On Security of RASP Data Perturbation for Secure Half-Space Queries in the Cloud

ABSTRACT

Secure data intensive computing in the cloud is challenging, involving a complicated tradeoff among security, performance, extra costs, and cloud economics. Although fully homomorphic encryption is considered as the ultimate solution, it is still too expensive to be practical at the current stage. In contrast, methods that preserve special types of data utility, even with weaker security, might be acceptable in practice. The recently proposed RASP perturbation method falls into this category. It can provide practical solutions for specific problems such as secure range queries, statistical analysis, and machine learning. The RASP perturbation embeds the multidimensional data into a secret higher dimensional space, enhanced with random noise addition to protect the confidentiality of data. It also provides a query perturbation method to transform half-space queries to a quadratic form and, meanwhile, preserving the results of half-space queries. The utility preserving property and wide application domains are appealing. However, since the security of this method is not thoroughly analyzed, the risk of using this method is unknown. The purpose of this paper is to investigate the security of the RASP perturbation method based on a specific threat model. The threat model defines three levels of adversarial power and the concerned attacks. We show that although the RASP perturbed data and queries are secure on the lowest level of adversarial power, they do not satisfy the strong indistinguishability definition on higher levels of adversarial power. As we have noticed, the indistinguishability definition might not be too strong to be useful in the context of data intensive cloud computation. In addition, the noise component in the perturbation renders it impossible to exactly recover the plain data; thus, all attacks are essentially estimation attacks. We propose a weaker security definition based on information theoretic measures to describe the effectiveness of estimation attacks, and then study the security under this weaker definition. This security analysis helps clearly identify the security weaknesses of the RASP perturbation and quantify the expected security under different levels of adversarial power.

1. INTRODUCTION

Data-driven (big data) approaches represent an emerging trend of scientific research and industrial development. With the rapid development of cloud infrastructures, conducting data intensive computing in the cloud has become the top choice for economical and scalable processing [4, 2]. Such computing tasks may include but not limited to querying, mining, and visualizing data. They share a commonality that users need to manipulate the data on top of the cloud infrastructures, not simply storing data in the cloud. However, once the data collections are exported to the cloud, the data owner loses the control over the data. As big data collections are becoming important properties, and cloud providers are normally considered as an untrusted party, data owners have reasonable concerns over data ownership, security, and privacy [22]. Unless we can find a way to conduct computation on "protected" data in the cloud, most data owners will not be likely to use public clouds to process their sensitive data.

Secure computation in the cloud has to put equal weights on both security and utility of the protected data. Traditional encryption approaches cannot be simply applied because they do not protect data utility (except for the simple case of storing data in the cloud). When considering the utility of a solution, we cannot ignore the cost of the scheme as well. Fully homomorphic encryption (FHE) [18] aims to implement the lowest level operations: addition and multiplication on the encrypted data without decrypting the data. Theoretically, any function can be built on addition and multiplication by using a FHE scheme. However, as most researchers noticed, at the current stage it is too expensive to be practical, even for a simple application like keyword search over encrypted database [34]. On the other hand, many approaches developed in the database community focus on performance, only providing very weak security. For example, Crypto-index [23, 25] and order-preserving encryption (OPE) [3, 5] depend on strong assumptions that attackers do not have sufficient prior knowledge about the data, which excludes many realistic attacks from investigation.

Instead of preserving the utility of low level operations on single dimensions, Chen et al. proposed the RASP perturbation approach [9] to preserve half-space queries [8] for multidimensional data. A half-space query works on multidimensional data space and retrieves data vectors satisfying the query condition. It uses a hyperplane to partition the whole space, and returns the result (normally statistics, like the count of vectors) about one of the half spaces. The RASP perturbation approach also allows the original half-space queries to be securely transformed and processed on the perturbed dataset to get the exact query results as they do on the original dataset. Half-space queries have many applications in data intensive analysis. For example,

- *Interactive Data Analysis.* Range queries are the most common queries in interactive data analysis, which are used to derive statistics on a specific portion of data that satisfies the

query conditions. A range query is an intersection of a number of half-space queries, describing "areas" in the multidimensional space. Thus, preserving half-space queries also preserves range queries.

- *Machine Learning.* Linear classifier is one of the most popularly used classification models. A linear classifier is represented as a half-space query. Given a record, the classifier checks whether the record is in the half-space query result set or not [24]. Linear classifiers can serve as the base classifiers in the bagging or boosting approach [15] to construct more sophisticated classification models.

Thus, we believe that the RASP perturbation has great potentials for many data analytics tasks in the cloud. However, the security of the RASP perturbation method has not been carefully studied yet.

Scope of the Paper. In this paper, we aim to systematically analyze the security of RASP perturbation based on a precisely defined threat model. The threat model describes the protected assets in the context of secure computation in the cloud, the concerned attacks (passive data disclosure attacks only), and the prior knowledge that attackers may have. The perturbed data and queries are studied under two types of security definitions: the strong indistinguishability definition [27], and the statistical-estimation based weakened definition, corresponding to the traditional distinguishing attacks and the more realistic estimation attacks in the context of cloud computing, respectively. Based on the threat model, we start our studies from the stronger security definition and then extend to the weaker definition. To evaluate the security to the estimation attacks, we also design a new measure - the loss of confidentiality (LOC) - based on statistical learning theory and information theory. Concretely, there are several unique contributions.

1. We define a precise and comprehensive threat model for applying the RASP perturbation approach in the cloud, which suggests studies on two types of attacks: distinguishing attacks and estimation attacks, based on three levels of adversarial power that attackers may obtain in the context of cloud computing.

2. We prove that the RASP perturbation is not indistinguishably secure under the Level 2 and 3 adversarial power.

3. We develop a weakened security definition for estimation attacks, and design the LOC measure for evaluating the security of RASP perturbation on estimation attacks.

4. We show how the LOC measure is applied to evaluate estimation attacks under the Level 3 adversarial power.

5. We also study the query privacy based on the same threat model. The result shows that perturbed queries are also not indistinguishably secure under the Level 2 and 3 adversarial power. We show that the information exposed by queries does not damage the security of the perturbed data.

The rest of the paper is organized as follows. Section 2 describes the framework and the thread model that the RASP perturbation works with. Section 3 precisely defines the RASP perturbation. Section 4 focuses on the security of the RASP perturbed data, and Section 5 focuses on the perturbed queries. Section 6 presents some relevant approaches for secure computation in the cloud.

2. FRAMEWORK AND THREAT MODELING

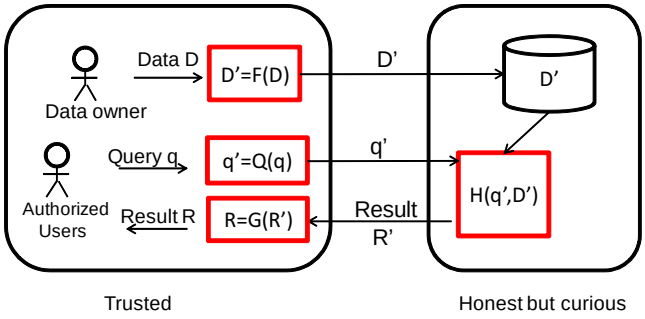

Figure 1: Illustration of the framework for perturbation-based secure computation in the cloud.

This section gives the framework that the RASP perturbation works with, which is actually general to all secure computation approaches in the cloud. Based on this framework, we precisely define a threat model for security analysis.

2.1 Framework

Figure 1 shows the typical scenario of hosting a secure interactive statistical analysis service in the cloud. The data owner exports the perturbed data to the cloud and then queries the protected data to learn statistics. Meanwhile, the authorized users can also submit queries to learn statistics. The database in the cloud serves as a statistical database [12], but it is only accessible to the authorized users. Thus, it is different from the privacy problems in traditional statistical databases, where the service provider is trusted and the users can be any person including malicious ones.

There are a number of basic procedures in this framework: (1) $F(D)$ is the RASP perturbation that transforms the original data D to the perturbed data D'; (2) $Q(q)$ transforms the original query q to the protected form q' that can be processed on the perturbed data; (3) $H(q', D')$ is the query processing algorithm that returns the result R', typically the number of records. When other statistics such as SUM or AVG of a specific dimension are needed, RASP can work with partial homomorphic encryption such as Paillier encryption [32] to compute these statistics on the encrypted data [17], which are then recovered with the procedure $G(R')$. For simplicity, we assume the queries return the simplest statistics: the number of records. Statistical analysis heavily depends on the this information to estimate the distributions. In addition, classification modeling can be achieved with only the count statistics [28], if the class label is presented unperturbed.

2.2 Threat Modeling

The application scenario is described as in Figure 1. The data owner stores and processes perturbed data in the cloud. The cloud environment is untrusted and out of the data owner's control. The aim of our design is to prevent attackers from precisely recovering or estimating the original data. We consider several aspects of the threat model [31] in our research.

Assets. The protected assets are the data stored and processed in the cloud. We assume these data are multidimensional vector databases. The privacy of query is also concerned.

Passive data disclosure attacks. Passive data disclosure is our major concern. Attackers can access the data at any of compromised virtual machines in the cloud. They might be interested in recovering or estimating the *original data records* or *distributional information*, based on the perturbed data. While distinguishing attacks are meaningful to traditional encryption systems, they are less

useful in our context. We identify that the main attacks are based on statistical estimation. Data tampering or dishonest cloud service providers is not addressed by our study, which can be covered by integrity preserving techniques [38, 35, 30].

Attacker modeling. Active attackers will try to obtain as much knowledge as possible to help recover the original data. To better analyze the security of the RASP perturbation, we define the adversarial power according to the levels of prior knowledge on the data.

- **Level 1**: the attacker observes only the perturbed data and the perturbed queries, without any other additional knowledge.
- **Level 2**: apart from the perturbed data, the attacker also knows the domain of the original data, such as the meaning of the attributes, the attribute domains, the attribute distributions (e.g., the probability density functions (PDF) or histograms), and the covariance between attributes. In practice, such distributional information is possibly exposed to the public via statistical database interfaces. Similarly, the query distribution might be leaked via side channels.
- **Level 3**: the attacker manages to obtain a small set of plain records and the corresponding perturbed data records, via some side channel. We call them known input-output pairs. This corresponds to the known-plaintext attack in cryptography. Similarly, they may gain known input-output query pairs.

Adversarial power in Level 3 is quite strong in the setting of secure cloud computing. Since the database is not open to the public, the attacker has to depend on unusual methods such as social engineering. Based on this thread model, we will analyze the security of the perturbed data and protected queries, respectively.

3. RASP DATA AND QUERY PERTURBATION

Chen et al. has presented the RASP perturbation method [9] for efficient secure range query processing in the cloud. In this section, we will precisely define the RASP perturbation algorithm with a more general setting. The query perturbation algorithm is also described to show how the query utility is preserved. We also describe some applications of the RASP perturbation to show its significance.

3.1 Perturbing Data

We assume the dataset exported to the cloud for processing is a set of d-dimensional column vectors. For the convenience of formal treatment, we assume all the values are n-bit real numbers. The i-th data vector is represented as a column vector $x_i = (x_{i1}, \ldots, x_{id})^T$. We also use $\mathbb{R}^d$ to represent the d-dimensional real space and thus $x_i \in \mathbb{R}^d$. The RASP perturbation method is defined in the following steps.

- **Gen:** Assume that we have a pseudo-random invertible matrix generator $\mathcal{K}_A$. Let's draw a $(d+2) \times (d+2)$ matrix A randomly, where each element has n bits.
- **PreP:** d randomly chosen non-linear strictly monotonic functions f_j, $j = 1 \ldots d$ are used to transform each dimension. Let y defined by the dimensional values $y_{.j} = f_j(x_{.j})$, where $x_{.j}$ and $y_{.j}$ represent the dimensional values. This step is to introduce non-linear components and change the distributions to increase the security of the perturbation.
- **Ext:** A random positive real value r, $r \in \mathbb{R}^+$, is selected from a pseudo-random positive real number generator $\mathcal{K}_r$, which is used to extend the vector y to $z = (y^T, 1, r)^T$, where the notation y^T represents vector transpose.
- **Pert:** On the matrix A, and the vector z, the final perturbed result is

$$p = Az, \tag{1}$$

  where Az is the multiplication between the matrix A and the vector z, which results in a $(d+2)$-dimensional vector p in the perturbed space.

In the RASP perturbation a set of parameters are fixed for the same dataset, including the functions $f_j()$ in the **PreP** step, and the parameter matrix A in the **Pert** step, while the positive random noise is individually generated for each vector. Note that the RASP perturbation does not have a decryption component. Because the perturbed data are only used to support statistical analysis, there is no need to recover the perturbed data. In the case that the original records need to be recovered, the vector p can be stored together with the encrypted vector $e = Enc(x)$, where Enc is an existing encryption algorithm.

Because of the random noise component in z, for any original vector x, there are possibly 2^{n-1} images (the total choices of the positive random numbers) in the perturbed space. For the same vector perturbed multiple times, each time we get a different image with high probability. However, any one of the perturbed vectors is uniquely mapped back to one of the original vectors. This guarantees the correctness of processing queries in the perturbed space.

In the **PreP** step, Chen et al. used an order-preserving encryption (OPE) scheme to transform the data distributions to Gaussian distributions [9]. According to the definition of OPE, OPE is one type of strictly monotonic function. Due to the **Ext** and **Pert** steps, the RASP approach does not preserves the dimensional value order, and it is thus not a variant of OPE schemes. Therefore, RASP is not subject to the dimensional-order based attacks on OPE schemes [3, 5].

3.2 Perturbing Half-space Queries

The typical half-space queries in the original space are like $x_{.j} < a$ as conditions presented in typical SQL queries [14], where $x_{.j}$ represents the dimension j, a is a constant, and the comparison operations can be any of $\{<, >, \geq, \leq\}$. For simplicity, we only discuss the case $x_{.j} < a$ while other cases are similar. In order to query the perturbed data, the original query needs to be represented in the perturbed data space, and the transformed query needs to return the same set of records as the original query does.

Let's look at the RASP transformations step by step and apply them to the query. The first step is to apply the strictly monotonic function to the query, which results in $y_{.j} < f_j(a)$ (or $y_{.j} > f_j(a)$ if f_j is monotonically decreasing), based on the strict monotonicity. Let u be a $(d+2)$-dimensional vector $(\ldots 1 \ldots - f_j(a), 0)$, where j-th dimension is 1, $(d+1)$-th dimension is $-f_j(a)$ and all other dimensions are 0. Then, the original half-space query is transformed to the query

$$\begin{aligned} y_{.j} - f_j(a) &= (\ldots, 1, \ldots, -f_j(a), 0)(y_{.1}, \ldots, y_{.d}, 1, r)^T \\ &= u^T z < 0, \end{aligned} \tag{2}$$

where z is as defined in the **Ext** step and $u^T z$ is dot product of the two vectors. It is then further transformed to $u^T A^{-1} p < 0$ according to the **Pert** step. Let $H(p) < 0$ denote the transformed query. It is straightforward to verify that $H(p) < 0$ is equivalent to

the original query $x_{.j} < a$, thus we can expect both to return the same set of records.

However, this initial version of query transformation is not resilient to attacks [9]. The problem is fundamentally rooted on the simplicity of the query function - essentially $H(p)$ is a linear function. The following method is used to expand the query function to an equivalent quadratic form by including the random noise dimension. Note that the $(d+2)$-th dimension of z is always positive and randomly chosen. Thus, the query function $z_{.,d+2}H(p) < 0$ is equivalent to the original query $H(p) < 0$. Similarly, $z_{.,d+2}$ can be represented in terms of the variable p. Let $\phi = (\ldots, 1)^T$ with all dimensions zero except for only the $d+2$ dimension set to 1. Thus, $z_{.,d+2}H(p) < 0$ can be represented as $p_i^T(A^{-1})^T\phi u^T A^{-1}p_i < 0$, which is simplified to the canonical quadratic form

$$p_i^T Q p_i < 0, \tag{3}$$

where Q is the *query parameter matrix*: $(A^{-1})^T\phi u^T A^{-1}$. Let α_j be the j-th row of A^{-1}. Q is in fact $(\alpha_j - f_j(a)\alpha_{d+1})^T\alpha_{d+2}$, which will be used in discussing the security of perturbed queries.

3.3 Sample Applications

Since the half-space queries are fully preserved, we can apply this technique to securely learn the range query results. A typical range query in the original space is a hyper-cube formed by multiple half-space queries. Figure 2 (Left) shows the enclosed area of a typical range query in the two dimensional space. Four half-space queries (the four sides) together define this range query. After the perturbation, the range query is transformed to irregular shapes in the perturbed data space, where traditional multidimensional indexing structures such as R-trees can be used to reduce the time of query processing to sub-linear [9].

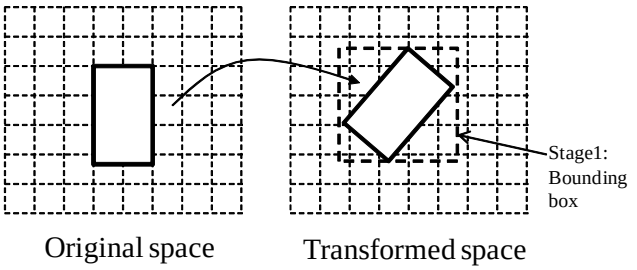

Figure 2: Illustration of range query processing on RASP-perturbed data.

This perturbed data can also be used for securely learning models from labeled data vectors. Without loss of generality, we work on the two-class case, where the labels are either '+' or '-'. And we assume the labels are insensitive and thus not perturbed. The user can thus submit random half-space queries to learn the numbers of '+' and '-' instances in the half-spaces, respectively. By applying techniques such as boosting [15] a sophisticated classification model can be learned (Figure 3).

4. SECURITY ANALYSIS ON RASP PERTURBED DATA

We analyze the data security in terms of the three levels of adversarial power. We will dedicate the next section to the security of perturbed queries.

4.1 Level 1 Security Analysis

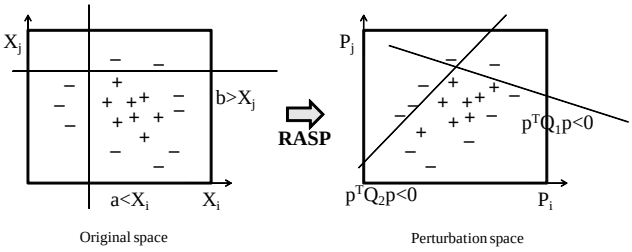

Figure 3: Illustration of learning classifiers.

The Level 1 assumption actually fits many application scenarios in the cloud computing context, where no additional information about the private data is leaked. We show that the RASP perturbation is indistinguishably secure under this assumption.

PROPOSITION 1. *RASP perturbed data is indistinguishably secure on the Level 1 assumption.*

PROOF. The proof is based on the assumption that the random invertible matrix is uniformly sampled from the whole set of $(d+2)$-dimension invertible matrices $\mathbf{\Omega}^{d+2}$. Let the perturbed vector drawn from a random variable P, and the vector generated in the **Ext** step drawn from another random variable Z. Let's just look at the difficulty of finding z_0 that generates a sample vector p of P, while finding the original vector x_0 will certainly not be easier.

We try to estimate the probability $Pr(Z = z|P = p)$, where z is *any* vector in the definition of Z. For a randomly selected z, there is an angle between z and p, denoted as θ. Thus, there exists a unique "rotation" (orthogonal) matrix R_θ and a scalar s, so that $p = sR_\theta z$, i.e., rotating z for an angle θ towards p and then scaling up/down z to the same length of p. Since R_θ is invertible, according to the definition of the **Pert**, sR_θ is a valid setting for A. Thus, the probability $Pr(Z = z|P = p)$ is $Pr(A = sR_\theta|P = p)$, which is in turn $Pr(A = sR_\theta)$ as A is randomly selected independent of P.

An n-bit representation can represent at most 2^n distinct values. According to the orthogonal group theory [33], the number of $(d+2) \times (d+2)$ orthogonal matrices in the field $\mathbb{F}_{2^n}$ is around $O(2^{dn}\prod_{i=1}^{d/2}(2^{nd} - 2^{2ni}))$. Based on the assumption of uniformly choosing the invertible matrix A from all invertible matrices that include the orthogonal group, we have $Pr(Z = z|P = p) = Pr(A = sR_\theta) < 1/O(2^{dn}))$. Thus, for any pair of vectors in Z, (z_0, z_1), the probability of distinguishing which one is perturbed to p is

$$|Pr(Z = z_0|P = p) - Pr(Z = z_1|P = p)| < 2/O(2^{dn}). \tag{4}$$

i.e., they cannot be distinguished. □

4.2 Level 2 and 3 Indistinguishability

We show that the RASP perturbed datasets do not satisfy the indistinguishability definition, if Level 2 or Level 3 assumption on the adversarial power is held.

4.2.1 ICA Attack with Level 2 Knowledge.

We show that if the attacker knows distributional information (Level 2 knowledge), with the **PreP** step some datasets can be possibly reconstructed with the Independent Component Analysis (ICA) method [26]. ICA is a fundamental problem in signal processing that has many applications such as blind source separation of mixed electro- encephalographic(EEG) signals, audio signals and the analysis of functional magnetic resonance imaging (fMRI)

data. Let matrix S composed by source signals, where row vectors represent source signals, and column vectors represent the values of different signals at certain time stamp. Suppose we can observe the mixed signals Y, which is generated by a linear transformation $Y = AS$. The ICA model is designed to reconstruct the independent components (the row vectors) of the original signals S from the mixed signals Y, if the following conditions are satisfied:

1. The source signals are independent to each other;
2. All source signals must have non-Gaussian distribution with possible exception of one signal;
3. The number of observed signals, i.e. the number of row vectors of Y, must be at least as large as the independent source signals;
4. The transformation matrix A must be of full column rank.

The existing ICA algorithms [26, 24] share the same idea that tries to find a matrix W where $\hat{S} = WY$ maximizes the non-Gaussianity and independency of the resultant row vectors, which are used to approximate the signals in S. However, the ICA algorithms will result in re-ordering and re-scaling of the signals, for which the user needs to use the signal characteristics such as signal distributions to precisely identify each reconstructed signal.

ICA reconstruction can be certainly used to attack RASP perturbed data. It has several implications. (1) If the noise dimension of z is non-Gaussian, the noise can be possibly reconstructed, which endangers the security of the perturbed data. (2) Without the **PreP** step, independent dimensions of the original data can be effectively reconstructed. With the Level 2 knowledge, the attacker can use the distributional information to correctly identify the reconstructed values and align them with the original data domain. (3) Without carefully designed **PreP**, which generates non-Gaussian independent dimensions, the independent and non-Gaussian dimensions of z in the **Ext** can be reconstructed, based on which more methods might be developed to crack the **PreP** step.

Thus, the best practice is to make all dimensions of z, except for the constant $(d+1)$-th dimension, transformed to Gaussian distribution. This can be done in the **PreP** step for the first d dimensions, and to use pseudo-random positive Gaussian noise generator for the $(d+2)$-th dimension.

4.2.2 *Plane Attack with Level 3 Knowledge.*

If attackers are able to submit half-space queries and obtain pairs of plain and corresponding perturbed vectors, we show that a *Plane Attack* can be used to distinguish the perturbation of any pair of chosen plain vectors.

PROPOSITION 2. *RASP is not indistinguishable to chosen plaintext attack, if the attacker can also submit half-space queries.*

PROOF. Let $c = F(m)$, where m is the plain vector and c is the perturbed vector. The distinguisher experiment is described as follows.

1. m_0 and m_1 are two vectors randomly sampled from the original data space $\mathbb{R}^d$.
2. The vector m_b, where $b \in \{1, 0\}$ is randomly selected, is perturbed to c_b with the RASP perturbation and given to the adversary.
3. In addition, the adversary can request a polynomial number of plain vectors $\{m_i, i > 1, m_i \in \mathbb{R}^d\}$ to be perturbed, where $m_i \neq m_0$ and $m_i \neq m_1$. With some attacking algorithm, the adversary finally outputs a bit b'.

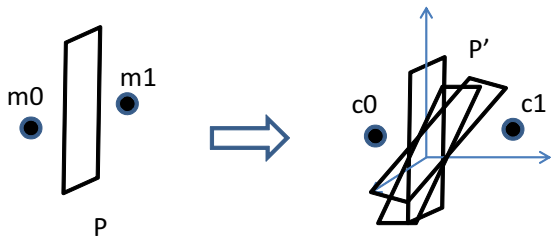

Figure 4: Plane Attack illustrated

If $|Pr(b' = b) - Pr(b' \neq b)| < 1/p(n)$, $p(n)$ is some polynomial function in terms of the bit-length n, we say the perturbed data vectors are computationally indistinguishable under chosen plaintext attack (IND-CPA). We show that the following "Plane Attack" allows the adversary to accurately predict b, i.e., $Pr(b' = b) = 1$. Thus, RASP is not IND-CPA.

If $m_0 \neq m_1$, there is always a plane on some dimension, say $x_{.j} = a$, separating these two vectors. Let's assume m_0 in the half-space $x_{.j} < a$, and m_1 in $x_{.j} > a$. Remember that the perturbed half-space queries always preserve the query results. If the attacker is able to submit the perturbed query of $x_{.j} < a$, say $p^T Q p < 0$, then the attacker can always precisely determine b - if $c_b^T Q c_b < 0$ then $b = 0$, otherwise, $b = 1$. Such planes can be many, and any one of them can serve as the distinguishing purpose. Figure 4 illustrates this attack. □

The above result shows that RASP perturbed datasets cannot defeat distinguishing attacks. However, distinguishing attacks in the context of RASP-based cloud computing appear not as important as in the secure communication. As described in the threat model, attackers could be more interested in estimating the original plain vectors. Such estimation attacks make us consider a weakened security definition.

4.3 Weakened Security Definition Based on Learning Theory and Information Theory

As the indistinguishability definition is not satisfied on the Level 2 and 3 adversarial power, we explore a weakened definition based on statistical estimation. This becomes necessary because the presence of the random noise component in the RASP perturbation makes all attacks essentially estimation attacks. Attacks cannot exactly discover the exact original vectors; the best they can do is statistical estimation - a unique feature distinguishing from encryption schemes. In this sense, attacks can be modeled as a learning problem: given the background knowledge (e.g., Level 2 and 3) and the perturbed data, what is the theoretical bound of the estimation accuracy?

As the accuracy is related to the data domain and distribution, we will need a measure to precisely describe it without being biased towards different domains and distributions. We adopt the idea proposed by Agrawal et al. [1] for the quantification of privacy in privacy-preserving data mining. The derived measure is explained and defined as follows. This measure considers the effect of original data distributions, assuming the Level 2 and 3 knowledge is available to the attackers. Let X represent a random variable generating a dimension of the dataset, and $\hat{X}$ represent the estimated values. The measure considers both the uncertainty of X and how accurately $\hat{X}$ can be used to model X. Intuitively, if most of the X population is in a narrow range, a random sample on that distribution will give an accurate estimation on an unknown value. We

call this uncertainty the inherent amount of confidentiality for that data dimension, which can be evaluated by the normalized entropy $2^{h(X)}$. Calculating this measure for a uniform distribution in the domain [0, a] can help explain the intuition: $2^{h(X)} = a$. Thus, the amount '1' of this measure has a specific implication: it is equivalent to the amount of uncertainty associated to a uniform distribution in [0, 1]. With $\hat{X}$, the X's uncertainty might be reduced, which is evaluated by the conditional differential entropy $h(X|\hat{X})$, similarly normalized to $2^{h(X|\hat{X})}$. Now we can define the normalized loss of confidentiality (LOC) as

$$L(X|\hat{X}) = (2^{h(X)} - 2^{h(X|\hat{X})})/2^{h(X)} = 1 - 2^{-I(X;\hat{X})}, \quad (5)$$

where $I(X;\hat{X}) = h(X) - h(X|\hat{X})$ is the mutual information between the two random variables. This measure has a lower bound 0, when the attack is no better than random sampling from the known distribution (with the Level 2 knowledge). It has a upper bound $1 - 2^{-h(X)}$, determined by the uncertainty of the original domain.

If this error E is also independent of the the estimation X, which is valid in the case of independent noise injection, then $h(X|\hat{X}) = h(E)$ for $X = \hat{X} + E$. Thus, the LOC measure is determined by $h(E)$ for a given data domain. The entropy of error represents the effectiveness of attack - the smaller $h(E)$, the more effective the attack has. However, the learning theory [29] says there is the theoretical lower bound of estimation error for any possible learners for the specific training data. That says $h(E)$ cannot be arbitrarily small; it has a lower bound. Therefore, we have the following weakened security definition.

DEFINITION 1. If the estimation error is independent of the the original data, the LOC measure $1 - 2^{h(E)-h(X)}$ defines the security of a specific dimension X for a RASP-perturbed dataset under the specific estimation attack. The theoretical lower bound of $h(E)$ defines the corresponding absolute security under all estimation attacks.

Sometimes, it might be inconvenient to directly analyze the entropy. If we further assume the error E has a Gaussian distribution, then there is a relationship between the variance and the entropy [10],

$$h(E) = (1/2)\ln(2\pi e Var(E)). \quad (6)$$

In this case, it will be equivalent to study the lower bound of error variance. This is important because it is easier to analyze the minimum error variance based on the statistical learning theory, and the error variance is also tightly related to the concept of mean-squared-error (MSE) in statistical learning [24].

Testing the Effectiveness of Attack. This measure has an extra benefit in evaluating a new type of estimation attacks. It can be used to assess how serious the attack can be based on attack simulation. First, we randomly sample the dataset to generate a subset. The simulated attack will generate an estimation on the subset, which is used to calculate the estimation error E. Then, the LOC measure can be calculated. Repeating this procedure multiple times on different random sample sets, we can get a robust estimation on the effectiveness of the new attack.

4.4 Revisiting Level 3 Security

Based on the discussion on the LOC security measure, we apply the weakened definition of security to analyze the effectiveness of estimation attacks under Level 3 adversarial power. In this section, we will first analyze the optimal estimation attack; then we derive the lower bound of error variance (equivalent to $h(E)$) for the estimation attack.

Assume the attacker knows a number of plaintext/perturbed vector pairs. Concretely, let $U_{d\times m}$ be the known m d-dimensional original records $(u_1, \ldots, u_m), m > d+2$ and $u_i \in \mathbb{R}^d$, that include $d+2$ linearly independent vectors. Let $W_{d+2\times m}$ be the corresponding $d+2$-dimensional vectors $(w_1, \ldots, w_m)$, $w_i \in \mathbb{R}^{d+2}$. Assume the noise dimension is drawn from a Gaussian distribution with the mean value μ_v and the variance σ_v^2.

We use the simpler version of RASP perturbation for easier manipulation, where the **PreP** step is not included. Note this exclusion will not increase the difficulty of attack. Thus, the derived lower bound will not be higher than the actual lower bound for the full version. Let the key matrix A decomposed into blocks $A = (A_1, A_2, A_3)$, where A_1, A_2 and A_3 have block sizes $(d+2)\times d$, $(d+2)\times 1$ and $(d+2)\times 1$, respectively.

Let X and P be the plain and perturbed datasets, respectively. Then, with the **Ext** step, the extended data is $\begin{pmatrix} X \\ \mathbf{1} \\ v \end{pmatrix}$ where $\mathbf{1}$ is the row vector with all '1' and v is a row vector with random positive values. According to the simpler version of RASP definition, the relationship between X and P is

$$P = (A_1, A_2, A_3)\begin{pmatrix} X \\ \mathbf{1} \\ \mathbf{v} \end{pmatrix} = A_1 X + A_2\mathbf{1} + A_3\mathbf{v}. \quad (7)$$

where $A_3\mathbf{v}$ is a random noise matrix, whose the elements follows a Gaussian distribution. At the first look, Eq. 7 is a standard affine transformation with a noise component.

We show that

PROPOSITION 3. *the lower bound of error variance for j-th dimension, in terms of the Level 3 estimation attack, is larger than $\gamma_j^2\sigma_v^2$, where γ_j is the j-th element of $(A_1^T A_1)^{-1}A_1^T A_3$, and σ_v^2 is the variance of the random noise dimension.*

PROOF. Let $v = \mu_v + \tilde{v}$, where $\tilde{v}$ has mean value zero and the same variance σ_v^2. Thus, the noise component can be decomposed to $A_3\mu_v + A_3\tilde{v}$. As the constant component $A_2\mathbf{1} + A_3\mu_v$ can be canceled by subtracting any pair of known plain/perturbed vectors from X and P, respectively. Let's denote the subtracted datasets as X' and P'. For easier manipulation, we transform the equation to the canonical regression problem that has the "responses" X' on the left side of the equation, with the constant items have being removed.

$$X' = (A_1^T A_1)^{-1}A_1^T P' - (A_1^T A_1)^{-1}A_1^T A_3\tilde{v}. \quad (8)$$

Let's consider the estimation on j-th dimension of X' only. Let x be the j-th row (i.e., j-th dimension) of X' to be estimated, and β be the j-th row of $(A_1^T A_1)^{-1}A_1^T$ and ϵ be j-th element of $-(A_1^T A_1)^{-1}A_1^T A_3\tilde{v}$. The equation is simplified to

$$x = \beta P' + \epsilon, \quad (9)$$

which is a canonical single-response regression problem. The standard method for the above problem is regression modeling. According to the Gauss-Markov theorem [24], the least square regression (LSR) method is also the minimum variance unbiased estimator, i.e., no other estimator gives lower variance than LSR.

Since the noise component ϵ is not recoverable, the best the attacker can do is to get the estimate of β: $\hat{\beta}$, which can be done with LSR and the known vector pairs in U and W, and then use $\hat{x} = \hat{\beta}P'$ to estimate x. By doing so, we can derive the variance of the estimation error $Var(x - \hat{x})$ is

$$Var(\beta P' + \epsilon - \hat{x}) = Var(\epsilon) + Var((\beta - \hat{\beta})C'). \quad (10)$$

The variance can be decomposed in such a way, because the noise generation is independent of the data distribution. The result shows that the variance of the estimation error is always larger than the variance of ϵ, regardless how small the β's estimation error is.

Let's look closer at the variance of ϵ to understand this relationship. Let γ_j be the j-th element of the vector $(A_1^T A_1)^{-1} A_1^T A_3$ (note this is a column vector). It follows that $var(\epsilon) = \gamma_j^2 \sigma_v^2$ immediately. □

Therefore, the lower bound variance $Var(\epsilon)$ is co-determined by the key matrix A and the variance of the original noise v. In particular, in order to get satisfactory lower bound, we can choose or tune A to make γ_j sufficiently large.

5. SECURITY ANALYSIS ON PERTURBED QUERIES

In this section, we briefly discuss the security of the perturbed queries and show whether the attacker can utilize it to damage the perturbed data. As described in Section 3.2, the perturbed query $p^T Q p < 0$ is used to query the perturbed data. Let α_j represent the j-th row of the matrix A^{-1}. We have $Q = (\alpha_j - f_j(a)\alpha_{d+1})^T \alpha_{d+2}$. Since the parameter A and the functions f_j are fixed for the perturbed dataset, this is a deterministic transformation. There are two questions related to the security analysis. (1) How much is the query privacy preserved? (2) Does the transformation leak the information of perturbation parameters?

Under the Level 1 assumption, the attacker sees only the query matrix Q. Without any other information, the attacker can gain nothing from the query matrices. We skip the detailed discussion here.

Level 2 Security and Distributional Attack. Since the query transformation is deterministic, the same query is always mapped to the same perturbed query. The attacker can keep track of the frequencies of the perturbed queries. With the Level 2 knowledge about the query distribution and counting a sufficiently large number of perturbed queries, the attacker can possibly build a mapping between the original queries and the perturbed queries. Thus, the privacy of some queries could be breached under the Level 2 adversarial power.

Level 3 Security and Eigen-Structure Attack. We also show that an eigen-structure based attack can work with the Level 3 knowledge to determine which dimension the query is about. Thus,

PROPOSITION 4. *with the Level 3 assumption only (excluding Level 2), the perturbed queries are not indistinguishably secure.*

PROOF. Let the known pairs of queries be $QS = \{(q_1, Q_1), \ldots, (q_m, Q_m)\}$, where q_i are original queries and Q_i are perturbed ones. We show that we can determine which dimension the new perturbed query Q is about based on the known pairs of queries. Each perturbed query is

$$Q = (\alpha_j - f_j(a)\alpha_{d+1})^T \alpha_{d+2} = S_j + f_j(a)S, \qquad (11)$$

where $S_j = \alpha_j^T \alpha_{d+2}$ and $S = \alpha_{d+1}^T \alpha_{d+2}$. For different dimensions S_j differs. Although the attackers cannot figure out the exact values in S_j and S, they are able to figure of the matrix structure (e.g., eigenvalue distribution and eigenvectors). Based on this knowledge, they can determine whether a pair of queries Q_i and Q_j are about the same dimension or not. Thus, the indistinguishability definition is not satisfied.

Assume two known queries, Q_1 and Q_2, are on the j-th dimension with different constants a_1 and a_2, we get $Q_1 - Q_2 = (f_j(a_2) - f_j(a_1))S$. Without knowing $(f_j(a_2) - f_j(a_1))$ and S we can still identify the structure of S via methods like eigenvalue decomposition. Note that with different constant r, the eigenvectors of rS will not change. If two known queries, e.g., Q_1 and Q_3, are on different dimensions, e.g., i-th and j-th, then $Q_1 - Q_3 = S_i - S_j + (f_i(a_2) - f_j(a_1))S$, which has high probability having different eigenvectors from $Q_1 - Q_2$. Therefore, we can use the following testing algorithm (Algorithm 1) to distinguish perturbed queries. □

Algorithm 1 Level 3 Query Distinguishing Attack.

```
Input: Q: the new perturbed query; QS: the known plain/perturbed query pairs. There is at least two distinct queries for some dimension, say Q_1 and Q_2;
eigenvectors E ← eigdecompose(Q_1 − Q_2);
for each dimension (i) do
    Q_(i) is the corresponding perturbed query;
    eigenvectors E_i ← eigdecompose(Q − Q_(i));
    if E_i matches E then
        Q is about dimension i;
        break;
    end if
end for
```

Parameter Security. Another important problem is whether the Level 3 knowledge will enable attackers to crack the perturbation parameters, i.e., discover the matrix A or some part of it. We informally show that it is impossible. $Q = (\alpha_j - f_j(a)\alpha_{d+1})^T \alpha_{d+2}$ involves $3(d+2)+1$ unknowns. With a pair of queries in the same dimension, say Q_1 and Q_2, we have $Q_1 - Q_2 = (f_j(a_2) - f_j(a_1))\alpha_{d+1})^T \alpha_{d+2}$, reducing unknowns to $2(d+2)+1$. It is clear that knowing more queries on the same dimension does not help further reduce the number of unknowns. However, knowing Q_1, Q_2, and even $(f_j(a_2) - f_j(a_1))$, is not enough for identifying the vectors α_{d+1} and α_{d+2}. In fact, there are an infinite number of solutions for α_{d+1} and α_{d+2}, because

$$\alpha_{d+1} = \frac{(Q_1 - Q_2)\alpha_{d+2}}{(f_j(a_2) - f_j(a_1))||\alpha_{d+2}||}, \qquad (12)$$

where $||.||$ means the vector length, as long as α_{d+1} and α_{d+2} are linearly independent.

On the other hand, knowing queries of different dimensions does not help reduce the unknowns. Thus, the quadratic query transformation helps protect the perturbation parameters.

6. RELATED WORK

The current research on secure computation in the cloud is still embryonic, requiring a balanced study on both utility and security. Fully homomorphic encryption [18, 19] in theory allows any operation on encrypted data that can be traced back to equivalent operations on the corresponding plaintexts. The current solutions focus on the basic operations: addition and multiplication, building an application on which is too expensive to be practical even for a simple application like encrypted keyword search. To achieve better performance, some researchers have applied partially homomorphic encryption schemes such as Paillier encryption [32] to data analysis [17] and matrix computation [37], which requires revealing partial data in the computation. However, the impact of the revealed data on the security was not fully analyzed.

Several methods emphasis more on data utility and performance than on security, such as Crypto-index [23, 25] and order-preserving encryption (OPE) [3, 5]. The order preserving encryption (OPE)

[3] preserves the dimensional value order after encryption. Thus, it can be used in most database operations, such as indexing and range query. Boldyreva et al. [5, 6] has formally analyzed the security of OPE. As widely understood, all OPE schemes are vulnerable to distributional attacks, if the attackers are aware of the distribution of the original data. Crypto-Index is based on column-wise bucketization. It assigns a random ID to each bucket; the values in the bucket are replaced with the bucket ID to generate the auxiliary data for indexing. However, the bucketization scheme leaks a lot of information. Thus, a bucket-diffusion scheme [25] was proposed to introduce noise records into the results to improve the security, which, however, has to sacrifice the precision of query results. Secure keyword search on encrypted documents [36, 21, 20, 7, 11] is another cluster of utility preserving encryption methods. They allow the server to scan each encrypted document in the database and find the documents containing the keyword. There have been rigid security analysis on this line of research [20, 11].

In the statistical database [12] setting, the trusted server (or the data owner) hosts sensitive databases and serves queries for possibly malicious users, who may want to figure out private information by submitting carefully designed queries. There are two settings: interactive and non-interactive. Traditional statistical database research focuses on the interactive setting and inference attacks [12]. Recent studies emphasis on the non-interactive setting, i.e., releasing micro data based on privacy definitions such as k-anonymity, l-diversity, t-closeness, and other improved definitions [16]. Among these definitions, differential privacy [13] has been a significant one for both interactive and non-interactive settings. Applications of statistical database are very similar to those we discussed for the RASP perturbation. However, their settings are totally different. In the cloud setting, the server in the cloud is untrusted and the users are authorized trusted users. The concerns are on the data security and query privacy.

7. CONCLUSION AND FUTURE WORK

The RASP perturbation technique was proposed to conduct half-space queries securely and efficiently on the data hosted in the cloud. The efficient range query processing algorithm has been proposed and evaluated in Chen et al. [9], but its security is not fully understood yet. In this paper we carefully analyze the security of RASP perturbed data and queries under the three-level adversarial assumptions. The initial analysis shows that the RASP perturbation does not satisfy the strong indistinguishability definition on Level 2 and 3 assumptions. We notice that the strong indistinguishability definition might not be necessary for the cloud computing setting and the perturbation techniques in general, where estimation-based attacks are the typical threats. Thus, we introduce a weakened definition on security. This definition is based on statistical learning theory and information theory, taking the Level 2 and 3 of adversarial knowledge into account. We then analyze a typical estimation attack based on the Level 3 assumption, the regression attack, under the new security definition. We will continue our study on the security of RASP perturbed data and queries, and explore more applications of the RASP perturbation for secure data intensive computing in the cloud.

8. REFERENCES

[1] AGRAWAL, D., AND AGGARWAL, C. C. On the design and quantification of privacy preserving data mining algorithms. In *Proceedings of ACM Conference on Principles of Database Systems (PODS)* (Madison, Wisconsin, 2002), ACM.

[2] AGRAWAL, R., AILAMAKI, A., BERNSTEIN, P. A., BREWER, E. A., CAREY, M. J., CHAUDHURI, S., DOAN, A., FLORESCU, D., FRANKLIN, M. J., GARCIA-MOLINA, H., GEHRKE, J., GRUENWALD, L., HAAS, L. M., HALEVY, A. Y., HELLERSTEIN, J. M., IOANNIDIS, Y. E., KORTH, H. F., KOSSMANN, D., MADDEN, S., MAGOULAS, R., OOI, B. C., O'REILLY, T., RAMAKRISHNAN, R., SARAWAGI, S., STONEBRAKER, M., SZALAY, A. S., AND WEIKUM, G. The claremont report on database research. *SIGMOD Record 37*, 3 (2008), 9–19.

[3] AGRAWAL, R., KIERNAN, J., SRIKANT, R., AND XU, Y. Order preserving encryption for numeric data. In *Proceedings of ACM SIGMOD Conference* (2004).

[4] ARMBRUST, M., FOX, A., GRIFFITH, R., JOSEPH, A. D., KATZ, R., KONWINSKI, A., LEE, G., PATTERSON, D., RABKIN, A., STOICA, I., AND ZAHARIA, M. Above the clouds: A berkeley view of cloud computing. *Technical Report, University of Berkerley* (2009).

[5] BOLDYREVA, A., CHENETTE, N., LEE, Y., AND O'NEILL, A. Order preserving symmetric encryption. In *Proceedings of EUROCRYPT conference* (2009).

[6] BOLDYREVA, A., CHENETTE, N., AND Oï£¡NEILL, A. Order-preserving encryption revisited: Improved security analysis and alternative solutions. In *CRYPTO* (2011).

[7] BONEH, D., CRESCENZO, G. D., OSTROVSKY, R., AND PERSIANO, G. Public-key encryption with keyword search. In *Proceedings of Advances in Cryptology, (EUROCRYPT)* (2004), Springer.

[8] BOYD, S., AND VANDENBERGHE, L. *Convex Optimization*. Cambridge University Press, 2004.

[9] CHEN, K., KAVULURU, R., AND GUO, S. Rasp: Efficient multidimensional range query on attack-resilient encrypted databases. In *ACM Conference on Data and Application Security and Privacy* (2011).

[10] COVER, T., AND THOMAS, J. *Elements of Information Theory*. Wiley, 1991.

[11] CURTMOLA, R., GARAY, J., KAMARA, S., AND OSTROVSKY, R. Searchable symmetric encryption: improved definitions and efficient constructions. In *Proceedings of the 13th ACM conference on Computer and communications security* (New York, NY, USA, 2006), ACM, pp. 79–88.

[12] DOMINGO-FERRER, J. *Inferance Control in Statistical Databases*. Springer, 2002.

[13] DWORK, C., AND LEI, J. Differential privacy and robust statistics. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing* (New York, NY, USA, 2009), ACM, pp. 371–380.

[14] ELMASRI, R., AND NAVATHE, S. *Fundamentals of Database Systems*. Addison Wesley, 2010.

[15] FREUND, Y., AND SCHAPIRE, R. E. A short introduction to boosting. In *In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence* (1999), Morgan Kaufmann, pp. 1401–1406.

[16] FUNG, B. C. M., WANG, K., CHEN, R., AND YU, P. S. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Survey 42* (June 2010), 14:1–14:53.

[17] GE, T., AND ZDONIK, S. Answering aggregation queries in a secure system model. In *Proceedings of the 33rd international conference on Very large data bases* (2007), VLDB '07, VLDB Endowment, pp. 519–530.

[18] GENTRY, C. Fully homomorphic encryption using ideal lattices. In *STOC '09: Proceedings of the 41st annual ACM symposium on Theory of computing* (New York, NY, USA, 2009), ACM, pp. 169–178.

[19] GENTRY, C., AND HALEVI, S. Implementing gentry's fully-homomorphic encryption scheme. In *EUROCRYPT* (2011), pp. 129–148.

[20] GOH, E.-J. Secure indexes. Cryptology ePrint Archive, Report 2003/216, 2003.

[21] GOLLE, P., STADDON, J., AND WATERS, B. Secure conjunctive keyword search over encrypted data. In *ACNS 04: 2nd International Conference on Applied Cryptography and Network Security* (2004), Springer-Verlag, pp. 31–45.

[22] GREENE, T. Survey: Most businesses haven't mastered cloud security. *NetworkWorld, http://www.infoworld.com/d/cloud-computing/survey-most-businesses-havent-mastered-cloud-security-280* (2009).

[23] HACIGUMUS, H., IYER, B., LI, C., AND MEHROTRA, S. Executing sql over encrypted data in the database-service-provider model. In *Proceedings of ACM SIGMOD Conference* (2002).

[24] HASTIE, T., TIBSHIRANI, R., AND FRIEDMAN, J. *The Elements of Statistical Learning*. Springer-Verlag, 2001.

[25] HORE, B., MEHROTRA, S., AND TSUDIK, G. A privacy-preserving index for range queries. In *Proceedings of Very Large Databases Conference (VLDB)* (2004).

[26] HYVARINEN, A., KARHUNEN, J., AND OJA, E. *Independent Component Analysis*. Wiley, 2001.

[27] KATZ, J., AND LINDELL, Y. *Introduction to Modern Cryptography*. Chapman and Hall/CRC, 2007.

[28] KEARNS, M. Efficient noise-tolerant learning from statistical queries. *Journal of ACM 45*, 6 (1998), 983–1006.

[29] KEARNS, M. J., AND VAZIRANI, U. V. *An Introduction to Computational Learning Theory*. MIT press, 1994.

[30] LI, F., HADJIELEFTHERIOU, M., KOLLIOS, G., AND REYZIN, L. Dynamic authenticated index structures for outsourced databases. In *Proceedings of ACM SIGMOD Conference* (2006).

[31] MYAGMAR, S., LEE, A. J., AND YURCIK, W. Threat modeling as a basis for security requirements. In *In Symposium on Requirements Engineering for Information Security (SREIS* (2005).

[32] PAILLIER, P. Public-key cryptosystems based on composite degree residuosity classes. In *EUROCRYPT* (1999), Springer-Verlag, pp. 223–238.

[33] RIEHM, C. R. *Introduction to Orthogonal, Symplectic and Unitary Representations of Finite Groups*. American Mathematical Society, 2011.

[34] SCHNEIER, B. Homomorphic encryption breakthrough. http://www.schneier.com/blog/archives/2009 /07/homomorphic_enc.html, 2009.

[35] SION, R. Query execution assurance for outsourced databases. In *Proceedings of Very Large Databases Conference (VLDB)* (2005).

[36] SONG, D. X., WAGNER, D., AND PERRIG, A. Practical techniques for searches on encrypted data. In *IEEE Symposium on Security and Privacy* (Washington, DC, USA, 2000), IEEE Computer Society, p. 44.

[37] WANG, C., REN, K., WANG, J., AND URS, K. M. R. Harnessing the cloud for securely solving large-scale systems of linear equations. In *Proceedings of ICDCS* (Washington, DC, USA, 2011), IEEE Computer Society, pp. 549–558.

[38] XIE, M., WANG, H., YIN, J., AND MENG, X. Integrity auditing of outsourced data. In *VLDB* (2007), pp. 782–793.