On Measuring Resistance to Linear Cryptanalysis

Serge Vaudenay

EPFL

Abstract. Linear cryptanalysis against cryptographic primitives C is known to rely on some $\mathrm{LP}^C_{\max}$ term. But most of studies so far are purely heuristic and only provide an argument on why linear cryptanalysis works. Other works provide an asymptotic bound without any clue where it is applicable for practical parameters. So there is still some doubt for the designer on whether making a low $\mathrm{LP}_{\max}$ term is enough or not.
In this paper we formally demonstrate that the efficiency of linear cryptanalysis is uniformly bounded, on average, by $\mathrm{MAXELP}(C)$ which is the maximum of the expected value of the linear probability LP^C. We further discuss on how pairwise independent random primitives can provably resist to these attacks.
This result provides insurance for the designer that making a primitive pairwise independent, or with a low MAXELP measure is enough to protect against linear cryptanalysis. It also provides a quantitative evaluation tool for security evaluation.

1 Introduction

Cryptographic primitives, such as symmetric encryption of message authentication codes, are cheap algorithms which are used in order to protect the confidentiality or the authenticity of digital information. They are initially set up by a secret key which is selected at random by authorized parties.

Some greedy schemes which use one key per operation, like the one-time pad (which is due to Vernam [35]) or the Wegman-Carter authentication code [36] provide a provable perfect security, but at an unreasonable cost in terms of key distribution. Since Shannon [28] proved that perfect secrecy is not possible in a cheaper way in an information theoretic sense, the only alternative is to base security on the ability limits for complexity: a scheme is secure if no attacker is able to mount an attack. Unfortunately, complexity lower bounds lead to too hard problems like the P vs NP problem. So security of cryptographic primitives seems to be bound to heuristic approaches.

Releasing the *Data Encryption Standard* (DES) [1] in the late 70's motivated researchers to work on cryptographic analysis. Real advances on the attack strategies on block ciphers have been made in the early 90's when Biham and Shamir invented the *differential cryptanalysis* and applied it against DES [3,4]. They later prove that DES developers actually knew this technique and designed

DES in order to resist to it. Matsui later developed the *linear cryptanalysis* which was more successful on DES [18,19]. This heuristic attack, which has been implemented, can recover the key with a 2^{43}-known plaintext attack. Since then, many researchers tried to generalize and to improve these attacks (see for instance [9,10,11,12,13,16,17,21,30,31]).

The basic idea of linear cryptanalysis is to use the probability $\Pr[a \cdot X = b \cdot C(X)]$ for two given constants a and b where $\cdot$ denotes the Boolean inner product (*i.e.* the parity of the bitwise AND). This probability should be close to $1/2$ if C were perfect. Linear cryptanalysis exploits the distance between this probability and $1/2$ when it is large enough. Indeed we define $\mathrm{LP}(a,b) = (2\Pr[a \cdot X = b \cdot C(X)] - 1)^2$. More precisely, linear cryptanalysis is an incremental one-known plaintext attack where we simply measure the correlation between the bits $a \cdot X$ and $b \cdot C(X)$. The complexity of this attack was heuristically proven to be $\Omega(1/\mathrm{LP}(a,b))$.

Inspired by the Nyberg [22] notion of resistance to differential cryptanalysis, Chabaud and Vaudenay formalized the notion of strength against linear cryptanalysis [5] by using $\mathrm{LP}_{\max}$ defined to be the maximum of $\mathrm{LP}(a,b)$ over all possible choices for a and b. However the link between the resistance and this quantity was always heuristic, so that it was not formally proven that having a successful attack and a very low $\mathrm{LP}_{\max}$ measure is impossible. In [33], Vaudenay proves an asymptotic bound, but it was only asymptotic and there was no clue how high the parameter had to be so that the bound was realistic.

We solve this problem in this paper. We define MAXELP as the maximum, over all a and b, of the expected value, over the distribution of the random key, of $\mathrm{LP}(a,b)$. We prove that the complexity of attacks are lower bounded by a function of $1/\mathrm{MAXELP}$. This demonstrates that if one wish to design a new cryptographic primitive which provably resists to linear cryptanalysis, it is enough to make sure that MAXELP is low. We further show that such a low MAXELP measure comes for free when we use pairwise independent random functions [6].

1.1 Related Work

Several researchers concentrated on the positive side of cryptanalysis: security arguments. Usually block cipher designers try to upper bound the probability of the best differential or linear characteristics in ad-hoc ways. Some results apply to multi-path characteristics like Nyberg-Knudsen [23,24], Aoki-Ohta [2], Keliher *et al.* [14,15], Park *et al.* [25,26].

1.2 Notations

In what follows we use the following notations:

$\mathcal{M}^d$**:** set of all sequences which consist of d elements of a set $\mathcal{M}$,
$\mathrm{Adv}_{\mathcal{A}}$**:** advantage of a distinguisher $\mathcal{A}$ (see Section 2.1),
1_P**:** variable which is set to 1 if the predicate P is satisfied or to 0 otherwise,
$\mathrm{LP}^c(a,b)$**:** linear probability of a function c with characteristic (a,b) (see Section 2).
$\mathrm{MAXELP}(C)$**:** maximum expected linear probability of a random function C (see Section 3.1).

We represent all random variables by capital letters. They are associated to a probability distribution which will be clear from the context. For instance, X may denote a random variable and $\Pr[X = x]$ may represent the probability that it takes a given value x.

Random functions or permutations will be considered. They will be represented by random variables, *e.g.* F or C.

2 Linear Cryptanalysis

2.1 Full Linear Cryptanalysis

Linear cryptanalysis has been invented by Matsui [18,19] based on the notion of statistical attacks which are due to Gilbert *et al.* [7,8,29]. Full linear cryptanalysis against an encryption process Enc rely on some distinguished property for an internal permutation C, following the idea of 1R, 2R or 3R attack of Biham and Shamir [3,4]. More precisely, the encryption can be written

$$\mathrm{Enc} = C_{\mathrm{post}} \circ C \circ C_{\mathrm{pre}}$$

where C_{pre} and C_{post} are some simple pre and post processing. It usually consists of a few rounds of encryption for which we can mount a dedicated attack as it will be discussed below. Note that the three components of the encryption here are random permutations defined by random (sub)keys.

Then, we use a distinguisher between the random permutation C and a perfect random permutation C^*. The distinguisher is an algorithm which sends queries to an oracle and eventually outputs either 0 or 1. We compute the probability p (resp. p^*) that the algorithm outputs 1 when the oracle implements C (resp. C^*). The power of the distinguisher is quantified by the *advantage* $\mathrm{Adv} = p - p^*$. Good distinguishers are characterized by a high values of $|\mathrm{Adv}|$.

In order to make the attack practical, the distinguisher needs to use a piece of information on the inputs and outputs of C which can be computed from

the plaintext and the ciphertext of Enc, and some small piece of information on the secret key through the pre and post encryption. (This is what we meant by "a few rounds of encryption for which we can mount a dedicated attack".) For linear cryptanalysis, we use linear distinguishers which relies on some statistical properties of the Boolean $a \cdot C_{\mathrm{pre}}(X) \oplus b \cdot C_{\mathrm{post}}^{-1}(\mathrm{Enc}(X))$ for random known plaintext X and we compute $a \cdot C_{\mathrm{pre}}(X)$ and $b \cdot C_{\mathrm{post}}^{-1}(\mathrm{Enc}(X))$ from X, $\mathrm{Enc}(X)$, and a piece of information $h(K)$ on the key K.

As explained in Matsui [18,19], the attack proceeds as follows. We exhaustively look for the value of $h(K)$. For every candidate, we make statistics on the Boolean information. We then sort all candidate according to the statistics. The attack works if the statistical behavior for wrong candidates looks like the statistical behavior of the perfect random permutation so that the distinguisher can isolate the right candidate from the others.

Therefore, the linear cryptanalysis cannot work for Enc if linear distinguishers have limited advantage against the internal permutation C. In the next sections we focus on linear distinguishers for C.

2.2 Linear Distinguishers

In this section we assume that $\mathcal{M} = \{0,1\}^m$. The inner dot product $a \cdot b$ in $\{0,1\}^m$ is the parity of the bitwise AND of a and b.

We call "basic linear distinguisher" the distinguisher characterized by a pair $(a,b) \in \mathcal{M}^2$ with $b \neq 0$ which is depicted on Fig. 1. We notice here that the attack depends on the way it accepts or rejects based on the final counter u value.

Parameters: a complexity n, a characteristic (a,b), a set $\mathcal{A}$
Oracle: a permutation c
1: initialize the counter value u to zero
2: **for** i from 1 to n **do**
3: pick a random X with a uniform distribution and query for $c(X)$
4: if $X \cdot a = c(X) \cdot b$, increment the counter u
5: **end for**
6: if $u \in \mathcal{A}$, output 1, otherwise output 0

Fig. 1. Linear Distinguisher.

As pointed out by Chabaud and Vaudenay [5], linear cryptanalysis against c is based on the quantity

$$\mathrm{LP}^c(a,b) = \left(2 \Pr_X[X \cdot a = c(X) \cdot b] - 1\right)^2.$$

(Here we use Matsui's notations taken from [20].)

2.3 Heuristic Analysis of Linear Distinguishers

When one wish to mount an attack against a cipher by using linear cryptanalysis techniques, one need to have a rough idea on how efficient it will be. For this a heuristic analysis is enough. Originally, Matsui [18,19] provided such an complexity estimate by using the Large Number Theorem. Indeed, if N_i is a Boolean random variable which is set to 1 if and only if the counter u is incremented in the distinguisher of Fig. 1, we notice that all N_i's are independent and with the same distribution defined by $z = \Pr[N_i = 1]$. The final value U of the counter is a random variable which can be approximated to a random variable with normal distribution of expected value $\mu = nz$ and standard deviation

$$\sqrt{n} \times \sqrt{\frac{1}{4} - \left(z - \frac{1}{2}\right)^2}.$$

When z is close to $\frac{1}{2}$ we can neglect terms of second order and approximate the variance to $\sigma = \frac{\sqrt{n}}{2}$. Hence we have

$$\Pr[U \leq x] \approx \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-nz)^2}{2\sigma^2}} dt.$$

As this is the case in all concrete examples, let us assume that C is a random cipher such that $z = \Pr_X[X \cdot a = C(X) \cdot b] = \frac{1}{2} + (-1)^{\kappa}\varepsilon$ where ε is constant, κ only depends on C and its parity is uniformly distributed. Note that $\mathrm{LP}^C(a,b) = 4\varepsilon^2$ is constant. When we use a uniformly distributed cipher C^* we can just do the same approximation with $\varepsilon \approx 0$. Hence, considering $\mathcal{A}$ as a continuous set, the advantage is

$$\begin{aligned} p - p^* &\approx \frac{1}{\sigma\sqrt{2\pi}} \int_{t\in\mathcal{A}} \left(\frac{e^{-\frac{\left(t-\frac{n}{2}-n\varepsilon\right)^2}{2\sigma^2}} + e^{-\frac{\left(t-\frac{n}{2}+n\varepsilon\right)^2}{2\sigma^2}}}{2} - e^{-\frac{\left(t-\frac{n}{2}\right)^2}{2\sigma^2}} \right) dt \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{t\in\mathcal{A}} e^{-\frac{\left(t-\frac{n}{2}\right)^2}{2\sigma^2}} \left(e^{-\frac{\alpha^2}{2}} \mathrm{ch}\frac{t-\frac{n}{2}}{\sigma}\alpha - 1 \right) dt \end{aligned}$$

where

$$\alpha = \sqrt{n.\mathrm{LP}^C(a,b)} = 2\varepsilon\sqrt{n}$$

from which we deduce that the advantage is optimal when $\mathcal{A}$ is the complement of $[\frac{n}{2} - \tau\sigma, \frac{n}{2} + \tau\sigma]$ where τ is such that $\mathrm{ch}(\alpha\tau) = e^{\frac{\alpha^2}{2}}$, i.e.

$$\tau = \frac{\alpha}{2} + \frac{1}{\alpha} \log\left(1 + \sqrt{1 - e^{-\alpha^2}}\right).$$

Note that the $u \in \mathcal{A}$ test is equivalent to $\left|u - \frac{n}{2}\right| > \tau\sigma$. Let us take the variable $x = \left(t - \frac{n}{2}\right)/\sigma$. We obtain

$$p - p^* \approx \frac{1}{\sqrt{2\pi}} \int_{-\tau}^{+\tau} \left(e^{-\frac{x^2}{2}} - \frac{1}{2} e^{-\frac{(x-\alpha)^2}{2}} - \frac{1}{2} e^{-\frac{(x+\alpha)^2}{2}} \right) dx$$

which is independent from n and ε. So with $n = \Theta\left(1/\mathrm{LP}^C(a,b)\right)$ we obtain a constant maximal advantage. As an illustration, here are a few values for α, τ, and $p - p^*$.

α	2^{-4}	2^{-3}	2^{-2}	$\frac{1}{2}$	1	2	4	2^3
τ	1.000	1.001	1.005	1.021	1.085	1.344	2.173	4.087
$p - p^*$	0.000945	0.00377	0.0150	0.0581	0.207	0.566	0.936	0.9999

Note that when $\alpha = o(1)$ we have $\tau \approx 1$ and $p - p^* = O(\alpha^2)$. So, in this situation, the best advantage is bounded by $O\left(n.E\left(\mathrm{LP}^C(a,b)\right)\right)$.

2.4 Analysis of Linear Distinguishers

In this section we concentrate on a fixed permutation c on $\{0,1\}^m$. Here is our main lemma.

Lemma 1. *For the distinguisher of Fig. 1 we let p^c be the probability that the output is 1 given an oracle c. We let p_0 be the probability that it outputs 1 when the counter is incremented with probability $\frac{1}{2}$ in every iteration instead of querying the oracle. We have*

$$|p^c - p_0| \leq 2\sqrt{n.\mathrm{LP}^c(a,b)}.$$

Furthermore, when n increases but $\mathrm{LP}^c(a,b) = o(\frac{1}{n})$, the maximum for $|p^c - p_0|$ is asymptotically equivalent to $\frac{1}{\sqrt{2\pi}}\sqrt{n.\mathrm{LP}^c(a,b)}$.

Proof. We first express the probability p^c that the distinguisher accepts c. Let N_i be the random variable defined as being 1 or 0 depending on whether or not we have $X \cdot a = c(X) \cdot b$ in the ith iteration. All N_i's are independent and with the same 0-or-1 distribution. Let z be the probability that $N_i = 1$. We also define $\theta = 2z - 1 = \sqrt{\mathrm{LP}^c(a,b)}$. We thus want to prove that $|p^c - p_0| \leq 2\theta\sqrt{n}$. We have

$$p^c = \sum_{u \in \mathcal{A}} \binom{n}{u} z^u (1-z)^{n-u}$$

thus

$$p^c - p_0 = \sum_{u \in \mathcal{A}} \binom{n}{u} \left(z^u (1-z)^{n-u} - \frac{1}{2^n} \right).$$

We would like to upper bound $|p^c - p_0|$ over all possible $\mathcal{A}$ depending on z. Since z and $1-z$ play a symmetric role we assume w.l.o.g. that $z \geq \frac{1}{2}$. For $z = \frac{1}{2}$, the result is trivially true, so from now on we assume that $z > \frac{1}{2}$. Since $z^u(1-z)^{n-u}$ is an increasing function in terms of u we have

$$\max_{\mathcal{A}} |p^c - p_0| = \sum_{u=k}^{n} \binom{n}{u} \left(z^u (1-z)^{n-u} - \frac{1}{2^n} \right)$$

where k is the least integer u such that the difference in parenthesis is non negative, *i.e.*

$$k = 1 + \left\lfloor n \frac{\log \frac{1}{2} - \log(1-z)}{\log z - \log(1-z)} \right\rfloor.$$

Replacing u by $\frac{n}{2}$ in the parenthesis we obtain a negative difference. Hence $k \geq \frac{n+1}{2}$. Similarly, replacing u by $n.z$, the parenthesis turns out to be an increasing function in terms of z which is 0 for $z = \frac{1}{2}$. Since $z > \frac{1}{2}$ we obtain that $k \leq \lceil n.z \rceil$. Therefore $\frac{n-1}{2} \leq k - 1 \leq (n-1)z + z$.

If $n = 1$, we have $k = 1$ thus $\max_{\mathcal{A}} |p^c - p_0| = z - \frac{1}{2}$ so the result holds. If $n = 2$, we have $k \geq \frac{3}{2}$ thus $k = 2$ and

$$\max_{\mathcal{A}} |p^c - p_0| = \left(z - \frac{1}{2} \right) \left(z + \frac{1}{2} \right) \leq \frac{3}{2} \left(z - \frac{1}{2} \right)$$

so the result holds as well. We now concentrate on $n \geq 3$.

We use the following identity taken from [27].[1]

$$\sum_{u=k}^{n} \binom{n}{u} z^u (1-z)^{n-u} = k \binom{n}{k} \int_0^z t^{k-1} (1-t)^{n-k} dt. \tag{1}$$

We obtain

$$\max_{\mathcal{A}} |p^c - p_0| = k \binom{n}{k} \int_{\frac{1}{2}}^z t^{k-1} (1-t)^{n-k} dt \tag{2}$$

thus

$$|p^c - p_0| \leq k \binom{n}{k} \left(z - \frac{1}{2} \right) \max_{t \in [0,1]} \left(t^{k-1} (1-t)^{n-k} \right).$$

The maximum is obtained for $t = \frac{k-1}{n-1}$ hence

$$|p^c - p_0| \leq k \binom{n}{k} \left(z - \frac{1}{2} \right) \frac{(k-1)^{k-1} (n-k)^{n-k}}{(n-1)^{n-1}}.$$

[1] We can easily prove it by derivating it in terms of z.

Let $x = 2\frac{k-1}{n-1} - 1$. We have $k - 1 = \frac{n-1}{2}(1+x)$ and $n - k = \frac{n-1}{2}(1-x)$. We have $0 \leq x \leq 1$ and

$$|p^c - p_0| \leq k \binom{n}{k} \left(z - \frac{1}{2}\right) \frac{1}{2^{n-1}} \left((1+x)^{1+x}(1-x)^{1-x}\right)^{\frac{n-1}{2}}.$$

By using $k\binom{n}{k} = n\binom{n-1}{k-1}$ and the Stirling approximation we obtain that this bound is asymptotically equal to $\frac{\theta\sqrt{n}}{\sqrt{2\pi}}$ so the bound we want to prove is note so loose.

We can easily prove that $(1+x)^{1+x}(1-x)^{1-x} \leq 2^{2x^2}$. Hence

$$|p^c - p_0| \leq k \binom{n}{k} \left(z - \frac{1}{2}\right) \frac{1}{2^{n-1}} 2^{(n-1)x^2}.$$

Since $k - 1 \leq (n-1)z + z$ we have $x \leq \theta + \frac{\theta}{n-1} + \frac{1}{n-1} = \frac{n\theta+1}{n-1}$. Thus

$$|p^c - p_0| \leq \theta \times \left[k \binom{n}{k} \frac{1}{2^n}\right] \times 2^{\frac{(n\theta+1)^2}{n-1}}.$$

For $n = 3$ we have $k\binom{n}{k}\frac{1}{2^n} \leq \frac{3}{4}$ thus

$$|p^c - p_0| \leq 2\theta\sqrt{n} \times \frac{1}{2\sqrt{3}} \times \frac{3}{4} \times 2^{\frac{(3\theta+1)^2}{n-1}}.$$

For $\theta \leq \frac{1}{2\sqrt{3}}$ we obtain $|p^c - p_0| \leq 2\theta\sqrt{n}$ and this remains true even for $\theta > \frac{1}{2\sqrt{3}}$. Let us now concentrate on $n \geq 4$.

The $\binom{n}{k}$ term is upper bounded by $\binom{n}{r}$ with $r = \lceil \frac{n}{2} \rceil$. Furthermore we have

$$\binom{n}{r} \frac{1}{2^n} \leq \prod_{i=1}^{r} \left(1 - \frac{1}{2i}\right)$$

with equality when n is even. Then

$$\begin{aligned}
\log\left(\binom{n}{r}\frac{1}{2^n}\right) &\leq \sum_{i=1}^{r} \log\left(1 - \frac{1}{2i}\right) \\
&\leq -\frac{1}{2}\sum_{i=1}^{r}\frac{1}{i} \\
&\leq -\frac{1}{2}\int_1^{r+1}\frac{dt}{t} \\
&\leq -\frac{1}{2}\log(r+1) \\
&\leq -\frac{1}{2}\log\frac{n}{2} + 1
\end{aligned}$$

therefore

$$\binom{n}{k}\frac{1}{2^n} \leq \sqrt{\frac{2}{n+2}}.$$

Now we have

$$k\binom{n}{k}\frac{1}{2^n} = n\binom{n-1}{k-1}\frac{1}{2^n} \leq \frac{n}{2}\sqrt{\frac{2}{n+1}} \leq \sqrt{\frac{n}{2}}.$$

We deduce

$$|p^c - p_0| \leq 2\theta\sqrt{n} \times 2^{\frac{(n\theta+1)^2}{n-1} - \frac{3}{2}}.$$

When $\theta\sqrt{n} < \frac{1}{2}$ and $n \geq 4$ we have $\frac{(n\theta+1)^2}{n-1} - \frac{3}{2} < 0$ so we obtain $|p^c - p_0| \leq 2\theta\sqrt{n}$. When $\theta\sqrt{n} \geq \frac{1}{2}$ this also holds since the right hand side of the inequality is greater than 1 and the left hand side is a difference between two probabilities. This proves the upper bound.

By definition of k we have $z^{k-1}(1-z)^{n-k} \geq \frac{1}{z2^n}$, so we have $t^{k-1}(1-t)^{n-k} \geq \frac{1}{2^{n-1}(1+\theta)}$ for any $t \in [\frac{1}{2}, z]$. From Equation (2) we deduce

$$\max_{\mathcal{A}} |p^c - p_0| \geq \frac{\theta}{1+\theta} \times \left[k\binom{n}{k}\frac{1}{2^n}\right].$$

If $\theta = o(\frac{1}{\sqrt{n}})$, we have $k = \frac{n}{2} + o(\sqrt{n})$ thus $\binom{n}{k} \sim \frac{2^{n+1}}{\sqrt{2\pi n}}$ from Stirling Formula. Hence $\max_{\mathcal{A}} |p^c - p_0|$ is asymptotically larger than $\frac{\theta\sqrt{n}}{\sqrt{2\pi}}$. Since it is also smaller, this is indeed an equivalent. □

3 Cipher Resistance to Linear Cryptanalysis

We now concentrate on a cipher C on $\{0,1\}^m$ which is defined by a secret key which is selected at random. We can thus consider C as a random permutation. We compare it to an ideal cipher C^* which is another random permutation with a uniform distribution. We call it the perfect cipher.

3.1 The MAXELP Measure

We focus on the expected value $E\left(\mathrm{LP}^C(a,b)\right)$ over the distribution of C and we define

$$\mathrm{MAXELP}(C) = \max_{b \neq 0, a} E\left(\mathrm{LP}^C(a,b)\right).$$

There is a linear expression of this mean value in terms of the pairwise distribution as expressed by the following result.

Lemma 2. *Given a random permutation C over $\{0,1\}^m$, for any a and b ,we have*

$$E(\mathrm{LP}^C(a,b)) = 2^{-2m} \sum_{\substack{x_1,x_2\\y_1,y_2}} (-1)^{(x_1\oplus x_2)\cdot a+(y_1\oplus y_2)\cdot b} \Pr\left[(x_1,x_2) \overset{C}{\mapsto} (y_1,y_2)\right]$$
$$= 1 - 2^{2-2m} \sum_{\substack{x_1\neq x_2\\y_1\neq y_2}} 1_{\substack{x_1\cdot a=y_1\cdot b\\x_2\cdot a\neq y_2\cdot b}} \Pr\left[(x_1,x_2) \overset{C}{\mapsto} (y_1,y_2)\right].$$

If C has a uniform distribution, $a \neq 0$ and $b \neq 0$, we have $E(\mathrm{LP}^C(a,b)) = \frac{1}{2^m-1}$. Note that $E(\mathrm{LP}^C(0,b)) = 0$ for $b \neq 0$.

Proof. In order to prove it, we first notice that $2\Pr_X[X\cdot a = C(X)\cdot b] - 1 = E\left((-1)^{X\cdot a+C(X)\cdot b)}\right)$, and we express $\mathrm{LP}^C(a,b)$ as

$$\mathrm{LP}^C(a,b) = E\left((-1)^{(X_1\oplus X_2)\cdot a+(C(X_1)\oplus C(X_2))\cdot b}\right)$$

where X_1 and X_2 are independent uniformly distributed random variables. We have

$$E(\mathrm{LP}^C(a,b)) = 2^{-2m} \sum_{\substack{x_1,x_2\\y_1,y_2}} (-1)^{(x_1\oplus x_2)\cdot a+(y_1\oplus y_2)\cdot b} \Pr\left[(x_1,x_2) \overset{C}{\mapsto} (y_1,y_2)\right].$$

The contribution of terms for which $x_1 = x_2$ is equal to 2^{-m}. Considering that C is a permutation we can concentrate on $x_1 \neq x_2$ and $y_1 \neq y_2$. Then we split the remaining sum into four groups depending on the two bits $(x_1\cdot a \oplus y_1\cdot b, x_2\cdot a \oplus y_2\cdot b)$. Let Σ_{b_1,b_2} be the sum of all probabilities for which the two bits are (b_1,b_2), $x_1 \neq x_2$, and $y_1 \neq y_2$. We have

$$E(\mathrm{LP}^C(a,b)) = 2^{-m} + 2^{-2m}\Sigma_{0,0} - 2^{-2m}\Sigma_{0,1} - 2^{-2m}\Sigma_{1,0} + 2^{-2m}\Sigma_{1,1}.$$

Due to symmetry we have $\Sigma_{0,1} = \Sigma_{1,0}$. Furthermore, the sum of the four sums is $2^m(2^m-1)$. Hence

$$E(\mathrm{LP}^C(a,b)) = 2^{-m} + 2^{-2m} \times 2^m(2^m-1) - 4 \times 2^{-2m}\Sigma_{0,1}$$

which leads to our second result. Computations when C is uniformly distributed are straightforward. □

3.2 Resistance to Linear Distinguishers

Theorem 3. *Let C be a cipher on $\mathcal{M} = \{0,1\}^m$. For any linear distinguisher (as depicted on Fig. 1) between C and the ideal cipher C^* we have*

$$\mathrm{Adv}_{\text{Fig. 1}} \leq 3\sqrt[3]{n.\mathrm{MAXELP}(C)} + 3\sqrt[3]{\frac{n}{2^m-1}}.$$

Proof. We first notice that the advantage is zero when $a = 0$ or $b = 0$ so the bound holds. Let us now assume that $a \neq 0$ and $b \neq 0$.

We now take a random permutation C with the corresponding Z and p^C as in Lemma 1. Let $\delta = E((2Z-1)^2)$. (Note that $\delta = E(\mathrm{LP}^C(a,b))$.) When $|2Z-1| \leq \alpha$, Lemma 1 says that

$$|p^C - p_0| \leq 2 \times \alpha\sqrt{n}.$$

Since $(2Z-1)^2$ is positive, the probability that $|2Z-1|$ is greater than α is less than $\frac{\delta}{\alpha^2}$. Hence

$$|p - p_0| \leq 2 \times \alpha\sqrt{n} + \frac{\delta}{\alpha^2}$$

for any α.

Let us now fix $\alpha = \left(\frac{\delta}{\sqrt{n}}\right)^{\frac{1}{3}}$. We obtain $|p - p_0| \leq 3 \times \sqrt[3]{\delta n}$.

We recall that $\delta = E\left(\mathrm{LP}^C(a,b)\right)$. We finally note that $E\left(\mathrm{LP}^{C^*}(a,b)\right) = \frac{1}{2^m-1}$ from Lemma 2 so we can have

$$|p^* - p_0| \leq 3\sqrt[3]{\frac{n}{2^m-1}}.$$

We finally use that $|p - p^*| \leq |p - p_0| + |p^* - p_0|$. □

3.3 Using Pairwise Independent Permutations

We recall the following definition.

Definition 4 (Carter-Wegman [6], Wegman-Carter [36]). *Let $\mathcal{M}$ be a finite sets. Let C be a random permutation over $\mathcal{M}$. We say that C is a (perfect) pairwise independent permutation if for any $x_1, x_2, y_1, y_2 \in \mathcal{M}$ such that $x_1 \neq x_2$, and $y_1 \neq y_2$, we have*

$$\Pr[C(x_1) = y_1, C(x_2) = y_2] = \frac{1}{\#\mathcal{M}(\#\mathcal{M}-1)}.$$

Due to Lemma 2, if C is a pairwise independent permutation, we have

$$E(\mathrm{LP}^C(a,b)) = E(\mathrm{LP}^{C^*}(a,b))$$

for any a and b. Hence $\mathrm{MAXELP}(C) = \frac{1}{2^m-1}$. We deduce the following result.

Theorem 5. *Let C be a cipher on $\mathcal{M} = \{0,1\}^m$ which is a perfect pairwise independent permutation. For any linear distinguisher (as depicted on Fig. 1) between C and the ideal cipher C^* we have*

$$\mathrm{Adv}_{\text{Fig. 1}} \leq 6\sqrt[3]{\frac{n}{2^m-1}}.$$

The notion of pairwise independent permutation extends as follows.

Definition 6. *Let $\mathcal{M}$ be a finite sets. Let C be a random permutation over $\mathcal{M}$. Let $\mathcal{M}_2$ be the set of all functions from $\mathcal{M}^4$ to the field $\mathbf{R}$ of real numbers. Let d be a distance over $\mathcal{M}_2$. We define We define $[C]^2 \in \mathcal{M}_2$ by*

$$[C]^2(x_1, x_2, y_1, y_2) = \Pr[C(x_1) = y_1, C(x_2) = y_2].$$

We similarly define $[C^*]^2$ for a uniformly distributed random permutation C^*. We say that C is an ε-d-almost pairwise independent permutation if we have $d([C]^2, [C^*]^2) \leq \varepsilon$.

Several distances are quite significant in cryptography, including the metric induced by the $|||.|||_\infty$ norm as defined in [32,34] by

$$|||f|||_\infty = \max_{(x_1,x_2)\in\mathcal{M}^2} \sum_{(y_1,y_2)\in\mathcal{M}^2} |f(x_1, x_2, y_1, y_2)|$$

We can also use the L_2 norm defined by

$$||f||_2 = \sqrt{\sum_{(x_1,x_2)\in\mathcal{M}^2} \sum_{(y_1,y_2)\in\mathcal{M}^2} f(x_1, x_2, y_1, y_2)^2}$$

We can thus conclude with the following result.

Theorem 7. *Let C be a cipher on $\mathcal{M} = \{0,1\}^m$ which is either an ε-$|||.|||_\infty$-almost pairwise independent permutation, or an ε-L_2-almost pairwise independent permutation. For any linear distinguisher (as depicted on Fig. 1) between C and the ideal cipher C^* of complexity n we have*

$$\mathrm{Adv}_{\text{Fig. 1}} \leq 3\sqrt[3]{n.\varepsilon + \frac{n}{2^m - 1}} + 3\sqrt[3]{\frac{n}{2^m - 1}}.$$

Proof. For $a \neq 0$ and $b \neq 0$, from Lemma 2 we have

$$E\left(\mathrm{LP}^C(a,b)\right) - \frac{1}{2^m - 1} =$$
$$2^{-2m} \sum_{\substack{x_1,x_2 \\ y_1,y_2}} (-1)^{f_{a,b}(x,y)} \left([C]^2(x_1, x_2, y_1, y_2) - [C^*]^2(x_1, x_2, y_1, y_2)\right)$$

for some function $f_{a,b}(x, y)$. We can thus deduce

$$\mathrm{MAXELP}(C) \leq \frac{1}{2^m - 1} + |||[C]^2 - [C^*]^2|||_\infty$$
$$\mathrm{MAXELP}(C) \leq \frac{1}{2^m - 1} + ||[C]^2 - [C^*]^2||_2.$$

We conclude by using Theorem 3. □

4 Discussion and Conclusion

We demonstrated that the advantage of any linear distinguisher is uniformly bounded by a function of the number of samples n multiplied by the MAXELP measure. It is further itself bounded by the pairwise independence according to several metrics. This shows that linear distinguishers cannot lead to significant attacks unless $n = \Omega(1/\mathrm{MAXELP})$, which corresponds to heuristic arguments that were given so far.

The practical consequence, for the designer of new cryptographic primitives, is that we need to make sure that MAXELP is small for a given number of rounds, either by using pairwise independence, or by any other construction. Then, the designer only needs to add a few rounds which could play the role of the pre and post processing in linear cryptanalysis. (These additional rounds are usually referred to as the "safety margin".) Our result formally demonstrates that no linear distinguisher will manage to distinguish the core rounds from an ideal primitive by linear cryptanalysis techniques.

We can also use our result for security evaluation purposes. If we can estimate the MAXELP measure of core rounds, we can have a fair idea on a security upper limit.

5 Acknowledgments

I would like to thank Thomas Baignères for many relevant comments, as well as Pascal Junod for interesting discussions and a pointer to Equation (1).

References

1. Data Encryption Standard. *Federal Information Processing Standard Publication 46*, U. S. National Bureau of Standards, 1977.
2. K. Aoki, K. Ohta. Strict Evaluation of the Maximum Average of Differential Probability and the Maximum Average of Linear Probability. *IEICE Transactions on Fundamentals*, vol. E80-A, pp. 1–8, 1997.
3. E. Biham, A. Shamir. Differential Cryptanalysis of DES-like Cryptosystems. In *Advances in Cryptology CRYPTO'90*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 537, pp. 2–21, Springer-Verlag, 1991.
4. E. Biham, A. Shamir. *Differential Cryptanalysis of the Data Encryption Standard*, Springer-Verlag, 1993.
5. F. Chabaud, S. Vaudenay. Links between Differential and Linear Cryptanalysis. In *Advances in Cryptology EUROCRYPT'94*, Perugia, Italy, Lecture Notes in Computer Science 950, pp. 356–365, Springer-Verlag, 1995.
6. J. L. Carter, M. N. Wegman. Universal Classes of Hash Functions. *Journal of Computer and System Sciences*, vol. 18, pp. 143–154, 1979.
7. H. Gilbert. *Cryptanalyse Statistique des Algorithmes de Chiffrement et Sécurité des Schémas d'Authentification*, Thèse de Doctorat de l'Université de Paris 11, 1997.

8. H. Gilbert, G. Chassé. A Statistical Attack of the FEAL-8 Cryptosystem. In *Advances in Cryptology CRYPTO'90*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 537, pp. 22–33, Springer-Verlag, 1991.
9. H. M. Heys. *The Design of Substitution-Permutation Network Ciphers Resistant to Cryptanalysis*, Ph.D. Thesis of Queen's University, Kingston, Ontario, Canada, 1994.
10. P. Junod. On the Complexity of Matsui's Attack. In *Selected Areas in Cryptography'01*, Toronto, Ontario, Canada, Lecture Notes in Computer Science 2259, pp. 199-211, Springer-Verlag, 2001.
11. P. Junod, S. Vaudenay. Optimal Key Ranking Procedures in a Statistical Cryptanalysis. In *Fast Software Encryption'03*, Lund, Sweden, Lecture Notes in Computer Science 2887, pp. 235–246, Springer-Verlag, 2003.
12. L. R. Knudsen. *Block Ciphers — Analysis, Design and Applications*, Aarhus University, 1994.
13. 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., Lecture Notes in Computer Science 839, pp. 26–39, Springer-Verlag, 1994.
14. L. Keliher, H. Meijer, S. Tavares. New Method for Upper Bounding the Maximum Average Linear Hull Probability for SPNs. In *Advances in Cryptology EUROCRYPT'01*, Innsbruck, Austria, Lecture Notes in Computer Science 2045, pp. 420–436, Springer-Verlag, 2001.
15. L. Keliher, H. Meijer, S. Tavares. Improving the Upper Bound on the Maximum Average Linear Hull Probability for Rijndael. In *Selected Areas in Cryptography'01*, Toronto, Ontario, Canada, Lecture Notes in Computer Science 2259, pp. 112–128, Springer-Verlag, 2001.
16. X. Lai. *On the Design and Security of Block Ciphers*, ETH Series in Information Processing, vol. 1, Hartung-Gorre Verlag Konstanz, 1992.
17. X. Lai, J. L. Massey, S. Murphy. Markov Ciphers and Differential Cryptanalysis. In *Advances in Cryptology EUROCRYPT'91*, Brighton, United Kingdom, Lecture Notes in Computer Science 547, pp. 17–38, Springer-Verlag, 1991.
18. M. Matsui. Linear Cryptanalysis Methods for DES Cipher. In *Advances in Cryptology EUROCRYPT'93*, Lofthus, Norway, Lecture Notes in Computer Science 765, pp. 386–397, Springer-Verlag, 1994.
19. M. Matsui. The First Experimental Cryptanalysis of the Data Encryption Standard. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 839, pp. 1–11, Springer-Verlag, 1994.
20. M. Matsui. New Structure of Block Ciphers with Provable Security against Differential and Linear Cryptanalysis. In *Fast Software Encryption'96*, Cambridge, United Kingdom, Lecture Notes in Computer Science 1039, pp. 205–218, Springer-Verlag, 1996.
21. S. Murphy, F. Piper, M. Walker, P. Wild. Likelihood Estimation for Block Cipher Keys. Unpublished.
22. K. Nyberg. Perfect Nonlinear S-Boxes. In *Advances in Cryptology EUROCRYPT'91*, Brighton, United Kingdom, Lecture Notes in Computer Science 547, pp. 378–385, Springer-Verlag, 1991.
23. K. Nyberg, L. R. Knudsen. Provable Security against a Differential Cryptanalysis. In *Advances in Cryptology CRYPTO'94*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 839, pp. 566–574, Springer-Verlag, 1994.
24. K. Nyberg, L. R. Knudsen. Provable Security against a Differential Cryptanalysis. *Journal of Cryptology*, vol. 8, pp. 27–37, 1995.
25. S. Park, S. H. Sung, S. Chee, E-J. Yoon, J. Lim On the Security of Rijndael-Like Structures against Differential and Linear Cryptanalysis. In *Advances in Cryptology ASIACRYPT'02*, Queenstown, New Zeland, Lecture Notes in Computer Science 2501, pp. 176–191, Springer-Verlag, 2002.

26. S. Park, S. H. Sung, S. Lee, J. Lim Improving the Upper Bound on the Maximum Differential and Maximum Linear Hull Probability for SPN Structures and AES. In *Fast Software Encryption'03*, Lund, Sweden, Lecture Notes in Computer Science 2887, pp. 247–260, Springer-Verlag, 2003.
27. A. Rényi. *Probability Theory*, Elsevier, 1970.
28. C. E. Shannon. Communication Theory of Secrecy Systems. *Bell system technical journal*, vol. 28, pp. 656–715, 1949.
29. A. Tardy-Corfdir, H. Gilbert. A Known Plaintext Attack of FEAL-4 and FEAL-6. In *Advances in Cryptology CRYPTO'91*, Santa Barbara, California, U.S.A., Lecture Notes in Computer Science 576, pp. 172–181, Springer-Verlag, 1992.
30. S. Vaudenay. *La Sécurité des Primitives Cryptographiques*, Doctoral Thesis from the University of Paris 7, Technical Report LIENS-95-10 of the Laboratoire d'Informatique de l'Ecole Normale Supérieure, 1995.
31. S. Vaudenay. An Experiment on DES — Statistical Cryptanalysis. In *3rd ACM Conference on Computer and Communications Security*, New Delhi, India, pp. 139–147, ACM Press, 1996.
32. S. Vaudenay. Provable Security for Block Ciphers by Decorrelation. In *STACS'98*, Paris, France, Lecture Notes in Computer Science 1373, pp. 249–275, Springer-Verlag, 1998.
33. S. Vaudenay. Resistance Against General Iterated Attacks. In *Advances in Cryptology EUROCRYPT'99*, Prague, Czech Republic, Lecture Notes in Computer Science 1592, pp. 255–271, Springer-Verlag, 1999.
34. S. Vaudenay. Decorrelation: a Theory for Block Cipher Security. *Journal of Cryptology*, vol. 16, pp. 249–286, 2003.
35. G. S. Vernam. Cipher Printing Telegraph Systems for Secret Wire and Radio Telegraphic communications. *Journal of the American Institute of Electrical Engineers*, vol. 45, pp. 109–115, 1926.
36. M. N. Wegman, J. L. Carter. New Hash Functions and their Use in Authentication and Set Equality. *Journal of Computer and System Sciences*, vol. 22, pp. 265–279, 1981.