ERROR DETECTING AND ERROR CORRECTING INDEXING SYSTEMS FOR LARGE SERIAL NUMBERS

R. C. Bose

University of North Carolina

Institute of Statistics Mimeo Series No. 426

March 1965

Given a set of serial numbers, they can be indexed by adding a few redundant digits, calculated according to certain rules. If X is the original serial number, then the number X* obtained after addition of the redundant digits is called the index of X. The purpose of adding the redundant digits is to enable us either to detect, or to both detect and correct the more common errors which are liable to be made in reporting X*. The most common errors are (i) wrongly reporting a single digit (error of type I) (ii) reversal of two adjacent digits (error of type II).

In this paper we describe two indexing systems. In the first system X is a six digit number, the digits being 0,1,2,3,4,5,6,7,8,9. Each X is indexed by adding no more than one redundant digit. Errors of type I and II in reporting X* can be detected. In the second system we consider an octal numerical system based on the 8 digits 0,1, ..., 7. We take X to be a seven digit serial number. It is indexed by adding three suitably chosen redundant digits. Errors of type I and II in reporting X* can be corrected.

This research was supported by the National Science Foundation Grant No. GP-3792 and the Air Force Office of Scientific Research Grant No. AF-AFOSR-760-65.

DEPARTMENT OF STATISTICS

UNIVERSITY OF NORTH CAROLINA

Chapel Hill, N. C.

Error Detecting and Error Correcting Indexing Systems
for Large Serial Numbers

R. C. Bose
University of North Carolina

1. Introduction. Given a set of serial numbers, they can be indexed by adding a few redundant digits, calculated according to certain rules. If X is the original serial number, then the number X* obtained after addition of the redundant digits is called the index of X. The purpose of adding the redundant digits is to enable us either to detect, or to both detect and correct the more common errors which are liable to be made in reporting X*.

An indexing system is useful when information has to be stored with respect to a large number of individuals. For example consider a firm with a large clientele. An index number is to be assigned to each client which he quotes in any correspondence. The clients'card on which the information relative to him is stored is identified by means of this index number. Now the index number may be wrongly reported. The most common errors are (i) wrongly quoting a single digit (error of Type I), (ii) reversal of two adjacent digits (error of type II). It would be useful to have an indexing system which is self-checking with respect to common errors, i.e., when a client makes such an error it is readily detected. Also it would be desirable to have a system in which when an error has been detected, it should be possible to correct it.

This research was supported by the National Science Foundation Grant No. GP-3792 and the Air Force Office of Scientific Research Grant No. AF-AFOSR-760-65.

In this paper we describe two indexing systems. In the first system X is a six digit number, the digits being 0,1,2,3,4,5,6,7,8,9. Each X is indexed by adding no more than one redundant digit. Errors of type I and II in reporting X* can be detected. In the second system we consider an octal numerical system based on the 8 digits 0, 1, ..., 7. We take X to be a seven digit serial number. It is indexed by adding three suitably chosen redundant digits. Errors of type I and II in reporting X* can be corrected.

2. <u>An Error Detecting Indexing System.</u> Suppose we have a set of six digit serial numbers (the digits being 0,1,2,3,4,5,6,7,8,9). We give here an example of an indexing system employing no more than one redundant digit and able to detect either a single error in one digit (including the redundant digit) or an error consisting of a reversal of two consecutive digits. These two types of error we shall call errors of type I and type II respectively. We shall also show that if an error has been detected then (under the assumption that the error belongs to one of the two types described above) there are only a few possibilities among which the true serial numbers must lie.

Let the serial number to be indexed be

$$X = x_6x_5x_4x_3x_2x_1 ,$$

where x_i $(i = 1,2,3,4,5,6)$ is one of the digits 0,1, ..., 9. Then there is a unique integer x_o such that

$$(2.1) \qquad x_o \equiv -(x_2+x_4+x_6) + (x_1+x_3+x_5) \pmod{11}, \; 0 \leq x_o \leq 10 .$$

If $0 \leq x_o \leq 9$, we say that the index of X is the seven digit number

$$X^* = x_6x_5x_4x_3x_2x_2x_0 .$$

If $x_0 = 10$, we say that the index of X is the same as X. Thus in this case $X^* = X$, is a six digit number.

For example if $X = 487293$, then $x_o \equiv -7 \pmod{11}$. Hence $x_o = 4$ and X is indexed by $X^* = 4872934$. Again if $X = 487299$, $x_o \equiv -1 \pmod{11}$. Hence $x_o = 10$, and X is indexed by $X^* = 487299$.

Now suppose we allow the possibility that in reporting X^* one of the two types of errorsconsidered has been committed. Let the reported value be $Y^* = y_6 y_5 y_4 y_3 y_2 y_1 y_0$ in case I, when X^* is a seven digit number and $Y^* = y_6 y_5 y_4 y_3 y_2 y_1$ in case II, when X^* is a six digit number. We can find a unique integer c such that $0 \leq c \leq 10$, and

$$(2.2) \qquad c \equiv -(y_0 + y_2 + y_4 + y_6) + (y_1 + y_3 + y_5), \pmod{11},$$

where y_0 is taken to be 10 in case II. Then if the index has been correctly reported $c = 0$.

Suppose now that in reporting X^* an error of type I has been committed, then there is an integer α such that $y_\alpha \neq x_\alpha$ whereas $y_i = x_i$ for $i \neq \alpha$. In case I, $0 \leq \alpha \leq 6$ and $i = 0,1,2, \ldots, 6$. In case II, $1 \leq \alpha \leq 6$ and $i = 1,2, \ldots, 6$. Then from (2.1) and (2.2)

$$(2.3) \qquad c \equiv (-1)^{\alpha+1} (y_\alpha - x_\alpha), \qquad \pmod{11}.$$

Since $0 < |y_\alpha - x_\alpha| \leq 9$, $c \neq 0$. Thus we will conclude that an error has been committed.

Again suppose that in reporting X^* an error of the second type has been committed. Then there is an integer α such that $y_\alpha = x_{\alpha+1}$, $y_{\alpha+1} = x_\alpha$ whereas $y_i = x_i$ for $i \neq \alpha$ or $\alpha + 1$. Also we can assume that $x_\alpha \neq x_{\alpha+1}$, otherwise there would be no error. In case I, $0 \leq \alpha \leq 5$ and

$i = 0,1,2, \ldots, 6$. In case II, $1 \leq \alpha \leq 5$ and $i = 0,1,2, \ldots, 6$. Then from (2.1) and (2.2) we have,

$$c \equiv (-1)^{\alpha+1} 2(x_{\alpha+1} - x_{\alpha}) , \quad (\text{mod } 11). \tag{2.4}$$

Since $0 < |x_{\alpha+1} - x_{\alpha}| \leq 9$, $c \neq 0$. Thus we will conclude that an error has been committed.

Notice that the non-vanishing of the check number c does not tell us what error has been committed.

Example 1. For the purpose of illustration let us suppose that X = 487293 then X* = 4872934. In reporting X* suppose 8 is wrongly reported as 5. Then Y* = 4572934. Now $0 \leqq c \leqq 10$

$$c \equiv -(4 + 9 + 7 + 4) + (3 + 2 + 5), \quad (\text{mod } 11).$$

Hence c = 8, and we conclude that an error has been committed. We can reconstruct the possible values of X* from Y* using the fact that c = 8, and assuming that the error is of type I or type II, i.e., there is an error in a single digit or there has been a reversal of two digits. If an error of type I has been committed in reporting x_α then from (2.3),

$$x_\alpha = y_\alpha + (-1)^\alpha c, \quad (\text{mod } 11).$$

We can make the following table:

Table 1

α	y_α	x_α	Correct X*
0	4	1	4 5 7 2 9 3 1
1	3	6	4 5 7 2 9 6 4
2	9	6	4 5 7 2 6 3 4
3	2	5	4 5 7 5 9 3 4
4	7	4	4 5 4 2 9 3 4
5	5	8	4 8 7 2 9 3 4
6	4	1	1 5 7 2 9 3 4

Again it is possible that the discrepancy is due to a reversal of two consecutive digits in reporting X*. In this case we can reconstruct X* by reversing two consecutive digits of Y* and check whether (2.1) holds. We then have

Table 2

Y*	Possible X*	x_o	
4 5 7 2 9 3 4	5 4 7 2 9 3 4	10	x
-	4 7 5 2 9 3 4	5	x
-	4 5 2 7 9 3 4	0	x
-	4 5 7 9 2 3 4	4	
-	4 5 7 2 3 9 4	2	x
-	4 5 7 2 9 4 3	2	x

Since x_o must be the same as the last digit in X* the only possibility is that the correct X* was 4579234. Thus altogether there are 8 different values of X* which could have lead to Y*.

Example 2. Let us next suppose that X = 487299. Then x_o = 10 and X* = 487299. Suppose there is an error of type II reversing the position of the digits 2 and 7. Then X* would be reported as Y* = 482799. Now $0 \leqq c \leqq 10$, and

$$c \equiv -(10+4+2+9)+(8+7+9) \qquad (\text{mod } 11).$$

Hence c = 10, which shows that an error has been committed. As in example 1, we find that any of the following 6 values of X* could have lead to Y*:

X* = 482789; 482899; 481799; 492799; 382799; 487299.

3. <u>An error correcting indexing system.</u>

Consider an octal system in which we use only the eight digits 0,1,2,3,4,5,6,7. A seven digit number in this system will take care of $8^7 = 2^{21}$ = 2,097,152 or roughly two million serial numbers. We shall now show that by adding 3 redundant digits, suitably calculated, one can correct any single error in a digit (error of type I), or any error consisting of a reversal of two adjacent digits (error of type II).

Consider the Galois field GF_8 of 8 elements. Let the elements be $\alpha_0 = 0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7$. It is well known [1,2,] that the non-zero elements can be written in two different forms:

(i) The multiplicative form where each non-zero element is a power of a primitive element t.

(ii) The additive form where each element is a quadric polynomial of t with coefficients 0 or 1.

Element		Multiplicative form		Additive form
α_0	=	0	=	0
α_1	=	t^0	=	1
α_2	=	t	=	t
α_3	=	t^2	=	t^2
α_4	=	t^3	=	$1 + t^2$
α_5	=	t^4	=	$1 + t + t^2$
α_6	=	t^5	=	$1 + t$
α_7	=	t^6	=	$t + t^2$

To add two elements we take the additive form and remember that the coefficients of the polynomials add (mod 2). Thus

$$\alpha_3 + \alpha_5 = t^2 + (1+t+t^2) = 1 + t = \alpha_6 .$$

To multiply two elements we take the multiplicative form and remember that $t^7 = t^0$. Thus

$$\alpha_0\alpha_5 = 0.t^4 = 0 = \alpha_0 , \ \alpha_6.\alpha_7 = t^5.t^6 = t^4 = \alpha_5 .$$

We can identify the digits 0,1,2,3,4,5,6,7 with the elements $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7$ respectively.

Then any seven digit octal number X can be made to correspond with a 7-vector $\underline{x}$ with elements from GF_8. Consider the vector

$$\underline{x} = (x_1, x_2, \ldots, x_7),$$

where the x_i's belong to GF_8. Let $\underline{x}$ be encoded to

$$\underline{x}^* = (x_1, x_2, \ldots, x_7, x_8, x_9, x_{10}),$$

so that x_8, x_9, x_{10} are determined by the relation

$$\underline{x}^* H^T = 0 ,$$

where H is the matrix given by

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & t & 0 & 0 \\ t & t^2 & t^3 & t^4 & t^5 & t^6 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & t^2 \end{bmatrix} .$$

This gives the encoding rules:

$$x_8 = t^6(x_1+x_2+x_3+x_4+x_5+x_6+x_7)$$
$$x_9 = t\,x_1 + t^2x_2 + t^3x_3 + t^4x_4 + t^5x_5 + t^6x_6 + t^7x_7$$
$$x_{10} = t^5(x_2 + x_4 + x_6)$$

Then the index of X is X* the octal number corresponding to the vector $\underline{x}^*$. For example to find the index of 2,401,546 we identify it with the vector $(\alpha_2, \alpha_4, \alpha_0, \alpha_1, \alpha_5, \alpha_4, \alpha_6)$ or $(t, t^3, 0, 1, t^4, t^3, t^5)$

$$\therefore\ x_8 = t^7 + t^2 + t^6 + t^3 + t^2 + t^4$$
$$= 1 + t^2 + (t+t^2) + (1+t^2) + t^2 + (1+t+t^2)$$
$$= 1+t^2$$
$$= t^3$$
$$= \alpha_4 .$$

$$x_9 = t^2 + t^5 + t^4 + t^2 + t^5$$
$$= t^2 + (1+t) + (1+t+t^2) + t^2 + t^2 + (1+t)$$
$$= 1 + t$$
$$= t^5$$
$$= \alpha_6 .$$
$$x_{10} = t^8 + t^5 + t^8$$
$$= t + (1+t) + t$$
$$= 1 + t$$
$$= t^5$$
$$= \alpha_6 .$$
$$\therefore \quad \underline{x}^* = (\alpha_2,\alpha_4,\alpha_0,\alpha_1,\alpha_5,\alpha_4,\alpha_6,\alpha_4,\alpha_6,\alpha_6) .$$

Hence the index of $X = 2,401,546$ is $X^* = 2,401,546,466$.

We shall now show that if an error of type I or type II is committed in reporting X^* it can be corrected. Let X^* be reported as Y^*. Let $\underline{x}^*$ and $\underline{y}^*$ be the vectors (with elements from GF_8 corresponding to X^* and Y^* respectively. The error vector e^* is defined by

$$\underline{e}^* = \underline{y}^* - x^*.$$

If an error of the type I has been committed, so that the i-th digit of X^* has been wrongly reported, then

$$\underline{e}^* = (0,0, \ldots, y_i-x_i, \ldots 0) .$$

Now $\underline{y}^*$ is known to us. We may therefore calculate the syndrome of $\underline{y}^*$ which may be defined as

$$\underline{y}^* H^T = (\underline{x}^* + e^*)H^T$$
$$= e^* H^T$$
$$= e_i\underline{h}_i ,$$

where $e_i = y_i - x_i$ and $\underline{h}_i$ is the i-th row vector of vector of H^T .

Again suppose there is an error of type II in reporting X* so that the i-th and (i+1)-th digits have been reversed. Then

$$\begin{aligned} e^* &= y^* - x^* \\ &= (0, 0, \ldots, x_{i+1} - x_i, x_i - x_{i+1}, \ldots, 0) . \end{aligned}$$

Since the field GF_8 is of characteristic 2, addition and subtraction are identical. Let

$$z_i = x_{i+1} - x_i = x_i - x_{i+1} .$$

Now the syndrome of y* is

$$\begin{aligned} y^* H^T &= (\underline{x}^* + e^*) H^T \\ &= e^* H^T \\ &= z_i (\underline{h}_i + \underline{h}_{i+1}) , \end{aligned}$$

where $\underline{h}_i$ and $\underline{h}_{i+1}$ are the i-th and (i+1)-th row vectors of H^T.

The sum of i-th and (i+1)-th columns of H is given by the i-th column of S, where

$$S = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & t^5 & t & 0 \\ t^6 & 1 & t & t^2 & t^3 & t^4 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & t^2 \end{bmatrix} .$$

Now it should be noted that the last three columns of S are proportional to

$$\begin{bmatrix} 1 \\ t^2 \\ 0 \end{bmatrix} , \begin{bmatrix} 1 \\ t^6 \\ 0 \end{bmatrix} , \begin{bmatrix} 0 \\ t^5 \\ 1 \end{bmatrix} .$$

It is now easy to check that no two of the 19 columns of H and S are proportional to one another. If there is one error in a digit or a single reversal error, the syndrome of y* will be proportional to just one row of H^T or S^T. In the first case we conclude that there is an error of type I, and in the second case we conclude that there is an error of type II.

The position of the column will give the position of the error and the constant of proportionality will give e_i or z_i as the case may be. The error can then be corrected.

<u>Example</u>. We shall now illustrate the procedure of error correction. We have seen that the index of X = 2,401,546 is X* = 2,401,546,466.

Suppose that there is an error of type I, so that the digit 4 in the 6-th position in X* has been reported as 2, then Y* = 2,401,526,466.

Hence

$$\underline{y}^* = (t, t^3, 0, 1, t^4, t, t^5, t^3, t^5, t^5) ,$$

$$\underline{y}^* H^T = (a_1, a_2, a_3) ,$$

where

$$a_1 = t + t^3 + 0 + 1 + t^4 + t + t^5 + t^4 + 0 + 0 = t^4 ,$$

$$a_2 = t^2 + t^5 + 0 + t^4 + t^2 + 1 + t^5 + 0 + t^5 + 0 = t^3,$$

$$a_3 = 0 + t^3 + 0 + 1 + 0 + t + 0 + 0 + 0 + t^7 = t^4 .$$

Hence the syndrome of $\underline{y}^*$ is

$$(t^4, t^3, t^4) = t^4(1, t^6, 1) .$$

Now $(1, t^6, 1)$ is the 6-th row of H^T. Hence we conclude that there is an error of type I in the 6th digit, and $e_6 = t^4$. Now $y_6^* = t$. Hence

$$x_6^* = y_6^* + e_6 = t + t^4 = t^3 = \alpha_4 .$$

Hence we correct the 6-th digit of Y* to 4, getting back the correct

X* = 2,401,546,466.

Now suppose there has been an error of type II, say the 5-th and 6-th digits of X* have been reversed giving Y* = 2,401,456,466. Then

$$\underline{y}^* = (t, t^3, 0, 1, t^3, t^4, t^5, t^3, t^5, t^5) ,$$

$$\underline{y}^* H^T = (b_1, b_2, b_3) ,$$

where

$$b_1 = t + t^3 + 0 + 1 + t^3 + t^4 + t^5 + t^4 + 0 + 0 = 0 ,$$

$$b_2 = t^2 + t^5 + 0 + t^4 + t + t^3 + t^5 + 0 + t^5 + 0 = t^4$$

$$b_3 = 0 + t^3 + 0 + 1 + 0 + t^4 + 0 + 0 + 0 + t^7 = t .$$

Hence the syndrome of $\underline{y}^*$ is

$$(0, t^4, t) = t(0, t^3, 1) .$$

We note that $0, t^3, 1$ is the 5-th row of S^T . Hence we conclude that there has been an interchange in the 5th and 6th digits. Hence interchanging the 5-th and 6-th digits in Y^* we get back the correct

$$X^* = 2,401,546,466 .$$

The value of z_5 is t. This should agree with $x_5 - x_6$. But

$$x_5 - x_6 = \alpha_5 - \alpha_4 = t^4 - t^3 = t ,$$

which checks.

References

1. R. C. Bose. On the application of the properties of Galois fields to the problem of construction of hyper-Graeco-Latin squares. Sankhya 3 (1938), pp. 323-338.
2. R. D. Carmichael. Introduction to the theory of groups of finite order. New York, Dover publications (1956), Chapter IX .