Security of Random Data Perturbation Methods

KRISHNAMURTY MURALIDHAR
University of Kentucky
and
RATHINDRA SARATHY
Illinois State University

Statistical databases often use random data perturbation (RDP) methods to protect against disclosure of confidential numerical attributes. One of the key requirements of RDP methods is that they provide the appropriate level of security against snoopers who attempt to obtain information on confidential attributes through statistical inference. In this study, we evaluate the security provided by three methods of perturbation. The results of this study allow the database administrator to select the most effective RDP method that assures adequate protection against disclosure of confidential information.

Categories and Subject Descriptors: H.2.7 [**Database Management**]: Database Administration—*Security, integrity, and protection*; H.2.8 [**Database Management**]: Database Applications—*Statistical databases*

General Terms: Security, Measurement

Additional Key Words and Phrases: Bias, covariance, noise addition, random data perturbation

1. INTRODUCTION

Organizations increasingly face the problem of protecting confidential information contained in their databases. Government agencies such as the Census Bureau, which are responsible for gathering and disseminating information, adopt many techniques including the masking of microdata, to limit the disclosure of confidential information [Fuller 1993]. Such agencies generally release static masked datasets so that undesirable inferences

Authors' addresses: K. Muralidhar, Carol Martin Gatton College of Business & Economics, University of Kentucky, Lexington, KY 40506-0034; email: kmura0@pop.uky.edu; R. Sarathy, Department of Accounting, Illinois State University, Normal, IL 61790–5520; email: rsarathy@ilstu.edu.

may be avoided. Limiting disclosure and maintaining confidentiality is more complicated in the context of organizations, where databases are dynamic and where online interactive queries are frequently permitted [Lunt 1992].

Database administrators have to balance requirements for confidentiality against the needs of legitimate users for easy access and analysis of organizational data. Using statistical databases, database administrators can allow users access to aggregate statistical information, but not information regarding a specific entity. However, even in statistical databases, a snooper[1] may be able to infer confidential information associated with a particular individual through a clever choice of queries, leading to disclosure [Adam and Wortmann 1989]. Hence, statistical databases often employ *random data perturbation*. Random data perturbation (RDP) involves the addition of random noise to confidential numerical attributes. Thus, even if a snooper is able to isolate an individual value of a confidential attribute, the true value is not disclosed.

While perturbation methods guarantee that complete or exact disclosure (i.e., the disclosure of the true value of a confidential attribute) will never occur, they are susceptible to partial disclosure. Partial disclosure occurs when the amount of information that a snooper is able to obtain about a confidential attribute through queries and statistical analysis, exceeds the amount that the database administrator intended to provide to users [Adam and Wortmann 1989]. In other words, the value of the perturbed attribute is "closer" to the true value of the attribute than intended. When a database administrator specifies a certain level of security, he/she is simultaneously specifying the maximum amount of disclosure about confidential attributes that should be available to any user, devious or otherwise. This, in turn, also means that the database administrator should be able to correctly *evaluate* the amount of disclosure (or conversely, the level of security) for a given situation. In this study, we evaluate the security provided by three methods of perturbation. These methods are described in the next section.

2. RANDOM DATA PERTURBATION METHODS

Following the setting used in Tendick and Matloff [1994], let the variables $\mathbf{A} = \{A_1, .., A_p\}$ represent the set of p attributes of the database. Further, as in Tendick and Matloff [1994], some of these attributes are confidential and will be perturbed using RDP (discussion will be limited to this set of confidential attributes). In some cases it may be necessary to perturb even nonconfidential attributes to avoid biased responses to queries. It is also assumed that $\mathbf{A}$ is a realization from a multivariate normal distribution with mean vector μ and covariance matrix Σ.

[1]A snooper is an authorized user of the database who attempts to obtain, through devious methods, unauthorized confidential information available in the databases.

In its simplest form, for a single confidential attribute A (with mean μ and variance σ^2) RDP involves the addition of random noise (ε) to result in the perturbed attribute $A^\#$ as follows:

$$A^\# = A + \varepsilon \tag{2.1}$$

where ε has a mean of 0 and variance of $d\sigma^2$. The perturbation level, d, represents the extent of protection against partial disclosure that the database administrator intends to provide. If both A and ε have a normal distribution, then the perturbed attribute $A^\#$ also has a normal distribution with mean μ and variance $(1 + d)\sigma^2$.

When there are multiple confidential attributes, one approach to perturbation is to apply RDP independently to each attribute [Traub et al. 1984]. The noise terms are not only independent of the original attributes, but are independent of each other as well. This method of perturbation is referred to as the "independent noise" (IN) method. The perturbed attributes can be represented as

$$\mathbf{A}^\# = \mathbf{A} + \varepsilon \tag{2.2}$$

where ε has a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $d\mathbf{D}$ where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the same as the diagonal elements in Σ (and whose off-diagonal elements are zeros). Hence, the resulting perturbed attribute $\mathbf{A}^\#$ has a multivariate normal distribution with mean vector μ and covariance matrix $(\Sigma + d\mathbf{D})$. The IN method results in bias to some query responses, and also changes the relationship between the confidential attributes in the database.

Kim [1986] and Tendick [1991] proposed a different correlation structure for ε, referred to as the "correlated noise" (CN) method. In this method, the noise term is specified so that ε has a mean vector of $\mathbf{0}$ and covariance matrix $d\Sigma$. The resulting perturbed attributes $\tilde{\mathbf{A}}$ has a mean vector of $\mathbf{0}$ and covariance matrix $(1 + d)\Sigma$. Since all the elements in the covariance matrix are altered by the same quantity $(1 + d)$, this perturbation method maintains the correlation structure of the perturbed attributes to be the same as that of the original attributes. This structure allows the CN method to eliminate one type of bias inherent in the IN method, namely, bias in measuring relationships between confidential attributes. However, since the variance of an individual perturbed attribute is $(1 + d)$ times that of the corresponding original attribute, the CN method results in bias in response to some queries. To eliminate this bias, a modification of the CN method was suggested by Kim [1986] and Tendick and Matloff [1994]. This method is referred to as the Bias Corrected Correlated Noise (BCCN) method.

The BCCN method performs a linear transformation of $\tilde{\mathbf{A}}$ as follows:

$$\mathbf{A}^* = (1/d_1)\tilde{\mathbf{A}} + (d_2/d_1)\mu \tag{2.3}$$

where $d_1 = (1 + d)^{0.5}$ and $d_2 = (d_1 - 1)$. The resulting $\mathbf{A}^*$ has a mean vector μ and covariance matrix Σ. Since the covariance matrix of both $\mathbf{A}$ and $\mathbf{A}^*$ are the same (as well as their means), under assumptions of multivariate normality, the distribution of $\mathbf{A}^*$ is the same as that of $\mathbf{A}$.

3. EVALUATING SECURITY OF RDP METHODS

In order to determine the efficacy of a perturbation method, it is necessary to consider the security provided by that method. Tendick and Matloff [1994] contend that the security provided by the BCCN method is the same as that provided by the CN method. This measurement of security is based on the assumption that a snooper would attempt to estimate a confidential value by using the best possible predictor of that attribute. Consider a single confidential attribute A. When the CN method is used to perturb the database, the best predictor of A, based on $\tilde{A}$, can be derived as[2]

$$f(\tilde{A}) = [d/(1 + d)]\mu + [\tilde{A}/(1 + d)] \tag{3.1}$$

The variance of the prediction error can be derived as

$$E[\{A - f(\tilde{A})\}^2] = [d/(1 + d)]\sigma^2 \tag{3.2}$$

where σ^2 represents the variance of A. The security expression above identifies the proportion of variability in the original attribute A that **cannot** be explained using the perturbed attribute $\tilde{A}$. Similarly, for the BCCN method, the best predictor of A using A^* can be derived as

$$g(A^*) = (d_2/d_1)\mu + (A^*/d_1) \tag{3.3}$$

The variance of the prediction error for the BCCN method can be derived as

$$E[\{A - g(A^*)\}^2] = [d/(1 + d)]\sigma^2 \tag{3.4}$$

In other words, the BCCN method provides the same level of protection against a snooper using A^* to predict A. The above security measures can be made scale-invariant by dividing the final expression in (3.2) and (3.4) by σ^2 to yield $d/(1 + d)$.

The derivations above deal with a single confidential attribute A. For the multivariate case, Tendick [1991], Eq. 3.8, p. 345, develops a measure to determine the proportion of variability that can be explained in a linear combination of the confidential attributes using a linear combination of the perturbed attributes. This measure can then be used to evaluate security, by considering the unexplained proportion of variability.

[2]For a single attribute, the prediction equation for both the CN and IN methods are the same.

When Tendick's measure is used, the security of both the CN and BCCN methods turns out to be equal to $d/(1 + d)$. Therefore, these two methods provide the same level of security in both the univariate and multivariate case. This, however, is not true for the IN method. Tendick [1991] has shown that the security provided by the IN method can vary depending on the characteristics of the specific database being perturbed. When the confidential attributes are completely uncorrelated, the IN method provides the same level of security as the other CN and BCCN methods. On the other hand, when there are a large number of independent variables that are very highly correlated, the security by the IN method is extremely low (see Tendick [1991], Table 1, p. 347). In most practical situations, the security provided by the IN method will vary between 0 and $d/(1 + d)$, and is likely to be lower than that of the CN and BCCN methods.

In evaluating security, one of the key assumptions thus far has been that the snooper has a strong motivation to get the best predictor of the confidential attribute(s) using perturbed attribute(s). We refer to such a user as a "professional" snooper. However, in an organizational context, there are likely to be other individuals who are either not as strongly motivated or lack the necessary statistical/computer skills to obtain the best predictor. Such "casual" snoopers may simply use the perturbed value of a confidential attribute as an estimate of the true value. It is necessary to perform a security evaluation for these snoopers as well. Also, it is reasonable to expect that the security provided by a perturbation method against such snoopers is much greater than that provided for a professional snooper.

The approach to evaluating security for casual snoopers is similar to that for professional snoopers. The security is evaluated as $E[(Actual - Predicted)^2]$, where the predicted value is simply the perturbed value of the attribute. The security provided by the IN method for casual snoopers for a given attribute A is

$$E[(A - A^{\#})^2] = E[(A - \{A + \varepsilon\})^2] = d\sigma^2 \tag{3.5}$$

where σ^2 represents the variance of attribute A. Dividing the expression above by σ^2 results in d, a scale-invariant measure of security, and indicates that a casual snooper is unable to explain d proportion of the variability in the confidential attribute by using the perturbed attribute as an estimate. In the multivariate case, since the diagonal elements in $\mathbf{D}$ are multiplied by a constant d, the level of security provided for each confidential attribute would also be d. The security provided by the CN method for any attribute for casual snoopers is

$$E[(A - \tilde{A})^2] \,/\sigma^2 = E[(A - \{A + \varepsilon\})^2] \,/\sigma^2 = d \tag{3.6}$$

which is the same as that of the IN method.

Table I. Covariance Matrix of Attributes

	A1	A2	A3	A4
A1	1.00	0.60	0.40	0.20
A2	0.60	1.00	0.30	0.10
A3	0.40	0.30	1.00	0.70
A4	0.20	0.10	0.70	1.00

Table II. Security Provided

	Type of Snooper	
Method	Professional	Casual
IN	0.32	1.00
CN	0.50	1.00
BCCN	0.50	0.58

Finally, the security provided by the BCCN method for casual snoopers can be derived as

$$E[(A - A^*)^2] / \sigma^2 = E[\{A - (1/d_1)\tilde{A} - (d_2/d_1)\mu\}^2] / \sigma^2$$

$$= 2[1 - (1 + d)^{-0.5}] \tag{3.7}$$

It can easily be shown that, for any value of $d > 0$, (3.7) is less than d. By contrast, the security provided by the IN and CN methods is d. In other words, the BCCN method provides less security than do the IN and CN methods against disclosure of information to casual snoopers.

The linear transformation used by the BCCN method (see Eq. 2.3) allows the casual snooper to gain estimates of A based on A^* that are closer than those based on $\tilde{A}$. As noted earlier, the professional snooper uses sophisticated techniques to achieve estimates of the confidential attribute, while the casual snooper uses only rudimentary techniques. However, if the BCCN method is implemented, it is possible that the estimates of the casual snooper are almost as accurate as that of a professional one. The following example illustrates this situation.

Consider four confidential attributes from a hypothetical database. Assume (for the sake of simplicity and without loss of generality) that the attributes have a multivariate normal distribution with a mean vector of **0**. The covariance matrix of this database is provided in Table I.

Assume that all three methods of perturbation have been applied to this data by specifying $d = 1$. The results of the security evaluation are provided in Table II.

Comparing the performance of the three methods with respect to the professional snooper, the IN method provides far lower security than do the CN and BCCN methods. Tendick [1991] also noted this result. As expected, the CN and BCCN methods maintain the level of security at $d/(1 + d)$ when perturbed attributes are used to predict the confidential attributes.

Comparing the results for casual snoopers, it is clear that, for a given value of d, the security provided by the BCCN method for this type of snooper is much lower than that provided by the other two methods. Further, the casual snooper is able to explain 0.42 $(1 - 0.58)$ of the variability in the confidential attribute, which is close to that of the professional snooper (0.50). Given the fact that a professional snooper has

far more information and knowledge than the casual snooper, it is reasonable to expect a much higher level of security for the casual snooper. The IN and CN methods provide such security. The BCCN method does not.

4. CONCLUSIONS

The objective of this study was to evaluate the level of security provided by three known RDP methods. The results indicate that, for casual snoopers, the BCCN method provides lower security than the IN and CN methods. This lower security is a direct consequence of the linear transformation employed by the BCCN method. This result has not been observed in earlier studies. In situations where it is desirable to have a high level of security against disclosure to casual snoopers, the BCCN method should be implemented with caution.

These results do not, however, imply that the BCCN method is not a good choice for perturbation. It eliminates bias arising from the difference in the distribution of the perturbed attributes from that of the original attributes. Only the BCCN method provides this advantage; the IN and CN methods do not. Further, in cases where all the attributes in the database are perturbed, the security provided by the BCCN method against casual snoopers can be increased by increasing the value of d. Hence, when a database administrator considers both bias and security, of the three methods in this study, the BCCN method may still be the preferred method of perturbation.

REFERENCES

ADAM, N. R. AND WORTMANN, J. C. 1989. Security-control methods for statistical databases: A comparative study. *ACM Comput. Surv. 21*, 4 (Dec. 1989), 515–556.

FULLER, W. A. 1993. Masking procedures for microdata disclosure limitation. *J. Official Stat. 9*, 2, 383–406.

KIM, J. 1986. A method for limiting disclosure in microdata based on random noise and transformation. In *Proceedings of the American Statistical Association on Survey Research Methods*, American Statistical Association, Washington, DC, 370–374.

LUNT, T. F. 1992. Security in database systems: A research perspective. *Comput. Secur. 11*, 1 (Mar. 1992), 41–56.

TENDICK, P. AND MATLOFF, N. 1994. A modified random perturbation method for database security. *ACM Trans. Database Syst. 19*, 1 (Mar. 1994), 47–63.

TENDICK, P. 1991. Optimal noise addition for preserving confidentiality in multivariate data. *J. Stat. Plan. Inference 27*, 2, 341–353.

TRAUB, J., YEMINI, Y., AND WOZNIAKOWSKI, H. 1984. Statistical security of a statistical database. *ACM Trans. Database Syst. 9*, 4 (Dec. 1984), 672–679.

Received: July 1995; revised: November 1998; accepted: August 1999