$ and the same length among other similar instances. Such intuition is helpful in further protecting the privacy of the patient.

In the situation where the estimated $band_i$ contains no solution, $RIMIDI_1$ iteratively searches in the lower bands: $band_{i-1}, band_{i-2}, \cdots$ and so forth until a solution is found or $n = 0$. Lines (27-31) in **Figure 5** show the approach by utilizing the *while* loop. To illustrate, consider the previous example in **Table 9** where $n = 5$ and $r = 4$. In this example, the total number of combinations returned by *binomial_coeffecient* function is $nCrTotal = 5$. Now the question is: out of these five combinations, what instances will have $Max\ w_i = 0.9$? The answer for $Max\ w_i = 0.9$ is $nCr\_i = 4$. Therefore, $nCrTotal = 5 - 4 = 1$. Since $nCrTotal \neq 0$, the *while* loop continues to calculate the frequencies for other objects. The question is, once again, asked: out of the remaining number of combinations, which is now $nCrTotal = 1$, what instances will have $Max\ w_i = 0.5$? and the answer is 1. The *while* loop terminates when $nCrTotal = 0$ and the obtained frequencies are used in calculating the mean, the standard deviation and the risk measures as in lines (15-27) in **Figure 5**.

Finally, two issues need to be solved. First, in the situation where $STDV_i = 0$, a situation usually arises when the trust level is very high such that $n = r$, then, as discussed in Section 4, the SD_{n-1} of the next largest band, $band_{n-1}$, is assumed as stated in lines (5-14) of **Figure 5**. *exFlag* is utilized to ensure that the calculated $STDEV_{n-1}$ belongs to the largest band, *i.e.* $band_n$. Furthermore, if the relevant data objects returned by DRM are all of the same weight, $STDV_i$ is equal to 0. Therefore, under such special and unusual situation, and to preserve the privacy of the patient, *RIMIDI* protects the data and requires patient consent prior to exposing the data objects as in lines (17-19) in **Figure 5**.

The following part implements both versions of *RIMIDI* algorithm and presents analysis and discussion regarding the obtained results.

5. Implementation, Experimental Results and Discussion

In this section, the Risk Mitigation Data Disclosure algorithms $RIMIDI_0$ and $RIMIDI_1$, which employ the Risk Measure formulas, ρ_{ai_0} and ρ_{ai_1}, are implemented. We developed our own simulator in C++. The results have been graphically illustrated using Matlab.

5.1 Experimental Results

To begin the evaluation process, let us consider the following medical scenario. Suppose that three patients, p_1, p_2 and p_3 are receiving healthcare services

from some healthcare professional u where p_1 is considered as a conservative patient, p_2 is moderate patient and p_3 is a non-conservative one. Furthermore, suppose that these three patients already have matching records of already diagnosed health issues. That is, they suffer from the same set of diseases. Moreover, when they came in for their health issue, doctor u, out of his responsibility to deliver quality healthcare services and to avoid potential repetitive tests and procedures, consulted the DRM to check for health issues that could possibly be relevant to his treatment effort. In response to his request, the DRM decided the set of relevant diseases and forwarded them to *RIMIDI* for further evaluation. Table 10 shows the set of diseases returned by the DRM with the privacy weights supplied by each patient.

Therefore, the DRM decided the same set of disease, $\{o_1, o_2, o_3, o_4, o_5\}$, are relevant to the doctor's ongoing medical effort. However, since *RIMIDI* requires the doctor's trust level, let us suppose that the trust level in every experiment is as shown in Table 11.

Controlled Experiments:

For demonstration purposes, we exhaustively generate the full set of data combinations and compute the risk measures for every one of them. The results are graphically presented in the figures and they show the data combination that has been selected for exposure.

To facilitate understanding, the $Max\ w_i$, *Risk Measure*, *Number of Diseases*, and *Trust Level* are illustrated. Based on the result of calculating the risk of exposure for every data combination, the *Revealed Data* is shown. Moreover, since the outcome of both algorithms for patient 1 and patient 3 are almost the same due to their selected privacy weights which are conservative and non-conservative, respectively. However, we present the results obtained by both algorithms for the moderate privacy weights of patient 2 since they show interesting results. As shown in Figure 7, for trust level $t = 0.8$, $RIMIDI_0$ returns the following set of disease objects: $\{o_2, o_3, o_4, o_5\}$ with corresponding weights: $\{0.6, 0.5, 0.4, 0.2\}$. That is because $RIMIDI_0$ found $MAX\,\rho_{a_{i_0}} = 0.772047$ of that combination satisfies $\leq t = 0.8$.

The risk of such combination is below the trust level of the doctor and holds the maximum possible number of data objects as per the risk measure value computed and when doctor u operates $RIMIDI_1$ the returned results are shown in Figure 8.

Table 10. The set of diseases returned by the DRM for each patient as well as the privacy preferences of each.

Diseases	Patient 1	Patient 2	Patient 3
o_1	0.9	0.2	0.3
o_2	0.7	0.6	0.4
o_3	0.6	0.5	0.5
o_4	0.5	0.7	0.2
o_5	0.1	0.4	0.1

Table 11. Ariable doctors trust levels.

Doctor's Trust Level	
t_1	0.8
t_2	0.5
t_3	0.3

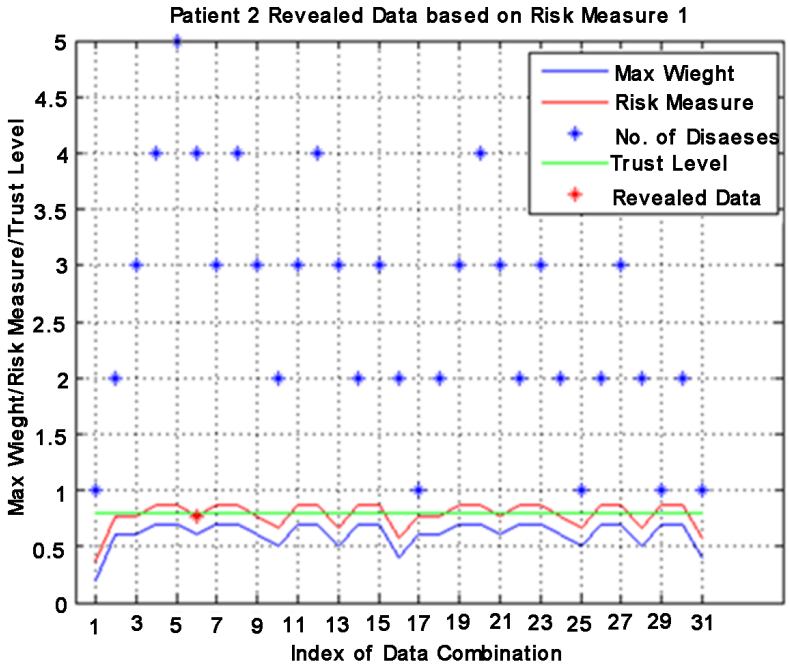

Figure 7. $RIMIDI_0$ results for p_2 and $t = 0.8$.

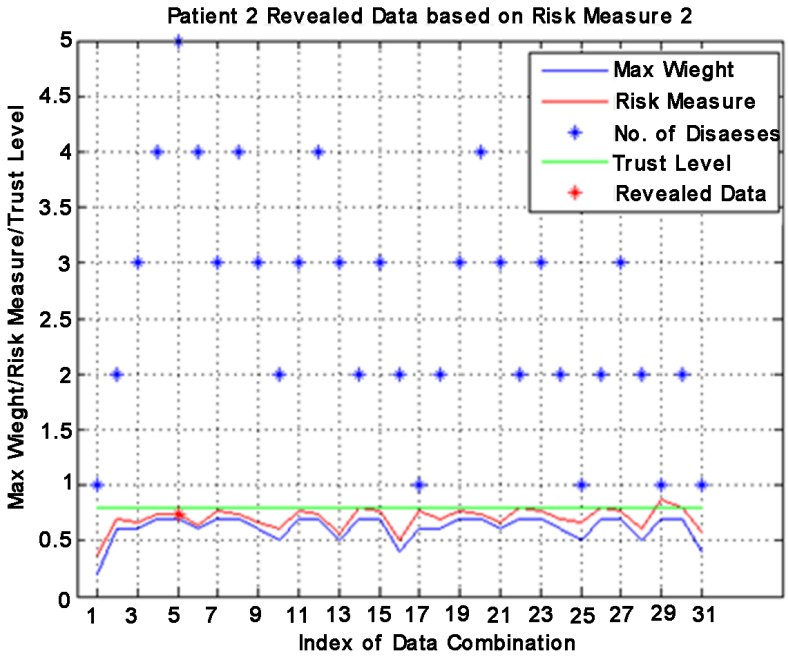

Figure 8. $RIMIDI_1$ results for p_2 and $t = 0.8$.

Figure 9 illustrates that, for trust level $t = 0.8$, $RIMIDI_1$ returns the following set of data objects: $\{o_1, o_2, o_3, o_4, o_5\}$ with corresponding weights: $\{0.7, 0.6, 0.5, 0.4, 0.2\}$, which is the complete set of data objects. That is because the computed risk measure, ρ_{ai_1}, for the combination having the complete set is $0.74 \le t = 0.8$. Recall that, for the instance of the complete set of data, of which belongs to a band of only one instance: itself, the standard deviation is equal to zero. As explained, $RIMIDI_1$ assumes the standard deviation of the immediately next lower band: $band_4$, the computed risk measure obtained concluded that the risk incurred of the complete set is tolerable when compared to trust level $t = 0.8$. Note the different results of risk measures, ρ_{ai_0} and ρ_{ai_1}, with values 0.772047 and 0.74, respectively. Such difference results from the nature of the two algorithms as explained in Section 4.

As shown in Figure 9, for trust level $t = 0.5$, $RIMIDI_0$ returns the following set of disease objects: $\{o_5\}$ with corresponding weights: {0.2}. That is because $RIMIDI_0$ found $MAX\,\rho_{ai_0}$ that satisfies $\le t = 0.5$, which is equal to 0.372047. On the contrary, as shown in Figure 10, when doctor u operates $RIMIDI_1$, the set of $\{o_4, o_5\}$ with corresponding weights $\{0.4, 0.2\}$ is returned. Such combination yields risk measure value of $\rho_{ai_1} = 0.5 \le t = 0.5$. Note that for this combination, $RIMIDI_0$ computed risk measure value of $\rho_{ai_0} = 0.572047$, which is obviously greater than trust level $t = 0.5$, and, therefore, has been blocked from being exposed.

Finally, for patient 2, when doctor u operates $RIMIDI_0$ and $RIMIDI_1$ for trust level $t = 0.3$, none of the algorithms returned a solution because the trust level is too low for the calculated risk measures in all data combinations. Note the red line, which illustrates the risk measure values computed, and the green line denoting the trust level; the risk incurred is very high for any data combination to be revealed with respect to this trust level. As a result, $RIMIDI_0$ and $RIMIDI_1$ returned *NULL*. According to the explanation above, Table 12 summarizes the results obtained from both algorithms for patient 2.

Randomized Experiments:

For demonstration purposes, the above controlled examples utilized a small number of diseases with various weight values and trust levels. Below are several randomized trials, with variable sizes, to show the effectiveness of the proposed algorithms. Unlike the controlled experiments which exhaustively generated the complete set of data combinations, the data combinations in the randomized experiments are generated according to the algorithms presented in Section 4; namely, $RIMIDI_0$ in Figure 3 and $RIMIDI_1$ in Figure 4. For each randomized trial, the experiment begins with specifying the total number of data objects which are assumed to be returned from the DRM as relevant. For this total number, a function that randomly generates the privacy weights is utilized to simulate the sensitivity weights. The algorithm used in the random number generator, *Rand*(), in C++ generates random numbers in an expected behavior. That is, the first sequence of generated values is always the same. To solve this issue, we seed such algorithm with values using the system's clock. Therefore,

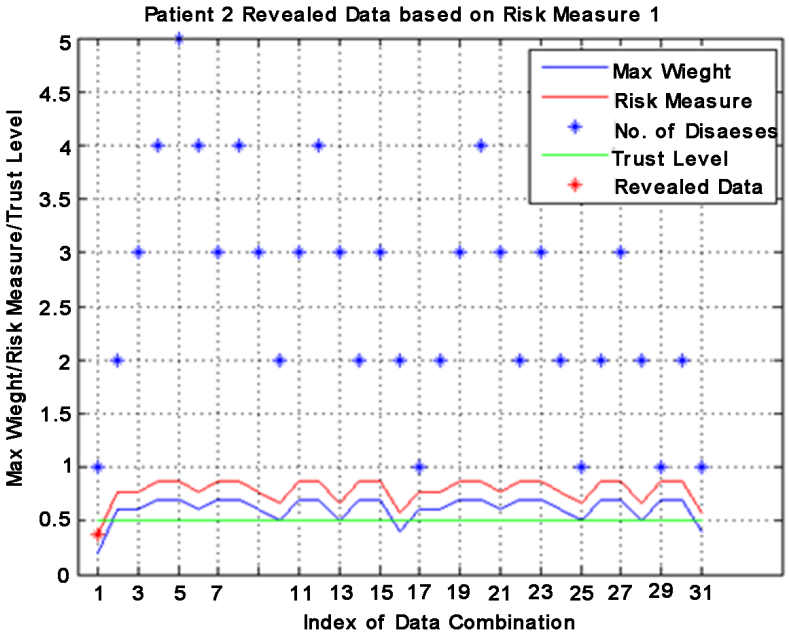

Figure 9. $RIMIDI_0$ results for p_2 and $t = 0.5$.

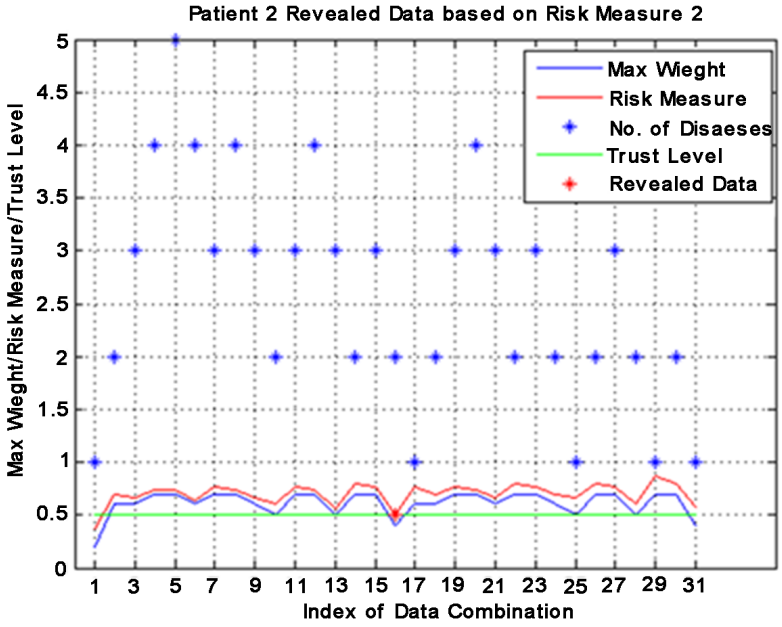

Figure 10. $RIMIDI_1$ results for p_2 and $t = 0.5$.

Table 12. IMIDI Results For Patient 2: The results are different between the two algorithms.

Trust Level (t)	Exposed Data	
	$RIMIDI_0$	$RIMIDI_1$
0.8	$\{o_2, o_3, o_4, o_5\} : \{0.6, 0.5, 0.4, 0.2\}$	$\{o_1, o_2, o_3, o_4, o_5\} : \{0.7, 0.6, 0.5, 0.4, 0.2\}$
0.5	$\{o_5\} : \{0.2\}$	$\{o_4, o_5\} : \{0.4, 0.2\}$
0.3	*NULL*	*NULL*

the values generated are always of different and unexpected values. Later, the generated data is fed to both versions of *RIMIDI* with various trust levels. Each set of the original data are summarized using the average and the standard deviation in order to give intuition about the shape of the random data.

Randomized Trial 1:

The number of data objects returned from DRM as relevant = 0. In this example, as shown in Table 16, the results obtained by both versions of *RIMIDI* are essentially the same. That is, the revealed set of data objects is identical. However, the risk measure incurred of either combination is different in each algorithm. As explained earlier in Section 4, each algorithm follows a different approach in computing the incurred risk of a combination.

Randomized Trial 2:

The number of data objects returned from DRM as relevant = 30. In this example, the results obtained by both algorithms are different. As shown in Table 13, $RIMIDI_1$ was more sensitive in computing the risk measures of data combinations as opposed to $RIMIDI_0$, of which is more conservative. Therefore, more data are revealed by $RIMIDI_1$ than those revealed by $RIMIDI_0$, as illustrated in Table 14.

Randomized Trial 3:

The number of data objects returned from DRM as relevant = 0. In this example, as shown in Table 15, similar to Randomized Trial 1, the results obtained by both algorithms are identical. However, the risk measures incurred of the revealed data combinations are different. Again, the reason is due to the differing behavior of both algorithms concerning the assumption of the distribution as well as the calculation of the standard deviation, as explained in Section 4.

Table 13. he DRM is assumed to return the set of $N = 10$ relevant data objects.

Relevant data objects									
o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}
0.9	0.9	0.7	0.7	0.6	0.6	0.5	0.3	0.2	0.1
MEAN:			*0.55*		***STDEV:***			0.261725	

Table 14. the DRM is assumed to return the set of $N = 30$ relevant data objects.

Relevant data objects									
o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}
0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.7
o_{11}	o_{12}	o_{13}	o_{14}	o_{15}	o_{16}	o_{17}	o_{18}	o_{19}	o_{20}
0.6	0.6	0.6	0.5	0.5	0.5	0.5	0.4	0.4	0.4
o_{21}	o_{22}	o_{23}	o_{24}	o_{25}	o_{26}	o_{27}	o_{28}	o_{29}	o_{30}
0.4	0.4	0.4	0.3	0.3	0.2	0.1	0.1	0.1	0.1
MEAN:			0.5		***STDEV:***			0.23094	

Table 15. The DRM is assumed to return the set of $N = 50$ relevant data objects.

Relevant data objects									
o_1	o_2	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}
0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.7
o_{11}	o_{12}	o_{13}	o_{14}	o_{15}	o_{16}	o_{17}	o_{18}	o_{19}	o_{20}
0.7	0.7	0.7	0.7	0.6	0.6	0.6	0.6	0.5	0.5
o_{21}	o_{22}	o_{23}	o_{24}	o_{25}	o_{26}	o_{27}	o_{28}	o_{29}	o_{30}
0.5	0.5	0.5	0.4	0.4	0.4	0.4	0.4	0.3	0.3
o_{31}	o_{32}	o_{33}	o_{34}	o_{35}	o_{36}	o_{37}	o_{38}	o_{39}	o_{40}
0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.2	0.2	0.2
o_{41}	o_{42}	o_{43}	o_{44}	o_{45}	o_{46}	o_{47}	o_{48}	o_{49}	o_{50}
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
MEAN:			0.438		***STDEV***:			0.256039	

5.2. Analysis and Discussion

Concerning the experimental objectives defined at the beginning of this part, and as shown in the above examples, the relationship between the risk incurred and privacy preferences is that privacy preferences have been protected under both versions of *RIMIDI* algorithms. That is, the algorithms never reveal data combinations, which have risk measures that exceed the trust level of the doctor requesting access. Furthermore, varying the trust level, for the same set of relevant diseases, have never undermined the patient's preferences, but rather, it protected the sensitive information from being unnecessarily exposed to the doctor; of which ultimately protected the patient's privacy.

Moreover, as in the examples of patients p_1, p_2 and p_3, the loading factor played a role in data selection. That is, since the assumption for the data distribution was different in each algorithm, the calculation of the standard deviation was different between the two and, hence, the loading factor varied accordingly. Inspection of the results generated by operating $RIMIDI_0$ and $RIMIDI_1$ for the three patients illustrates the impact of the different loading factors. In $RIMIDI_0$, the loading factor, which is directly affected by the fixed standard deviation, increased the riskiness of a combination by a fixed amount among all combination instances. On the contrary, the loading factors in $RIMIDI_1$ increased the riskiness of combinations with regards to the band to which they belong. That is, the loading factor, π_i, was variable among the different instances and, hence, generated a more sensitive risk measure as opposed to the former as shown in Table 16. Consequently, in various situations, the resulting combination of $RIMIDI_1$ contained more data objects as compared to the results of $RIMIDI_0$. Such result show that $RIMIDI_1$ is more sensitive for calculating the risk measures of the data objects and less conservative as opposed to $RIMIDI_0$, which is more conservative but with simpler and faster algorithm as compared to $RIMIDI_1$. Additionally, the algorithms operated and delivered

Table 16. The results returned from both versions of *RIMIDI* as well as the risk measure values.

Trust Level (t)	$RIMIDI_0$									$RIMIDI_1$								
	ρ_{ai_0}	Data Combination								ρ_{ai_1}	Data Combination							
0.9	0.799998	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}	0.729481	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}
		0.7	0.7	0.6	0.6	0.5	0.3	0.2	0.1		0.7	0.7	0.6	0.6	0.5	0.3	0.2	0.1
0.8	0.799998	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}	0.729481	o_3	o_4	o_5	o_6	o_7	o_8	o_9	o_{10}
		0.7	0.7	0.6	0.6	0.5	0.3	0.2	0.1		0.7	0.7	0.6	0.6	0.5	0.3	0.2	0.1
0.5	0.399998	o_8	o_9	o_{10}						0.40000001	o_8	o_9	o_{10}					
		0.3	0.2	0.1							0.3	0.2	0.1					
0.3	0.299998	o_9	o_{10}							0.30000007	o_9	o_{10}						
		0.2	0.1								0.2	0.1						

results regardless of the number of the data objects returned by the DRM. That is, increasing the number of data objects never undermined the algorithms from producing results promptly.

In $RIMIDI_0$ the complexity of the algorithm is $O(n\log n)$ in the worst case. That is because the most complex part in the algorithm is sorting the data with *MergeSort*, as in line (3) in **Figure 3**, of which has complexity of $O(n\log n)$. On the other hand, $RIMIDI_1$ has a worst-case complexity of $O(n^2 * k)$. That is because the most complex part in the algorithm is in the *while* loop of *sdValues* function, lines (6-14) in **Figure 5**, which iteratively makes system calls to the *binomial_coefficient* function in **Figure 6**, which has $O(n*k)$ complexity due to the utilization of Dynamic Programming. In the worst case, $RIMIDI_1$ runs in $O(n^2k)$. Since the total number of generated combinations for n elements taken r elements at a time is $n!/r!(n-r)!$ with complexity of $O(n^r)$ [46], this is considered as a great improvement as opposed to the run time of exhaustively generating and testing *all* data combinations, which produces and tests 2^n-1 combination instances. Nonetheless, the difference in complexity, between $RIMIDI_0$ and $RIMIDI_1$, is of low impact. We argue that, the number of health issues returned from the DRM is usually with small number. That is because, even if the patient has a large record of health issues, the set of relevant diseases could be small.

One situation could rise when the decided set of relevant data are of the same privacy weights. In this special situation, the calculation of the risk measures, ρ_{ai_0} and ρ_{ai_1}, will be of the same value for every data combination due to the fact that the standard deviation is equal to zero and the $Max\ w_i$ for every instance is equal. Therefore, both versions of *RIMIDI* trap this unusual situation and requests patient consent prior to revealing data combinations.

Furthermore, comparisons between the two algorithms show distinguishable differences. First, with regards to the assumption of the distribution, $RIMIDI_0$ follows a simple and direct approach of assuming the distribution over the set of n data objects. In contrast, $RIMIDI_1$ assumes the distribution over the in-

stances belonging to the same band. Such assumption has a direct impact on the calculations of risk measures. Therefore, while $RIMIDI_1$ had a more complicated approach, it produced a more sensitive risk measure and, hence, more data objects were decided to be safe for exposure as opposed to $RIMIDI_0$.

Comparison between our approach and the other ones show some distinguishable differences. First, with regards to the formula utilized to assess the risk, the approaches [17] [19] [23] [38] utilize the formula devised by NIST, which calculates the risk as the product of the likelihood of vulnerability exploitation multiplied by the potential impact of the occurrence of such event. In contrast, our approach proposes a novel Risk Measure formula that utilizes the standard deviation as a measure of variability and riskiness multiplied by a scaling factor to optimize the formula for the intended purpose. The result of such multiplication is added to the Max weight as preferred by the patient so that no data combination containing such high-weight item is exposed or disclosed; a situation which could undermine the privacy preferences of the patient. Furthermore, the developed Risk Measure is mathematically proven to be *coherent* which means that it can manage the risk effectively.

Moreover, in the other works, when risk is assessed, risk mitigation approaches are often not defined. However, a subset of works proposes risk tolerating intervals such that each access request is evaluated to a risk interval. Each interval is mapped to an associating risk mitigation plan, such as anonymizing the dataset, increasing security measures, obtaining security clearances from system administrators and so forth. In contrast, we propose a risk mitigation approach based on data disclosure according to the risk incurred of data exposure. In this regard, two access requests with the same trust level as well as the same set of relevant data, can have different data exposure because of the unique privacy preferences of each patient. In this regard, our approach provides tailored access control to each health record based on the riskiness of data exposure. This approach protects the privacy of the patient and discloses data in compliance with HIPAA of which can bring forth quality healthcare services. Finally, our approach considers the unique privacy preferences of each patient unlike using majority voting for considering the privacy preferences or using standardized classifications of sensitivity weights. In this way, patients can provide their privacy weights according to their unique preferences. Also, since our approach uses trust level evaluations, access control schemes that use trust calculations can be extended with our risk mitigation approach to control the risk incurred of an access request.

6. Conclusions

When access control schemes employ risk assessment elements, they become dynamic and flexible. Risk assessment can be utilized before or after access is granted to the system resource. When access control allows tolerable risky access to the system resources, risk mitigation approaches can be exploited to lower down the risk incurred of such access. In this research, a risk mitigation ap-

proach for the healthcare systems, utilizing the two developed novel risk measures, is proposed. The risk measures calculate the risk incurred of a data combination. That is, when access is granted to the patient's medical data, the risk measures mitigate the risk accompanying such access using controlling the exposure of the patient's data to the requesting entity.

The risk measures are mathematically proven to effectively manage the risk and, therefore, are coherent. Furthermore, two algorithms, $RIMIDI_0$ and $RIMIDI_1$, which employ the risk measures, ρ_{ai_0} and ρ_{ai_1}, respectively are proposed. Experimental results show the feasibility and effectiveness of the proposed approaches. Specifically, $RIMIDI_0$ holds a slight advantage in terms of algorithm complexity but delivers a more conservative result. In contrast, $RIMIDI_1$ is more sensitive to calculate the risk incurred of data exposure and, therefore, outperforms $RIMIDI_0$ of which can be considered as a benchmark for comparing $RIMIDI_1$.

Future directions include adding relevance factor into consideration for calculating the risk measure. One should note that, in situations where the relevant data objects, returned from the DRM, are of the same relevance value, then this reduces to our problem in this research. Furthermore, in the situation where the returned data objects are of equal relevance and privacy values, one could provide a solution for *RIMIDI* to follow instead of trapping and requesting explicit consent. Finally, the available access control schemes can be extended with the risk measures in this research and compared to other access control schemes that do not employ risk assessment.

References

[1] Ambinder, E.P. (2005) Electronic Health Records. *Journal of Oncology Practice*, **1**, 57. https://doi.org/10.1200/jop.2005.1.2.57

[2] Rindfleisch, T.C. (1997) Privacy, Information Technology, and Health Care. *Communications of the ACM*, **40**, 92-100. https://doi.org/10.1145/257874.257896

[3] (1996) Health Insurance Portability and Accountability Act of 1996. 104-191.

[4] Yang, J.-J., Li, J.-Q. and Niu, Y. (2015) A Hybrid Solution for Privacy Preserving Medical Data Sharing in the Cloud Environment. *Future Generation Computer Systems*, **43**, 74-86. https://doi.org/10.1016/j.future.2014.06.004

[5] Sweeney, L. (2002) k-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **10**, 557-570. https://doi.org/10.1142/S0218488502001648

[6] Stallings, W. and Brown, L. (2014) Computer Security: Principles and Practice. Pearson Education, The United State of America.

[7] Ferraiolo, D.F., Sandhu, R., Gavrila, S., Kuhn, D.R. and Chandramouli, R. (2001) Proposed NIST Standard for Role-Based Access Control. *ACM Transactions on Information and System Security* (*TISSEC*), **4**, 224-274. https://doi.org/10.1145/501978.501980

[8] Lampson, B.W. (1974) Protection. *ACM SIGOPS Operating Systems Review*, **8**, 18-24. https://doi.org/10.1145/775265.775268

[9] Sandhu, R., Ferraiolo, D. and Kuhn, R. (2000) The NIST Model for Role-Based

Access Control: Towards a Unified Standard. In ACM Workshop on Role-Based Access Control, 47-63.

[10] Hu, V.C., Ferraiolo, D., Kuhn, R., Friedman, A.R., Lang, A.J., Cogdell, M.M., Schnitzer, A., Sandlin, K., Miller, R. and Scarfone, K. (2013) Guide to Attribute Based Access Control (ABAC) Definition and Considerations (Draft). *NIST Special Publication*, **800**, 162.

[11] Wang, Q. and Jin, H. (2011) Quantified Risk-Adaptive Access Control for Patient Privacy Protection in Health Information Systems. *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, Hong Kong, 22-24 March 2011, 406-410. https://doi.org/10.1145/1966913.1966969

[12] Jøsang, A., Ismail, R. and Boyd, C. (2007) A Survey of Trust and Reputation Systems for Online Service Provision. *Decision Support Systems*, **43**, 618-644. https://doi.org/10.1016/j.dss.2005.05.019

[13] Kandala, S., Sandhu, R. and Bhamidipati, V. (2011) An Attribute Based Framework for Risk-Adaptive Access Control Models, *Sixth International Conference on Availability, Reliability and Security* (*ARES*), Vienna, 22-26 August 2011, 236-241. https://doi.org/10.1109/ARES.2011.41

[14] Stoneburner, G., Goguen, A.Y. and Feringa, A. (2002) Sp 800-30. Risk Management Guide for Information Technology Systems.

[15] Chen, L. and Crampton, J. (2011) Risk-Aware Role-Based Access Control. In: Meadows, C. and Fernandez-Gago, C., Eds., *Security and Trust Management*, *STM* 2011. *Lecture Notes in Computer Science*, Vol. 7170, Springer, Berlin, Heidelberg, 140-156.

[16] Dorri Nogoorani, S. and Jalili, R. (2016) TIRIAC. *Future Generation Computer Systems*, **55**, 238-254. https://doi.org/10.1016/j.future.2015.03.003

[17] Cheng, P.C., Rohatgi, P., Keser, C., Karger, P.A., Wagner, G.M. and Reninger, A.S. (2007) Fuzzy Multi-Level Security: An Experiment on Quantified Risk-Adaptive Access Control. *IEEE Symposium on Security and Privacy*, Berkeley, CA, 20-23 May 2007, 222-230. https://doi.org/10.1109/SP.2007.21

[18] Dimmock, N., Belokosztolszki, A., Eyers, D., Bacon, J. and Moody, K. (2004) Using Trust and Risk in Role-Based Access Control Policies. *Proceedings of the Ninth ACM Symposium on Access Control Models and Technologies*, Yorktown Heights, New York, 2-4 June 2004, 156-162. https://doi.org/10.1145/990036.990062

[19] Ni, Q., Bertino, E. and Lobo, J. (2010) Risk-Based Access Control Systems Built on Fuzzy Inferences. *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, Beijing, 13-16 April 2010, 250-260.

[20] Shaikh, R.A., Adi, K. and Logrippo, L. (2012) Dynamic Risk-Based Decision Methods for Access Control Systems. *Computers & Security*, **31**, 447-464.

[21] Burnett, C., Chen, L., Edwards, P. and Norman, T.J. (2014) TRAAC: Trust and Risk Aware Access Control. *Twelfth Annual International Conference on Privacy, Security and Trust* (*PST*), Toronto, 23-24 July 2014, 371-378. https://doi.org/10.1109/PST.2014.6890962

[22] Pontual, M., Chowdhury, O., Winsborough, W.H., Yu, T. and Irwin, K. (2011) On the Management of User Obligations. *Proceedings of the 16th ACM Symposium on Access Control Models and Technologies*, Innsbruck, 15-17 June 2011, 175-184. https://doi.org/10.1145/1998441.1998473

[23] Díaz-López, D., Dólera-Tormo, G., Gómez-Mármol, F. and Martínez-Pérez, G. (2016) Dynamic Counter-Measures for Risk-Based Access Control Systems: An Evolutive Approach. *Future Generation Computer Systems*, **55**, 321-335.

https://doi.org/10.1016/j.future.2014.10.012

[24] Taneja, H. and Singh, A.K. (2015) Preserving Privacy of Patients Based on Re-Identification Risk. *Procedia Computer Science*, **70**, 448-454. https://doi.org/10.1016/j.procs.2015.10.073

[25] Armando, A., Bezzi, M., Metoui, N. and Sabetta, A. (2015) Risk-Aware Information Disclosure. In: Garcia-Alfaro, J., *et al*., Eds., *Data Privacy Management*, *Autonomous Spontaneous Security*, *and Security Assurance*, *Lecture Notes in Computer Science*, Vol. 8872, Springer, Berlin, Heidelberg, 266-276.

[26] WHO. (2017) International Classification of Diseases. http://www.who.int/classifications/icd/en/

[27] Stine, K.M., Kissel, R., Barker, W.C., Lee, A., Fahlsing, J. and Gulick, J. (2008) SP 800-60 Rev. 1. Volume I: Guide for Mapping Types of Information and Information Systems to Security Categories; Volume II: Appendices to Guide for Mapping Types of Information and Information Systems to Security Categories.

[28] Hypertension and Cardiovascular Disease. http://www.world-heart-federation.org/cardiovascular-health/cardiovascular-disease-risk-factors/hypertension/

[29] Wang, Q. and Jin, H. (2012) An Analytical Solution for Consent Management in Patient Privacy Preservation. *Proceedings of the* 2*nd ACM SIGHIT International Health Informatics Symposium*, Miami, FL, 28-30 January 2012, 573-582. https://doi.org/10.1145/2110363.2110427

[30] Smari, W.W., Clemente, P. and Lalande, J.-F. (2014) An Extended Attribute Based Access Control Model with Trust and Privacy: Application to a Collaborative Crisis Management System. *Future Generation Computer Systems*, **31**, 147-168. https://doi.org/10.1016/j.future.2013.05.010

[31] J. P. Office. (2004) Horizontal Integration: Broader Access Models for Realizing Information Dominance. MITRE Corporation Technical Report JSR-04-132.

[32] McGraw, R. (2009) Risk-Adaptable Access Control (radac). Privilege (Access) Management Workshop. NIST, National Institute of Standards and Technology, Information Technology Laboratory.

[33] Zhang, L., Brodsky, A. and Jajodia, S. (2006) Toward Information Sharing: Benefit and Risk Access Control (BARAC). *Seventh IEEE International Workshop on Policies for Distributed Systems and Networks*, London, 5-7 June 2006, 9-53. https://doi.org/10.1109/POLICY.2006.36

[34] Molloy, I., Cheng, P.-C. and Rohatgi, P. (2009) Trading in Risk: Using Markets to Improve Access Control. *Proceedings of the* 2008 *Workshop on New Security Paradigms*, Lake Tahoe, CA, 22-25 September 2008, 107-125.

[35] Crossley, M.L. (2000) The Desk Reference of Statistical Quality Methods. ASQ Quality Press, Milwaukee, Wisconsin.

[36] Zhou, L., Varadharajan, V. and Hitchens, M. (2015) Trust Enhanced Cryptographic Role-Based Access Control for Secure Cloud Data Storage. *IEEE Transactions on Information Forensics and Security*, **10**, 2381-2395. https://doi.org/10.1109/TIFS.2015.2455952

[37] Kamwan, C. and Senivongse, T. (2016) Risk of Privacy Loss Assessment of Cloud Storage Services. 18*th International Conference on Advanced Communication Technology* (*ICACT*), Pyeong Chang, 31 January-3 February 2016, 105-111.

[38] Khambhammettu, H., Boulares, S., Adi, K. and Logrippo, L. (2013) A Framework for Risk Assessment in Access Control Systems. *Computers & Security*, **39**, 86-103.

[39] Crampton, J. and Morisset, C. (2010) An Auto-Delegation Mechanism for Access Control Systems. In: Cuellar, J., Lopez, J., Barthe, G. and Pretschner, A., Eds., *Security and Trust Management*, *STM* 2010. *Lecture Notes in Computer Science*, Vol. 6710, Springer, Berlin, Heidelberg, 1-16.

[40] Krautsevich, L., Martinelli, F., Morisset, C. and Yautsiukhin, A. (2012) Risk-Based Auto-Delegation for Probabilistic Availability. In: Garcia-Alfaro, J., Navarro-Arribas, G., Cuppens-Boulahia, N. and de Capitani di Vimercati, S., Eds., *Data Privacy Management and Autonomous Spontaneous Security*, *Lecture Notes in Computer Science*, Vol. 7122, Springer, Berlin, Heidelberg, 206-220.

[41] Ardagna, C.A., De Capitani di Vimercati, S., Foresti, S., Paraboschi, S. and Samarati, P. (2012) Minimising Disclosure of Client Information in Credential-Based Interactions. *International Journal of Information Privacy*, *Security and Integrity*, **1**, 205-233.

[42] Anciaux, N., Nguyen, B. and Vazirgiannis, M. (2011) Miminum Exposure in Classification Scenarios. INRIA Research Report, 2012. http://www-smis.inria.fr/~anciaux/MinExp/

[43] Ferson, S., Kreinovich, V., Hajagos, J., Oberkampf, W. and Ginzburg, L. (2007) Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty. Sandia National Laboratories, Report SAND2007-0939, 162.

[44] Hogg, R., Craig, A. and McKean, J. (2005) Introduction to Mathematical Statistics. Prentice Hall, Upper Saddle River, New Jersey.

[45] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999) Coherent Measures of Risk. *Mathematical Finance*, **9**, 203-228. https://doi.org/10.1111/1467-9965.00068

[46] Cormen, T.H. (1993) Introduction to Algorithms. MIT Press, MIT Press, Cambridge, MA.