On the Design of Linear Transformations for Substitution Permutation Encryption Networks

A.M. Youssef, S. Mister and S.E. Tavares

Department Of Electrical and Computer Engineering

Queen's University, Kingston, Ontario, Canada, K7L 3N6

E-mail: {amr_y, misters and tavares}@ee.queensu.ca

http://adonis.ee.queensu.ca:8000

***Abstract*—** In this paper we study the security of Substitution Permutation Encryption Networks (SPNs) with randomly selected bijective substitution boxes and a randomly selected invertible linear transformation layer. In particular, our results show that for such a *64*–bit SPN using 8×8 s-boxes, the number of s-boxes involved in any 2 rounds of a linear approximation or a differential characteristic is equal to 8 with probability exceeding 0.8. For these SPNs the number of plaintext/ciphertext pairs that are required for the basic linear and differential cryptanalysis exceeds 2^{64} within 6 rounds. We also provide two construction methods for involution linear transformations based on Maximum Distance Separable Codes.

1 Introduction

Heys and Tavares [3][4][5] showed that replacing the permutation layer of Substitution Permutation encryption Networks (SPNs) with a diffusive linear transformation improves the avalanche characteristics of the cipher and increases the cipher's resistance to differential and linear cryptanalysis . Linear [8] and differential [1] cryptanalysis are two of the most powerful attacks on block ciphers. In particular it was shown [3][4] that with such a linear transformation we can develop upper bounds on the differential characteristic probability [1] and on the probability of a linear approximation [9] as a function of the number of rounds of substitution. These bounds are achieved by choosing the linear transformation in such a way that we can have a lower bound on the number of s-boxes involved in any 2 rounds of a differential characteristic or linear approximation expression. Letting N represent the block size of an SPN consisting of R rounds of $n \times n$ s-boxes (M per round), a simple example of an SPN with $N = 16,\ n = 4, M = \frac{N}{n} = 4,$ and $R = 3$ is illustrated in Figure 1.

An interesting class of linear transformations is the one based on Maximum Distance Separable (MDS) codes [7]. The use of such linear transformations was first proposed by Vaudenay in [13] and then utilized in the cipher SHARK [12] and later in the cipher SQUARE [2]. This class of linear transformations has the advantage that the number of s-boxes involved in any 2 rounds of a linear approximation or in any 2 rounds of a differential characteristic is equal to $M + 1$ which is the maximum theoretically possible number.

In this paper we study the security of SPNs with randomly selected *n*-bit bijective substitution boxes and a randomly selected linear transformation layer over $GF(2^n)$. We also provide two construction methods for involution linear transformations based on Maximum Distance

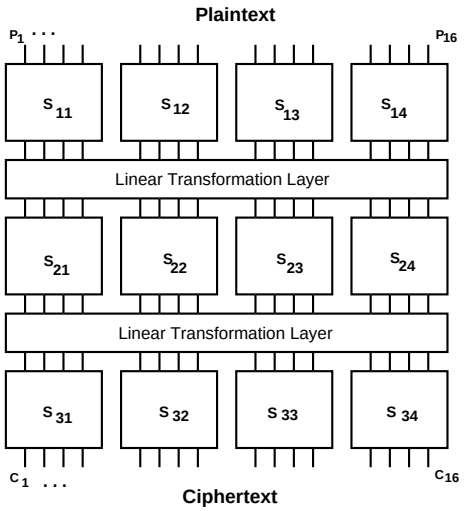

Figure 1 SPN with $N = 16$, $n = 4$, and $R = 3$.

Separable Codes. Involution linear transformations have the advantage that the resulting network can be used to perform both the encryption and the decryption operations [16].

Rijmen *et al* [12] noted that the framework of linear codes over $GF(2^n)$ provides an elegant way to construct the linear transformation layer. More details about the theory of error correcting codes can be found in [7].

Let C be a $(2M, M, d)$ code over $GF(2^n)$. Let $G = [I|A]$ be the generator matrix in echelon form where A is a nonsingular $M \times M$ matrix and I is the $M \times M$ identity matrix. Then A defines an invertible linear mapping

$$GF(2^n)^M \rightarrow GF(2^n)^M : X \rightarrow Y = AX. \tag{1}$$

If the matrix A is used in the implementation of the linear transformation of the SPN, then it is easy to see that the number of s-boxes involved in any 2 rounds of a differential characteristic or linear approximation expression is lower bounded by d, the minimum distance of the code [12]. The minimum distance of the code is equal to the minimum number of linearly dependent columns in its null matrix (also known as the parity-check matrix). For an MDS code with parameters $(2M, M, d)$, the minimum distance d is equal to $M + 1$. Throughout this paper we assume that M is an even number.

2 Randomly Selected Linear Transformations

Lemma 1

Let $G = [I|A]$ be the generator matrix of a code in echelon form where A is a randomly selected $M \times M$ nonsingular matrix and I is the $M \times M$ identity matrix with elements over $GF(q), q = 2^n$. Then the probability that this code has a minimum distance $d \geq r$, $2 \leq r \leq M + 1$, is lower

bounded by

$$\frac{1}{\Psi(M,q)}\prod_{i=1}^{M}\left(q^M-\sum_{j=0}^{r-2}\binom{M+i-1}{j}(q-1)^j-\sum_{j=r-1}^{i-1}\binom{i-1}{j}(q-1)^j\right), \tag{2}$$

where

$$\Psi(M,q)=\prod_{i=0}^{M-1}\left(q^M-q^i\right) \tag{3}$$

is the number of nonsingular $M\times M$ matrices over $GF(q)$.

Proof: If $G=[I|A]$ then the null matrix H is given by

$$H=\left[-A^T|I\right]=\left[A^T|I\right] \tag{4}$$

since we are working over $GF(2^n)$. It is clear that as A varies over all possible nonsingular matrices, A^T varies over the same set. We construct the matrix A^T column by column to meet our criterion.

The columns of A^T must not equal any linear combination of up to $r-2$ of the other columns of H, and, for A^T to be invertible, no column of A^T should be a linear combination of the other columns of A^T.

Suppose we have already assigned $i-1$ columns of A^T. We may choose any of the q^M possibilities for column i except the

$$\sum_{j=0}^{r-2}\binom{M+i-1}{j}(q-1)^j \tag{5}$$

linear combinations of up to $r-2$ of the $M+i-1$ assigned columns of H and the

$$\sum_{j=r-1}^{i-1}\binom{i-1}{j}(q-1)^j \tag{6}$$

linear combinations of known columns of A^T not counted in (5).

Note that the combinations counted in (5) and (6) may not be distinct. Thus, the number of choices available for column i is at least

$$q^M-\sum_{j=0}^{r-2}\binom{M+i-1}{j}(q-1)^j-\sum_{j=r-1}^{i-1}\binom{i-1}{j}(q-1)^j \tag{7}$$

and hence the number of choices of A is at least

$$\prod_{i=1}^{M}\left(q^M-\sum_{j=0}^{r-2}\binom{M+i-1}{j}(q-1)^j-\sum_{j=r-1}^{i-1}\binom{i-1}{j}(q-1)^j\right). \tag{8}$$

The lemma follows by dividing the expression above by the total number of nonsingular $M \times M$ matrices over $GF(q)$. □

O'Connor [11], and Youssef and Tavares [15], [14] studied the XOR distribution table and the Linear Approximation Table (LAT) properties of randomly selected bijective s-boxes. From the analysis in [11], [15] and [14] the expected value of the maximum XOR table entry of an 8×8 randomly selected bijective mapping $\triangle$ is less than or equal to 12 and the expected nonlinearity $\mathcal{NL}$ is greater than 92.

Using an approach similar to the analysis in [4], it is possible to establish upper bounds on the most likely differential characteristic and linear approximation expression using a randomly selected SPN for which the number of s-boxes involved in any 2 rounds of a differential characteristic is greater than or equal to d. The results are obtained by assuming that all the round keys are independent.

The number of chosen plaintext/ciphertext pairs required for differential cryptanalysis of an R round SPN (based on the best *characteristic* and not the best *differential* [10], [6]) may be approximated by [1], [4]

$$N_D \geq \frac{1}{(P_\delta)^\alpha}, \tag{9}$$

where $P_\delta = \frac{\triangle}{2^n}$ and

$$\alpha \geq d\left(\frac{R}{2} - 1\right) + 1. \tag{10}$$

Similarly, the number of known plaintexts required for the *basic* linear cryptanalysis (algorithm 1 in [9]) may be approximated by [4]

$$N_L \geq \frac{1}{|2^{\alpha-1} P_\epsilon^\alpha|^2} \tag{11}$$

where

$$P_\epsilon = \frac{2^{n-1} - \mathcal{NL}}{2^n}, \tag{12}$$

and

$$\alpha \geq \frac{dR}{2}. \tag{13}$$

Letting R_L and R_D denote the minimum even number of rounds required so that N_L and N_D are greater than 2^{64}, Table 1 shows R_L and R_D as a function of d for $n = 8$, $\triangle = 12$ and $\mathcal{NL} = 92$.

d	4	5	6	7	8
R_L	10	10	8	8	6
R_D	10	8	8	6	6

Table 1 R_L and R_D as a function of d ($n = 8$, $\triangle = 12$ and $\mathcal{NL} = 92$)

Table 2 shows the theoretical lower bound (equation (2)) as well as the experimental result (sample size $= 10^5$) for the probability of picking a random invertible linear transformation, with $n = M = 8$, for which d is lower bounded by r, $4 \leq r \leq 8$.

r	4	5	6	7	8
Theoretical bound (eqn. 2)	$1 - 1.58 \times 10^{-12}$	$1 - 1.51 \times 10^{-9}$	$1 - 9.78 \times 10^{-7}$	$1 - 4.66 \times 10^{-4}$	0.839
Experimental (Random)	1.0	1.0	1.0	$1 - 4.6 \times 10^{-4}$	0.844
Experimental (Involution)	1.0	1.0	1.0	$1 - 1.18 \times 10^{-3}$	0.922

Table 2 Lower Bounds for $P(d \geq r)$ for a Randomly Chosen Linear Transformation ($n = M = 8$)

3 Involution Linear Transformations based on MDS codes

In general, SPNs need two different modules for the encryption and the decryption operations. In an SPN, decryption is performed by running the data backwards through the inverse network (i.e., applying the key scheduling algorithm in reverse and using the inverse s-boxes and the inverse linear transformation layer). In [16] the authors proposed a special class of SPNs that has the advantage that the same network can be used to perform both the encryption and the decryption operations. The basic idea is to use involution substitution layers and involution linear transformations. In this section we study two construction methods for involution linear transformations based on MDS codes.

For a linear (n, k, d) code over any field, $d \leq n - k + 1$. Codes with $d = n - k + 1$ are called Maximum Distance Separable Codes, or MDS codes for short [7].

Lemma 2[7]:

An (n, k, d) code with generator matrix $G = [I|A]$, where A is a $k \times (n - k)$ matrix, is MDS if and only if every square submatrix (formed from any i rows and any i columns, for any $i = 1, 2, \cdots, min\{k, n - k\}$) of A is nonsingular.

3.1 Random Construction

One way to obtain an involution matrix A which satisfies the above constraint is to pick a random involution matrix and test it for the above constraint.

Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{14}$$

be an $M \times M$ random matrix where A_{11}, A_{12}, A_{21} and A_{22} are nonsingular $\frac{M}{2} \times \frac{M}{2}$ matrices. An involution matrix is one which satisfies $A^2 = I$, and thus A is an involution iff

$$A_{11}A_{12} \oplus A_{12}A_{22} = 0, \tag{15}$$

$$A_{11}^2 \oplus A_{12}A_{21} = I, \tag{16}$$

$$A_{21}A_{11} \oplus A_{22}A_{21} = 0, \tag{17}$$

$$A_{21}A_{12} \oplus A_{22}^2 = I. \tag{18}$$

If we let $A_{22} = A_{11}$ then equation (15) is satisfied iff A_{11} and A_{12} commute with each other. To achieve this we let $A_{12} = A_{11}^{-1}$. For these choices of A_{12} and A_{22}, equations (16), (17) and (18) are linearly dependent with the solution $A_{21} = A_{11}^3 \oplus A_{11}$.

Thus the $M \times M$ matrix

$$A = \begin{bmatrix} A_{11} & A_{11}^{-1} \\ A_{11}^3 \oplus A_{11} & A_{11} \end{bmatrix}, \tag{19}$$

where A_{11} is a random nonsingular $\frac{M}{2} \times \frac{M}{2}$ matrix, is an involution over $GF(2^n)$.

For $n = 8$, a random search for a matrix A, with the structure in equation (19), that satisfies the condition in lemma 2, terminates within a few seconds for even values of M, $M \leq 6$. For $M = 8$ we were unable to obtain any matrix that satisfies the conditions in lemma 2 by random search. Table 2 shows the experimental results for 10^5 randomly chosen involution linear transformations in the form of equation (19) for $M = n = 8$.

3.2 Algebraic Construction

In this section we show how to obtain an involution matrix satisfying lemma 2 by a simple algebraic construction.

Lemma 3[7]:

Given $\mathbf{x}_0, \cdots, \mathbf{x}_{n-1}$, and $\mathbf{y}_0, \cdots, \mathbf{y}_{n-1}$ the matrix $A = [\mathbf{a}_{ij}], 0 \leq i, j \leq n-1$ where $\mathbf{a}_{ij} = \frac{1}{\mathbf{x}_i + \mathbf{y}_j}$ is called a Cauchy matrix. It is known that

$$det(A) = \frac{\prod\limits_{0 \leq i < j \leq n-1} (\mathbf{x}_j - \mathbf{x}_i)(\mathbf{y}_j - \mathbf{y}_i)}{\prod\limits_{0 \leq i,j \leq n-1} (\mathbf{x}_i + \mathbf{y}_j)}. \tag{20}$$

Hence, provided the $\mathbf{x}_i$ are distinct, the $\mathbf{y}_i$ are distinct, and $\mathbf{x}_i + \mathbf{y}_j \neq \mathbf{0}$ for all i, j, it follows that any square submatrix of a Cauchy matrix is nonsingular over any field.

Let

$$\begin{aligned} \mathbf{x}_i &= \mathbf{i}, \\ \mathbf{y}_i &= \mathbf{i} \oplus \mathbf{r}, \end{aligned} \tag{21}$$

where

$$\mathbf{i} = (00\cdots 0 i_\tau \cdots i_1 i_0) \in GF(2^n), \sum_{l=0}^{\tau} 2^l i_l = i, \tau = \lceil log_{2M} \rceil - 1, \tag{22}$$

and the least significant $log_2(M)$ bits of $\mathbf{r} \neq \mathbf{0}$ are zeros.

For $A^2 = H = [\mathbf{h}_{ij}]$ we have

$$\mathbf{h}_{ij} = \bigoplus_{k=0}^{M-1} \frac{1}{(\mathbf{i} \oplus \mathbf{k} \oplus \mathbf{r})(\mathbf{j} \oplus \mathbf{k} \oplus \mathbf{r})} = \begin{cases} \bigoplus_{k=0}^{M-1} \frac{1}{(\mathbf{k} \oplus \mathbf{r})^2}, & i = j \\ \mathbf{0}, & i \neq j, \end{cases} \tag{23}$$

where $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ are evaluated as in equation (22). Thus the matrix A will satisfy $A^2 = \mathbf{c}^2 I$, $\mathbf{c} = \bigoplus_{i=1}^{n} \mathbf{a}_{1i}^2$ over $GF(2^n)$. Dividing (division over $GF(2^n)$) each element of A by

$$\sqrt{\mathbf{c}} = \bigoplus_{k=0}^{M-1} \frac{1}{(\mathbf{k} \oplus \mathbf{r})} = \bigoplus_{i=1}^{n} \mathbf{a}_{1i}, \tag{24}$$

we obtain an involution matrix for which every square submatrix is nonsingular over $GF(2^n)$. Figure 2 shows an example for $M = n = 8$, using the irreducible polynomial 11d†.

93	13	57	da	58	47	c	$1f$
13	93	da	57	47	58	$1f$	c
57	da	93	13	c	$1f$	58	47
da	57	13	93	$1f$	c	47	58
58	47	c	$1f$	93	13	57	da
47	58	$1f$	c	13	93	da	57
c	$1f$	58	47	57	da	93	13
$1f$	c	47	58	da	57	13	93

Figure 2 Involution Linear Transformation Based on MDS Codes ($M = n = 8$, Irreducible Polynomial = 11d)

† All numbers are in hexadecimal format

Conclusions

In this paper we studied SPNs with randomly selected s-boxes and a randomly selected invertible linear transformation layer. The results of our analysis show that SPNs with good cryptographic properties can be obtained using this random construction approach. Although this random construction can be used to implement an actual cipher, the analysis in this paper was aimed to prove the robustness of the SPN model.

We also provided two construction methods for involution linear transformations based on MDS codes. Involution linear transformations have the advantage that the resulting network can be used to perform both the encryption and the decryption operations, which enhances the practical aspects of this the class of SPN ciphers.

References

[1] E. Biham and A. Shamir. Differential cryptanalysis of DES-like cryptosystems. *Journal of Cryptology, vol. 4, no. 1, pp. 3–72,* 1991.

[2] J. Daemen, L. Knudsen, and V. Rijmen. The block cipher SQUARE. *Proc. of Fast Software Encryption (4), LNCS , Springer-Verlag,* 1997.

[3] H.M. Heys and S.E. Tavares. The design of substitution-permutation networks resistant to differential and linear cryptanalysis. *Proceedings of 2nd ACM Conference on Computer and Communications Security, Fairfax, Virginia, pp. 148–155,* 1994.

[4] H.M. Heys and S.E. Tavares. The design of product ciphers resistant to differential and linear cryptanalysis. *Journal of Cryptology, Vol. 9, no. 1, pp. 1–19,* 1996.

[5] H.M. Heys and S.E. Tavares. Avalanche characteristics of substitution-permutation encryption networks. *IEEE Trans. Comp., Vol. 44, pp.1131–1139,* Sept. 1995.

[6] X. Lai, J.L. Massey, and S. Murphy. Markov ciphers and differential cryptanalysis. *Advances in Crytology: Proc. of EUROCRYPT '91, Springer-Verlag, pp.17–38,* 1992.

[7] F.J. MacWilliams and N.J.A. Sloane. *The theory of error correcting codes*. North-Holland Publishing Company, 1977.

[8] M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. *Advances in Cryptology: Proc. of CRYPTO '94, Springer-Verlag, Berlin, pp. 1–11,* 1994.

[9] M. Matsui. Linear cryptanalysis method for DES cipher. *Advances in Cryptology: Proc. of EUROCRYPT '93, Springer-Verlag, Berlin, pp. 386–397,* 1994.

[10] K. Nyberg and L.R. Knudsen. Provable security against differential cryptanalysis. *Advances in Crytology: Proc. of CRYPTO '92, Springer-Verlag, pp. 566–574,* 1993.

[11] L.J. O'Connor. On the distribution of characteristics in bijective mappings. *Advances in Cryptology: Proc. of EUROCRYPT '93, Springer-Verlag, Berlin, pp. 360–370,* 1994.

[12] V. Rijmen, J. Daemen, B. Preneel, A. Bosselaers, and E. De Win. The cipher SHARK. *Fast Software Encryption, LNCS 1039, D. Gollmann, Ed., Springer-Verlag, pp. 99-112,* 1996.

[13] S. Vaudenay. On the need for multipermutations: Cryptanalysis of MD4 and SAFER. *Proc. of Fast Software Encryption (2), LNCS 1008, Springer-Verlag, pp. 286–297,* 1995.

[14] A.M. Youssef. Ph.D. thesis, under preparation.

[15] A.M. Youssef and S.E. Tavares. Resistance of balanced s-boxes to linear and differential cryptanalysis. *Information Processing Letters, 56(1995), pp. 249-252,* 1995.

[16] A.M. Youssef, S.E. Tavares, and H.M. Heys. A new class of substitution-permutation networks. *Workshop on Selected Areas in Cryptography, SAC '96, Workshop Record, pp.132-147,* 1996.