WP. 38
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.

USING FIXED INTERVALS TO PROTECT SENSITIVE CELLS INSTEAD OF CELL SUPPRESSION

Supporting Paper

Submitted by the Bureau of Labor Statistics, United States of America[1]

[1] Prepared by Steve Cohen (cohen_steve@bls.gov) and Bogong T. Li (li_t@bls.gov).

Using Fixed Intervals to Protect Sensitive Cells instead of Cell Suppression

Steve Cohen and Bogong T. Li *

*Bureau of Labor Statistics, 2 Massachusetts Ave. NE., Washington, D.C. 20212, e-mail: cohen_steve@bls.gov and li_t@bls.gov,

1 BLS QCEW Proposed Publication Change

BLS Quarterly Census of Employment and Wages (QCEW) is a census that collects data under a cooperative program between BLS and the State Employment Security Agencies. The data contain broad employment and wage information for all U.S. workers covered by state unemployment insurance laws and federal workers covered by the Unemployment Compensation for Federal Employee program. Tabulations of QCEW outcomes are available by 6-digit NAICS industry, by county, by ownership sectors and by size groups, in the form of print, automatic e-mail, fax or plain text file directly from BLS Internet ftp servers. The detailed coverage and readily availability of the QCEW tabular data make it especially vulnerable to confidentiality disclosure risks. Cell suppression is used as for the tabular data confidentiality protection schema.

Since cell suppression methods currently implemented suppress a large number of cells in order to protect QCEW publication tables, an alternative method is sought. Using QCEW data analyzed in this paper, following the BLS confidentiality sensitivity measures, we found for this data set containing employment of five major industry sectors (2-digit NAICS sectors) within a medium-sized U.S. State, 9979 or 59% of 16,878 publication cells have to be completely suppressed using network method, 10631 or 62% of all cells using the hypercube method (for a description of the hypercube method see Repsilber (1994)). The level of employment represented by the suppressed cells is relatively small in comparison to the number of cells suppressed, ranging from 10% to 15% of the total value. Similar results of this magnitude for cell suppression have been also reported by other researchers. Much detail on industry employment distribution at various geographic levels and other cross-classifications is lost due to confidentiality protection

One alternative to complete suppression considered by QCEW would be to publish primary cells in pre-defined, fixed intervals (FIs). Instead of suppressing the value of the sensitive cells, this method would publish all primary suppression cells in FIs which contain the exact value of the sensitive cell value. The consistency of the definition of these pre-defined intervals is kept across tables so that the users can compare values between various industries, geographic locations and other classifications by establishment characteristics, by just looking at the intervals. This method of publication can be used for employment and earnings data, though our discussion in this paper will only focus on employment level data.

Similar to the issues surrounding the cell suppression problem (CSP), if QCEW data is published with FIs replacing primary suppression cells, to prevent outside intruders gaining identifiable information of individual contributors to a cell, additional protecting cells (PCs) may have to be published in FIs. Otherwise an intruder may be able to utilize this additional information and the additive relationships existing in the table to estimate the value of primary cells now in FIs and therefore the value of some contributors to the cell. Intruders can produce better estimates now than

before with the added information of published FI bounds. The problem of minimizing the amount of cell values now expressed in FIs by selecting the right set of PCs while still preserving the protection of primary cells is what we call the fixed interval publication problem (FIPP). We will use the following fixed interval ranges for employment levels: 0-19, 20-99, 100-299, 250-499, 500-999, 1000-2499, 2500-4999, 5000-9999, 10000-24999, 25000-49999, 50000-99999, 100000 or more.

Since this risk arises from the additive relationships in the table and is similar to CSP solutions that have been implemented in some BLS survey programs, we start searching solutions made to solve CSP. Our current knowledge indicates CSP problem has been established by researchers as a MILP problem, see Kelly (1990). Exact solution to MILP model belongs to the class of the strong *NP*-hard problem. Heuristic solution procedures such as the network flow method, see Cox (1980 and 1995), for 2-dimensional tables, multi-commodity network flow method for n-dimensional tables, see Castro and Nabona (1996) and hypercube method by Repsilber (1994) and Giessing (2001) have been proposed. These heuristic methods only provide sub-optimal solutions as pointed by Castro (2001). Fischetti and Salazar (1999) proposed a solution using branch-and-cut algorithm as one of the mathematical programming techniques to reach a solution with proven optimality on 2-dimensional tables with up to 500 rows and 500 columns. The problem is solved in a few minutes on a standard PC. Fischetti and Salazar-Gonzales (2000) extended their work to other tabular data including k-dimensional table with k>2, hierarchical tables, linked tables etc., using branch-and-cut based procedures. Alternatively, instead of completely suppressing table cells, Salazar (2001); Fischetti and Salazar (2003) proposed a "partial cell suppression" method that will publish a subset of table cells with variable estimation intervals. Though FIPP and CSP shares the same MILP model, *unfortunately*, so far we think all of the above mentioned secondary cell selection methods do not apply directly to selecting protecting cells (PCs) that are to be published in FIs, neither optimally nor heuristically. The reason is that these models can not accommodate the knowledge of the FI bounds.

In this research we will propose an iterative "selection-improvement" algorithm, which improves cell selection upon each previous suppression pattern until all primary cells are sufficiently protected. All of the selection-improvement steps begin with procedures already implemented in BLS QCEW program. Though no claim of optimality is made in this paper, this method does make publication of tables with FIs realistic, and, as the evaluation at the end shows, there aren't significantly more cells published as FIs than the number of cells completed suppressed. After describing our procedure, we will provide an evaluation study using actual employment data from a U.S. state. We will compare the results with current suppression methods, look into convergence rates, level of information loss and computer programming difficulties associated with various cell selection methods.

2 The Selection-Improvement Algorithm

The iterative selection-improvement algorithm has two stages at each iteration, (1) selecting PCs and (2) conducting an audit on the publication table with the newly selected PCs in FIs. If the audit finds any primary cell is still at risk, the algorithm re-iterates by selecting more PCs and conducting another audit until all primary cells are protected. The initial set of PCs is the set of cells selected through one of the CSP methods. In case the iterations fail at the end, i.e. no candidate PCs available for selection while there are still unprotected cells, the method defaults back to the usual CSP solutions targeting only the remaining exposed cells. The steps of the algorithm are summarized as follows:

Step 1. Identify primary and secondary cells in a table via a CSP method and publish them in pre-defined FIs.
Step 2. Apply linear constrained optimization to identify those primary cells with disclosure risks.
Step 3. For those primary cells at risk, select additional cells that have not been selected previously from the publication table and publish them in FIs. Three specific methods are proposed for this research and will be briefly described in following paragraph and sections. This is the 'selection step".
Step 4. Apply linear constrained optimization again to check if any primary cell in the original table is still at risk. If yes, return to step 3; otherwise EXIT the algorithm, the table is successfully protected. This is the "audit step".
Step 5. If the step 2 – 4 iteration fails to protect every primary cells, i.e. no further unsuppressed cells available for selection while there are still disclosed primary cells, use any solution method to CSP, i.e. completely suppress these exposure primary and corresponding secondary cells.

There are several alternative methods can be used to select additional PCs in Step 3. We can randomly select cells that are within the same row or column of the exposed primary cells, or we can select through more complex MILP models and mathematical programming techniques. We would like to minimize either the number of cells to be selected or the total value of the selected cells. In this paper we studied the following three methods in the selection step: the Systematic, Single-Source Shortest Path (SSSP) and the Random Selection methods. The methods are briefly described below with a detailed discussion in section 2.1.

1. Systematic Method. To minimize values published in intervals, this method selects the smallest cell among all cells that form additive relationship with two selected exposure cells that need further protection that has not been suppressed during the previous iteration(s). This cell is published as a pre-defined FI. Default to Random Selection Method (see 3 next) at the end if this method fails.

2. Single-Source Shortest Path (SSPS) Method. This method models the table as a network similar to Traveling Salesman's Problem (TSP), treat all primary exposure cells on a table as destinations of a traveling map. The method aims to find the shortest path through these destinations, to minimize the total cell values expressed in FIs. To make this TSP solvable for all tables, the method fixes the order of the destinations or vertices on the table network. The method only needs to find the shortest path connecting the order-fixed set of vertices to form a closed "loop" with minimized path. Publish all cells that are not already selected in previous iterations on the chosen loop in FIs. Default to Random Selection Method if this method fails at the end.

3. Random Selection Method. This method randomly selects a cell among all cells that form additive relationship with the primary exposure cells. The candidate cells are cells that are either in the same row or column as the primary cell. If all cells forming additive relationships are already selected during previous iteration(s), or it by itself is the only decent from the higher hierarchy, go one hierarchy step higher until additional protecting cells can be found through additive relationships. Randomly select protecting cells among the candidates, publish these and all cells along the hierarchical searching path as FIs.

2.1 Methods to Selection Additional PCs at the Selection Step

We denote the set of cells with exposure risk by their indices as $E = \{i_1, \cdots, i_p\}$, where i's are cell indices in a publication table. We need to select additional PCs among cells that are not published previously in intervals, and publish them in intervals in the next iteration(s). We denote in the k^{th} iteration the additional subset cells of E that are still at risk as E_k, k =0, 1, 2, …, K, $K \in N$, such that $E \supseteq E_0 \supseteq E_1 \supseteq E_2 ... \supseteq E_K$. We say a publication table is "safe" if and only if $E_K = \mathrm{O}$ (O is the empty set) and K is finite. Notice we restrict our selection of additional set of PCs targeting only E_k at step k, i.e. within previous risk cell subsets which are subsets of E_k, though audit at k^{th} iteration may indicate an exposure risk cell(s) that are outside of E_k. We also denote the set of PCs selected at k^{th} iteration as F_k, k=0, 1, 2, …, K, $K \in N$. F_ks are mutually exclusive given the set indices are different. F_0 is an empty set, F_1 is the first PC set etc., F_K is the last set of PCs selected when the $E_K = \mathrm{O}$ condition is met etc. $\bigcup F_k$ is the final set of all PCs selected for the entire table.

Systematic Method. To select additional PC to protect all exposure cells found in the last iteration, Systematic method begins by randomly selecting a pair of cells in E_k, say (i_p, i_q), where $p \neq q$, then identify all cells i such that they form additive relationships with both i_q and i_q. The cell with the smallest cell value among these cells is to be selected and put into F_k - we can alternatively select the cell with the smallest number of establishments contributing to the cell. In this way one additional PC is selected for every pair of exposure cells in E_k. If this is not possible, for example in the case that no candidate PC is available, we need to resort to Random Selection method which we discuss below. In case that the number of cells in E_k is odd, i.e. there will be one cell does not find a pair within E_k for it, choose the smallest cell that forms an additive relationship with that cell. The cells in E_k which were not sufficiently protected are put into the set E_{k+1}, which will require additional PCs during the next iteration. Notice this process will continue if and only if $E_k \neq 0$, otherwise the table is declared "safe". It is possible however $E_k \neq \mathrm{O}$ but $F_k = \mathrm{O}$, i.e. no further PC is available for selection while the table is still not fully protected. This happens because all available PCs have already been selected in previous iterations. In order to carry the process to completion, Systematic method uses the Random Selection method which selects additional PCs in a way that guarantees a full protection at the end of all iterations.

SSSP Method. This method differs from Systematic method only in selecting the additional PC set F_k in iteration k. This method utilizes an algorithm similar to the single-source shortest path (SSSP) algorithm, as explained in Huo (2004, p.308) that finds the shortest path between two vertices on a network connected by weighted edges. We need to view the publication table in the form of closed networks, where table margins form the vertices and the cells the connecting edges, for more discussion on statistical table's network conversion see Cox (1995). Cells in E_k are ordered on the network in a fixed fashion. The ends of the cells in E_k -- the cells are edges in the connected network. They form the set of vertices we desire to find the shortest paths to connect them. The path of edges that along this shortest path forms the set of cells in F_k, the PC set in iteration k. The order of the vertices in E_k is fixed to avoid turning the problem into a Travelling Salesman's Problem that

in theory is NP-hard, see Chartrand (1977), though by doing this we are not guaranteed to find the truly shortest path running through all vertices in E_k. During the selection step in iteration, we use an optimized algorithm to search through all paths joining the vertices formed by cells in E_k and select the path with the smallest cell value combined. These cells then will be selected to form F_k. After the selection step, an audit step will proceed to check if any cell in E_k besides the ones are already protected is still at exposure risk. Next iteration will proceed if and only if $E_K = \mathrm{O}$. It is possible $E_k \neq 0$ but $F_k = 0$, as explained earlier. To complete the process, SSSP method uses the Random Selection method which will guarantee a full protection at the end of the iterations.

Random Selection Method. Random Selection method differs from the previous two methods only in how it selects the additional PC set F_k. Systematic method aims to minimize the number of cells in F_k in each iteration, while the SSSP method aims to minimize total cell value in F_k. Both the Systematic and SSSP method do not guarantee a solution, because there could be no more table cells available for selection before all cells in E are completely protected. The Random Selection method is primarily designed to complete this process, though it could be used alone from start. Random Selection method selects one additional PC randomly among all cells that are not previously selected and also form additive relationships with one cell in E_k that needs protection. Usually this cell in E_k is the cell fails the previous two methods. The auditing step will indicate if the table is "safe" at the end of step k, i.e. whether $E_K = \mathrm{O}$, otherwise the selection-audit iterative process continues. When all cells forming additive relationship with cell in E_k are already selected before E_k is protected, this is very likely for cells in E_k that are the only decedent in a hierarchical structure, Random Selection method chooses to step one or more steps up in the hierarchy to find PCs where are available, treating cells along the path as additional primary risky cells put into E_k. Random Selection is performed targeting these new primary risky cells as well as the original exposure cells in E_k. In practice we found in many instances that a cell with one hierarchy higher than an exposure cell is also an exposure cell therefore we have to use Random Selection method quite often in order to carry the whole process to finish.

In addition to providing a valid solution, the FIPP algorithm introduced here is easy to implement in production, since it simply combines separate existing confidentiality protection procedures, such as the complementary cell suppression techniques and auditing of tabular data through linear programming. It requires less software changes in the survey production environment because the only change to current complementary cell selection procedures is the addition of auditing cycles. The difficulty of selecting additional PCs could be simple, for the Random Selection Method, or modestly complex, for the SSSP. The auditing of a table during any stage of the process can be done through available table auditing software tools. Programming work for selecting PCs is only need to be done once and be reused later. More importantly, this method does not alter the actual micro data behind the tabular publication, as that of methods like adding noise to the micro data, which may add unwelcome noise to even safe cells.

3 Evaluation of the Method Using a Subset of Actual QCEW Data

We used actual QCEW employment publication tables for evaluating our stated FIPP procedure. This subset of QCEW data contains eight major 2-digit NAICS industry sectors in a medium-sized U.S. State. Table 1 displays the distribution of these industries and their establishment composition. In actual BLS publications, these data are published in tabular form separately in multi-dimensional table format classified by county and 6-digit NAICS industry, as well as by establishment size group, metropolitan statistical area (MSA) and ownership types. We used only the 2-dimensional employment table classified by county and hierarchical NAICS code, from 2 to 6-digit, to demonstrate our algorithm. Uses of 2-diminensional table may limit our evaluation conclusion, since multi-dimensional publication tables are "connected", or in other words there are more additive relationships existing than what we considered. Nevertheless these additional additive relationships are identifiable. Once they are correctly identified, we can always add them in the model. Therefore we believe with some modification our method applies to tables with any dimensions and we should expect the number of cells in FIs somewhat more than we report here. Table 2a displays a portion of this publication table currently a user sees in BLS publications. In this table the cells marked with "x" are suppressed cells due to primary and secondary suppressions. In this evaluation, we will apply our FIPP procedure to the data and compare their performance with that of the complete suppression. Table 2b shows the results treated with our FIPP procedures.

NAICS code	Industry	# of Establishments	Employment
31-33 (34)	Manufacturing	3,847	158,398
44-45	Retail Trade	15,563	288,980
48-49	Transportation and Warehousing	3,004	61,016
51	Information	704	16,805
52	Finance and Insurance	6,942	138,400
53	Real Estate and Rental and Leasing	4,504	45,839
54	Professional- Scientific- and Technical-Services	14,955	191,343
62	Healthcare and Social Assistance	11,326	265,607
	Total	*60,845*	*1,166,388*

Table 1. Study industry establishment population distribution

For a quick note about how we process the data through some computing tools: we first put the raw micro data through primary and secondary suppression selection using software tool Tau-Argus, see Hundepool, Willenborg et al. (2004). The suppressed table is then formatted to lp format in S-plus® to be used in Matlab®. In Matlab® we called solver lp_solve to conduct the audit of the table, lp_solve is a MILP solver available from the Internet community. If the iteration is not finished, the audited table is passed back to S-plus® and again we select additional PCs with the three methods we stated earlier in this paper. S-plus® and Matlab® were used to convert the publication table between publication tables and LP model input formats. Additional PCs are selected within S-plus® where additive relationship of the entire publication table is kept. Unless the cycles successfully

protected all primaries, the cycle should reiterate itself continuously. The convergence is guaranteed through the Random Selection method.

4 Summary of Evaluation Results

Each method was carried out to the end without having to apply the Random Selection method. For cell suppression, there are 9979 (59% of total) cells, or 4,535 (7.5% of total) establishments and 162,368 (14% of total) of employment value that are completely suppressed. With the Systematic Selection method, the entire publication table is successfully protected at the end with only two additional iterations beyond the traditional secondary suppression stage. However the number of cells published in FIs is quite large, not surprisingly since the procedure incorporates the existing secondary suppression methods. For the Systematic Selection method, 10,199 or 60% of all publication cells are selected for FI publication, they account for 6,337 establishments or 10% of all establishments in the table and 180,742 or 15% of total employment in the table. However, if taking into consideration of the number of publication cells suppressed, the number of establishments and total values in FIs, the difference between FIPP solution and complete suppression solution is not very large.

Separately for SSSP method and Random Selection method there are about 64% and 69% of all cells in FIs respectively. In terms of the number of iterations required to reach complete protection of the publication table, Systematic Selection takes 2, SSSP takes 3 and Random Selection takes 5. The reason for the difference in the number of iterations could be attributed to the relatively inefficient methods of picking addition PC cells used by the latter two methods. In particular, the Random Selection method does not taken into consideration of the magnitude of all qualified cells.

See Table 3 for summaries and comparisons of the total number of cells and values contained in FIs under each of the three different selection methods as compared to cell suppression.

5 Conclusions

We developed this FIPP solution with the goal to minimize either the total number of cells selected or the total value contained in the cells selected to be FIs. However since the initial step is built upon secondary cells suppression through CSP solutions and subsequent ad hoc PC selection steps, we probably do not achieve this goal truly. The good news is that all confidentiality rules imposed on the table are well preserved and the number of cells released as FIs is reasonable at the conclusion of the algorithm. The last audit step on the table clearly demonstrates all primary cells on the table are well protected. With reasonable effort a feasible solution can be found to a seemingly unsolvable optimization problem. The success of these methods relies on the assumption that the number of iteration cycles is not large, since current CSP solutions tend to over suppress in the first place. Even with the least inefficient selection method, the Random Selection method, only a maximum of five cycles are needed. The complexity of programming, computer usage time and manual intervention varies depending on selection method used. We found the Random Selection takes the least amount of programming time and manual intervention, SSSP takes longer to run on computer and needs more overhead programming effort, and Systematic method requires more manual interaction during the process than any of the other two methods, therefore is the most cumbersome to use. It is possible with more effort put into the computer programming in the future, we can integrated various parts of

NAICS code	Counties of a U.S. State								
	Total	County 1	County 2	County 3	County 4	County 5	County 6	County 7	etc.
…	…	…	…	…	…	…	…	…	
…	…	…	…	…	…	…	…	…	
451	13940	113	1758	2691	111	X	241	64	
4511	9070	82	1121	1699	x	X	166	x	
45111	4187	26	703	773	89	-	51	51	
451110	4187	26	703	773	89	-	51	51	
45112	2648	x	274	451	x	X	x	-	
451120	2648	x	274	451	x	X	x	-	
45113	1237	x	110	302	-	X	x	x	
451130	1237	x	110	302	-	X	x	x	
45114	998	x	35	173	-	-	38	x	
451140	998	x	35	173	-	-	38	x	
4512	4870	31	637	992	x	-	75	x	
45121	3415	x	504	444	x	-	x	x	
451211	3193	x	x	438	x	-	x	x	
451212	222	x	x	6	-	-	-	x	
45122	1455	x	133	548	x	-	x	-	
451220	1455	x	133	548	x	-	x	-	
…	…	…	…	…	…	…	…	…	
…	…	…	…	…	…	…	…	…	
Total	1166388	15589	98129	190226	7524	5018	22485	12171	etc.

"x" are nondisclosable data due to primary and secondary suppressions

Table 2a. A sample evaluation data set as published perturbed for confidentiality

NAICS code	Counties of a U.S. State								
	Total	County 1	County 2	County 3	County 4	County 5	County 6	County 7	etc.
…	…	…	…	…	…	…	…	…	
…	…	…	…	…	…	…	…	…	
451	13940	113	1758	2691	111	0-19	241	64	
4511	9070	82	1121	1699	20-99	0-19	166	20-99	
45111	4187	26	703	773	89	-	51	20-99	
451110	4187	26	703	773	89	-	51	20-99	
45112	2648	0-19	274	250-499	0-19	0-19	0-19	-	
451120	2648	0-19	274	250-499	0-19	0-19	0-19	-	
45113	1237	0-19	110	302	-	0-19	20-99	0-19	
451130	1237	0-19	110	302	-	0-19	20-99	0-19	
45114	998	20-99	20-99	173	-	-	38	0-19	
451140	998	20-99	20-99	173	-	-	38	0-19	
4512	4870	31	637	992	0-19	-	75	0-19	
45121	3415	20-99	504	444	0-19	-	20-99	0-19	
451211	3193	20-99	250-499	438	0-19	-	20-99	0-19	
451212	222	0-19	20-99	6	-	-	-	0-19	
45122	1455	0-19	133	548	0-19	-	20-99	-	
451220	1455	0-19	133	548	0-19	-	20-99	-	
…	…	…	…	…	…	…	…	…	
…	…	…	…	…	…	…	…	…	
Total	1166388	15589	98129	190226	7524	5018	22485	12171	etc.

Table 2b. The same section of the evaluation data set as it is published under FIPP method

	Systematic	*SSSP*	*Random*	*Cell Suppression*
Number of iterations to reach convergence	2	3	5	NA
Total number of cells in Fis or completely suppressed	10,199 (60%)	10,772 (64%)	11,615 (69%)	9979 (59%)
Total employment level in FIs or completely suppressed	180,724 (15%)	184289 (17%)	188,955 (16%)	162,368 (14%)
Total number of establishments in Fis or completely suppressed	6,337 (10%)	7,362 (12%)	7,971 (13%)	4535 (7.5%)

Table 3. Cells published as FIs by three difference selection methods compared to CSP method

software tasks into a single program. This is necessary if our proposal is to be adopted in regular publication production environment

One other advantage of our method is that a user can specify cells that he or she does not want to be published in FIs. Once specified, these cells will be treated as if they are constants in the model. The method also allows a user do global coding, i.e. combining categorical variables such that the result will be a table with fewer unsafe cells, though this may need to be done before the selection-audit cycles begin.

We also noticed the following problems with our methods during evaluation of the test data:

1. For the Systematic and SSSP selection methods, the order of the exposure primary cells during each iteration affect the additional PCs selected. In other words, the final set of FI cells could possibly be different if the process is run more than once, since the order of exposure primary cells entered the local protection cycle may be in a different order. Unless the order of cell entry is fixed, which is possible, the process is not repeatable.
2. The Random Selection method produces a different set of selection cells every time it runs, due to the random nature of its selection of PCs in local cycles. Setting the random seeds during iterations will be intractable. Therefore the PC selection process is not repeatable.
3. Though in theory the methods apply to table with any dimensions and hierarchical structures, as long as the additive relationships in the table is expressible, the time allowed us to conduct the study so far limit ourselves to only 2-dimentional tables with hierarchical structure in one dimension. Higher dimensional tables require us decompose the table into lower dimensional tables and process lower dimensional tables separately then “back-track” separate results at the end. We chose not to experiment that in this study.

Since the test data we used in this study are in reality published as multi-dimensional tables, i.e. there are other additive relationships in the table we actually did not take into consideration, indubitably more cells will be published in FIs and the programming working will be more demanding if the multi-dimensionality is taken into consideration. This is stated in the limitation 3 above. It is convincible that with other practical issues surrounding publishing sensitive cells in FIs, more works have to be done before we can adopt this method for QCEW regular publication.

Nevertheless, this paper demonstrates with experiments on actual data that our method provides one feasible solution to a seemingly difficult problem. We think our SDC method has good potential for future use in tabular statistical publications.

References

Castro, J. (2001). "Using Modeling Languages for the Complementary Suppression Problem Through Network Flow Models." Joint ECE/Eurostat Work Session on Statistical Data Confidentiality.

Castro, J. and N. Nabona (1996). "An Implementation of Linear and Nonlinear Multi-commodity Network Flows." European Journal of Operational Research **92**: 37-53.

Chartrand, G. (1977). Introductory Graph Theory, Dover Publications, Inc.

Cox, L. H. (1980). "Suppression Methodology and Statistical Disclosure Control." Journal of the American Statistical Association **75**: 377-385.

Cox, L. H. (1995). "Network Models for Complementary Cell Suppressions." Journal of the American Statistical Association **90**: 1453-1462.

Fischetti, M. and J. J. Salazar-Gonzales (2000). "Models and Algorithms for Optimizing Cell Suppression Problems in Tabular Data with Linear Constraints." Journal of the American Statistical Association **95**: 916-928.

Fischetti, M. and J. J. Salazar (1999). "Models and Algorithms for the 2-Dimensional Cell Suppression Problem in Statistical Disclosure Control." Mathematical Programming **84**: 283-312.

Fischetti, M. and J. J. Salazar (2003). "Partial Cell Suppression: A New Methodology for Statistical Disclosure Control." Statistics and Computing **13**: 13-21.

Giessing, S. (2001). Nonperturbative Disclosure Control Methods for Tabular Data. Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies. Doyle, Lane, Theeuwes and Zayatz, North-Holland.

Hundepool, A. J., L. C. R. J. Willenborg, et al. (2004). Tau-Argus User's Manual, Version 3.2.

Huo, H. W. (2004). Exercises & Solutions on Algorithms. Beijing, China, China Higher Education Press.

Kelly, J. P. (1990). Confidentiality Protection in Two and Three-Dimensional Tables. College Park, Maryland, University of Maryland, College Park, Maryland. Ph.D. Thesis.

Repsilber, R. D. (1994). Preservation of Confidentiality in Aggregated Data. Second International Seminar on Statistical Confidentiality. Luxembourg.

Salazar, J. J. (2001). "Improving Cell Suppression in Statistical Disclosure Control." Joint ECE/Eurostat Work Session on Statistical Data Confidentiality.

Note that the views expressed in this paper are those of the authors and do not necessary represent the policy of the Bureau of Labor Statistics.