

Cryptanalysis of the Hwang-Rao Secret Error-Correcting Code Schemes

Kencheng Zeng¹, Chung-Huang Yang², and T.R.N. Rao³

¹ SKLOIS, Graduate School of Chinese Academia Sinica, P.O.Box 3908, Beijing 100039, Peoples Republic of China

² National Kaohsiung First University of Science and Technology, 1 University Road, Yenchao, Kaohsiung 824, Taiwan, chyang@computer.org

³ Center for Advanced Computer Studies, University of Louisiana, Lafayette, Louisiana 70504-4330, U.S.A., trao@cacs.louisiana.edu

Abstract. In this paper, the cryptanalytic strength of two Hwang-Rao Secret Error-Correcting Code (SECC) schemes is examined under a known-plaintext attack. In particular, we found the existence of key information redundancy in all SECCs used in the electronic codebook (ECB) mode. Also, our investigations indicate the existence of *synergism* in the SECC schemes, that is, the security of SECC (containing three transformations, Ψ and \mathbf{E} and \mathbf{P}) is much stronger than the individual strength of either Ψ or \mathbf{E} or \mathbf{P} .

1 Introduction

Using error-correcting codes as cryptosystems was introduced by McEliece [1–4]. McEliece’s proposal was to use a Goppa code as the underlying basis of an ingenious public-key scheme. Rao and Nam [5–7] subsequently introduced a new approach to the private-key algebraic-coded cryptosystems requiring simple error-correcting codes (distance ≥ 6 codes). Hwang and Rao [8] then devised a class of private key cryptosystems, called the Secret Error-Correcting Codes (SECCs).

A SECC provides both data security and data reliability while retaining the full error-correcting capability of the introduced code for possible channel errors. Also in a SECC scheme, any unauthorized user would find it hard to correct channel errors without the decoding keys and the presence of channel errors introduces additional level of security to the system. In this research, we will examine the cryptographic strength of SECCs used in the electronic codebook (ECB) mode. Figure 1 illustrates the three transformations, Ψ and \mathbf{E} and \mathbf{P} , involved in the SECC scheme operating in ECB mode.

The ciphertext C is given by

$$C = \mathbf{E}(\Psi(M)) \cdot \mathbf{P}$$

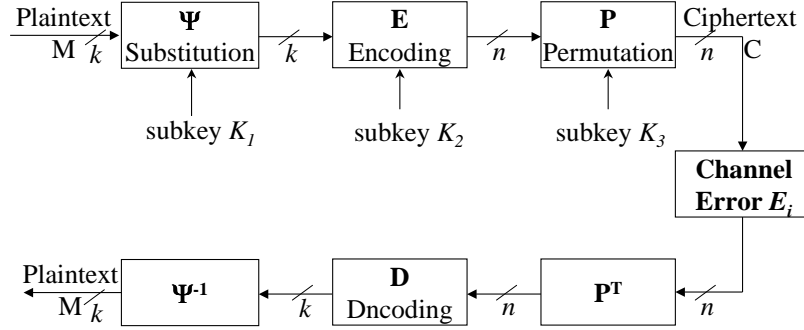


Fig. 1. Hwang-Rao Secret Error-Correcting Code (SECC)

where Ψ is a cryptographic transformation selected by the subkey K_1 , \mathbf{E} is the encoding of an (n, k) nonlinear code selected by the subkey K_2 , and \mathbf{P} is a random $n \times n$ permutation matrix selected by the subkey K_3 . The cryptographic transformation Ψ is installed with the purpose of withstanding chosen plaintext attacks like those done in [6] and the Preparata code [9] was mentioned as the underlying basis of \mathbf{E} .

Preparata codes [9] are a family of $(n = 2^r - 1, k = 2^r - 2r, 5)$ optimal double error-correcting codes, for even $r \geq 4$. The representation of Preparata codes in terms of polynomials over $GF(2)$ modulo $(x^{2^{r-1}-1} + 1)$ is as follows. Let α denote a primitive element of $GF(2^{r-1})$; $g(x)$ the minimum polynomial for the α ; $g_3(x)$ the minimum polynomial for the α^3 ; $\{r(x)\}$ a $\{2^{r-1}-1, 2^{r-1}-r, 3\}$ Hamming code generated by $g(x)$; $\{s(x)\}$ a $[2^{r-1}-1, 2^{r-1}-2r, 6]$ BCH code generated by $(1+x)g(x)g_3(x)$; $f(x)$ the nonzero code polynomial of the dual code of $\{m(x)\}$ such that $f^2(x) = f(x)$; $u(x) = 1 + x + x^2 + \dots + x^{2^{r-1}-2}$; $q(x) \in \{0, 1, x, x^2, \dots, x^{2^{r-1}-2}\}$; $b \in \{0, 1\}$. Then the 3-block binary vectors of the form $w = [m(x) + q(x), b, m(x) + (m(1) + b)u(x) + q(x) \cdot f(x) + s(x)]$ are the codewords of the Preparata codes.

We will examine the security of SECC using Preparata code in the presence of a random error vector of weight ≤ 1 , as shown in Fig. 2, we called it SECC Scheme I in this research.

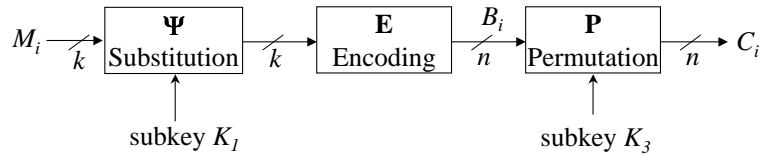


Fig. 2. Hwang-Rao SECC Scheme I

In all that follows, we assume that the Preparata code of interest is fixed and known to the cryptanalyst, for the number of possible Preparata codes of a given code length $n = 2^r - 1$ is $\frac{\Phi(2^r-1)}{r-1}$, which is very small (less or equal to 48 for code length $n \leq 1023$). In order to increase the error-correcting capability of SECC using Preparata codes, Hwang and Rao also suggest to use the $|u| |u + v|$ code construction method [10, p. 76]. This scheme, shown in Figure 3, will be called SECC Scheme II.

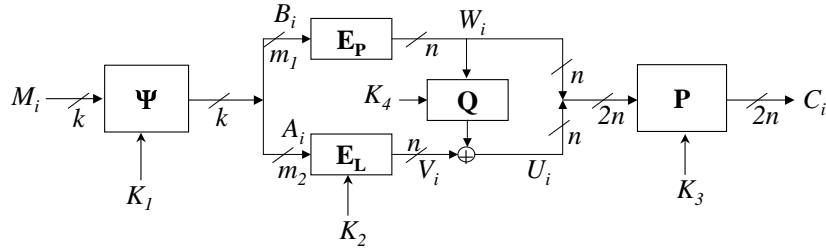


Fig. 3. SECC Scheme II

For the SECC Scheme II, the ciphertext C_i is given by

$$\begin{aligned} C_i &= (W_i, U_i) \cdot \mathbf{P} \\ &= (W_i, W_i \cdot \mathbf{Q} + V_i) \cdot \mathbf{P} , \end{aligned}$$

where $W_i = \mathbf{E}_P(B_i)$, $V_i = \mathbf{E}_L(A_i)$, $(B_i, A_i) = \Psi_{K_1}(M_i)$, \mathbf{E}_P denote the encoding a (n, m_1) Preparata code; \mathbf{E}_L is the generator matrix of a (n, m_2) linear code selected by the subkey K_2 ; $k = m_1 + m_2$; \mathbf{P} is a $2n \times 2n$ permutation matrix, \mathbf{Q} is an $n \times n$ permutation matrix, both randomly selected by the subkeys K_3, K_4 respectively.¹

In the following sections, the cryptanalytic strength of above two SECC schemes is examined under a known-plaintext attack. In particular, we found the existence of key information redundancy in all SECCs used in the electronic codebook (ECB) mode. That is, we could perform exhaustive searching applied to the subkey K_1 . and recover other subkeys. Since exhaustive searching has been applied, the conclusions obtained do not mean that the schemes considered here are cryptographically insecure, rather they show that under a known-plaintext attack all the additional functions introduced into the schemes do not contribute substantially to their cryptographic strength. To counterattack, key expansion might be used, where a short key is stretched into a long one [11].

2 A Known-Plaintext Attack to the SECC Scheme I

Suppose we have s plaintext-ciphertext pairs, denoted by $(M_1, C_1), (M_2, C_2), \dots, (M_s, C_s)$, then our attack can be described as the following two steps.

¹ We also use the symbol \mathbf{E}_L to denote the linear code selected by K_2

2.1 Determining the subkey K_1 by exhaustive searching

First, we will try to find the unknown subkey K_1 by a brute-force approach. Let $\overline{K_1}$ denote the subkey in trial, then for each given plaintext M_i , $1 \leq i \leq s$, compute $\overline{B_i} = \mathbf{E}_{\overline{K_1}}(\Psi(M_i))$ under the control of the trial key $\overline{K_1}$. It is clear that the permutation \mathbf{P} is a (Hamming) weight preserving transformation. Therefore, we can use the condition.

$$\text{weight}(\overline{B_i}) = \text{weight}(C_i), \quad 1 \leq i \leq s$$

to search for the correct subkey K_1 . Let p_i denote the probability for a randomly chosen codeword of the Preparata code to have weight i , then we have $p_i < \frac{1}{2}$ for any $4 \leq i \leq n-4$, as can be seen from the symmetry of the weight enumerator of the Preparata codes [10, p.473]. Thus, the probability p for a false $\overline{K_1}$ to pass the test on s plaintext-ciphertext pairs is $p < \frac{1}{2^s}$. The correct subkey K_1 , in general, will be uniquely determined if we have $s > |K_1|$ (length of the subkey K_1 in bits) such pairs.

2.2 Determining the permutation \mathbf{P}

Once the subkey K_1 has been found, we note that

$$\mathbf{B} \cdot \mathbf{P} = \mathbf{C},$$

where \mathbf{B} is the $s \times n$ matrix

$$\mathbf{B} = \begin{bmatrix} B_1 \\ B_2 \\ \dots \\ B_s \end{bmatrix}$$

and \mathbf{C} is the $s \times n$ matrix

$$\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \\ \dots \\ C_s \end{bmatrix}$$

Now the permutation \mathbf{P} can be determined by comparing the n columns in the matrix \mathbf{B} with the n columns in the matrix \mathbf{C} .

3 A Known Plaintext Attack to the SECC Scheme II

Our attack is based on utilizing the linearity of \mathbf{E}_L , \mathbf{Q} , and \mathbf{P} and we will again assume that s plaintext-ciphertext pairs, $(M_1, C_1), (M_2, C_2), \dots, (M_s, C_s)$, are given.

3.1 Determining the subkey K_1

Let $\overline{K_1}$ denote the subkey in trial and $(\overline{B_i}, \overline{A_i}) = \Psi_{\overline{K_1}}(M_i)$. We begin by finding a maximal set of linearly independent solutions $\lambda^{(1)}, \lambda^{(2)}, \dots, \lambda^{(N)}$ of the system of homogeneous linear algebraic equations

$$\lambda \overline{\mathbf{A}} = \mathbf{0},$$

where $\overline{\mathbf{A}}$ is the $s \times n$ matrix with $\overline{A_j}$, $1 \leq j \leq s$ as rows and

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_s)$$

denotes the s -dimensional vector of unknowns. We know from a well-known theorem in linear algebra (see, for example, [12, p. 76]) that $N \geq s - m_2$, and it follows from the linearity of the (linear) code $\mathbf{E_L}$ that

$$\sum_{j=1}^s \lambda_j^{(i)} \mathbf{E_L}(\overline{A_j}) = \mathbf{E_L} \left(\sum_{j=1}^s \lambda_j^{(i)} \overline{A_j} \right) = \mathbf{0}, \quad i = 1, 2, \dots, N.$$

Therefore, if $\overline{K_1}$ is the right choice, then

$$\begin{aligned} \sum_{j=1}^s \lambda_j^{(i)} C_j &= \left(\sum_{j=1}^s \lambda_j^{(i)} \mathbf{E_P}(\overline{B_j}), \sum_{j=1}^s \lambda_j^{(i)} (\mathbf{E_L}(\overline{A_j}) + \mathbf{E_P}(\overline{B_j}) \cdot \mathbf{Q}) \right) \cdot \mathbf{P} \\ &= \left(\sum_{j=1}^s \lambda_j^{(i)} \mathbf{E_P}(\overline{B_j}), \sum_{j=1}^s \lambda_j^{(i)} \mathbf{E_P}(\overline{B_j}) \cdot \mathbf{Q} \right) \cdot \mathbf{P} \end{aligned}$$

and hence we shall have

$$2 \text{ weight} \left(\sum_{j=1}^s \lambda_j^{(i)} \mathbf{E_P}(\overline{B_j}) \right) = \text{weight} \left(\sum_{j=1}^s \lambda_j^{(i)} C_j \right), \quad 1 \leq i \leq N. \quad (1)$$

We then determine the unknown K_1 by exhaustive searching and use (Eq. 1) as our key identification criterion. This is a good criterion, for we know that under some general conditions the probability that a pair of randomly generated l -vectors will have the same (Hamming)weight is

$$p = \frac{1}{2^{2l}} \sum_{r=0}^l (C_l^r)^2 = \frac{1}{2^{2l}} C_{2l}^l < \frac{1}{4},$$

so one can expect to determine K_1 uniquely if $s > m_2 + \frac{|K_1|}{2}$. We emphasize that the basic idea in designing the key identification criterion (Eq. 1) is to eliminate the influence of the subkeys K_2, K_3, K_4 in trying to determine the subkey K_1 .

3.2 Determining the Image Set $\mathbf{T} = \mathbf{P}(i) \mid n < i \leq 2n$

Once K_1 has been found, we will determine the image set \mathbf{T} of the half interval $[n + 1, 2n]$ under the permutation \mathbf{P} . We start with finding a maximal set of linearly independent solution vectors

$$\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(L)}, \quad L \geq s - m_1,$$

of the system of homogeneous linear algebraic equations

$$\mu \mathbf{W} = 0,$$

where \mathbf{W} is the $s \times n$ matrix with W_j , $1 \leq j \leq s$, as rows and

$$\mu = (\mu_1, \mu_2, \dots, \mu_s)$$

is the s -vector of unknowns. Then we shall have for each $1 \leq i \leq L$,

$$\Phi^{(i)} = \sum_{j=1}^s \mu_j^{(i)} C_j = \left(\mathbf{0}^{(n)}, D^{(i)} \right) \cdot \mathbf{P},$$

where

$$D^{(i)} = \sum_{j=1}^s \mu_j^{(i)} \mathbf{E}_{\mathbf{L}}(A_j), \quad 1 \leq i \leq L,$$

are codewords in $\mathbf{E}_{\mathbf{L}}$. Since the dimension of the linear code $\mathbf{E}_{\mathbf{L}}$ is m_2 , the probability that there exists among these codewords a linear basis for $\mathbf{E}_{\mathbf{L}}$ will be [13]

$$p = \prod_{L-m_2+1}^L \left(1 - \frac{1}{2^i} \right),$$

which is nearly equal to 1 when s is sufficiently large, say, $s \geq m_1 + m_2 + 4$. Now suppose such a basis does exist, then we can determine \mathbf{T} according to the rule:

$$k \in \mathbf{T} \iff k\text{-th component of } \Phi^{(i)} \text{ is 1 for some } i \in [1, L].$$

This criterion for determining \mathbf{T} is based on the simple observation that if we fix a certain linear basis of the code $\mathbf{E}_{\mathbf{L}}$, then for any $1 \leq k \leq n$ there is a codeword belonging to that basis such that its k -th component will be 1. For otherwise it would mean that the k -th location is redundant for the code $\mathbf{E}_{\mathbf{L}}$.

3.3 Decrypting ciphertexts without knowing K_2, K_3, K_4

Once the subkey K_1 is determined from the given s plaintext-ciphertext pairs, we could decrypt any other ciphertexts without knowing K_2, K_3, K_4 . To recover

the plaintexts, we arrange the numbers of the two sets \mathbf{T} and $\mathbf{S} = \mathbf{I}^{(2n)} - \mathbf{T}$ in some fixed, say increasing, order as

$$\mathbf{S} = \{s_1, s_2, \dots, s_n\}, \quad \mathbf{T} = \{t_1, t_2, \dots, t_n\}$$

and for any $2n$ -vector

$$C = (c_1, c_2, \dots, c_{2n})$$

write

$$C(\mathbf{S}) = (c_{s_1}, c_{s_2}, \dots, c_{s_n}), \quad C(\mathbf{T}) = (c_{t_1}, c_{t_2}, \dots, c_{t_n}).$$

Since $\Psi(M) = (B, A)$, we will first find B and A from the received error-free ciphertext C , then recover the plaintext M .

To find B , let r be the dimension of the linear closure $\langle \mathbf{E} \rangle$ of the Preparata code and find from among the codewords W_i , $1 \leq i \leq s$, a linear basis

$$W_{\alpha_1}, W_{\alpha_2}, \dots, W_{\alpha_r}$$

for $\langle \mathbf{E} \rangle$. As given in the above discussion, the success probability of doing this is nearly 1. Then we shall have

$$C(\mathbf{S}) = \sum_{k=1}^r \xi_k C_{\alpha_k}(\mathbf{S})$$

and after having computed the coefficients ξ_k by solving the corresponding system of non-homogeneous linear algebraic equations, we shall have

$$W = \sum_{k=1}^r \xi_k W_{\alpha_k}$$

and

$$B = \mathbf{E}_P^{-1}(W).$$

To find A , we replace $C(\mathbf{T})$ by

$$C^*(\mathbf{T}) = C(\mathbf{T}) + \sum_{k=1}^r \xi_k C_{\alpha_k}(\mathbf{T}),$$

and find from among the vectors $\Phi^{(i)}(\mathbf{T})$, $1 \leq i \leq L$, a set of m_2 linearly independent ones

$$\Phi^{(\beta_1)}(\mathbf{T}), \Phi^{(\beta_2)}(\mathbf{T}), \dots, \Phi^{(\beta_{m_2})}(\mathbf{T}).$$

Once such a set has been found, $C^*(\mathbf{T})$ can also be expressed as a linear combination of them:

$$C^*(\mathbf{T}) = \sum_{k=1}^{m_2} \eta_k \Phi^{(\beta_k)}(\mathbf{T}).$$

The combination coefficients η_k can be computed by solving the corresponding system of n equations in m_2 unknowns and it follows from the linearity of \mathbf{E}_L that

$$A = \sum_{k=1}^{m_2} \eta_k \sum_{j=1}^s \mu_j^{(\beta_k)} A_j + \sum_{k=1}^r \xi_k A_{\alpha_k} .$$

Plaintext M can then be obtained by

$$M = \Psi^{-1}(A, B) .$$

3.4 Determining the subkeys K_2, K_3, K_4

We shall illustrate how to determining the subkeys K_2, K_3, K_4 by considering the case where \mathbf{E}_L is an (n, m_2) cyclic code generated by a polynomial

$$g(x) = x^d + g_{d-1}x^{d-1} + \dots + g_0, \quad d = n - m_2$$

with coefficients $g_i, 0 \leq i \leq d-1$ to be specified by the subkey K_2 .

First, consider the generator matrix

$$\mathbf{G} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & g_{d-1} & \cdots & g_1 & g_0 \\ 0 & 0 & \cdots & 1 & g_{d-1} & g_{d-2} & \cdots & g_0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & g_{d-1} & \cdots & g_2 & g_1 & g_0 & \cdots & 0 & 0 \end{pmatrix} \quad (2)$$

together with the $m_2 \times m_2$ non-singular matrix

$$\mathbf{Y} = \begin{bmatrix} Y^{\beta_1} \\ Y^{\beta_2} \\ \cdots \\ Y^{\beta_{m_2}} \end{bmatrix}$$

with rows

$$Y^{\beta_k} = \sum_{j=1}^s \mu_j^{(\beta_k)} A_j$$

and the $m_2 \times n$ matrix

$$\mathbf{\Phi} = \begin{bmatrix} \Phi^{(\beta_1)}(\mathbf{T}) \\ \Phi^{(\beta_2)}(\mathbf{T}) \\ \cdots \\ \Phi^{(\beta_{m_2})}(\mathbf{T}) \end{bmatrix} .$$

Both \mathbf{Y} and $\mathbf{\Phi}$ can be computed from the known plaintext-ciphertext pairs and we have

$$\mathbf{G} = \mathbf{Y}^{-1} \mathbf{\Phi} \mathbf{R} ,$$

where \mathbf{R} is an $n \times n$ permutation matrix. This means the matrix \mathbf{G} , i.e., the subkey K_2 can be determined by reducing the product $\mathbf{Y}^{-1} \cdot \mathbf{\Phi}$ to the canonical form (Eq. 2) through column permutation.

Once \mathbf{E}_L has been found, we can determine the value of $\mathbf{P}(n + i) \in \mathbf{T}$, $1 \leq i \leq n$, by comparing the columns of the matrices

$$\begin{bmatrix} D^{(1)} \\ D^{(2)} \\ \dots \\ D^{(N)} \end{bmatrix}, \quad \begin{bmatrix} \Phi^{(1)}(\mathbf{T}) \\ \Phi^{(2)}(\mathbf{T}) \\ \dots \\ \Phi^{(N)}(\mathbf{T}) \end{bmatrix}$$

and determine the values $\mathbf{P}(i) \in \mathbf{S}$ by comparing the columns of

$$\begin{bmatrix} W_1 \\ W_2 \\ \dots \\ W_s \end{bmatrix}, \quad \begin{bmatrix} C_1(\mathbf{S}) \\ C_2(\mathbf{S}) \\ \dots \\ C_s(\mathbf{S}) \end{bmatrix}.$$

Finally, by comparing the first n columns with the last n columns in the matrix with the rows

$$R_i = C_i \cdot \mathbf{P}^{-1} + (\mathbf{0}^{(n)}, \mathbf{E}_L(A_i)), \quad 1 \leq i \leq s,$$

we can easily determine the permutation \mathbf{Q} .

4 Conclusions

We have shown that, in the presence of sufficient plaintext-ciphertext pairs, both the SECC Scheme I and Scheme II can be attacked by exhaustive searching applied to the subkey K_1 . Since exhaustive searching has been applied, the conclusions obtained do not mean that the SECC schemes considered here are cryptographically insecure, rather they show that under a known-plaintext attack all the additional devices introduced into the schemes do not contribute substantially to their cryptographic strength. To counterattack, key expansion might be used, where a short key is stretched into a long one.

The above investigations also indicate the existence of *synergism* in the SECC schemes, that is, the security of SECC using nonlinear codes (containing three transformations, Ψ and \mathbf{E} and \mathbf{P}) is much stronger than the individual strength of either Ψ or \mathbf{E} or \mathbf{P} .

References

1. R.J. McEliece, "A Public-Key Cryptosystem Based on Algebraic Coding Theory," *DSN Progress Report*, Jet Propulsion Laboratory, Calif., Jan. & Feb. 1978, pp. 114-116.
2. A. Canteaut and N. Sendrier, "Cryptoanalysis of the Original McEliece Cryptosystems," *Proc. Asiacrypt'98*, 1998, pp. 187-199.
3. P. Loidreau, "Strengthening McEliece Cryptosystem," *Proc. Asiacrypt'2000*, 2000, pp. 585-598.

4. P. Loidreau and N. Sendrier, "Weak keys in the McEliece public-key cryptosystem," *IEEE Transactions on Information Theory*, Vol. 47, No. 3, March 2001, pp.1207-1211.
5. T.R.N. Rao and K.H. Nam, "A Private-Key Algebraic-Coded Cryptosystem," *Proc. Crypto'86*, 1986, pp. 35-48.
6. R. Struik and J. van Tilburg, "The Rao-Nam Scheme is Insecure Against a Chosen-Plaintext Attack," *Proc. Crypto'87*, 1987, pp. 445-457.
7. T.R.N. Rao and K.H. Nam, "Private-Key Algebraic-Code Encryptions," *IEEE Trans. Info. Theory*, 1989, pp. 829-833.
8. T. Hwang and T.R.N. Rao, "Secret Error-Correcting Codes (SECC)," *Proc. Crypto'88*, 1988, pp. 540-563.
9. F. P. Preparata, "A Class of Optimum Nonlinear Double-Error-Correcting Codes," *Information and Control*, Vol. 13, 1968, pp. 378-400.
10. F.J. MacWilliams and J.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.
11. C.P. Schnorr, "On the construction of random number generators and random function generators," *Proc. Eurocrypt'88*, 1988, pp. 225-232.
12. A. Adrian Albert, *Fundamental Concepts of Higher Algebra*, University of Chicago Press, 1956.
13. K.C. Zeng, C.H. Yang and T.R.N. Rao, "On the Linear Consistency Test (LCT) in Cryptanalysis with Applications," *Proc. Crypto'89*, 1989, pp. 164-174.