ECOLE NORMALE SUPERIEURE

An Experiment on DES Statistical Cryptanalysis

Serge VAUDENAY

LIENS - 95 - 29

Département de Mathématiques et Informatique

CNRS URA 1327

An Experiment on DES Statistical Cryptanalysis

Serge Vaudenay

LIENS - 95 - 29

November 1995

Laboratoire d'Informatique de l'Ecole Normale Supérieure
45 rue d'Ulm 75230 PARIS Cedex 05

Tel : (33)(1) 44 32 00 00

Adresse électronique : vaudenay @dmi.ens.fr

An Experiment on DES Statistical Cryptanalysis

Serge Vaudenay*

Ecole Normale Supérieure — DMI
45, rue d'Ulm
75230 Paris Cedex 5 France
Serge.Vaudenay@ens.fr

Abstract

Linear cryptanalysis and differential cryptanalysis are the most important methods of attack against block ciphers. Their efficiency have been demonstrated against several ciphers, including the Data Encryption Standard. We prove that both of them can be considered, improved and joined in a more general statistical framework. We also show that the very same results as those obtained in the case of DES can be found without any linear analysis and we slightly improve them into an attack with theoretical complexity $2^{42.9}$.

We can apply another statistical attack — the χ^2-cryptanalysis — on the same characteristics without a definite idea of what happens in the encryption process. It appears to be roughly as efficient as both differential and linear cryptanalysis. We propose a new heuristic method to find good characteristics. It has found an attack against DES absolutely equivalent to Matsui's one by following a distinct path.

1 Introduction

Since the proposal of the Data Encryption Standard by the U.S. government, the scientific community concentrated a significant part of its efforts on its cryptanalysis [1]. This well-known function encrypts a 64-bits plaintext into a 64-bits ciphertext using a 56-bits secret key, so that the best attack is expected to have complexity 2^{56} (2^{55} if we take into account the complementation property of DES as in [6]).

A first significant result, obtained by Biham and Shamir, gave a general method for chosen plaintext attacks — the *differential cryptanalysis* [2, 3]. Using a deep analysis of the internal framework of the function, they try to control a

* *Laboratoire d'Informatique de l'Ecole Normale Supérieure*, research group affiliated with the CNRS

correlated piece of information on several particular plaintexts and recover it by statistical attacks. The correlated piece of information used is simply a chosen bit-wise exclusive *or* difference between two texts. The main result of Biham and Shamir proves, using heuristic arguments, that it is possible to mount an attack with 2^{47} chosen plaintexts.

A second result gave also a general method, called *linear cryptanalysis*, for known plaintext attacks. It has been discovered by Matsui who proved that it is possible to implement an attack against DES with 2^{43} known plaintexts [8]. Using another deep analysis of the function, this attack tries to trace a correlation between one bit of information on the plaintext and one bit of information on the ciphertext. One more time, the information is obtained linearly with respect to the exclusive *or*.

Both methods are bottom-up approachs based on the concept of *characteristic*. This is a scenario of the propagation of the correlated piece of information. It is associated to a probability, which has to be as biased as possible. The goal of the heuristic arguments consists in finding *efficient* characteristics, first analyzing the linear properties of the substitution boxes, then plugging them into one another in a such a way that a linear information is leaked throughout the encryption process. Once this analysis has led to an efficient characteristic, we only need to keep which information on the plaintext and the ciphertext is required for the upper level of the attack.

The success of those methods have focused the attention on the linear properties of the boxes. In this paper, we try to prove that the linear properties are not so important. We propose another heuristic approach based on statistics. We show how to recover an attack similar to Matsui's one without any linear consideration. We also propose a top-down approach which unifies linear and differential cryptanalysis. We prove that a simple χ^2 test can get the similar results without knowing precisely what happens (for instance on a black box which implements a secret encryption function). Those results have already partially been presented in [9].

Throughout this paper, we use the following notations:

- k is a secret key in the domain $\mathcal{K}$;
- P is a plaintext in the domain $\mathcal{P}$;
- C is a ciphertext in the domain $\mathcal{C}$;
- Enc_k is an encryption function which maps P to C using key k;
- $x \wedge y$ denotes the bitwise *and* of the bit-strings x and y;
- $W(x)$ is the Hamming weight of the bit-string x;

- $x \cdot y$ is the dot-product of x and y, that is the parity of $W(x \wedge y)$;
- $1_{predicate}$ is 1 if *predicate* is true, 0 otherwise.

2 Heuristic using projection

2.1 Transition matrix of a projected cipher

In the encryption process, intermediate results of the encryption function can be arbitrarily ignored and supposed to be uniformly independent of the rest of the computation. We call this operation *projection*. After projection, assuming that the removed inputs are random, each box becomes stochastic, transforming the leftover inputs into the remaining outputs. Thus, it is possible to compute the *transition matrix* of the projected boxes.

For instance, if a and b are the masks of all remaining inputs and outputs of an S-box S (that is to say, that we only know the value $x \wedge a$ from the input x, and that we are only interested in the value $y \wedge b$ from the output y), we compute the matrix of all

$$T_{i,j} = \Pr_{X \text{ uniform}} [S(X) \wedge b = j / X \wedge a = i] .$$

We use tools from tensorial algebra to compute the transition matrix of a network of S-boxes: the transition matrix of $(x, y) \mapsto (F(x), G(x))$ is the tensorial product (also called the Kronecker product) of the transition matrix of F and the transition matrix of G, and the transition matrix of $F \circ G$ is the matrix-product of the transition matrices of F and G whenever the output mask of G is the input mask of F. Assuming that M^i and M^o are the masks of the remaining inputs and outputs of the whole encryption function, we obtain all values

$$v_k^x = \Pr_{P \in \mathcal{P}, C = \mathrm{Enc}_k(P)} \left[(P \wedge M^i, C \wedge M^o) = x\right] - \frac{1}{q}$$

for all q possible values of x under the heuristic assumption, depending on the key k. This forms the *bias vector* $V_k = (v_k^{x_1}, v_k^{x_2}, \ldots)$ of the distribution of $X = (P \wedge M^i, C \wedge M^o)$ with respect to the uniform distribution. In the following sections, we consider a more general X of the form

$$X = h_3(h_1(k), h_2(P, C))$$

depending on a small piece of information $h_1(k)$ on k and a small piece of information $h_2(P, C)$ on the pair (P, C).

2.2 Projection of DES

DES is based on the Feistel scheme illustrated on Figure 1. There are two 32-bits registers L and R modified at each round by a process which depends on a subkey k_i depending on the master key k (see [1]).

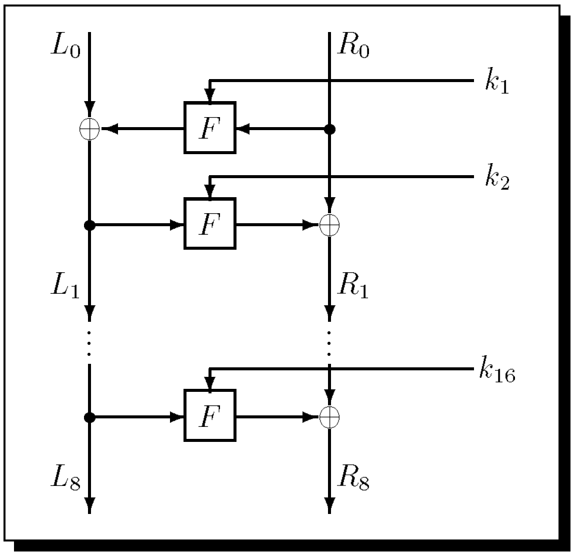

Figure 1: Feistel's scheme with 16 rounds

As an example, we ignore the same 27 bits of the left register of DES and the same 31 bits of the right one between the second and the fifteenth round (the reason why we use only the 14 middle rounds will appear in the next Sections). More precisely, for all masks m_L and m_R such that $W(m_L) = 5$ and $W(m_R) = 1$, we only keep the information $(L_i \wedge m_L, R_i \wedge m_R)$ in each round, that is 6 bits of information. We computed the $2^6 \times 2^6$ bias vector $V_k(m_L, m_R)$ of

$$X = (L_1 R_0 \wedge m_L m_R, L_8 R_7 \wedge m_L m_R).$$

Experiments shows that the norm $||V_k(m_L, m_R)||_2$ which we call *deviation* does not depend significantly on k provided it is large. Thus, trying all the possible positions of the 6 bits kept, we have found the best choices of the bit positions the first of are:

m_L	m_R	$\log_2 \|\|V_k\|\|_2$
$\mathtt{21040081}_{16}$	$\mathtt{00008000}_{16}$	-25.580
$\mathtt{21040082}_{16}$	$\mathtt{00008000}_{16}$	-25.583
$\mathtt{21040084}_{16}$	$\mathtt{00008000}_{16}$	-25.583
$\mathtt{21040088}_{16}$	$\mathtt{00008000}_{16}$	-25.583
$\mathtt{21040090}_{16}$	$\mathtt{00008000}_{16}$	-25.583
$\mathtt{210400a0}_{16}$	$\mathtt{00008000}_{16}$	-25.583
...	...	...

This shows that the best choices are exactly those which contain the pattern

$$m_L = \mathtt{21040080}_{16} \quad m_R = \mathtt{00008000}_{16}$$

(for which $\log_2 ||V_k||_2 = -24.583$) that is the bits used in Matsui's attack [8]. Trying all the 4 and 2 positions achieved an analogous result. Trying all the 3 and 3 positions did not provide any larger deviation.

The best other choice which does not contain Matsui's characteristic is:

$$m_L = \texttt{04010104}_{16} \quad m_R = \texttt{00c00000}_{16}$$

for which $\log_2 ||V_k||_2 = -30.768$.

2.3 Information on the key leaked

To see how much $V_k(m_L, m_R)$ depends on k, we studied how many different vectors we get with different k. Linear cryptanalysis only consider a one-bit long value X. Thus, the bias vector V_k has the form $(-\delta, \delta)$ and there are only two different vectors $(\mp\delta, \pm\delta)$, depending on one bit of information on k. Moreover, keys which produce the same bit of information are in the same affine space with codimension 1.

More generally, when a characteristic defined by (m_L, m_R) contains c linear characteristics, the vector $V_k(m_L, m_R)$ depends on c bits of information on k. Thus, there are 2^c different vectors, and keys which produce the same one are in the same affine space with codimension c. To study the nature of the information on k which influences the vector, we compute all the affine spaces spanned by random keys which produce the same vector. For instance, with the characteristic defined by

$$m_L = \texttt{21040080}_{16} \quad m_R = \texttt{00008000}_{16}$$

for which $\log_2 ||V_k||_2 = -24.583$ the experiment shows 4 different vectors $V_k(m_L, m_R)$. Thus, the key space is partitioned into 4 classes, and we can prove that each class spans an affine space with codimension 2. We already know that Matsui's linear characteristic defines one bit of information on k which is computed linearly:

$$\text{Parity}\left(\bigoplus_{i\in\{3,5,7,9,11,13,15\}} k_i \wedge \texttt{0000000020000000}_8 \quad \oplus \right.$$
$$\left. \bigoplus_{i\in\{4,8,12\}} k_i \wedge \texttt{0400000000000000}_8 \right).$$

We have found another bit of information which influences the vector:

$$\text{Parity}\left(\bigoplus_{i\in\{2,4,6,8,10,12,14\}} k_i \wedge \texttt{0400000000000000}_8\right).$$

Using Matsui's notations, those bits are respectively

$$k_3[22] \oplus k_4[44] \oplus k_5[22] \oplus k_7[22] \oplus k_8[44] \oplus k_9[22]$$
$$\oplus k_{11}[22] \oplus k_{12}[44] \oplus k_{13}[22] \oplus k_{15}[22]$$

and

$$k_2[44] \oplus k_4[44] \oplus k_6[44] \oplus k_8[44] \oplus k_{10}[44] \oplus k_{12}[44] \oplus k_{14}[44]$$

With the characteristic defined by

$$m_L = \texttt{04010104}_{16} \quad m_R = \texttt{00c00000}_{16}$$

we observed 16 different classes. We observed that keys in the same class spanned an affine space with codimension 4. Thus, this characteristic uses 4 linear bits on k.

3 Statistical cryptanalysis

3.1 Model of the attack

In the model of the attack[1], the concept of *characteristic* defines three hash functions:

- $h_1 : \mathcal{K} \rightarrow \mathcal{L}$ where $\mathcal{L}$ is a small space with cardinality ℓ (the aim of the cryptanalysis is to obtain probabilistic information on $k' = h_1(k)$);
- $h_2 : \mathcal{P} \times \mathcal{C} \rightarrow \mathcal{S}$ where $\mathcal{S}$ is the *sample space* with cardinality s which only contains useful information for the analysis;
- $h_3 : \mathcal{L} \times \mathcal{S} \rightarrow \mathcal{Q}$ where $\mathcal{Q}$ is a space with cardinality q.

For a random sample $S = h_2(P, C)$ coming from random P and $C = \mathrm{Enc}_k(P)$, we let $X = h_3(k', S)$, $k' = h_1(k)$. Basically, X is a piece of information depending on the intermediate results in the encryption. For the purpose of the cryptanalysis, X should be both

- computable with small pieces of information on (P, C) and k, namely S and k',
- and sufficiently biased for $X = h_2(k', S)$ (where $k' = h_1(k)$) to be statistically distinguishable from the distribution of $h_2(K, S)$ comming from a wrong guess $K \neq k'$.

The principle of the attack consists in seeking for the good k' which makes the distribution of all the observed X deviate significantly from a smooth distribution. In the example of DES, h_3 is a mask over messages obtained after the first round and before the last round, and h_1 and h_2 give the information required to compute it.

We assume that we can use several independent samples $S = h_2(P, C)$, given that P follows a given distribution H in the domain $\mathcal{P}$ and such that $C = \mathrm{Enc}_k(P)$

[1] This model appears to be similar to Harpes's partitioning cryptanalysis [5].

with the unknown k. The attack is a known or chosen plaintext attack depending on whether H corresponds to an available real plaintext distribution or not. It may be a ciphertext only attack when S is computable from C. For all candidates K to $k' = h_1(k)$, we can compute a candidate $X = h_3(K, S)$ to $h_3(k', S)$. The main idea of the attack consists in assuming that we can distinguish $K = k'$ from $K \neq k'$ by a statistical measurement Σ on the observed distribution. In most cases, for $K = k'$, this distribution will look *less* regular than for $K \neq k'$. The attack proceeds in four phases:

- **Counting Phase.** Collect several random samples $S_i = h_2(P_i, C_i)$, $i = 1, \ldots, n$. This consists in counting all occurrences of all the possible values of S in s counters.
- **Analysis Phase.** For each of the ℓ candidates K, count all the occurrences in all $X_i = h_3(K, S_i)$ and give it a mark M using the statistic $\Sigma(X_1, \ldots, X_n)$. Hereafter n_x denotes (for a given K) the number of samples such that $h_3(K, S_i) = x$.
- **Sorting Phase.** Sort all the candidates K using their marks M_K.
- **Searching Phase.** Exhaustively try all keys following the sorted list of all the candidates.

The space complexity is $O(s+\ell)$ since we need s counters for all $S = h_2(P, C)$ and ℓ registers for all candidates K. The time complexity is $O(n)$ for the Counting Phase, $O(s\ell)$ for the Analysis Phase and $O(\ell \log \ell)$ for the Sorting Phase. The average complexity of the Searching Phase, which depends on the expected rank of the good candidate in the sorted list, will be discussed below. Typically, the bottleneck computations are the Counting Phase and the Searching Phase, and we need to study the trade-off between them: we need many samples to expect the good candidate to have a high rank, but not too many to be able to count them.

3.2 Analysis of the attack

We make several approximations which might be justified by heuristic arguments in concrete examples. We recall that H denotes the distribution of the random plaintext source.

Approximation 1. *If $K \neq h_1(k)$, the distribution $h_3(K, H)$ of*

$$X = h_3(K, h_2(P, \mathrm{Enc}_k(P)))$$

is a distribution D which does not depend on K.

ϵ	0	2^{-2}	2^{-1}	1	2	2^2	2^3
$\Phi\left(-\epsilon/\sqrt{2}\right)$	50%	43%	36%	24%	$8\% = 2^{-3.7}$	$2^{-8.7}$	$2^{-27.0}$

Figure 2: Decreasing of the normal law

Approximation 2. *If $K = h_1(k)$, the distribution of X is a distribution D' which is independent on D.*

Typically, D is the uniform distribution in the domain $\mathcal{Q}$ with cardinality q. We call *deviation* between D and D' the value

$$d_2(D, D') = \sqrt{\sum_x \left(\Pr_{X \in D'}[X = x] - \Pr_{X \in D}[X = x]\right)^2}.$$

The accurate analysis depends on the choice of the statistic Σ, but we give here the outline of the analysis. We denote μ and σ (resp. μ' and σ') the mean and the standard deviation of $\Sigma(X_1, \ldots, X_n)$ all X_i following the distribution D (resp. D'). In the rest of this paper, we make another Approximation.

Approximation 3. *We have $\sigma \approx \sigma'$ and $\mu \not\approx \mu'$.*

The mark M_K of K is defined to be

$$M_K = \frac{\Sigma(h_3(K, S_1), \ldots, h_3(K, S_n)) - \mu}{\sigma}$$

so, the standard deviation of any mark is 1, the expected mark of a wrong candidate is 0, and the expected mark of the good candidate is $\epsilon = \frac{\mu' - \mu}{\sigma}$ which will be called the *efficiency* of the attack. In real applications, Approximation 3 may corresponds to a first order approximation.

Let

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{t^2}{2}} dt.$$

be the normal distribution function. In the following, the sentence "*the distribution of M is asymptotically normal*" means that

$$\Pr\left[\frac{M - E(M)}{\sigma(M)} < t\right] \to \Phi(t)$$

when the number of samples is large.

Theorem 1. *Under the Approximations and if the distribution of all M_K are asymptotically normal, the average complexity of the Searching Phase tends to*

$$\frac{N}{\ell} + \left(N - \frac{N}{\ell}\right) \Phi\left(-\epsilon/\sqrt{2}\right)$$

where N is the number of keys k.

This will be practically approximated by $N.\Phi\left(-\epsilon/\sqrt{2}\right)$.

Proof. The mark of the good candidate $k' = h_1(k)$ with samples $S_1, \ldots, S_n$ is

$$M_{k'} = \frac{\Sigma(h_3(k', S_1), \ldots, h_3(k', S_n)) - \mu}{\sigma}$$

which is approximately normal, with mean ϵ and standard deviation 1. The mark of any wrong candidate $K \neq h_1(k)$ is

$$M_K = \frac{\Sigma(h_3(K, S_1), \ldots, h_3(K, S_n)) - \mu}{\sigma}$$

which is approximately standardized and normal. $M_{k'}$ and M_K are independent by Approximation 2, so $\frac{M_{k'} - M_K - \epsilon}{\sqrt{2}}$ is standardized and normal. Thus, the probability that the $M_{k'}$ is less than M_K is $\Phi\left(-\epsilon/\sqrt{2}\right)$. The rank of k' is

$$1 + \sum_K 1_{M_{k'} < M_K}$$

so the expected rank of k' is $1 + (\ell - 1).\Phi\left(-\epsilon/\sqrt{2}\right)$ and the average complexity of the Searching Phase is obtained multiplying this by $\frac{N}{\ell}$. $\square$

The decreasing of $\Phi\left(-\epsilon/\sqrt{2}\right)$ is illustrated on the table on Figure 2.

3.3 The use of several characteristics

It is possible to use several characteristics C_i (or the same one several times) with efficiency ϵ_i for $i = 1, \ldots, c$ using a trick analog to the one analyzed by Kaliski and Robshaw [7]. We get lists of several candidates so that each full key K have marks M_K^i. We let

$$\bar{\epsilon} = \sqrt{\sum_{i=1}^{c} {\epsilon_i}^2} \qquad (1)$$

and we define the general mark

$$M_K = \sum_{i=1}^{c} \frac{\epsilon_i}{\bar{\epsilon}} M_K^i.$$

We can do the exhaustive search following the general marks. It is easy to prove that the Theorem 1 remains valid if we replace ϵ by $\bar{\epsilon}$ when all M_K^i are independent. Thus, it is possible to slightly improve the best known linear attack on DES collecting a huge number of less efficient characteristics.

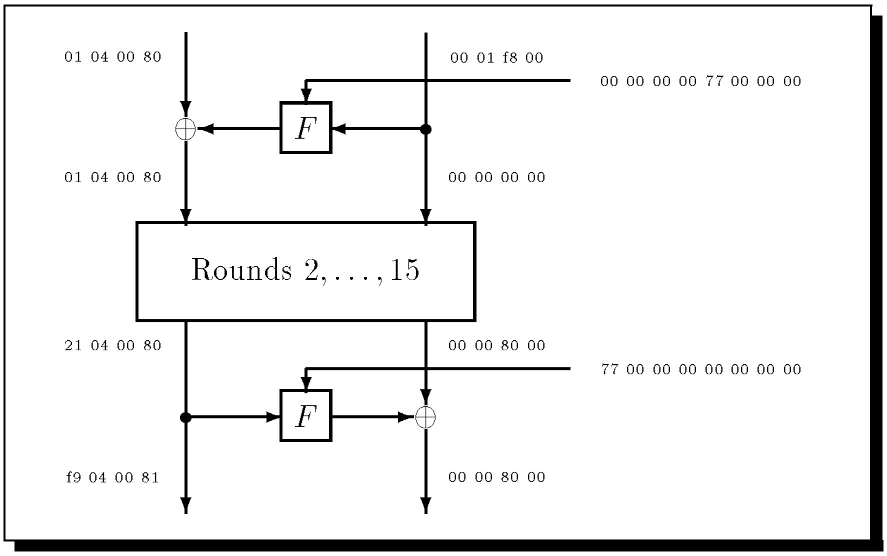

Figure 3: Matsui's characteristic

4 Differential approach

For a given nonzero a, the statistic Σ_{diff} counts the number of sample pairs (X_i, X_j) such that $X_i \oplus X_j = a$:

$$\Sigma_{\text{diff}}(X_1, \ldots, X_n) = \sum_{i,j=1}^{n} 1_{X_i \oplus X_j = a} = \sum_{x \oplus y = a} n_x n_y.$$

For vectors a coming from a *differential characteristic*, a heuristic analysis from Biham and Shamir enables to approximate (for $i \neq j$)

$$\Pr_{X_i, X_j \in D'}[X_i \oplus X_j = a] - \Pr_{X_i, X_j \in D}[X_i \oplus X_j = a] = \delta.$$

Theorem 2. *If D is uniform over $\mathcal{Q}$, the efficiency of the attack using Σ_{diff} is*

$$\epsilon \approx n \frac{q}{\sqrt{2(q-1)}} \delta \leq n \frac{q}{\sqrt{2(q-1)}} (d_2(D, D'))^2.$$

Proof. We have $\mu' - \mu = n(n-1)\delta$ and

$$\sigma = \frac{\sqrt{n(n-1)}\sqrt{2(q-1)}}{q}$$

so we get ϵ. Using Cauchy-Schwarz's Inequality, we have

$$|\delta| = \left| \sum_{x \oplus y = a} v_k^x v_k^y \right| \leq (d_2(D, D'))^2$$

where v_k^x is defined in Section 2. □

5 Linear approach

5.1 Linear cryptanalysis

For a given nonzero a, the statistic Σ_{lin} counts the number of samples X_i such that the dot product $X_i \cdot a$ is zero:

$$\Sigma_{\mathrm{lin}} = \sum_{i=1}^{n} 1_{X_i \cdot a = 0} = \sum_{x \cdot a = 0} n_x.$$

For vectors a coming from a *linear characteristic*, a heuristic analysis from Matsui enables to approximate

$$\Pr_{X_i \in D'}[X_i \cdot a = 0] - \Pr_{X_i \in D}[X_i \cdot a = 0] = \delta.$$

Theorem 3. *If D' is uniform over $\mathcal{Q}$, the attack using Σ_{lin}, which is asymptotically normal, is*

$$\epsilon = \sqrt{n}.2\delta \leq \sqrt{nq}.d_2(D, D').$$

This Theorem will be proved in a more general form below.

5.2 Matsui's attack against DES

As illustrated by Figure 3, Matsui's characteristic is defined by

$$\begin{aligned} k' &= \begin{pmatrix} k_1 & \wedge & \mathtt{0000000077000000}_8 \\ k_{16} & \wedge & \mathtt{7700000000000000}_8 \end{pmatrix} \\ S &= \begin{pmatrix} L_0R_0 & \wedge & \mathtt{010400800001f800}_{16} \\ L_8R_8 & \wedge & \mathtt{f904008100008000}_{16} \end{pmatrix} \\ X &= \begin{pmatrix} L_1 & \wedge & \mathtt{01040080}_{16} \\ L_8R_7 & \wedge & \mathtt{2104008000008000}_{16} \end{pmatrix} \end{aligned}$$

with the notations used in Section 2 and where L_1, L_8 and R_7 are computed from $P = (L_0, R_0)$ and $C = (L_8, R_8)$ using k [8]. It is easy to see that k' and S are sufficient to compute X. For reasons related to the structure of F, the masks on the subkeys are coded in octal while the masks on the message registers are coded in hexadecimal. We have $\ell = 2^{12}$ (this is the number of candidates K), $s = 2^{19}$ (number of possible samples S) and $q = 2^8$. The bias is approximated by tricky heuristic arguments to $|\delta| = 1.19 \times 2^{-21}$. Using $n = 2^{43}$, we have $\epsilon = 3.37$. Therefore, using two such characteristics as Matsui did (that is using it together with its reversed characteristic obtained by exchanging the left and the right masks), the global efficiency is given by the Equation (1) and the exhaustive search gets its complexity improved by a factor $\Phi(-3.37) = 2^{-11.4}$. The complexity of the Searching Phase is evaluated to $2^{44.6}$. (Matsui's experiment would have yielded complexity 2^{43}, so Theorem 1 may be a little pessimistic, but we notice that the approximation $\epsilon \approx \sqrt{nq}d_2(D, D') = 3.78$ yields complexity $2^{42.38}$.)

Experiment	1	2	3	4	5	6	7	8	9	10
Matsui's attack	1	280	1	2	59	10	12	35	1	4
linear mark #1	1	46	1	2	205	48	8	58	2	4
linear mark #2	1	46	1	2	204	81	11	59	2	4
linear mark #3	1	100	1	2	205	48	8	60	2	6
linear mark #4	1	100	1	2	206	80	11	57	2	6
linear mark #5	1	45	1	2	103	79	8	58	2	4
linear mark #6	1	44	1	2	104	48	11	58	2	6
linear mark #7	1	101	1	2	104	48	11	57	2	6
linear mark #8	1	100	1	2	103	80	8	57	2	4
$\sum$(linear marks)^2	1	68	1	2	140	57	10	55	2	6
χ^2 attack	100	2221	516	197	435	1294	3667	2389	335	1320

Figure 4: Experiment of attacks on 8-rounds DES

5.3 Generalized linear test

We can generalize Matsui's statistic by any linear one using suitable a_x:

$$\Sigma_{\text{glin}} = \sum_{i=1}^{n} a_{X_i} = \sum_x a_x n_x.$$

Theorem 4. *Σ_{glin} is asymptotically normal. The best efficiency is obtained with $a_x = v_k^x$. If D is uniform over $\mathcal{Q}$, it is*

$$\epsilon = \sqrt{nq}.d_2(D, D').$$

Proof. We have

$$\mu' - \mu = n \sum_x a_x v_k^x$$

and

$$\sigma = \sqrt{n}\sqrt{\sum_x (a_x - \bar{a})^2 \Pr_{X\in D}[X = x]}$$

where $\bar{a} = \sum_x a_x \Pr_{X\in D}[X = x]$. Σ_{glin} is asymptotically normal, due to the central limit Theorem [4]. Thus, we have

$$\epsilon^2 = n \frac{\left(\sum_x a_x v_k^x\right)^2}{\sum_x (a_x - \bar{a})^2 \Pr_{X\in D}[X = x]}.$$

The Theorem comes from Cauchy-Schwarz's Inequality in the particular case where we have $\Pr_{X\in D}[X = x] = \frac{1}{q}$. $\square$

The problem of using the best linear statistic is similar to the problem of linear cryptanalysis, where we have to guess a vector a coming from a linear characteristic. Here, we have to bet on the transition matrix to get all a_x. If there are only few possible transition matrices, we use the sum of the squares of all the linear marks as a new statistic, which turns out to be almost as efficient as the best one. (The reason why we use the sum-of-squares is that the different marks are linearly dependent, so a linear mean would be suject to strange cancellations.)

5.4 A slight improvement of Matsui's attack

For Matsui's symmetrized characteristic which is defined by

$$\begin{aligned} h_1(k) &= \begin{pmatrix} k_1 & \wedge & \texttt{0000000077000000}_8 \\ k_{16} & \wedge & \texttt{7700000000000000}_8 \end{pmatrix} \\ h_2(P,C) &= \begin{pmatrix} L_0R_0 & \wedge & \texttt{210400800001f800}_{16} \\ L_8R_8 & \wedge & \texttt{f904008100008000}_{16} \end{pmatrix} \\ h_3(k',S) &= \begin{pmatrix} L_1R_0 & \wedge & \texttt{2104008000008000}_{16} \\ L_8R_7 & \wedge & \texttt{2104008000008000}_{16} \end{pmatrix} \end{aligned}$$

we have $s = 2^{20}$, $\ell = 2^{12}$ and $q = 2^{10}$. Using the heuristic with projections, the deviation has been approximated to $d_2(D, D') \approx 2^{-24.58}$. Hence, using $2^{42.93}$ known plaintext/ciphertext couples (instead of $2^{43.00}$, which is 5% larger), we obtain $\epsilon = 3.69$. With two such characteristics, the exhaustive search is improved by a factor $\Phi(-3.69) = 2^{-13.14}$ and the Searching Phase gets a complexity $2^{42.86}$. Since off-line exhaustive search is cheaper than getting a new sample, we can afford 2^{42} known plaintexts which gives $\epsilon = 3.78$ then $2^{56}.\Phi(-3.78) = 2^{47.93}$: 2^{42} known plaintexts enables to find the key within a 2^{48} average complexity.

For eight-rounds DES, Matsui announced 1.49×2^{17} knwon plaintexts, but it was to get the same complexity than for sixteen-rounds DES in the exhaustive search, that is 2^{43}. Here, we have $d_2(D, D') \approx 2^{-11.86}$, so, with 2^{17} known plaintexts, we have $\epsilon = 3.11$ and the exhaustive search has complexity $2^{56}.\Phi(-3.11) = 2^{45.93}$ with two characteristics. (With 2^{18} known plaintexts, the same computation yields complexity $2^{38.50}$.) This attack has been implemented.

Experiments show there are only eight kinds of bias vector V_K. We use as a statistic the sum-of-the-squares of the eight marks obtained with the eight corresponding linear statistics. With the only characteristic defined in this Section, we have $\ell = 2^{12}$ candidates and the rank of the good candidate in the sorted list should be $1 + \ell.\Phi(-\epsilon/\sqrt{2})$ on average. For $n = 2^{17}$ samples, we have $\epsilon = 3.11$ so the average rank should be 57.86. Ten random experiments yielded ranks illustrated on Figure 4. We put ranks obtained by Matsui's mark, by each of the eight linear marks, by the sum-of-squares of the linear marks, and by the χ^2 mark we will present on next Section. This shows the use of the best linear statistic slightly improves Matsui's attack. It also confirms that the χ^2 attack is a little less efficient than the other attacks, as we will prove.

6 χ^2 cryptanalysis

The deviation from the uniform distribution in a domain with cardinality q can be tested using the χ^2 test [4]:

$$\Sigma_{\chi^2} = \frac{q}{n} \sum_x \left(n_x - \frac{n}{q} \right)^2 = \frac{q}{n} \sum_x {n_x}^2 - n.$$

Theorem 5. *Under the hypothesis, the efficiency of the attack using Σ_{χ^2} is*

$$\epsilon \approx n \frac{q}{\sqrt{2(q-1)}} (d_2(D, D'))^2.$$

Proof. For bad candidates, the statistic Σ_{χ^2} tends to the χ^2 distribution with $q-1$ degrees of freedom: $\mu = q - 1$ and $\sigma = \sqrt{2(q-1)}$. When the degree of freedom is large, this distribution can be approximated by a normal one.

Let

$$\Sigma'_{\chi^2} = \frac{q}{n} \sum_x \left(n_x - \frac{n}{q} - n v_k^x \right)^2.$$

Σ'_{χ^2} is a kind of χ^2 statistic such that

$$E(\Sigma'_{\chi^2}) = q - 1 - q(d_2(D, D'))^2.$$

So, we have

$$\Sigma_{\chi^2} = \Sigma'_{\chi^2} + 2d_2(D, D').\sqrt{qn}\Sigma_{\text{glin}} - nq(d_2(D, D'))^2$$

where Σ_{glin} is the best standardized linear test (*i.e.* with $E(\Sigma_{\text{glin}}) = 0$ and $\sigma(\Sigma_{\text{glin}}) = 1$). So, we have

$$\mu' = q - 1 + (n-1)q(d_2(D, D'))^2$$

which allows to compute ϵ. □

A straightforward consequence of this Theorem is that with the same characteristic and the same number a plaintext/ciphertext couples, the χ^2 cryptanalysis is more efficient than the differential cryptanalysis (which is a quadratic statistic).

Using this statistic, we do not need to have a precise idea of which information is leaked throughout Enc, such as what would have been done in linear or differential cryptanalysis using a particular vector a. Here, we use a characteristic, and if there exists a powerful subcharacteristic according to linear of differential cryptanalysis, the χ^2 test is able to detect it and to use it to distinguish the good k'.

For instance, we can try Matsui's symmetrized characteristic with the χ^2 cryptanalysis. We have $q = 2^{10}$ and $d_2(D, D') \approx 2^{-24.58}$. Using $2^{46.2}$ known plaintexts (9 times as Matsui does), we get $\epsilon = 2.90$. Here the χ^2 variable is approximately normal. So, using two such characteristics, we get the average complexity $2^{56}.\Phi(-2.90) = 2^{46.9}$.

7 Conclusion

We have shown that differential and linear cryptanalysis can be viewed in a more statistical approach. It is possible to join the efforts of several characteristics to improve them. Both attacks can be improved using an additional information, that is the vector of all v_k^x. Conversely, with less knowledge about the characteristic (that is without the precise knowledge of which bits of the input and the output play a role and what happens in between), the χ^2 cryptanalysis performs an attack which is roughly as efficient.

To prove that the linear aspects of differential or linear cryptanalysis are not unavoidable, we presented a new heuristic method which has produced the same attack than Matsui's. This leads to new directions in cryptanalysis. We hope that this new approach and the experiments presented in this paper will motivate further investigations in the use of statistic experiments in cryptanalysis.

Acknowledgment

We wish to thank Eli Biham, Don Coppersmith, Carlo Harpes, Lars Knudsen and Jacques Stern for fruitful discussions.

References

[1] U. S. National Bureau of Standards. Data Encryption Standard. Federal Information Processing Standard Publication 46, 1977.

[2] E. Biham, A. Shamir. Differential Cryptanalysis of the full 16-round DES. In *Advances in Cryptology CRYPTO'92*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 740, pp. 487–496, Springer-Verlag, 1993.

[3] E. Biham, A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.

[4] H. Cramer. *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[5] C. Harpes. Partitioning Cryptanalysis. Post-diploma thesis, ISI, ETH Zurich, 1994.

[6] M. E. Hellman, R. Merkle, R. Schroeppel, L. Washington, W. Diffie, S. Pohlig, P. Schweitzer. *Results of an Initial Attempt to Cryptanalyze the NBS Data Encryption Standard*, Stanford University, September 1976.

[7] B. R. Kaliski Jr., M. J. B. Robshaw. Linear Cryptanalysis using Multiple Approximations. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 839, pp. 26–39, Springer-Verlag, 1994.

[8] M. Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lectures Notes in Computer Science 839, pp. 1–11, Springer-Verlag, 1994.

[9] S. Vaudenay. *La sécurité des primitives cryptographiques.* Thèse de Doctorat de l'Université de Paris 7, Technical Report LIENS-95-10 of the Laboratoire d'Informatique de l'Ecole Normale Supérieure, 1995.