BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES
SRD Research Report Number: Census/SRD/RR-87/06

Research at the Census Bureau into Disclosure Avoidance Technique for Tabular Data

by

Lawrence H. Cox
James T. Fagan
Brian Greenberg
Robert Hemmig

Bureau of the Census
Washington, D.C. 20233

Recommended by: Nash Monsour

Report completed: June 30, 1986

Report issued: February 16, 1987

RESEARCH AT THE CENSUS BUREAU INTO DISCLOSURE AVOIDANCE TECHNIQUES FOR TABULAR DATA

1. INTRODUCTION

This paper describes recent results obtained by the Census Bureau Confidentiality Staff in its research into disclosure avoidance methods for publicly released tabular data. Tabular data can be in the form of frequency counts where a population is cross-classified by specified characteristics, for example, age by sex. Each cell contains the number of individuals (or households, etc.) belonging to that cell. Tabular data can also be in the form of amounts where each cell contains the cross-classified aggregate total of some variable, such as total payroll at the state level displayed for SIC by county. A major goal of the research described here is the development of improved disclosure avoidance procedures for frequency count data for the 1990 Decennial Censuses. Procedures developed for frequency count data can be applied to tables of amounts (and conversely). However, the notions of what constitutes (1) a disclosure and (2) adequate protection are quite different in each instance. The discussion in this paper will be couched in terms of frequency count data with the understanding that the basic structures can be applied to tables of amounts as appropriate.

Data tables can be one-dimensional (e.g., county populations summing to a state total), two-dimensional (e.g., age by sex), or of three or more dimensions (e.g., age by race by sex). We will report on rigorous new procedures which have been successfully developed for rounding, perturbation, and cell suppression in two-dimensional tables, with a focus on their common underlying structure.

Each of the three procedures will be described in terms of a common mathematical structure, circuits in a graph. Every two-way table of m internal rows and n internal columns gives rise to a bipartite graph of $(m+1) + (n+1)$ nodes in which nodes correspond to marginal positions and edges correspond to nonzero

table cells. This common conceptual framework highlights the similarities and differences among these three procedures, suggests ways for extending them, and sheds light on why methods successfully employed on two-dimensional tables fail in three dimensions.

We begin by establishing the notation to be used throughout. A two-way table, A, is represented as:

$$A = \begin{array}{c|c} (a_{0,0})_{1x1} & (a_{0,j})_{1xn} \\ \hline (a_{i,0})_{mx1} & (a_{i,j})_{mxn} \end{array}$$

where a_{ij} $(0<i<m,\ 0<j<n)$ are non-negative integers. The vectors $(a_{i,0})$ and $(a_{0,j})$ $(1<j<m,\ 1<j<n)$ are row and column totals, respectively, of A, and a_{00} is the grand total. Thus, A is an additive table. Disclosure occurs in a frequency count table when small counts are released or can be narrowly estimated. If releasing A would result in disclosure, one creates a masked table

$$B = \begin{array}{c|c} (b_{0,0})_{1x1} & (b_{0,j})_{1xn} \\ \hline (b_{i,0})_{mx1} & (b_{i,j})_{mxn} \end{array}$$

from A which is suitable for public release. A major objective is that the information loss in releasing B rather that A is as low as possible subject to the restriction that the risk of disclosing confidential data is at an acceptable level. Under rounding and perturbation, each b_{ij} will be an integer close to a_{ij}. Under suppression, some cells in B will not be released, while those released will be unchanged. For an extensive discussion of these three disclosure avoidance techniques and policy issues in releasing masked tables, see Cox, et al [6].

In this report we present new techniques developed by the Census Bureau Confidentiality Staff for unbiased controlled rounding and unbiased controlled perturbation. In addition we

present new methods to audit protection under a cell suppression methodology. Computer code has been developed to implement each of these procedures (running on the Sperry mainframe and on an IBM/AT under Ryan-McFarland Fortran), and these programs have been successfully tested using data from the 1980 Decennial Censuses. We begin by developing the mathematical structure which served as a unifying framework in the design of the methodologies presented here.

2. UNIFYING MATHEMATICAL STRUCTURE

2.1 Circuits in a table

Let A be an arbitrary additive table as defined earlier. A path of length n is a sequence of distinct table cells;

$$Q = \{(i_1,j_1),(i_2,j_2),\ldots,(i_n,j_n)\},$$

such that:

(1) $a_{i_k j_k} \neq 0$, $k=1,\ldots,n$

(2) any two consecutive cells are in the same row or column, but

(3) no three cells are in the same row or column.

A circuit of length n is a path of length n such that

(4) if a row or column has at least one cell in Q it has exactly two.

For each circuit cell define a signature, $\sigma_{i_k j_k} = (-1)^{k+1}$, and note that the sum of signatures along any row or column equals zero. As we show below, one can add or subtract an integer from each cell in a circuit, yet maintain table additivity. The range of values by which we can alter each cell

while maintaining non-negative values is called the circuit flow. Under a rounding or perturbation strategy one masks a positive cell by embedding the target cell in a circuit and adding or subtracting around the circuit within the limits of the flow. Under a cell suppression strategy, a necessary condition for a table to be disclosure protected is that every disclosure cell is contained in a circuit of suppressed cells; a sufficient condition is that the collection of such containing circuits allows sufficient flow to adequately mask each disclosure cell. Rounding and perturbation methods are discussed in Section 3 and suppression is discussed in Section 4. We continue this section with a description of procedures for altering cell values along a circuit.

Let **C** be a circuit and let α be an arbitrary integer. For each $(i,j) \in \mathbf{C}$, let

$$\tau_{ij} = \begin{cases} \sigma_{ij} & \text{for } 1 \le i \le m, \quad 1 \le j \le n \\ -\sigma_{ij} & \text{for } i=0 \text{ and } 1 \le j \le n \text{ or } 1 \le i \le m \text{ and } j=0 \\ \sigma_{ij} & \text{for } i=0 \text{ and } j=0 \end{cases}$$

and let $\tau_{ij} = 0$ for $(i,j) \notin \mathbf{C}$.

The array B, where $b_{ij} = a_{ij} + \alpha\tau_{ij}$ for $(0 \le i \le m, \ 0 \le j \le n)$ is additive and differs from A only for those cells in **C**. By selecting α in the range of the flow of **C**, each entry in B will be non-negative; hence B will be an additive table. If, in addition, α is chosen to provide sufficient disclosure protection, then we say that B is an additive masked table for A. If **C** consists only of interior cells of A, B will have the same marginal values as A, and if the expected value of each α (through our selection probabilities) equals zero, then each cell in B will be an unbiased estimate of the corresponding cell in A. We seek additive unbiased rounding and perturbation procedures and show how the perturbation procedure can be restricted to change the fewest cells possible.

Example 1: Let Table 1 be our initial table, and let us focus on cell (1,1).

54	13	9	21	11
12	3	7	0	2
18	4	0	8	6
11	0	2	9	0
13	6	0	4	3

Table 1

Two circuits containing cell (1,1) are:

$C_1 = \{(1,1),(1,2),(3,2),(3,3)(4,3),(4,1)\}$

$C_2 = \{(1,1),(2,1),(2,0),(1,0)\}$.

For C_1 we have flow $F_1 = [-2,6]$ and for C_2 we have flow $F_2 = [-3,4]$. If we form the masked table using C_1 and $\alpha = -1\varepsilon F_1$ we get Table 2, and using C_2 with $\alpha = 3\varepsilon F_2$ we get Table 3.

54	13	9	21	11
12	2	8	0	2
18	4	0	8	6
11	0	1	10	0
13	7	0	3	3

Table 2

54	13	9	21	11
15	6	7	0	2
15	1	0	8	6
11	0	2	9	0
13	6	0	4	3

Table 3

Although it might be a desirable objective to form circuits consisting only of internal cells so that a masked table would retain the marginal values of the original, this is not always possible, e.g., in Table 4 there is no circuit containing cell (1,4) consisting of interior cells.

42	5	11	10	16
12	3	5	0	4
8	2	6	0	0
7	0	0	3	4
15	0	0	7	8

Table 4

It is important to note, however, that every non-zero table cell is contained in at least one circuit (which may include marginal positions).

A revised (and feasible) objective is to find a circuit consisting entirely of interior cells when such a circuit exists and to include marginals in a circuit only when necessary. To do this, we define a length for each non-zero cell of a table and define the length of a circuit to be the sum of the lengths of cells it contains. Each positive internal cell is initialized at length one, row and column marginal cells are initialized at a large length M, and the grand total cell is initialized at length N>>M. Given an arbitrary positive internal cell (i,j) by forming a circuit of minimal length containing cell (i,j) we will obtain circuits consisting only of internal cells if any exist, and include as few marginal cells as feasible when they are needed. In using minimal length circuits to alter table values, we may increase the length of a cell once it has been perturbed to minimize the possibility of multiple changes to a single cell.

2.2 Graph Theoretic Framework For Two-Dimensional Tables and Cycles

An arbitrary table, A, can be represented by an undirected bipartite graph, G, in which the edges correspond to positive cells and nodes correspond to rows or columns. That is, let G be the bipartite graph whose node sets are:

$$N_R = r_1, r_2, \ldots, r_m, c_0 \qquad N_C = c_1, c_2, \ldots, c_n, r_0 ,$$

and having the edge (r_i, c_j) if and only if cell $a_{ij} \neq 0$ $(0 \leq i \leq m,\ 0 \leq j \leq m)$. The graph representing Table 1 is shown in Figure 1.

In an arbitrary directed graph, an elementary path of length m is a sequence of arcs

$$P = e_1, e_2, \ldots, e_m \quad \text{with}$$
$$e_1 = (n_0, n_1),\ e_2 = (n_1, n_2), \ldots,\ e_m = (n_{m-1}, n_m)$$

such that each node is reached at most once when traversing P. An elementary circuit is an elementary path such that $n_0 = n_m$. We omit repeating the term "elementary" in discussing paths and circuits with the understanding that all paths and circuits discussed here will be elementary. If the graph is not directed, we replace the term "arc" by "edge" and the definitions above still prevail. Our reference for graph and network theoretic information is Gondran and Minoux [8] and we conform to the terminology therein.

If G is the bipartite graph representing table A, there is a one-to-one, onto correspondence between circuits in A and circuits in G. For example, the circuit C_1 in Table 1 is shown by the darkend edges in Figure 1.

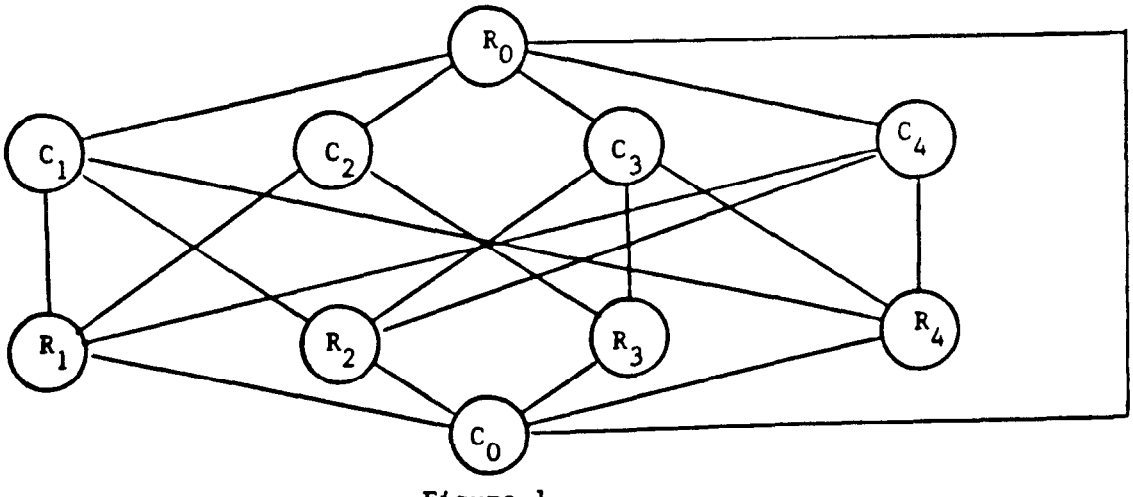

Figure 1.

By assigning a length to each edge in an (undirected) graph, one defines the length of a path to be the sum of the lengths of edges it contains. If two arbitrary nodes are connected by at least one path, there exists a path of minimal length connecting them. Every edge in G is contained in a circuit, and after removing an arbitrary edge from G, its end points are connected by at least one path of minimal length. Thus, to find a minimal length circuit containing an arbitrary edge (x,y): (1) remove edge (x,y) from G, (2) find a minimal length path between nodes x and y, and (3) adjoin the edge (x,y) forming a minimal length circuit.

In Section 2.1, we made use of table circuits (of minimal length) to alter cell values; by expressing table circuits in terms of circuits within graphs, we can exploit graph-theoretic methods to find them.

3. UNBIASED CONTROLLED ROUNDING AND UNBIASED CONTROLLED PERTURBATION

3.1 Controlled Rounding

Let A be an additive table and let b be a positive integer. A table, B, is called a rounding of A to base b if: (1) $b_{ij} = b[a_{ij}/b]$ or $b([a_{ij}/b]+1)$ (where [x] denotes the integer part of x).

Rounding techniques traditionally have treated each cell independently (including marginals) and round values up or down based on some random process, see Nargundkar and Saveland [11]. (Fellegi, [7], rounded cells additively, but his method is applicable only to one-way tables.) Accordingly, rounded two-way tables may fail to be additive.

If, in addition, we have that: (2) B is additive, we say B is a controlled rounding of A, see Cox and Ernst [5]. If, furthermore: (3) $E(b_{ij}) = a_{ij}$, we say the controlled rounding is unbiased. A simple unbaised controlled rounding procedure has been developed by Cox [3] based on circuits in a table; and we

report on some of this work below. If (1)-(3) hold, it follows that $|b_{ij} - a_{ij}| < b$, so if a_{ij} is a multiple of b, then $b_{ij} = a_{ij}$.

Starting with a table A, and a base b, one employs circuits in A to create a masked table which is a controlled rounding. One crucial observation is that each unrounded cell is contained in a circuit consisting exclusively of unrounded cells (Cox, [3]. Thus, if all the marginal values of A are multiples of b, we can confine our attention to circuits and adjustments of interior cells. If some marginals are not multiples of b, they too will be adjusted. The procedure is as follows. If at least one cell in A is not rounded, form a circuit, **C**, consisting of unrounded cells. For each cell in **C** let

$$s_{ij} = \begin{cases} a_{ij} - b[a_{ij}/b] & \text{for } \tau_{ij} = -1 \\ b[a_{ij}/b] + b - a_{ij} & \text{for } \tau_{ij} = 1, \end{cases}$$

$$t_{ij} = \begin{cases} b[a_{ij}/b] + b - a_{ij} & \text{for } \tau_{ij} = -1 \\ a_{ij} - b[a_{ij}/b] & \text{for } \tau_{ij} = 1. \end{cases}$$

Letting

$$s = \min_{(i,j)\in \mathbf{C}} \{s_{ij}\} \quad \text{and} \quad t = \max_{(i,j)\in \mathbf{C}} \{-t_{ij}\},$$

and setting α equal to <u>either</u> s <u>or</u> t we have

$$b[a_{ij}/b] < a_{ij} + \alpha\tau_{ij} < b[a_{ij}/b] + b$$

for each $(i,j) \in \mathbf{C}$. For at least one $(i,j) \in C$, $a_{ij} + \alpha\ \tau_{ij}$ will be a multiple of b. Thus, one selects a value for α (either s or t) adds or subtracts α from each circuit cell as appropriate, and obtains a revised table having at least one more multiple of b than did A. If every element of the revised table, B, is a multiple of b, then B is a controlled rounding of A. If not, repeat this procedure, noting eventually it will terminate yielding a controlled rounding of A. Selecting

$$\alpha = \begin{cases} s \text{ with probability } \dfrac{-t}{s-t} \\ t \text{ with probability } \dfrac{s}{s-t}, \end{cases}$$

then $E(\alpha)=0$ and the controlled rounding will be unbiased (Cox, [3]).

3.2 Unbiased Controlled Perturbation

Given a table A, by a random perturbation of A, one usually means a masked table each of whose entries differ from A by a small randomly selected value, see Newman [12]. One forms a random perturbation of A by selecting a positive integer called the perturbation base, k, and a family of probabilities, $P=\{p_\alpha \mid \alpha\varepsilon[-k,k]\}$, (where $[-k,k]$ is the set of integers between $-k$ and k, inclusive) such that:

$$(1) \quad \sum_{\alpha\varepsilon[-k,k]} p_\alpha = 1 \qquad (2) \quad \sum_{\alpha\varepsilon[-k,k]} \alpha p_\alpha = 0.$$

(Although a symetric interval, $[-k,k]$, is often chosen, any interval satisifying (1) and (2) will suffice.)

For each interior cell one randomly selects a value α according to the distribution P, and lets

$$b_{ij} = \begin{cases} a_{ij} + \alpha & \text{if } a_{ij}+\alpha > 0 \\ 0 & \text{otherwise.} \end{cases}$$

One may sum interior cells to obtain marginal values, $b_{i,0}$ and $b_{0,j}$ for $0<i<m$ and $0<j<m$ to form the masked table B. Note that B is additive, but neither interior nor marginal cells of B are unbiased estimates of their counterparts in A, and the marginal values can differ from their counterparts in A by a value exceeding k. We use a different procedure below which achieves additivity and unbiasedness within a cell perturbation framework.

We first show how random perturbation can be made unbiased. Start with an additive table A, a perturbation base k, and distribution P as before. To perturb cell (i,j) we let $h = \min[a_{ij},k]$ and choose the value to be added to a_{ij} from the interval $[-h,h]$. Let the probability of selecting $\alpha\varepsilon[-h,h]$ be $q_\alpha(h)$ which satisfies:

$$\sum_{\alpha\varepsilon[-h,h]} q_\alpha(h) = 1 \qquad \sum_{\alpha\varepsilon[-h,h]} \alpha q_\alpha(h) = 0.$$

For example, one can let

$$q_\alpha(h) = p_\alpha/\beta \text{ where } \beta = \sum_{\alpha\varepsilon[-h,h]} p_\alpha .$$

After selecting α , form $b_{ij} = a_{ij} + \alpha$ for each interior cell. Zero values are not perturbed and each b_{ij} is an unbiased estimate of the corresponding a_{ij} (including marginals). Note, as in the biased procedure above, revised marginals can differ from their counterparts by a value greater than k.

If the masked table B is released to the public and cell (i,j) is observed to be b_{ij}, different inferences can be drawn about the corresponding value a_{ij} under these two perturbation strategies. Under the usual (biased) procedure one can say that:

$$\text{Max}\{0, b_{ij}-k\} \leq a_{ij} \leq b_{ij}+k ,$$

whereas under the unbiased procedure one has that:

$$\text{Max}\{[(b_{ij}+1)/2],\ b_{ij}-k\} \leq a_{ij} \leq b_{ij}+k.$$

Note that for $a_{ij} > k$ the two procedures perform the same.

Our next objective is to maintain table additivity and alter marginals as infrequently as feasible. To this end we introduce the notion of <u>controlled perturbation</u>. Start with a table A, a perturbation base k, and a distribution $P=\{p_\alpha \mid \alpha\varepsilon[-k,k]\}$. To perturb a_{ij} we form a circuit containing cell (i,j) and let F denote the circuit flow. We select $\alpha\varepsilon F$ $[-k,k]$ by any specified random process such that $E(\alpha) = 0$ and add or subtract α from each cell in the circuit as discussed above for rounding (Greenberg, [9]).

3.3 Restrictive Controlled Perturbation

In tables of frequency counts, cells containing large values do not pose a direct disclosure risk. It will suffice to perturb cells with small values, the disclosure cells, and such cells will be called primary perturbation cells. It will often be necessary to perturb cells other than primary perturbation cells to ensure table additivity -- and such cells will be referred to as complementary perturbations. We can implement an unbiased restricted controlled perturbation using the framework established above.

One begins by assigning length one to all primary perturbation cells, length two to all other positive interior cells, and length M and N to marginal cells as earlier. Given a table A with at least one primary perturbation cell; (1) form a circuit of minimal length containing that cell, (2) choose the value to be added or subtracted from each cell in the circuit by some unbiased random process, (3) form the revised table, and (4) update cell lengths. If no cell has length one in the revised table, we are done. If any cell has length one, repeat the process as often as necessary, noting that this process will terminate.

4. SUPPRESSION METHODS

4.1 Introduction

A primary suppression set for a table, A, is a set of cells, P, whose values will be suppressed when A is released. Because of linear relations along rows and columns of a table, one can always find the range of a suppressed cell (see Cox, [2]). To prevent disclosure of sensitive information, data releasing agencies must ensure that a suppressed cell cannot be estimated too closely. For any suppressed cell its level of protection is related to the circuits consisting of suppresssed cells to which it belongs. In fact, a suppressed cell can be estimated exactly

if and only if it exists in no circuit consisting of suppressed cells.

72	16	11	15	30
20	5*	6*	0	9*
13	2*	3*	2	6
15	3	0	4*	8*
24	6	2	9*	7*

Table 5

72	16	11	15	30
20	5*	6*	0	9*
13	2*	3*	2	6*
15	3	0	4	8*
24	6	2	9*	7*

Table 6

If Table 5 were released with starred cells suppressed, one could determine that the value in cell (1,4) must be 9. Note that cell (1,4) is contained in no circuit consisting of suppressed cells.

On the other hand, consider Table 6 in which starred cells are to be suppressed. Forming the circuit (1,4), (2,4), (2,1), (1,1) we can add 5 units to cell (1,4) obtaining the Table 7 and subtract 2 units from cell (1,4) obtaining Table 8.

72	16	11	15	30
20	0	6	0	14
13	7	3	2	1
15	3	0	4	8
24	6	2	9	7

Table 7

72	16	11	15	30
20	7	6	0	7
13	0	3	2	8
15	3	0	4	8
24	6	2	9	7

Table 8

Forming the circuit (1,4), (2,4) (2,2), (1,2) we can add 1 unit to cell (1,4) in Table 7 and subtract 3 units from cell (1,4) in Table 8 yielding, respectively, Tables 9a and 9b.

72	16	11	15	30
20	0	5	0	15
13	9	4	2	0
15	3	0	4	8
24	6	2	9	7

Table 9a

72	16	11	15	30
20	7	9	0	4
13	0	0	2	11
15	3	0	4	8
24	6	2	9	7

Table 9b

We can no longer form circuits of suppressed cells to either add or subtract from cell (1,4). Thus if table 9c were released, one can only say that cell (1,4) lies in the interval [4,15].

72	16	11	15	30
20	D	D	0	D
13	D	D	2	D
15	3	0	D	D
24	6	2	D	D

Table 9c

67	15	16	24	12
9	1*	2*	2	4
31	5*	6	17*	3*
27	9	8*	5*	5*

Table 10

4.2 Auditing Protection Using Flows In A Network

Based on the pattern of suppressions and released cell values, one can find the interval $[m_{pq}, M_{pq}]$ containing the true value of suppressed cell (p,q) by solving a family of linear equations, Cox, [2]. In this section we show how an agency releasing data can derive this interval, and thereby *audit protection*, by employing the concept of a *capacitated network flow*. Afterwards we couch the process in terms of circuits in a table thus coming full cycle in our analysis of this problem in terms of circuits.

Given a table A and primary suppression set, one constructs the following capacitated network. The underlying graph has the same bipartite structure as G defined earlier, however arcs correspond to suppressed cells. For each *suppressed cell*, (i,j), there are two *directed* arcs; (r_i,c_j) and (c_j,r_i). Thus, for Table 10 whose starred cells correspond to suppressed positions, the associated network is shown in Figure 2. To find the amount by which we can *increase* the value in an arbitrary suppressed cell (p,q) we form the capacitated network where: (1) the capacity in arc (r_i,c_j) equals a_{ij}, (2) the capacity in arc (c_j,r_i) is infinite, (3) the arcs (r_p,c_q) and (c_q,r_p) are deleted and (4) a *source*, s, is added along with arc (s,r_p) of infinite capacity and a sink, t, and arc (c_q,t) with infinite capacity. The value M_{pq} equals a_{pq} *plus* the maximum flow from *s* to *t*, i.e.,

the maximum we can increase a_{pq} without disturbing the relationship between the sum of interior cells and marginals. To find $M_{2,3}$ for Table 10, we use the capacitated network in Figure 2, where finite capacities are shown on each arc. The maximum flow is equal to 5 units, and the flow along each arc is indicated alongside the arcs in Figure 3. Thus $M_{2,3} = 22$.

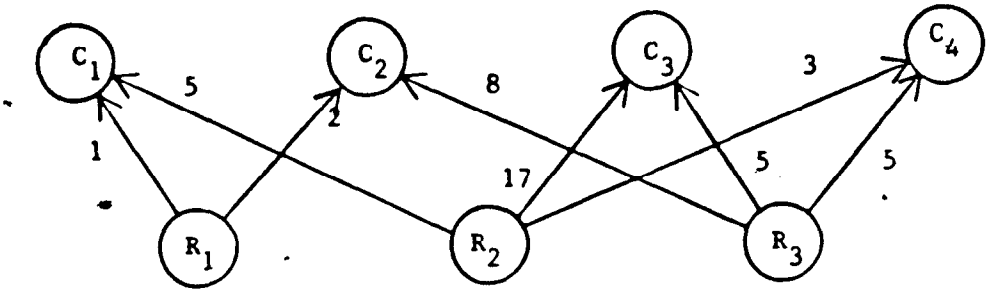

Figure 2. Capacities are alongside arcs. Each arc has a counterpart in the reverse direction with infinite capacity (not drawn).

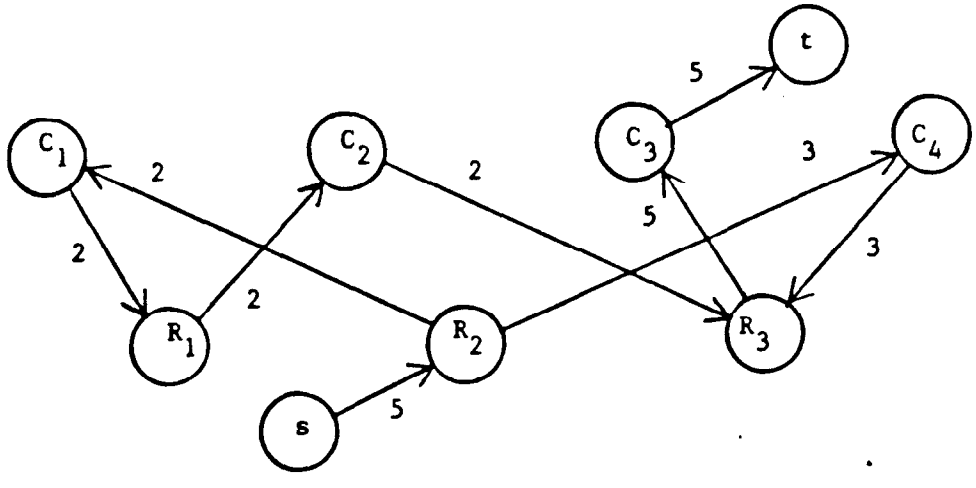

Figure 3. The flow is shown along each arc.

To find the value m_{pq}, we alter the network above so that (1) arcs (s,r_p) and (c_q,t) have capacity a_{pq}, and (2) arc (r_i,c_j) has infinite capacity and arc (c_j,r_i) has capacity a_{ij}. As before we compute the maximum flow from s to t. The value m_{pq} equals a_{pq} minus the maximal flow on the revised network. The maximum flow along this revised network is 6 units, so $m_{pq} = 17-6 = 11$.

Using networks to obtain M_{pq} one finds a flow from s to t. When one unit moves from s to t it determines a path from r_p to c_q, and (along with the arc (r_p,c_q)) a circuit containing the arc (r_p,c_q). That is, finding M_{pq} as outlined has a direct counterpart when viewing the problem in terms of circuits in a table. Consider Table 11a with circuit as noted which was obtained by moving one unit along path (s,r_2), (r_2,c_4), (c_4,r_3), (r_3,c_3), (c_3,t). Add or subtract the value 3 as appropriate yielding Table 11b.

67	15	16	24	12
9	1	2	2	4
31	5	6	17^+	3^-
27	9	8	5^-	5^+

Table 11a

67	15	16	24	12
9	1^+	2	2	4
31	5^-	6	20^+	0
27	9	8^+	2^-	8

Table 11b

Adding and subtracting 2 units from the circuit in Table 11b obtained by moving one unit along path (s,r_2), (r_2,c_1), (c_1,r_1), (r_1,c_2), (c_2,r_3), (r_3,c_3), (c_3,t), yields Table 12.

67	15	16	24	12
9	3	0	2	4
31	3	6	22	0
27	9	10	0	8

Table 12

67	15	16	24	12
9	0	3	2	4
31	6	6	11	8
27	9	7	11	0

Table 13

By circuiting only on suppressed cells in Table 12, one cannot add any more to cell (2,3), and as above, we see that $M_{2,3} = 22$. Similar considerations show $m_{2,3} = 11$ if each circuit containing cell (2,3) in the revised network is used to subtract

from the (2,3) position. The final table one would obtain is shown in Table 13.

4.3 Complementary Suppressions

If cell (p,q) is a primary suppression, and the interval $[m_{pq}, M_{pq}]$ is not sufficiently large to provide adequate protection, other table cells must be suppressed; called complementary suppressions. Complementary perturbation cells for restricted controlled perturbation and complementary suppression cells play a similar role. A complementary perturbation is introduced in order to complete a circuit containing a primary perturbation cell. One introduces complementary suppressions when the flow through a primary suppression cell is too little to offer adequate protection. In essence, new circuits are created through the introduction of complementary suppression cells, and these new circuits allow a greater flow through the primary suppressions. This is related to the network flow analysis by observing that each complementary suppression introduces a new pair of arcs in the underlying network, allowing for a greater flow from source to sink.

Methods for introducing complementary cells differ for controlled perturbation and cell suppression. Under controlled perturbation, the process is local to the extent that for each primary perturbation, complementary perturbations are introduced as needed. Their number is controlled by forcing a minimal length circuit. In contrast, when finding complementary suppression cells the process is global to the extent that generally one seeks a minimal set of complementary suppressions to protect all primary cells.

It is beyond the scope of this paper to present techniques for finding complementary suppressions. Techniques using a combination of linear analysis and branch-and-bound techniques have been developed by Cox [2] and have been successfully employed at the Census Bureau for the 1977 and 1982 Economic Censuses. Recent, promising results of Gusfield [10] couch the

search for complementary suppressions as a graph augmentation problem. Gusfield's results extend some of Cox's methods in that a comprehensive approach is offered to problems such as that illustrated in Table 5.

5. TABLES IN THREE DIMENSIONS

The procedures for forming and analizing masked two-dimensional tables fail in three-dimensions basically because three-dimensional tables lack the underlying graph and associated network structure (Cox, [4]). We can define circuits in a three-dimensional table traversing only positive cells, and in fact show that each positive cell is contained in such a circuit. To that extent, (restricted) controlled perturbations do exist and can be found. However, in the absence of the underlying graph one does not have an efficient procedure for finding requisite minimal length circuits for perturbing non-zero cells. For controlled rounding, the situation is worse. The crucial result for two-dimensional tables is that every unrounded cell is contained in a circuit consisting of unrounded cells. This result is not true in three-dimensional tables, and indeed an unbiased, controlled rounding of an arbitrary three-dimensional table does not always exist (Causey, Cox, Ernst, [1]).

A common thread running through this paper focuses on the role of circuits in creating masked tables. In two dimensions, circuits, along with the underlying graph and network structures are available and are used to full advantage. In three dimensions, required circuits do not exist for unbiased controlled rounding nor are they readily accessible for controlled perturbation due to the absence of the underlying graph structure.

An optimal strategy for masking three-dimensional tables may be to (1) design efficient and effective three-dimension heuristic counterparts to the two-dimension exact procedures, or (2) resolve each two-dimension cross-section and integrate the masked two-dimension faces to form a three-dimensional masked

table. The Census Bureau Confidentiality Staff is actively pursuing research into rigorous techniques for masking three-dimensional tables and also into the area of intertable consistency.

REFERENCES

[1] Causey, Beverley, Cox, Lawrence, and Ernst, Lawrence (1985), "Applications of Transportation Theory to Statistical Problems," Journal of the American Statistical Association, **80**, 392, 903-909.

[2] Cox, Lawrence H. (1980), "Suppression Methodology and Statistical Disclosure Control," Journal of the American Statisical Association, **75**, 377-385.

[3] _____ (1985) "A Constructive Procedure for Unbiased Controlled Rounding", manuscript dated November 18, 1985, submitted to Journal of the American Statistical Association.

[4] _____ (1986), "Rounding and Perturbing Frequency Counts in 3-dimensional Tables", unpublished manuscript dated April 17, 1986.

[5] _____ and Ernst, Lawrence R. (1982), "Controlled Rounding," INFOR, **20**, 423-432. Reprinted in Some Recent Advances in the Theory, Computation and Application of Network Flow Models, University of Toronto Press, 1983, 139-148.

[6] _____, McDonald, Sarah-Kathryn, Nelson, Dawn, (1986), "Confidentiality Issues at the U.S. Bureau of the Census," Journal of Official Statistics, 2,2, to appear.

[7] Fellegi, Ivan P. (1975), "Controlled Random Rounding," Survey Methodology, **1**, Statistics Canada, 123-135.

[8] Gondran, Michel and Minoux, Michel (1984), Graphs and Algorithms, John Wiley and Sons, New York.

[9] Greenberg, Brian (1986), "An Additive, Unbiased Procedure for Random Perturbation," Unpublished Manuscript.

[10] Gusfield, Dan (1984), "A Graph Theoretic Approach to Statistical Data Security," Department of Computer Science, Yale University, New Haven (31 pp. + figures).

[11] Nargundkar, M.S. and Saveland, W. (1972), "Random Rounding to Prevent Statistical Disclosures, "American Statistical Association -- Proceedings of the Social Statistics Section, Washington, DC, 382-385.

[12] Newman, Dennis (1975), "Techniques for Ensuring the Confidentiality of Census Information in Great Britain, "Proceedings of the 40th Session of the International Statistical Institute, Warsaw.