Watermarking for Multilevel Access to Statistical Databases*

Josep Domingo-Ferrer, Josep M. Mateo-Sanz and Francesc Sebé
Dept. of Computer Engineering and Mathematics
Universitat Rovira i Virgili
ETSE-Autovia de Salou, s/n
E-43006 Tarragona, Catalonia, Spain
e-mail {jdomingo,jmateo,fsebe}@etse.urv.es

Abstract

Increased corporate, government and academic demand has prompted official statistics to release individual respondent data (microdata) in addition to the traditional tabular data. Microdata must be masked by a statistical disclosure control (SDC) method before being published, because otherwise the statistical confidentiality of respondents would be compromised. A novel application of watermarking is discussed in this paper which allows multilevel access to numerical microdata: depending on her clearance, the data user can remove more or less of the masking. Non-privileged users just see the published data, but as the clearance of a user increases she can get a data set which is closer and closer to the original one.

Keywords: *Statistical database protection, Watermarking applications, Microdata.*

1 Introduction

In [2], the following applications of watermarking were identified: broadcast monitoring, owner identification, proof of ownership, authentication, fingerprinting, copy control and covert communication. Multilevel access to statistical databases is presented as a novel application of watermarking in this paper.

Increased corporate, government and academic demand has prompted official statistics to release individual respondent data, in addition to the traditional tabular data. Released data must be masked (*i.e.* distorted) by a statistical disclosure control (SDC) method [18, 1, 4] before being published, because otherwise the statistical confidentiality of respondents would be compromised. The main problem in disclosure control is to provide sufficient protection without seriously damaging the information contained in the original data. Masked data being published should not disclose information on specific respondents, but statistical computations performed on them should yield results similar to those that would be obtained on original data. To understand the tradeoff, consider the two extreme cases between which SDC methods lie:

- If masking consists of encrypting original data, then no disclosure is possible, but no information at all is released.
- If no masking is performed and the original data are released, users can perform fully accurate computations, but disclosure of individual respondent data is very likely, especially when releasing individual respondent data (called microdata from now on).

We concentrate in what follows on describing an approach allowing multilevel access to masked *numerical* microdata —categorical microdata taking values in a set of categories, such as name, eye color, etc. will not be considered here—. *Depending on her clearance, the data user can remove more or less of the masking*. Non-privileged users just see the published data, but as the clearance of a user increases (perhaps as a result of signing non-disclosure agreements) she can get a data set which is closer and closer to the original one. Section 2 summarizes relevant SDC masking methods for numerical microdata. Section 3 describes partially removable masking. Section 4 describes a watermarking solution for multilevel access. Requirements on the watermarking algorithm are identified in Section 5, and Section 6 justifies the choice of LSB oblivious watermarking. Section 7 is a conclusion.

*Work partly funded by the U. S. Bureau of the Census contract no. OBLIG-2000-29158-0-0, by the European Union contract no. IST-2000-25069 "CASC" and by the Spanish CICYT under grant no. TEL98-0699-C02-02

2 Relevant masking methods for numerical microdata

We assume that the information of a microdata file is represented as a two-dimensional table where one dimension corresponds to the set of objects (*i.e.* elements, individuals, persons) and the other is the set of attributes (*i.e.* variables). The microdata file contains a value for each object-attribute pair, so that it can be modelled as a function

$$\mathbf{V} : \mathbf{O} \rightarrow D(V_1) \times D(V_2) \times \cdots \times D(V_m)$$

where $\mathbf{O}$ denotes the set of objects, $V_1, V_2, \cdots V_m$ denote the attributes and $D(V_i)$ refers to the domain of attribute V_i. Without loss of generality, the m-dimensional function V can be assumed to be of the form:

$$\mathbf{V}(\cdot) = (V_1(\cdot), V_2(\cdot), \cdots, V_m(\cdot))$$

where $V_i(\cdot) : \mathbf{O} \rightarrow D(V_i)$ is a one-dimensional function assigning a value for attribute V_i to a given object.

Microdata masking methods are based on one of two principles: sampling (*i.e.* the masked data set is simply a fraction of the original one) and perturbation (the masked data set is a perturbed version of the original one). Formally speaking, sampling consists of publishing the values of a function $\mathbf{V}'$ which is $\mathbf{V}$ restricted to a subset $S \subset O$.

We argue that, in a general disclosure scenario, sampling methods are not suitable for protecting numerical microdata. The reason is that such methods leave a numerical attribute V_i unperturbed for all objects in S. Thus, if attribute V_i is present in an external administrative public file, unique matches with $\mathbf{V}'$ are very likely, since for a numerical attribute (even a digitally represented one) it is highly unlikely that $V_i(o_1) = V_i(o_2)$ if $o_1 \neq o_2$.

Therefore, only perturbative methods will be considered. Some of the main alternatives for protecting numerical data are reviewed below:

- *Additive noise*. If V_i is an attribute to be protected, then it is replaced by

  $$V_i' = V_i + \varepsilon_i$$

  where the ε_i are independent random noise variables with the same covariance as the V_i[12].

- *Data distortion by probability distribution*. This method involves three steps[13]. First, identify the underlying probability distribution function of the original data set $\mathbf{V}$. Second, generate a data set $\mathbf{V}'$ by drawing from the estimated distribution function. Third, map and replace the original data set with the generated one.

- *Rank swapping*. Although originally described only for ordinal attributes [15], this method can be used for any numerical attribute. First values of attribute V_i are ranked in ascending order; then each ranked value of V_i is swapped with another ranked value randomly chosen within a restricted range (*e.g.* the rank of two swapped values cannot differ by more than p% of the total number of objects).

- *Microaggregation*. Objects are clustered into small aggregates or groups (which should be as homogeneous as possible). Rather than publishing an original attribute V_i for a given object, the average of the values of V_i over the group to which the object belongs is published [3, 8].

- *Resampling*. Originally proposed for protecting tabular data [11, 5], resampling can also be used for microdata. Take t independent samples $X_1, \cdots, X_t$ of the values of an original attribute V_i. Sort all samples using the same ranking criterion. Build the masked attribute V_i' as $\bar{v}_1, \cdots, \bar{v}_n$, where n is the number of objects and $\bar{v}_j$ is the average of the j-th ranked values in $X_1, \cdots, X_t$.

- *Lossy compression*. This method is new and proposed by these authors [7]. The idea is to regard a numerical microdata file as an image (with rows being objects and columns being attributes). Lossy compression (*e.g.* JPEG) is then used on the image, and the compressed image is interpreted as a masked microdata file. Depending on the lossy compression algorithm used, appropriate mappings between attribute ranges and color scales will be needed.

For more comprehensive surveys on SDC methods for microdata, including information loss and security comparisons, see [7, 1].

3 Partially removable masking

Assumption 1 We assume that a perturbative masking method can be expressed as a masking algorithm F which takes as inputs the original microdata file $\mathbf{V}$ and the outputs of r pseudorandom number generators $PRNG_i$ seeded by s_i, for $i = 1, \cdots, r$. The output of F is the masked microdata file $\mathbf{V}'$. Formally speaking,

$$\mathbf{V}' = F(\mathbf{V}, \{s_1, \cdots, s_r\}) \quad (1)$$

F and $PRNG_i$, for $i = 1, \cdots, r$ are assumed to be public, so the only secret parameters of masking are s_i, for $i = 1, \cdots, r$.

Assumption 2 We assume that each $PRNG_i$ is used to independently mask a part of the microdata file, so that *knowledge of the random numbers generated by $PRNG_i$ should allow to retrieve the original values from the corresponding masked values in that part of the microdata file*. Formally speaking, given a subset $S \subset \{s_1, \cdots, s_r\}$, we can compute

$$\mathbf{V}'(S) = F^{-1}(\mathbf{V}', S) \quad (2)$$

where $\mathbf{V}'(S)$ is a microdata file resulting from removing the masking of $\mathbf{V}'$ that was produced using generators seeded by elements in S. In particular $\mathbf{V}'(\{s_1, \cdots, s_r\}) = \mathbf{V}$.

For some masking methods mentioned in Section 2, it is easy to meet Assumptions 1 and 2, because they make explicit use of random number generation and knowledge of the generated random numbers suffices to undo the masking. Such is the case for rank swapping and additive noise; regarding Assumption 2, the following applies to those two methods:

- If the masking algorithm masks attributes independently, like rank swapping, each attribute could be masked using a different $PRNG_i$.
- If the masking algorithm masks individual values independently, like additive noise, a different $PRNG_i$ can be used for each value (yet the scheduling of the various $PRNG_i$ should be public).

For methods which do not directly meet both assumptions above, consider the sequence of differences between the masked and the original data

$$V_i'(o_j) - V_i(o_j) \text{ for } i = 1, \cdots, m \text{ and } j = 1, \cdots, n \quad (3)$$

where m is the number of attributes and n is the number of objects. Now, the Berlekamp-Massey algorithm [14] can be used to synthesize a Linear Feedback Shift Register (LFSR) generating the sequence (3). More generally, r LFSRs can be synthesized such that their interleaved outputs yield the sequence (3) (the i-th LFSR generates integers in positions j of the sequence such that $j \bmod r = i$). This construction reduces any perturbative method to a variant of additive noise ($\mathbf{V}'$ can be computed by adding the Sequence (3) to $\mathbf{V}$), which thus meets Assumptions 1 and 2.

Now, if a user is revealed a subset S of the seeds, by Assumption 2 she can remove the masking in those parts of $\mathbf{V}'$ masked using generators seeded by values in S, so that the user obtains a partially unmasked file $\mathbf{V}'(S)$. In particular, if the user is revealed all seeds, she can retrieve the original file $\mathbf{V}'(\{s_1, \cdots, s_r\}) = \mathbf{V}$.

4 Watermarking solutions for multilevel access

From the previous section, we can see that, the larger the subset S of seeds known by a user, the more masking the user can remove, *i.e.* the closer is the unmasked file $\mathbf{V}'(S)$ to the original $\mathbf{V}$. This suggests the following algorithm to implement multilevel access to the masked file $\mathbf{V}'$:

Algorithm 1 *1. Let H be a clearance hierarchy comprising u user categories (for example, "statistician", "researcher", "civil servant", "other users"). For each category j, let k_j be a secret key known only to users in that category (the user does not actually need to know k_j, which can reside in her smart card).*

2. *For $i = 1, \cdots, r$ and $j = 1, \cdots, u$, encrypt s_i with some redundancy r_i under k_j to get $E_{k_j}(s_i||r_i)$ if s_i should be revealed to user category j.*
3. *Use a watermarking algorithm to embed $E_{k_j}(s_i||r_i)$, for $i = 1, \cdots, r$ and $j = 1, \cdots, u$, into the masked file $\mathbf{V}'$ to get a watermarked file $\hat{\mathbf{V}}'$.*

From $\hat{\mathbf{V}}'$, a user can retrieve the subset S of seeds her category is entitled to know, and thus retrieve $\mathbf{V}'(S)$. Redundancy r_i encrypted with s_i allows the user to check that s_i was correctly decrypted. We next discuss which features the watermarking algorithm should offer.

5 Watermarking requirements

Unlike in most usual watermarking applications (see [2]), watermarking in the application described here is *positive for the user*. In the worst case, the user with no clearance just gets $\mathbf{V}'$, but the user with some clearance gets a better file. Therefore, there is no reason to expect a malicious behaviour by the user to destroy the watermark. In addition, the correct protected file $\mathbf{V}'$ is normally publicly and easily available (*e.g.* on a Web site), so watermarking *tamper-proofness* is *not* really a requirement in this application.

Robustness should provide for those normal accidental alterations that may occur during the life cycle of a numerical file. These are basically rounding errors, mostly due to the software used to manipulate the data (*e.g.* when importing an ASCII version of $\mathbf{V}'$ into a spreadsheet which rounds to two decimal positions). The rest of processing manipulations an image watermark should resist (see [16]) do not make much sense on a microdata file, because nobody is really interested in a cropped, scaled or compressed version of $\mathbf{V}'$.

The *capacity* of the watermarking scheme should be sufficient to allow embedding of $E_{k_j}(s_i||r_i)$, for $i = 1, \cdots, r$ and $j = 1, \cdots, u$. It must be noticed that a numerical microdata file is usually smaller than multimedia files: just in one color 512×512 RGB image, we have 3×2^{18} pixel values, which is more than the number of values in a typical microdata file. Thus capacity should be medium to high in comparison to standard multimedia watermarking schemes.

Regarding *obliviousness* and *imperceptibility*, there is an interesting tradeoff in this multilevel access application:

- An oblivious watermarking scheme does not require $\mathbf{V}'$ to recover the watermark from $\hat{\mathbf{V}}'$. In principle, distribution of $\mathbf{V}'$ is thus unnecessary, which saves storage and communication. However, this means that the user will remove masking from $\hat{\mathbf{V}}'$ rather than from $\mathbf{V}'$; so unless both files are very similar (*i.e.* the watermark is very imperceptible), unmasking $\hat{\mathbf{V}}'$ will not yield analytically valid results, because masking was performed on $\mathbf{V}'$.
- A non-oblivious watermarking scheme assumes that $\mathbf{V}'$ is available when recovering the mark from $\hat{\mathbf{V}}'$. So there is no problem if $\hat{\mathbf{V}}'$ differs significantly from $\mathbf{V}'$, because the user will be able to perform the unmasking on $\mathbf{V}'$. Thus, for a non-oblivious watermarking scheme, imperceptibility is not a requirement.

6 Choice of a watermarking algorithm

As noted in Section 2 when discussing SDC methods based on lossy compression, a numerical microdata file can be regarded as an image. Therefore, all image watermarking algorithms are potentially useable. However, from the requirements analysis of Section 5, we can conclude the following:

- Robust oblivious schemes like [10] or [9] cannot yield enough capacity while preserving a good level of imperceptibility. For example, for a microdata file with 1080 records and 13 variables, 10% distortions are needed to embed a 60-bit mark using [10]; it can be empirically seen that the amount of bits that can be embedded grows linearly with the percent distortion being used. In spite of 10% being already a distinctly perceptible distortion, a 60-bit mark can hardly accomodate a single encrypted seed (whereas embedding several encrypted seeds would be desirable).
- Robust non-oblivious schemes like [17, 6] offer a good level imperceptibility and good capacity, but require distributing $\mathbf{V}'$ along with $\hat{\mathbf{V}}'$.
- Oblivious methods, like Least Significant Bit (LSB) embedding, which are not robust enough for image watermarking, may be successfully adapted for the purposes of the application discussed here. Basically, the only manipulation LSB methods should survive in our case is quantization (due to rounding errors): this can be achieved by embedding one bit in a *group* of least significant bits rather than in the least significant bit. LSB methods offer high imperceptibility and allow embedding one bit in each numerical value of the microdata file, so they offer high capacity as well.

7 Conclusion

We have presented a novel application of watermarking which allows multilevel access to disclosure-protected numerical microdata sets. The higher the clearance of a user, the closer to the original one is the data set she can get. Unlike for most other applications, in the application presented here watermarking is *positive* for the user. Thus, the requirements on the watermarking algorithm are substantially different from standard requirements for multimedia content protection. What is needed is high embedding capacity, robustness only against rounding errors and obliviousness if compatible with imperceptibility. Interestingly enough, oblivious algorithms usually regarded as weak, like LSB watermarking, turn out to be the most suitable ones.

References

[1] N. R. Adam and J. C. Wortmann. Security-control methods for statistical databases: a comparative study. *ACM Computing Surveys*, vol. 21, 1989, pp. 515-556.

[2] I. J. Cox, M. L. Miller and J. A. Bloom. Watermarking applications and their properties. In *Proceedings of ITCC'2000*. Los Alamitos CA: IEEE Computer Society, 2000, pp. 6-10.

[3] D. Defays and P. Nanopoulos. Panels of enterprises and confidentiality: the small aggregates method. In *Proc. of 92 Symposium on Design and Analysis of Longitudinal Surveys*. Ottawa: Statistics Canada, 1993, pp. 195-204.

[4] D. E. Denning, *Cryptography and Data Security*, Reading, MA: Addison-Wesley, 1982.

[5] J. Domingo-Ferrer and J. M. Mateo-Sanz. On resampling for statistical confidentiality in contingency tables. *Computers & Mathematics with Applications*, no. 38, 1999, pp. 13-32.

[6] J. Domingo-Ferrer and J. Herrera. Simple collusion-secure fingerprinting schemes for images. In *Proceedings of ITCC'2000*. Los Alamitos CA: IEEE Computer Society, 2000, pp. 128-132.

[7] J. Domingo-Ferrer and V. Torra. Disclosure protection methods and information loss for microdata. In *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, North-Holland (to appear).

[8] J. Domingo-Ferrer and J. M. Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Transactions on Knowledge and Data Engineering* (to appear).

[9] J. Fridrich. Visual hash for oblivious watermarking. In *Proc. of SPIE Photonic West Electronic Imaging 2000. Security and Watermarking of Multimedia Contents*. San José CA: January 24-26, 2000.

[10] F. Hartung and B. Girod. Digital watermarking of raw and compressed video. In *Proceedings of SPIE*, vol. 2952, 1996, pp. 205-213.

[11] G. R. Heer. A bootstrap procedure to preserve statistical confidentiality in contingency tables. In *Proceedings of the International Seminar on Statistical Confidentiality*. Luxembourg: Eurostat, 1993, pp. 261-271.

[12] J. J. Kim. A method for limiting disclosure in microdata based on random noise and transformation. In *Proceedings of the ASA Section on Survey Research Methods*, 1986, pp. 303-308.

[13] C. K. Liew, U. J. Choi and C. J. Liew. A data distortion by probability distribution. *ACM Transactions on Database Systems*, vol. 10, 1985, pp. 395-441.

[14] J. Massey. Shift-register synthesis and BCH decoding. *IEEE Transactions on Information Theory*, vol. IT-15, pp. 122-127.

[15] R. A. Moore Jr. *Controlled data-swapping techniques for masking public use microdata sets*. Research Report, Statistical Research Division, U. S. Bureau of the Census, 1996.

[16] F. A. P. Petitcolas, R. J. Anderson and M. G. Kuhn. Attacks on copyright marking systems. In *2nd International Workshop on Information Hiding*, LNCS 1525. Berlin: Springer-Verlag, 1998, pp. 219-239.

[17] F. Sebé, J. Domingo-Ferrer and J. Herrera. Spatial-domain image watermarking robust against compression, filtering, cropping and scaling. In *Information Security Workshop'2000*, LNCS. Berlin: Springer-Verlag (to appear).

[18] L. Willenborg. *Statistical Disclosure Control in Practice*. New York: Springer-Verlag, 1996.