WP. 34
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

Joint UNECE/Eurostat work session on statistical data confidentiality
(Geneva, Switzerland, 9-11 November 2005)

Topic (v): Confidentiality aspects of tabular data, frequency tables, etc.

PROTECTING TABLES WITH CELL PERTURBATION

Invited Paper

Submitted by the University of La Laguna, Tenerife, Spain[1]

[1] Prepared by Juan-José Salazar-González (jjsalaza@ull.es).

Protecting tables with Cell Perturbation

Juan-José Salazar-González*

* DEIOC, University of La Laguna, Tenerife, Spain. (jjsalaza@ull.es)

Abstract. This paper presents a new methodology to protect sensitive information in tabular data. It is named Cell Perturbation and can be modelled as a linear programming problem suitable to be solved through a cutting-plane approach. The solutions will satisfy all the protection level requirements, as in other classical approaches like Cell Suppression or Controlled Rounding. Additionally optimal solutions of Cell Perturbation will have smaller loss of information. The paper concludes with computational results on benchmark instances publicly availably for comparisons with other approaches.

1 Introduction

Statistical agencies collect data to make reliable information available to the public. This information is typically made available in the form of tabular data (i.e., a table), defined by cross-classification of a small number of variables. By law, the agencies are obliged to preserve the confidentiality of data pertaining to individual entities such as persons or businesses. There are various methodologies to preserve confidentiality. We refer the reader to Willenborg and de Waal [16] for a wider introduction to this area, called *Statistical Disclosure Limitation*.

In this area, experts typically distinguish two different problems. The *primary problem* concerns the problem of identifying the sensitive data, i.e., the cell values corresponding to private information that cannot be released within a prescribed exactitude. In this problem also the set of potential attackers and their a-priori knowledge must be identify. The *secondary problem* (also named the *complementary problem*) consists in applying methods to guarantee protection requirements on each sensitive cell against each attacker, while minimizing the overall loss of information. This paper concerns only the secondary problem. The most popular methodologies for solving the secondary problem are variants of the well-known Cell Suppression and Controlled Rounding methods. These two fundamental methodologies will be described next. In practice, some implementations cannot inherently guarantee the protection requirements and great computational effort must be applied to check the proposed output before publication. This checking is called the *Disclosure Auditing* phase and basically consists in computing lower and upper bounds on the original value for each sensitive cell; in the literature there are several techniques to perform this third phase, including linear and integer programming, the Frechet and Bon-

ferroni bounds, and the Buzzigoli and Giusti's shuttle algorithm (see, e.g., Duncan, Fienberg, Krishnan, Padman and Roehrig [6] for references).

Cell Suppression is a methodology that allows the practitioner to do not publish the values in some cells while publishing the original values of the others. In particular, once the primary problem was solved, the cells containing sensitive information must be also not published and they are the *primary suppressions*. Due to the existence of the total marginals in a table, other cells must be also unpublished to guarantee protection of the values under the primary cells, leading to the *secondary suppressions*. They must be identified by solving the so-called *Cell Suppression Problem*, which is a very interesting combinatorial problem widely addressed in the literature. Apart from satisfying the protection requirements, the output of the problem must have a minimum loss of information, which for this methodology could be considered as the sum of the unpublished cell values. See, e.g., [16] for more details on this methodology.

Controlled Rounding is an alternative classical methodology that has not been extensively analyzed in the literature. When applying a rounding procedure the experts are given a base number and they are allowed to modify the original value of each cell by rounding it up or down to a near multiple of the base number. An output pattern must be associated with the minimum loss of information, which for this methodology can be considered as the distance between the original and the modified tables. In the *Random Rounding* version the experts decide to round up or down each cell by considering a probability that depends on its original cell value, without taking care of the marginal cell values. Therefore, the Random Rounding produces output tables where the marginal values are not the sum of their internal cells, which is a disadvantage of this rounding version. Another version is the so-called *Controlled Rounding*, where probabilities are not considered and the expert should round up or down all cell values such that all the equations in the table hold in the published table. In the so-called *zero-restricted Controlled Rounding* the original values which are already multiple of the base number cannot be modified. Even not considering protection level requirements, a Controlled Rounding solution may not exist for a given table (e.g., Causey, Cox and Ernst [1] showed a simple infeasible 3-dimensional instance). Kelly, Golden and Assad [12] proposed a branch-and-bound procedure for the case of 3-dimensional tables, and Fischetti and Salazar [9] extended this procedure to 4-dimensional tables. Heuristic methods for finding solutions of this problem on multi-dimensional tables have been proposed by several authors, including Kelly, Golden and Assad [12, 13]. The problem was first introduced in a statistical context by Cox and Ernst [2].

Salazar [14] presents a common framework to apply Controlled Rounding and Cell Suppression. In addition, two other closely related techniques are described. One tecnique is named *Interval publication*, and it is (in a sense) the linear programming variant of the Cell Suppression method. More details and computational

experiments are presented in Fischetti and Salazar [10]. The other technique is named *Cell Perturbation*, which similarly can be seen as the linear programming variant of the Controlled Rounding method. We present in this paper more details and computational results.

Section 2 introduces the main concepts of the Statistical Disclosure Limitation problem. These concepts are fundamental for comparing Cell Perturbation with other similar approaches, like for example the *Controlled Tabular Adjustment* introduced by Cox and Dandekar (see, e.g., Cox, Kelly and Patil [3]). Section 3 describes the *Cell Perturbation* method, with a cutting-plane procedure to find the optimal solution which runs in polynomial time. Results from computational experiments using the proposed methods are analyzed in Section 4.

2 Basic concepts and notation

A statistical agency is provided with a set of n values a_i for $i \in I := \{1, \ldots, n\}$. Vector $a = [a_i : i \in I]$ is known as "nominal table" and satisfies a set of m equations $\sum_{i \in I} m_{ji} y_i = b_j$ for $j \in J := \{1, \ldots, m\}$. For convenience of notation the linear system will be denoted by $My = b$, thus $Ma = b$ holds. Each solution y of $My = b$ is called *congruent table*.

Statistical tables typically contain sensitive data. We denote the subset of sensitive cells by P. In a general situation, all the sensitive cells in a table must be protected against a set K of *attackers*. The attackers are the intruders or data snoopers that will analyze the final product data and will try to disclose confidential information. They can also be coalitions of respondents who collude and behave as single intruders. The aim of the Disclosure Limitation Methods is to reduce the risk of them succeeding. Each attacker knows the set of linear system $My = b$ plus extra information that bound each cell value. For example, the simplest attacker is the so-called *external intruder* knowing only that unknown cell values are, say, nonnegative. Other more accurate attackers know tighter bounds on the cell values, and they are called *internal attackers*. In general, attacker k is associated with two bounds lb_i^k and ub_i^k such that $a_i \in [lb_i^k \ldots ub_i^k]$ for each cell $i \in I$. The literature on statistical disclosure control (see, for example, Willenborg and de Waal [16]) typically addresses the situation where $|K| = 1$, thus protecting the table against the external intruder with only the knowledge of the linear system and some external bounds; nevertheless this is a simplification of the real problem in Disclosure Limitation and statistical offices are interested in protecting tables against several intruders (see, for example, Jewett [11]).

To protect the sensitive cell p containing value a_p in the input table, the statistical office is interested in publishing an output containing several congruent tables, including not only the original nominal table but also others so that no attacker can disclose the private information a_p (neither a narrow approximation). The output

of a Disclosure Limitation Method is generally called a *pattern*, and it can assume a particular structure depending on the methodology considered.

The congruent tables associated to a pattern must differ so that each attacker analyzing the pattern will not compute the original value of a sensitive cell within a narrow approximation. For each potential intruder, the idea is to define a protection range for p and to demand that the a posteriori protection be such that any value in the range is potentially the correct cell value. To be more precise, by observing the published pattern, attacker k will compute an interval $[\underline{y}_p^k \ldots \overline{y}_p^k]$ of possible values for each sensitive cell p. The pattern will be considered *valid* to protect cell p against attacker k if the computed interval is "wide enough". To set up the definition of "wide enough" in a precise way, the statistical office gives three input parameters for each attacker k and each sensitive cell p with nominal value a_p:

- Upper Protection Level: it is a number UPL_p^k representing a desired lower bound for $\overline{y}_p^k - a_p$;
- Lower Protection Level: it is a number LPL_p^k representing a desired lower bound for $a_p - \underline{y}_p^k$;
- Sliding Protection Level: it is a number SPL_p^k representing a desired lower bound for $\overline{y}_p^k - \underline{y}_p^k$.

The values of these parameters can be defined by using common-sense rules (see, e.g, Sande [15]). In all cases, the protection levels are assumed to be unknown to the attackers. An elementary assumption is that

$$lb_p^k \leq a_p - LPL_p^k \leq a_p \leq a_p + UPL_p^k \leq ub_p^k$$

and

$$ub_p^k - lb_p^k \geq SPL_p^k,$$

for each attacker k and each sensitive cell p. For notational convenience, let us also define

$$lpl_p^k := a_p - LPL_p^k, \quad upl_p^k := a_p + UPL_p^k, \quad LB_i^k := a_i - lb_i^k, \quad UB_i^k := ub_i^k - a_i.$$

Given a pattern, the mathematical problems of computing values $\underline{y}_p^k$ and $\overline{y}_p^k$ are known as *attacker problems* for cell p and attacker k. The overall problem of solving the attacker problems for all cells is called *Disclosure Auditing Problem*. This should not be confused with the Disclosure Auditing Phase mentioned in Section 1 and which is an unnecessary phase for the methodologies proposed in this paper since they will implicitly guarantee the protection requirements on the output pattern.

Finally, among all possible valid patterns, the statistical office is interested in finding one with minimum information loss. The *information loss* of a pattern is

intended to be a measure of the number of congruent tables in the pattern. A valid pattern must always allow the nominal table to be a feasible congruent table, but it must also contain other different congruent tables so as to keep the risk of disclosure controlled. In practice, since it is not always easy to count the number of congruent tables in a pattern from the point of view of an intruder k, the loss of information of a pattern is replaced by the sum of the loss of information of its cells. In this case, the individual cost for cell p is generally proportional to the difference between the worse-case situations (i.e., to $\overline{y}_p^k - \underline{y}_p^k$), it is proportional to the number of respondents contributing to the cell value a_p, or it is simply a positive fixed cost when a_p is not published (i.e., when $\overline{y}_p^k - \underline{y}_p^k > 0$).

It is very important to observe that these concepts do not always coincide with the one used in other articles in the literature. This observation is fundamental to compare the methodology introduced in this paper with the methodology introduced by other authors. For example, in the Controlled Tabular Adjustment described in Cox, Kelly and Patil [3] the concept of "protected output" is different. In our framework an output is protected if, for each sensitive cell and each value in its protected range, an attacker must deduce the existence of a congruent table assuming this value in this cell. When considering different attackers, this congruent table may not be the same for all the attackers. Also when considering one attacker, different sensitive cells and different values may show different tables. In the framework used in [3] an output is protected if there is a congruent table valid which satisfy *one* of the two protection levels for *all* the sensitive cells. Of course, the reader should not understand from this words that the basic concepts used in [3] is wrong, but only different than the concepts for which Cell Perturbation has been proposed.

3 Cell Perturbation Methodology

The main disadvantage of the Controlled Rounding methodology is that a protected pattern does not always exist due to the tight requirement of rounding each cell value either down or up. A wat of ensuring the existence of protected patterns is to relax this requirement in the Controlled Rounding model and to look for a congruent table $v = [v_i : i \in I]$ such that

$$v_i \in [\lfloor a_i \rfloor \ldots \lceil a_i \rceil]. \tag{1}$$

where $\lfloor a_i \rfloor$ and $\lceil a_i \rceil$ are given in advance from the statistical office such that $\lfloor a_i \rfloor \leq a_i \leq \lceil a_i \rceil$. These extreme values can be defined as the nearest numbers to a_i which are multiples of a given number (i.e., defined as in the standard Controlled Rounding methodology from a given base number), but they can also be the two values within a given difference with respect to a_i (i.e., $\lfloor a_i \rfloor := a_i - t_i$ and $\lceil a_i \rceil := a_i + t_i$ for a given base number $t_i > 0$). Table v is then a pattern in the *Cell Perturbation methodology* and the novelty with respect to the Controlled Rounding is that now v_i can be any

value between the two extremes of the interval $[\lfloor a_i \rfloor \ldots \lceil a_i \rceil]$. As in the Controlled Rounding methodology, the loss of information of a cell i could be defined to be proportional to $|v_i - a_i|$, and the "loss of information" of a pattern is the sum of the loss of information of all the cells.

Obviously, if the requirement of rounding up or down is removed for all the cells, and no new one is added to the continuous relaxation of a model minimizing the non-linear function $\sum_{i \in I} |v_i - a_i|$, then the valid pattern with minimum loss of information is the nominal table a. A way to avoid this disappointing solution is to keep some requirements (for example, concerning the sensitive cells) or simply require that the published values in each sensitive cell must be equal to some given values (for example, $v_p = \lceil a_i \rceil$ for all $p \in P$). Still these additional constraints may lead to infeasible problems. Practitioners in statistical offices prefer another way of avoiding the nominal table as published table: it consists in defining a different objective function. Indeed, by considering the objective as the distance between each published value v_i and the value in $\{\lfloor a_i \rfloor, \lceil a_i \rceil\}$ closest to a_i we get the same criteria used in the classical Controlled Rounding methodology, and allow the objective function to be linear on the variables x_i.

Let $r_i := \lceil a_i \rceil - \lfloor a_i \rfloor$ a (possibly) known information for attackers. Then the attacker problems associated with attacker k are now exactly the same as in the Controlled Rounding methodology, i.e.

$$\begin{aligned} My &= b \\ v_i - r_i \le y_i &\le v_i + r_i \qquad && \text{for all } i \in I \\ lb_i^k \le y_i &\le ub_i^k && \text{for all } i \in I. \end{aligned}$$

As in the Controlled Rounding methodology, a necessary (but not sufficient) condition for feasibility is that $\max_{k \in K}\{SPL_i^k, UPL_i^k + LPL_i^k\} \le 2r_i$ for all $i \in I$.

In the literature there are several methodologies to protect tables by data perturbation (see, for example, Evans, Zayatz and Slanta [8], Duncan and Fienberg [7]) but, as far as we know, they all concern the direct modification of the microdata and, therefore, there is less control on the final protection interval of each cell in the published pattern.

To write a first model for the Cell Perturbation model, it is convenient to introduce two continuous variables z_i^- and z_i^+ for each cell i, with the following meaning:

$$z_i^- := \max\{0, a_i - v_i\}$$

$$z_i^+ := \max\{0, v_i - a_i\}.$$

Note that $v_i = a_i + z_i^+ - z_i^-$. Let w_i^- be the given cost for each unit of z_i^-, and w_i^+ be the given cost for each unit of z_i^+. Hence the objective function is

$$\sum_{i \in I} w_i^+ z_i^+ + w_i^- z_i^-$$

as in the Controlled Rounding methodology. One way to write the protection level requirements is to introduce additional variables f^{kp} and g^{kp} for each attacker k and each sensitive cell p.

It is again possible to avoid the explicit introduction of the auxiliary variables f^{kp} and g^{kp} ($k \in K$ and $p \in I$) along with the associated linking constraints, by using the standard LP Duality Theory. See Salazar [14] for a full technical description of two LP models for Cell Perturbation. Briefly, a first model is a compact formulation using a large (but still polynomial) number of variables. More precisely, the first model is a linear program using the auxiliary variables f^{kp} and g^{kp}. The second model replaces these variables by an exponential number (but polynomially separable) of linear inequalities. Although the two model are in equivalent in theory, in practice the second one is preferred. The reason is because it works with an small number of variables (z_i^+ and z_i^-, two for each cell), while the inequalities are generated on-the-fly only when required through an iterative procedure. In practice the number of iterations is small. Section 4 empirically supports this claim. What is more, under some hypothesis on the magnitude of the protection levels, the number of linear inequalities in this second model can be strongly reduced. An example is when the table is a frequency table, and the sensitive cells, the external bounds and the protection levels are set in accordance with the criteria proposed in [5]. This is a situation produced when protecting a frequency table with τ-ARGUS. In other words, a cell i with $\lfloor a_i \rfloor < \lceil a_i \rceil$ in a table generated by τ-ARGUS always satisfies

$$lb_i \leq \lfloor a_i \rfloor \leq lpl_i \text{ and } upl_i \leq \lceil a_i \rceil \leq ub_i$$

because $lb_i = \lfloor a_i \rfloor = lpl_i = 0$ and $upl_i = \lceil a_i \rceil < ub_i$. Under these hypothesis the external bounds and the protection levels are useless, and therefore the rounder called by τ-ARGUS will have the task of finding a fractional solution satisfying all the $|J|$ equations, optimal according to the objective function.

Cell Perturbation has some similarities with the Partial Cell Suppression introduced in Fischetti and Salazar [10]. Both methodologies can be formulated as a Linear Programming (LP) model with an exponential number of constraints that can be efficiently separated in a cutting-plane approach. They are closely related to the LP relaxations of two standard methodologies —Cell Suppression and Controlled Rounding— and they differ in several aspects. For example, an important requirement in the Cell Perturbation is the additivity of the output data, which should be a congruent table. This requirement is not present in Partial Cell Suppression, where the output is a table of intervals. Another requirement in the Cell Perturbation methodology is that each cell value cannot be modified by more than a given base number, which is not an input parameter of the Partial Cell Suppression methodology. The used base numbers, released to the public together with the output data when using Cell Perturbation, have a large impact in the utility of this data. From the practical point of view, it is preferred to use small base numbers

Name	Type	$\|I\|$	$\|J\|$	$\|P\|$	nzeros
bts4	hierarchical from $54 \times 54 \times 4 \times 4$	36570	36310	2260	136912
hier13	hierarchical from $13 \times 13 \times 13$	2020	3313	112	11929
hier16	hierarchical from $16 \times 16 \times 16$	3564	5484	224	19996
nine12	linked from $10 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6$	10399	11362	1178	52624
nine5d	linked from $4 \times 29 \times 3 \times 4 \times 5 \times 6 \times 5 \times 4 \times 6$	10733	17295	1661	58135
ninenew	linked from $10 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6 \times 6$	6546	7340	858	32920
two5in6	linked from $6 \times 4 \times 16 \times 4 \times 4 \times 4$	5681	9629	720	34310

Table 1: Short description of the benchmark instances

Name	mult	down	frac	up	distance	cost	max	time
bts4	226	11138	14146	11060	1825973.4	2725081.6	247.0	85.8
hier13	10	142	1749	119	33599.2	161975.1	106.6	0.8
hier16	10	255	3050	249	69988.1	291378.0	105.1	7.2
nine12	57	1852	6547	1943	383235.9	804652.1	177.0	164.3
nine5d	73	1736	7221	1703	355980.7	844324.4	204.1	31.3
ninenew	36	1119	4229	1162	224053.2	506292.6	171.0	34.5
two5in6	67	806	3971	837	167860.3	432765.1	141.9	5.8

Table 2: Results applying Cell Perturbation with $r_i = 250$

subject to the existence of a solution. This consideration does not apply to Partial Cell Suppression.

4 Computational Results

We have implemented the cutting-plane algorithm for solving the Cell Perturbation Problem. The implementation has been done in ANSI C using the Microsoft Visual C 6.0 compiler and the branch-and-cut framework of CPLEX 9.0. The experiments have been executed on a PC Pentium IV 2.5 Ghz. under Microsoft Windows XP.

For benchmarking purposes, we have run our codes on a collection of artificial instances close to being realistic. It consists of seven test cases of magnitude data created by Ramesh Dandekar (U.S. Department of Energy), described in Dandekar [4] and available through the webpage `http://webpages.ull.es/users/casc`. Three of the seven instances ("bts4", "hier13" and "hier16") are hierarchical tabulations, while the remaining four are linked tabulations. For each instance of the collection, Table 1 gives the name, the number of cells, the number of equations, the number of sensitive cells, and the number of non-zero elements in M. The protection against one attacker is assumed, i.e. $|K| = 1$.

Table 2 shows the results of applying the Cell Perturbation approach on this collection of data. We considered $r_i = 250$ and $s = 0$. The results are very similar

to the ones obtained when solving the LP relaxation of the Controlled Rounding model.

We have also run the algorithms on a benchmark instance provided by Anco Hundepool (Statistics Netherlands), available at `http://webpages.ull.es/users/casc`. It is a frequency table described by 30886 cells and 39800 equations in a 6-level hierarchical structure. There are 10680 sensitive cells and 120819 non-zero elements in M. An optimal Controlled Rounded solution was found in 279.7 seconds using our computer after exploring 842 nodes, and the optimal objective value is 42600. The solution of the LP relaxation at the end of the root node had 2772 variables with fractional values and the objective value is 42545.5. The root node was solved in 3.6 seconds and had a heuristic solution with objective value 42744. An optimal Cell Perturbation solution was found in 4.2 seconds, with 4121 fractional values and with objective value 42253.2.

Each equation in these instances determines a marginal cell by adding a subset of other cells, thus all the cells of each instance can be considered linked in a hierarchical structure. Since the cell values are assumed to be non-negative ($lb_i^k = 0$), starting from the grand total (the cell with the largest value) one can automatically assign levels of the hierarchical structure to the cells. The grand total cell is assigned to level 0. We have prioritized the cell variables so the branching phase selects a variable associated with a cell with higher level first. This consideration improved the performance of the algorithm in our experiments.

Acknowledgement

Work supported by the European project IST-2000-25069 (CASC) and by the Spanish project TIC2002-00895 (TEIDE).

References

[1] Causey, B.D., Cox, L.H. and Ernst, L.R. (1985) "Applications of Transportation Theory to Statistical Problems", *Journal of the American Statistical Association*, **80**, 903–909.

[2] Cox, L. H. and Ernst, L. R. (1982) "Controlled Rounding", *INFOR*, **20**, 423–432.

[3] Cox, L.H., Kelly, J.P. and Patil, R. (2004) "Balancing Quality and Confidentiality for Multivariate Tabular Data", 87–98, in J. Domingo-Ferrer and V. Torra (Eds.) *Privacy in Statistical Databases*, Lecture Notes in Computer Science 3050, Springer.

[4] Dandekar, R. A. (2004) "Maximum Utility-Minimum Information Loss Table Server Design for Statistical Disclosure Control of Tabular Data", 121–135 in Domingo-Ferrer, J. (editor) *Privacy in Statistical Databases*, Lecture Notes in Computer Science 3050, Springer.

[5] J.J. Salazar, P. Lowthian, C. Young, G. Merola, S. Bond, D. Brown, "Getting the Best Results in Controlled Rounding with the Least Effort", *Privacy in Statistical Databases* (ed. J. Domingo-Ferrer) LNCS 3050 (2004) 58–72.

[6] Duncan, G. T., Fienberg, S. E., Krishnan, R., Padman, R. and Roehrig, S. F. (2001) "Disclosure Limitation Methods and Information Loss for Tabular Data", 135–166 in Doyle, P., Lane, J., Theeuwes, J. and Zayatz, L. (editors) *Confidentiality, Disclosure and Data Access: Theory and Practical Applications for Statistical Agencies*, Elsevier Science.

[7] Duncan, G.T. and Fiendberg, S.E. (1998) "Obtaining information while preserving privacy: a Markov perturbation method for tabular data", Proceedings of the *Statistical Data Protection* conference, 351–362.

[8] Evans, T., Zayatz, L. and Slanta, J. (1998) "Using Noise for Disclosure Limitation of Establishment Tabular Data", *Journal of Official Statistics*, **14/4**, 537–551.

[9] Fischetti, M. and Salazar, J. J. (1998) "Computational Experience with the Controlled Rounding Problem in Statistical Disclosure Control", *Journal of Official Statistics*, **14/4**, 553–565.

[10] Fischetti, M. and Salazar, J. J. (2003) "Partial Cell Suppression: a New Methodology for Statistical Disclosure Control", *Statistics and Computing*, **13**, 13–21.

[11] Jewett, R. (1993) "Disclosure Analysis for the 1992 Economic Census", internal report, U.S. Bureau of the Census, Washington.

[12] Kelly, J. P., Golden, B. L. and Assad, A. A. (1990) "Using Simulated Annealing to Solve Controlled Rounding Problems", *ORSA Journal on Computing*, **2**, 174–185.

[13] Kelly, J. P., Golden, B. L. and Assad, A. A. (1993) "Large-Scale Controlled Rounding Using TABU Search with Strategic Oscillation", *Annals of Operations Research*, **41**, 69–84.

[14] Salazar, J. J. (2003) "A Unified Mathematical Programming Framework for Different Statistical Disclosure Limitation Methods", to appear in *Operations Research*, 2005.

[15] Sande, G. (1984) "Automated Cell Suppression to preserve confidentiality of business statistics", *Statistical Journal of the United Nations ECE*, **2**, 33–41.

[16] Willenborg, L. C. R. J. and de Waal, T. (2001) *Elements of Statistical Disclosure Control*. Lecture Notes in Statistics 155, Springer.