Some Applications of High-Speed Computers to the Case $n = 2$ of Algebraic Cryptography

By Jack Levine

1. Introduction. In 1929 Hill [1] proposed the use of simultaneous linear congruences as a method of encipherment (see also [2], [3]). If the number, n, of such congruences be 5 or more this results in a cryptographic system of unusual security. In this article it is shown that high-speed computers can be used in the problem of the decipherment of the simplest case $n = 2$. We give first a brief description of the system using this value of n.

The 26 letters of the alphabet are assigned numerical values according to some arrangement of the numbers 0, 1, 2, $\cdots$, 24, 25. For example:

(1.1)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
19	2	21	0	4	7	6	9	17	24	11	15	14	13	12	16	18

R	S	T	U	V	W	X	Y	Z
1	25	20	3	22	5	8	23	10

A 2 x 2 involutory matrix, mod 26, is selected to form the congruences

$$\begin{aligned} C_1 &\equiv aP_1 + bP_2, \\ C_2 &\equiv cP_1 + dP_2 \end{aligned} \quad \text{mod } 26 \tag{1.2}$$

where the matrix is

$$H \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad H^2 \equiv I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \text{mod } 26. \tag{1.3}$$

As an illustration we use

$$H = \begin{bmatrix} 4 & 7 \\ 9 & 22 \end{bmatrix}. \tag{1.4}$$

A given plain-text, say CRYPTOGRAPHY, is then divided into two-letter groups, *CR YP TO GR AP HY*; each pair of letters is selected as the P_1P_2 of (1.2), and C_1, C_2 calculated, using the numerical equivalents of (1.1). Thus, as $C\,R = P_1\,P_2 = 21\ 1$, we find, using (1.4),

$$C_1 \equiv 4(21) + 7(1) \equiv 13 = N,$$

$$C_2 \equiv 9(21) + 22(1) \equiv 3 = U,$$

so *CR* is enciphered by *NU*. The converse is also true, since matrix H is involutory. The complete encipherment becomes

(1.5)

CR	*YP*	*TO*	*GR*	*AP*	*HY*
NU	*VN*	*XB*	*WJ*	*GU*	*LL*

The decipherment, knowing the matrix, is performed in an identical manner.

Received November 30, 1960.

With this in mind we may state the problem to be solved in the following way. Given a cipher-text, obtained as in (1.5), determine the corresponding plain-text, assuming as known the numerical alphabetic values as in (1.1). This amounts to determining the matrix H which is to be considered as the unknown quantity. As mentioned above, this soon becomes a complex problem with increasing n. For the present case of $n = 2$, however, it can be readily solved by machine, an IBM 650 actually being used for this purpose. Two methods will be explained. The first method has the advantage that it can be applied to extremely short messages, the second that it can be extended with some changes to higher values of n.

2. First Method. Since the only unknowns are the four elements a, b, c, d of the matrix H, the most direct method is to test in succession all such possible matrices. If $n > 2$ this becomes impractical, but if $n = 2$ there are only 740 matrices, a relatively small number. It is found from (1.3) that the elements must satisfy the conditions

$$(2.1) \quad a^2 + bc \equiv 1, \quad d^2 + bc \equiv 1, \quad b(a + d) \equiv 0, \quad c(a + d) \equiv 0, \quad \mod 26$$

and it is easily shown that these imply two types of matrices,

$$\text{Type 1.} \quad H = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}$$

$$\text{Type 2.} \quad H = \begin{bmatrix} a & b \\ c & a \end{bmatrix}$$

with

$$(2.2) \qquad a^2 + bc \equiv 1 \mod 26$$

in both cases.
Type 2 contains only 8 matrices.

Type 2

a	b	c	a	b	c
1	0	0	14	13	13
1	0	13	25	0	0
1	13	0	25	0	13
12	13	13	25	13	0

The remaining 732 of Type 1 can be obtained from the basic set listed below. If we place

$$(a, b, c) = \begin{bmatrix} a & b \\ c & -a \end{bmatrix},$$

then associated with (a, b, c) we have the set of eight

(a, b, c)	$(-a, b, c)$
(a, c, b)	$(-a, c, b)$
$(a, -b, -c)$	$(-a, -b, -c)$
$(a, -c, -b)$	$(-a, -c, -b)$

A complete list of Type 1 matrices can be exhibited by writing one of each associated set of 8 (or 4 or 2 in case of duplicates). Such a basic list is given below.

TYPE 1. Basic List.

$a = 0$		$a = 1$		$a = 2$		$a = 3$		$a = 4$		$a = 5$		$a = 6$	
b	c	b	c	b	c	b	c	b	c	b	c	b	c
1	1	0	c*	1	23	1	18	1	11	1	2	1	17
3	9	2	13	5	15	2	9	3	21	2	14	3	23
5	21	4	13	7	7	3	6	7	9	3	18	5	19
7	15	6	13	9	17	4	11			4	7	11	11
		8	13			4	24			4	20		
		10	13			5	14			5	16		
		12	13			6	16			6	9		
		(*c = 0, 1,				7	10			8	10		
		$\cdots$, 13)				8	12			11	12		

$a = 7$		$a = 8$		$a = 9$		$a = 10$		$a = 11$		$a = 12$		$a = 13$	
b	c	b	c	b	c	b	c	b	c	b	c	b	c
1	4	1	15	1	24	1	5	1	10	1	13	1	14
2	2	3	5	2	12	3	19	2	5	3	13	2	7
2	15	7	17	3	8	9	15	2	18	5	13	2	20
3	10			4	6			3	12	7	13	3	22
4	14			4	19			4	9	9	13	4	10
5	6			5	10			4	22	11	13	4	23
6	18			6	17			6	6	13	13	5	8
7	8			8	16			6	19			6	11
9	12			11	14			8	11			8	18
10	16							10	14			9	16
												12	12

Cases $a = 0$ and $a = 13$ could also be considered as Type 2, but it is convenient to place them here. A basic Type 2 list is given below.

TYPE 2. Basic List.

a	b	c
1	0	0
1	0	13
12	13	13

3. Machine Procedure. A card is punched for each of the 109 entries (matrices) in the two basic lists, and each of these in turn generates the remaining seven associated with it, duplicates being retained. Each matrix of a set of 8 then performs a deciphering operation indicated by the matrix congruence $P = HC$, or

$$\begin{bmatrix} P_1 & P_3 & P_5 & P_7 & P_9 \\ P_2 & P_4 & P_6 & P_8 & P_{10} \end{bmatrix} \equiv \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} C_1 & C_3 & C_5 & C_7 & C_9 \\ C_2 & C_4 & C_6 & C_8 & C_{10} \end{bmatrix} \quad \text{mod } 26 \tag{3.1}$$

where the columns of the C-matrix represent the first five pairs of cipher letters, and the columns of the P-matrix the corresponding pairs of the deciphered "plaintext," these last 10 letters being printed. When all 8 matrices of a set have been used, the next matrix of the basic lists is read and the entire process repeated. There results $8 \times 109 = 872$ decipherments, each of 10 letters. An inspection of these easily locates the correct decipherment and the corresponding matrix which is also

printed along with the decipherment. The entire procedure requires approximately 30 minutes.

4. Second Method. We assume here a Type 1 matrix has been used for H, all Type 2 having been tested by hand.

Let a cipher-text be represented as

$$A_1B_1\ A_2B_2 \cdots A_iB_i\ A_{i+1}B_{i+1} \cdots A_mB_m\ . \tag{4.1}$$

Suppose the plain-text contains the 3 consecutive letters $P_jQ_jR_j$ and it is desired to locate their position in (4.1). The 3 letters must be divided as

(a) $$P_jQ_j\ R_j\cdot \quad \text{or}$$

(b) $$\cdot P_j\ Q_jR_j\ .$$

Assume for case (a) the cipher-text location being tested is given by

(4.1a) $$\begin{matrix} A_iB_i\ A_{i+1}B_{i+1} \\ P_jQ_j\ R_j\ . \end{matrix}$$

Then from (1.2), in the case of a Type 1 matrix, we must have

(4.2a) $P_j \equiv aA_i + bB_i\ ,$ (4.2c) $Q_j \equiv cA_i - aB_i\ ,$

(4.2b) $A_i \equiv aP_j + bQ_j$ (4.2d) $B_i \equiv cP_j - aQ_j$

(4.2e) $R_j \equiv aA_{i+1} + bB_{i+1}\ .$

If a, b be eliminated between (4.2a, b, e) we obtain

$$E_{ij} = \begin{vmatrix} A_i & B_i & P_j \\ A_{i+1} & B_{i+1} & R_j \\ P_j & Q_j & A_i \end{vmatrix} \equiv 0 \mod 26 \tag{4.3}$$

In case (b), if the location be at

(4.1b) $$\begin{matrix} A_iB_i\ A_{i+1}B_{i+1} \\ \cdot P_j\ Q_jR_j \end{matrix}$$

we have similarly,

(4.4a) $P_j \equiv cA_i - aB_i\ ,$ (4.4c) $Q_j \equiv aA_{i+1} + bB_{i+1}\ ,$

(4.4b) $B_{i+1} \equiv cQ_j - aR_j\ ,$ (4.4d) $B_{i+1} \equiv cQ_j - aR_j\ ,$

(4.4e) $R_j \equiv cA_{i+1} - aB_{i+1}\ .$

Eliminating a, c from (4.4a, b, e) results in

$$F_{ij} = \begin{vmatrix} A_i & B_i & P_j \\ A_{i+1} & B_{i+1} & R_j \\ Q_j & R_j & B_{i+1} \end{vmatrix} \equiv 0 \mod 26. \tag{4.5}$$

We may state these results in the following theorem:

THEOREM. *A necessary condition that three consecutive letters $P_jQ_jR_j$ occur as*

plain-text in positions (4.1a) *or* (4.1b) *of a cipher text* (4.1) *is that* $E_{ij} \equiv 0$ *or* $F_{ij} \equiv 0$, *mod* 26, *respectively.*

Equations (4.2a, b, c, d) can be solved to give

$$(4.6)\quad Da \equiv \begin{vmatrix} A_i & Q_j \\ P_j & B_i \end{vmatrix}, \quad Db \equiv \begin{vmatrix} P_j & A_i \\ A_i & P_j \end{vmatrix}, \quad Dc \equiv \begin{vmatrix} B_i & Q_j \\ Q_j & B_i \end{vmatrix}, \quad D \equiv \begin{vmatrix} P_j & Q_j \\ A_i & B_i \end{vmatrix}$$

In this case (4.3) is equivalent to

$$(4.7)\qquad D(aA_{i+1} + bB_{i+1} - R_j) \equiv 0, \qquad \text{mod } 26,$$

and in addition,

$$(4.8)\qquad D^2(a^2 + bc - 1) \equiv 0, \qquad \text{mod } 26.$$

It follows that for this case, (a), *if D is prime to* 26, then a, b, c are determined uniquely by (4.6), with $a^2 + bc \equiv 1$, and all equations (4.2) are satisfied for these values of a, b, c.

However, *if D is even, mod* 26, (*but not* 0), several solutions of (4.6) are possible, and it is best to solve them mod 13 and mod 2. If a, b, c is a solution mod 13, then $a + 13k_1$, $b + 13k_2$, $c + 13k_3 (k_i = 0, 1)$ may be solutions mod 26. To pick the correct choices, we must solve (4.6) mod 2. There are just four solutions mod 2 satisfying $a^2 + bc = 1$. These are

$$(4.9)\quad M_1 = (0\ 1\ 1), \qquad M_2 = (1\ 0\ 0), \qquad M_3 = (1\ 0\ 1), \qquad M_4 = (1\ 1\ 0),$$

which when taken in conjunction with the mod 13 solutions of (4.6) will give all mod 26 solutions. In this case, D even, it is necessary to check (4.2e) for each such solution.

If the seven numerical values under consideration in (4.1a) be reduced mod 2,

TABLE 1. (Used when $E_{ij} \equiv 0$, and D is even, mod 26)

	PQR							
ABAB	000	001	010	011	100	101	110	111
0000	1, 2 3, 4	x	x	x	x	x	x	x
0001	2, 3	1, 4	x	x	x	x	x	x
0010	1	2, 3 4	x	x	x	x	x	x
0011	4	1, 2 3	x	x	x	x	x	x
0100	x	x	2, 3	x	1	x	4	x
0101	x	x	2, 3	x	x	1	x	4
0110	x	x	x	2, 3	1	x	x	4
0111	x	x	x	2, 3	x	1	4	x
1000	x	x	1	x	2, 4	x	3	x
1001	x	x	x	1	2	4	3	x
1010	x	x	1	x	x	2, 4	x	3
1011	x	x	x	1	4	2	x	3
1100	x	x	4	x	3	x	1, 2	x
1101	x	x	x	4	3	x	2	1
1110	x	x	x	4	x	3	1	2
1111	x	x	4	x	x	3	x	1, 2

then Table 1 gives the possible M_i solutions of (4.2), and also all contradictions (these resulting from a wrong location of the $P_jQ_jR_j$). The entries 1, 2, 3, 4 represent M_1, M_2, M_3, M_4 of (4.9) respectively, and x represents an impossible setting. From Table 1 we see, e.g., that the (4.1a) setting

$$18 \quad 5 \quad 22 \quad 5$$
$$20 \quad 9 \quad 4 \quad \cdot$$

converted mod 2 is 0101 010, giving M_2 and M_3 as $(a\ b\ c)$ solutions, mod 2, whereas the setting

$$18 \quad 5 \quad 22 \quad 5$$
$$20 \quad 10 \quad 4 \quad \cdot$$

or 0101 001 is impossible mod 2, and therefore mod 26 also.

Return now to the case of a (4.1b) setting. From (4.4abcd) we find

$$D'a \equiv \begin{vmatrix} A_{i+1} & R_j \\ Q_j & A_{i+1} \end{vmatrix}, \qquad D'b \equiv \begin{vmatrix} Q_j & A_{i+1} \\ A_{i+1} & Q_j \end{vmatrix},$$
$$D'c \equiv \begin{vmatrix} B_{i+1} & R_j \\ R_j & B_{i+1} \end{vmatrix}, \qquad D' \equiv \begin{vmatrix} Q_j & R_j \\ A_{i+1} & B_{i+1} \end{vmatrix}. \tag{4.10}$$

If D' is prime to 26, a, b, c are uniquely determined, (4.4e) is satisfied as is $a^2 + bc \equiv 1$. *If D' is even mod* 26 (*not* 0), we solve (4.10) mod 13 and mod 2. In this case, Table 2 gives the various mod 2 solutions for $(a\ b\ c)$. For each $(a\ b\ c)$ mod 26 so obtained, (4.4e) must be tested.

5. Machine Procedure. If an assumed trigraph $P_jQ_jR_j$ is actually present in the plain-text of (4.1) and is tested in its correct position, say $A_iB_i\ A_{i+1}B_{i+1}$, (called

TABLE 2. (Used when $F_{ij} \equiv 0$ and D' is even, mod 26)

	PQR							
ABAB	000	001	010	011	100	101	110	111
0000	1, 2 3, 4	x	x	x	x	x	x	x
0001	x	2, 3	1	4	x	x	x	x
0010	x	1	2, 4	3	x	x	x	x
0011	x	4	3	1, 2	x	x	x	x
0100	1	x	x	x	2, 3, 4	x	x	x
0101	x	x	1	x	x	2, 3	x	4
0110	x	1	x	x	x	x	2, 4	3
0111	x	x	x	1	x	4	3	2
1000	2, 4	x	x	x	1, 3	x	x	x
1001	x	2	x	4	x	3	1	x
1010	x	x	2, 4	x	x	1	x	3
1011	x	4	x	2	x	x	3	1
1100	3	x	x	x	1, 2, 4	x	x	x
1101	x	3	x	x	x	2	1	4
1110	x	x	x	3	x	1	2, 4	x
1111	x	x	3	x	x	4	x	1, 2

a causal setting), one of the $(a\ b\ c)$ solutions obtained as outlined above (assuming D or $D' \not\equiv 0$ or 13) will produce the correct decipherment. If then, a list of the highest frequency (English) trigraphs be tested in every cipher-text position the great majority of the assumed settings will be rejected. Most of those that survive (produce an $(a\ b\ c)$ solution) will be accidental settings, and the few remaining will be the causal settings. If the first few cipher pairs be deciphered using each $(a\ b\ c)$ solution thus derived, the correct $(a\ b\ c)$ matrix will be immediately evident by inspection, since this will be the only one yielding English text. This correct matrix will be produced from the causal settings only.

Suppose the trigraphs

$$P_1Q_1R_1\ ,\ \cdots\ ,\ P_jQ_jR_j\ ,\ \cdots\ ,\ P_kQ_kR_k$$

are to be tested. Such a list would include, for example, *THE*, *AND*, *ING*, *ENT*, *HER*, *ION*, *NTH*, *OFT*. For a given (i, j), $E_{ij}(F_{ij})$, $(i = 1, \cdots, m-1; j = 1, \cdots, k)$, is evaluated and if 0 mod 26 the corresponding $AB\ AB$, PQR mod 2 entries in Table 1 (2) are located. If these are present (the M_i solutions) then $D(D')$ is computed, and all $(a\ b\ c)$ solutions obtained using (4.6) or (4.10) respectively (and using the Table 1 (2) entries in case of multiple solutions). In case $D(D')$ is even $(\not\equiv 0)$ mod 26, (4.2e) or (4.4e), respectively is tested. Finally, the first five cipher groups are deciphered with each $(a\ b\ c)$ solution. These decipherments, together with the corresponding $(a\ b\ c)$, are printed. If no decipherment gives plausible plaintext, the procedure is repeated using further trigraphs, but this would occur only very rarely.

In an example containing 100 cipher groups ($m = 100$), 50 trigraphs ($k = 50$) were tested, giving some 10,000 trials to be examined. Of these, 280 gave apparent settings with 320 $(a\ b\ c)$ possibilities. There were 20 causal settings.

In general, the longer the cipher-text the fewer the number of trigraphs that need be tested. Discarding cases where D or D' is 0 or 13 mod 26 causes no difficulty since sufficient trigraphs are used. The time for testing the 10,000 trials was approximately 30 minutes.

North Carolina State College
Raleigh, North Carolina

→ L. S. Hill, "Cryptography in an algebraic alphabet," *Amer. Math. Monthly*, v. 36, 1929, p. 306–312.
→ L. S. Hill, "Concerning certain linear transformation apparatus of cryptography," *Amer. Math. Monthly*, v. 38, 1931, p. 135–154.
→ Jack Levine, "Variable matrix substitution in algebraic cryptography," *Amer. Math. Monthly*, v. 65, 1958, p. 170–179.