On Linear Cryptanalysis with Many Linear Approximations (full version)

Benoît Gérard and Jean-Pierre Tillich

INRIA project-team SECRET, France
{benoit.gerard,jean-pierre.tillich}@inria.fr

Abstract. In this paper we present a theoretical framework to quantify the information brought by several linear approximations of a block-cipher without putting any restriction on these approximations. We quantify here the entropy of the key given the plaintext-ciphertext pairs statistics which is a much more accurate measure than the ones studied earlier. The techniques which are developed here apply to various ways of performing the linear attack and can also been used to measure the entropy of the key for other statistical attacks. Moreover, we present a realistic attack on the full DES with a time complexity of 2^{48} for 2^{41} pairs what is a big improvement comparing to Matsui's algorithm 2 ($2^{51.9}$).

Keywords : linear cryptanalysis, multiple linear approximations, information theory.

1 Introduction

Related work

Linear cryptanalysis is probably one of the most powerful tools available for attacking symmetric cryptosystems. It was invented by Matsui [1, 2] to break the DES cipher building upon ideas put forward in [3, 4]. It was quickly discovered that other ciphers can be attacked in this way, for instance FEAL [5], LOKI [6], SAFER [7]. It is a known plaintext attack which takes advantage of probabilistic linear equations that involve bits of the plaintext $\mathbf{P}$, the ciphertext $\mathbf{C}$ and the key $\mathbf{K}$

$$\mathbf{Pr}(<\pi, \mathbf{P}> \oplus <\gamma, \mathbf{C}> \oplus <\kappa, \mathbf{K}>= b) = \frac{1}{2} + \epsilon. \tag{1}$$

Usually, ϵ is called the *bias* of the equation, π, γ and κ are linear masks and $<\pi, \mathbf{P}>$ denotes the following inner product between $\pi = (\pi_i)_{1 \leq i \leq m}$ and $\mathbf{P} = (P_i)_{1 \leq i \leq m}$, $<\pi, \mathbf{P}> \overset{\text{def}}{=} \bigoplus_{i=1}^{m} \pi_i P_i$. There might be several different linear approximations of this kind we have at our disposal and we let n be their number. We denote the corresponding key masks by $\kappa_i = (\kappa_i^j)_{1 \leq j \leq k}$ and the corresponding biases by ϵ_i for $i \in \{1, \dots, n\}$.

Such an attack can be divided in three parts:
- *Distillation phase:* It consists in extracting from the available plaintext-ciphertext pairs the relevant parts of the data. Basically, for each linear approximation, the attacker counts how many times $< \pi, \mathbf{P} > \oplus < \gamma, \mathbf{C} >$ evaluates to zero.
- *Analysis phase:* It consists in extracting from the values taken by the counters some information on the key and testing whether some key guesses are correct or not by using the linear approximation(s) (1) as a distinguisher. Typically, the output of this phase is a list of all possible subkey guesses sorted relatively to their likelihood.
- *Search phase:* It typically consists in finding the remaining key bits by exhaustive search.

In [1] Matsui used only one approximation to distinguish wrong last round keys from the right one. One year later, he refined his attack by using a second approximation obtained by symmetry [2] and by also distinguishing with them the first round key. Later Vaudenay [8] has presented a framework for statistical cryptanalysis where Matsui's attack is presented as a particular case. With Junod, he has also studied the optimal way of merging information from two (or more) approximations [9]. This kind of attack can use several approximations but the key masks must have disjoint supports. A second approach of using multiple equations is given by Kaliski and Robshaw [10]. They improved Matsui's first attack using several approximations which have the same key mask κ. Biryukov and al. suggest in [11] a way of using multiple linear approximations without putting any restriction on them. They present a theoretical framework to compute the expected rank of the good subkey guess. This framework has been used for SERPENT cryptanalysis [12, 13]. Recent works by Hermelin and al. [14] give a way to compute the good subkey ranking probability law in the case of *multidimensional linear cryptanalysis*. More details on this work are given later to compare it to ours.

All these improvements have a common goal: reducing the amount of messages needed for the attack. Clearly, using several approximations should give more information than a single one.

Our contribution

The purpose of this paper is to study how much multiple linear approximations may benefit linear cryptanalysis. We aim at quantifying accurately how much information is gained on the key from the knowledge of statistical data derived from linear characteristics of type (1).

Several statistics have been proposed to study how many plaintext-ciphertext pairs we need in order to carry out successfully a linear cryptanalysis. This includes for instance the probability of guessing incorrectly a linear combination of key bits by Matsui's Algorithm 1 [2], the ranking of the right subkey in the ordered list of candidates [15, 16] or the expected size of the number of keys which are more likely than the right key [11]. Some of these statistics are either not relevant for multilinear cryptanalysis or are extremely difficult to compute (such as for instance the ranking statistics of [15, 16] when we do not allow restrictions on the approximations used). This is not the case of the expected size of the number L of keys which are more likely than the right key considered in [11]. However, this kind of statistics also leads to pessimistic predictions concerning the number of plaintext-ciphertexts which are needed. To be more specific, it turns out that its prediction of the number of plaintext/ciphertext pairs ensuring that the most likely key is indeed the right key is in many cases twice the number of plaintext/ciphertexts which are really needed ! This is detailed in Proposition 3.1. We obtain the right amount by our analysis. It consists in studying instead of the expectation of L, the *entropy* $\mathcal{H}(\mathbf{K}|\mathbf{Y})$ of the key $\mathbf{K}$ (or more generally $\mathcal{H}(\mathbf{K}'|\mathbf{Y})$, where $\mathbf{K}'$ is a certain subkey of $\mathbf{K}$- for instance it can be the part of the key involved in a distinguisher attack) given the statistics $\mathbf{Y}$ we have derived from the plaintext-ciphertext pairs.

The fact that the entropy is a much better statistic than the expecation of L is is related to the following probabilistic phenomenon : this expectation is in a rather wide range of amount of plaintext-ciphertext pairs exponential in the key size k, while for most plaintext-ciphertext pairs the most likely key is the right one. This comes from the fact that rare events (of exponentially small probability) yield values of L which are exponentially large in k. In other words, while for typical plaintext-ciphertext pairs L is equal to zero, for some rare occurrences of the plaintext-ciphertext pairs L is very large, and this accounts quite heavily in the expectation of L. The entropy behaves here much better. In a certain sense, it is related to the expectation of the logarithm $\log_2(L)$. The logarithm of L varies much less than L and this why the typical size of $\log L$ coincides quite well with the expectation.

Despite the fact that it is much more desirable to estimate the entropy than the expectation of L, it might seem that this quantity is much harder to calculate. Our main result is to give here a lower bound on this quantity (see Subsections 3.1 and 3.2) which is quite sharp. The sharpness of the bound is illustrated by the results of Subsection 3.3. We apply this bound in three different scenarios: (i) the linear attack which recovers only the linear combination of the key bits, (ii) the usual linear distinguishing attack which recovers some linear combinations of the key bits of the first (and/or) last round, and (iii) the algorithm MK2 in [11]. We wish to emphasize the fact that the technique to derive the lower bound is quite general and applies in a very wide range of situations, and not only in the case where $\mathbf{Y}$ corresponds to a function of the counters of linear approximations (see Subsection 3.1). A second useful property of this lower bound on the entropy is that it gives an upper bound on the information we gain on the $\mathbf{K}$ when we know $\mathbf{Y}$ which is independent of the algorithm we use afterwards to extract this information.

Complementarity with [14]
The work of Hermelin, Nyberg and Cho gives a framework for multidimensional linear cryptanalysis that does not require statistical independence between the approximations used. A set of m linearly independent approximations is chosen and the correlations of the linear combinations of those approximations are computed. For each plaintext/ciphertext pair, the m bits vector corresponding to the m base approximation evaluations is extracted. Hence, the attacker gets an empirical probability distribution for the m bits vector. Actually, this distribution depends on the key used for encryption (usually it depends on m bits of this key). Using the correlations of the 2^m approximations, the probability distribution of the m bits vector can be computed for each possible key. Using enough pairs, the empirical distribution is likely to be the closest to the distribution provided by the correct key. The guessed key is the one with the maximum log-likelihood ratio (LLR) to the uniform distribution.

As the statistical independence hypothesis for linear approximations may not hold for many ciphers, this is an important theoretic improvement. Nevertheless, some of the results are not tight because of some other conjectures or simplifications. For instance, saying that the LLR of a wrong keys has a mean of 0 gives very pessimistic results as supposing statistical independence of LLRs does (for 8-round DES at least). Moreover, this method may not apply to some cryptanalyses (the one presented in this paper for instance). Using m base approximations leads to a time complexity of $2^m 2^d$ in the analysis phase (where d is the number of information bits to recover). In the case of the presented attack, 32968 approximations are used to recover 42 key bits. This set of approximation has a dimension of 54. Hence, the analysis time complexity is about 2^{96} what is much greater that exhaustive search. The approach presented in this paper is based on a statistical independence hypothesis. Thus, it is an orthogonal and complementary approach to the one of [14]. This approach leads to an attack with a better complexity than Matsui's algorithm 2 as soon as less than 2^{42} pairs are available (see Section 4). Using the same approximations in the framework of Hermelin and al. leads to an attack with higher complexity.

Actually, our method is based on some decoding techniques that are easily practicable in case of statistical independence of the approximations. That is why our theoretical framework seems to be the more suitable in that case. In the other hand, the work of Hermelin and al. is the more suitable when no assumption is made on statistical independence up to now.

2 The probabilistic model

It will be convenient to denote by $\tilde{\mathbf{K}} \overset{\text{def}}{=} (\tilde{K}_i)_{1\leq i\leq n}$ the vector of linear combinations of the key bits induced by the key masks, that is

$$\tilde{K}_i \overset{\text{def}}{=} \bigoplus_{j=1}^{k} \kappa_i^j K_j.$$

A quantity will play a fundamental role in this setting : the dimension (what we will denote by d) of the vector space generated by the κ_i's. It can be much smaller than the number n of different key masks.

We denote by Σ the set of N plaintext-ciphertext pairs. The information available after the distillation phase is modeled by

Model 1 — *The attacker receives a vector* $\mathbf{Y} = (Y_i)_{1\leq i\leq n}$ *such that:*

$$\forall\, i \in \{1,\ldots,n\},\quad Y_i = (-1)^{\tilde{K}_i} + N_i \quad,\quad N_i \sim \mathcal{N}(0,\sigma_i^2), \tag{2}$$

where $\sigma_i^2 \overset{def}{=} \frac{1}{4N\epsilon_i^2}$ *(N is the number of available plaintext/ciphertext pairs).*

We denote by $f(\mathbf{Y}|\tilde{\mathbf{K}})$ *the density function of the variable* $\mathbf{Y}$ *conditioned by the value taken by* $\tilde{\mathbf{K}}$ *and* $f_i(Y_i \mid \tilde{K}_i)$ *denotes the density of the variable* Y_i *conditioned by* $\tilde{K}_i$*.*
These conditional densities satisfy the independence relation

$$f(\mathbf{Y} \mid \tilde{\mathbf{K}}) = \prod_{i=1}^{n} f(Y_i \mid \tilde{K}_i) \tag{3}$$

The vector $\mathbf{Y}$ is derived from Σ as follows. We first define for every i in $\{1,\ldots,n\}$ and every j in $\{1,\ldots,N\}$ the following quantity

$$D_i^j \overset{\text{def}}{=} < \pi_i, \mathbf{P}^j > \oplus < \gamma_i, \mathbf{C}^j > \oplus b_i,$$

where the plaintext-ciphertext pairs in Σ are indexed by $(\mathbf{P}^j, \mathbf{C}^j)$ and b_i is the constant appearing in the i-th linear approximation. Then for all i in $\{1,\ldots,n\}$ we set up the counters D_i with $D_i \overset{\text{def}}{=} \sum_{j=1}^N D_i^j$ from which we build the vector of counters $\mathbf{D} = (D_i)_{1\leq i\leq n}$. D_i is a binomial random variable which is approximately distributed as a normal law $\mathcal{N}((1/2 - \epsilon_i(-1)^{\tilde{K}_i})N, (1/4 - \epsilon_i^2)N)$. This explains why the vector $\mathbf{Y} = (Y_i)_{1\leq i\leq n}$ is defined as:

$$Y_i \overset{\text{def}}{=} \frac{N - 2D_i}{2N\epsilon_i} \tag{4}$$

and why Equation (2) holds. There is some debate about the independence relation (3). This point is discussed by Murphy in [17] where he proves that even if some key masks are linearly dependent, the independence relation (3) holds asymptotically if for a fixed key the covariances $\text{cov}(D_{i_1}^j, D_{i_2}^j)$ are negligible. We have checked whether this holds in our experimental study. We had 129 linear approximations on 8-round DES with biases in the range $[1.45.10^{-4}, 5.96.10^{-4}]$ and we found empirical covariances in the range $[-2.10^{-7}, 2.10^{-7}]$ for 10^{12} samples. This corroborates the fact that the covariances are negligible and that the independence relation (3) approximately holds.

3 Bounds on the required amount of plaintext-ciphertext pairs

3.1 An information-theoretic lower bound

The purpose of this subsection is to derive a general lower bound on the amount of uncertainty $\mathcal{H}(\mathbf{K}|\mathbf{Y})$ we have on the key given the statistics $\mathbf{Y}$ derived from the plaintext-ciphertext pairs. We recall that the (binary) *entropy* $\mathcal{H}(X)$ of a random variable X is given by the expression:

$$\begin{aligned}\mathcal{H}(X) &\overset{\text{def}}{=} -\sum_x \mathbf{Pr}(X = x)\log_2 \mathbf{Pr}(X = x) \text{ (for discrete } X) \\ &\overset{\text{def}}{=} -\int f(x)\log_2 f(x)dx \text{ (for continuous } X \text{ of density } f)\end{aligned} \tag{5}$$

For a couple of random variables (X, Y) we denote by $\mathcal{H}(X|Y)$ the *conditional entropy of X given Y*. It is defined by

$$\mathcal{H}(X|Y) \overset{\text{def}}{=} \sum_{y} \mathbf{Pr}(Y = y)\mathcal{H}(X|Y = y),$$

where $\mathcal{H}(X|Y = y) \overset{\text{def}}{=} -\sum_x \mathbf{Pr}(X = x|Y = y) \log_2 \mathbf{Pr}(X = x|Y = y)$ when X and Y are discrete variables and when $\mathbf{Y}$ is a continuous random variable taking its values over $\mathbb{R}^n$ it is given by

$$\mathcal{H}(X|\mathbf{Y}) = \int_{\mathbb{R}^n} \mathcal{H}(X|\mathbf{Y} = \mathbf{y}) f(\mathbf{y}) d\mathbf{y},$$

where $f(\mathbf{y})$ is the density of the distribution of $\mathbf{Y}$ at the point $\mathbf{y}$. A related quantity is the *mutual information $\mathcal{I}(X; Y)$ between* X and Y which is defined by

$$\mathcal{I}(X; Y) \overset{\text{def}}{=} \mathcal{H}(X) - \mathcal{H}(X|Y). \tag{6}$$

It is straightforward to check [18] that this quantity is symmetric and that

$$\mathcal{I}(X; Y) = \mathcal{I}(Y; X) = \mathcal{H}(Y) - \mathcal{H}(Y|X). \tag{7}$$

Since K is a discrete random variable and Y is a continuous one, it will be convenient to use the following formula for the mutual information where the conditional distributions of Y given K has density $f(Y|K)$.

$$\mathcal{I}(K; Y) = \sum_k \mathbf{Pr}(K = k) \int f(y|k) \log \frac{f(y|k)}{\sum_k f(y|k)} dy. \tag{8}$$

We will be interested in deriving a lower bound on $\mathcal{H}(\mathbf{K}'|\mathbf{Y})$ when $\mathbf{K}' = (K'_1, \dots, K'_n)$ is a subkey derived from $\mathbf{K}$ which satisfies:
(i)(conditional independence assumption)

$$f(\mathbf{Y} \mid \mathbf{K}') = \prod_{i=1}^{n} f(Y_i \mid K'_i), \tag{9}$$

where $f(\mathbf{Y}|\mathbf{K}')$ is the density function of the variable $\mathbf{Y}$ conditioned by the value taken by $\mathbf{K}'$ and $f_i(Y_i \mid K'_i)$ denotes the density of the variable Y_i conditioned by K'_i.
(ii) The subkey $\mathbf{K}'$ may take $2^{k'}$ values and all are equally likely.

With these assumptions we have the following result

Lemma 1.

$$\mathcal{I}(\mathbf{K}';\mathbf{Y}) \leq \sum_{i=1}^{n} \mathcal{I}(K'_i;Y_i) \tag{10}$$

$$\mathcal{H}(\mathbf{K}'|\mathbf{Y}) \geq k' - \sum_{i=1}^{n} \mathcal{I}(K'_i;Y_i). \tag{11}$$

The proof of this lemma can be found in the appendix. It will be used in what follows in various scenarios for linear attacks, but it can obviously be used to cover many other cryptographic attacks. This lower bound is in general quite sharp as long as it is non-trivial, i.e when $k' \geq \sum_{i=1}^{n} \mathcal{I}(K'_i;Y_i)$. We will prove this for Attack 1 in what follows but this can also be done for the other cases.

3.2 Application to various scenarios

Attack 1 : In this case, we do not use the linear equations as distinguishers but only want to recover the $< \kappa_i, \mathbf{K} >$'s. This corresponds in the case of a single equation to Matsui's attack 1 and in the case of multiple equations to the attack MK1 in [11]. We have here

$$\begin{aligned} K'_i &= \tilde{K}_i =< \kappa_i, \mathbf{K} > \\ Y_i &= \frac{N - 2D_i}{2N\epsilon_i}. \end{aligned}$$

Variables $\mathbf{K}'$ and $\mathbf{Y}$ satisfy the required conditional independence assumption (see Equation 3) and a straightforward calculation using Formula (8) yields

$$\mathcal{I}(K'_i;Y_i) = \mathbf{Cap}(\sigma_i^2)$$

where

$$\mathbf{Cap}(\sigma^2) \overset{\text{def}}{=} 1 - \frac{\sigma e^{-\frac{1}{2\sigma^2}}}{\sqrt{8\pi}} \int_{-\infty}^{\infty} e^{-\frac{u^2\sigma^2}{8}} e^{\frac{u}{2}} \log_2\left(1 + e^{-u}\right) du.$$

and therefore by applying Lemma 1 we obtain

$$\mathcal{H}(\mathbf{K}'|\mathbf{Y}) \geq d - \sum_{i=1}^{n} \mathbf{Cap}(\sigma_i^2) \tag{12}$$

Attack 2: This attack corresponds to cryptanalyses using a distinguisher such as [12, 19]. Approximations of the form (1) are applied to a reduced cipher say the cipher peeled off by the first and the last round what is usually the case. Here, we focus on the subkeys used for the first ($\mathbf{K}_{\text{first}}$) and the last rounds ($\mathbf{K}_{\text{last}}$). The idea is to encrypt and decrypt the pairs with each possible value for $\mathbf{K}_{\text{first}}$ and $\mathbf{K}_{\text{last}}$ and then to observe the bias obtained. The candidate that gives the greater bias is then choosen. Notice that we do not take care of the information given by the $< \kappa_i, \mathbf{K} >$. This may be the case when cryptanalyzing ciphers for which it is difficult to find the key masks of linear approximations.

The $< \pi_i, \mathbf{P} >$'s and the $< \gamma_i, \mathbf{C} >$'s might not depend on all the bits of $\mathbf{K}_{\text{first}}$ and $\mathbf{K}_{\text{last}}$. We denote by $\hat{\mathbf{K}}_i$ the vector composed of the bits of $\mathbf{K}_{\text{first}}$ and $\mathbf{K}_{\text{last}}$ on which the $< \pi_i, \mathbf{P} >$'s and the $< \gamma_i, \mathbf{C} >$'s depend on. We define $\mathbf{K}'$ by the vector $(\hat{\mathbf{K}}_i)_{i=1}^n$ and assume that it may take $2^{\hat{k}}$ values. The aim is to recover $\mathbf{K}'$ based on the values of the counters D_i^z for i in $\{1, \ldots, n\}$ and z ranging over all possible values for $\mathbf{K}'$. These counters are defined similarly as in Section 2 with the difference being that we use the value $\mathbf{K}' = z$ for deriving the relevant couples $(\mathbf{P}, \mathbf{C})$. The statistics $\mathbf{Y} = (Y_i)_{1 \le i \le n}$ we consider in this case is given by $Y_i \overset{\text{def}}{=} (Y_i^z)_z$ with

$$Y_i^z = \frac{|N - 2D_i^z|}{2N\epsilon_i}.$$

The conditional independence relation (3) is also satisfied in this case. With the help of Lemma 1, we can write $\mathcal{H}(\mathbf{K}'|\mathbf{Y}) \ge \hat{k} - \sum_{i=1}^n \mathcal{I}(K'; Y_i)$. We can again use Lemma 1 and obtain $\mathcal{I}(K'; Y_i) \le \sum_z \mathcal{I}(K'; Y_i^z)$. The variable Y_i^z has density r_i if z corresponds to the right choice for K' and w_i otherwise, where $r_i(t) = \varphi_i^1(t) + \varphi_i^{-1}(t)$, $w_i(t) = 2\varphi_i^0(t)$ for nonnegative t with $\varphi_i^\alpha(t) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(t-\alpha)^2}{2\sigma_i^2}\right]$ being the density of a normal variable of expectation α and variance σ_i^2. A straightforward application of Formula (8) gives

$$\mathcal{I}(K_i'; Y_i^z) = \int_0^\infty \frac{r_i(t)}{2^{\hat{k}}} \log\left(\frac{r_i(t)}{s_i(t)}\right) dt + \int_0^\infty (1 - 2^{-\hat{k}}) w_i(t) \log\left(\frac{w_i(t)}{s_i(t)}\right) dt, \tag{13}$$

with $s_i(t) \overset{\text{def}}{=} 2^{-\hat{k}} r_i(t) + (1 - 2^{-\hat{k}}) w_i(t)$. We denote this quantity by I_i and we finally obtain

$$\mathcal{H}(\mathbf{K}'|\mathbf{Y}) \ge \hat{k} - 2^{\hat{k}} \sum_{i=1}^n I_i.$$

Attack 3: This corresponds to the attack MK2 in [11] which is a variation of the previously seen distinguisher attack. In this case, we wish to find simultaneously the $\hat{\mathbf{K}}_i$'s defined in Attack 2 and the vector $\tilde{\mathbf{K}}$ defined in

Attack 1. In this case, we let $\mathbf{K}'_i = (\hat{\mathbf{K}}_i, \tilde{K}_i)$ and define $\mathbf{K}' \stackrel{\text{def}}{=} (\mathbf{K}'_i)_{1 \le i \le n}$. We assume that $2^{k'}$ is the number of all possible values for $\mathbf{K}'$ and that $2^{\hat{k}}$ is the number of all possible values for $\hat{\mathbf{K}}$. Here, we define the relevant statistics $\mathbf{Y} = (\mathbf{Y}_i)_{1 \le i \le n}$ by $\mathbf{Y}_i = (Y_i^z)_z$ where z ranges over all possible values for $\hat{\mathbf{K}}$ and where

$$Y_i^z = \frac{N - 2D_i^z}{2N\epsilon_i}.$$

We have again the desired independence relation (3) and as in the previous example we can use Lemma 1 twice to obtain

$$\mathcal{H}(\mathbf{K}'|\mathbf{Y}) \ge k' - \sum_{i=1}^{n} \mathcal{I}(\mathbf{K}'_i; \mathbf{Y}_i) \ge k' - 2^{\hat{k}} \sum_{i=1}^{n} \mathcal{I}(\mathbf{K}'_i; Y_i^z)$$

A straightforward application of Formula (8) yields

$$\mathcal{I}(\mathbf{K}'_i; Y_i^z) = \int_{-\infty}^{\infty} \frac{\varphi_i^1(t)}{2^{\hat{k}}} \log\left(\frac{\varphi_i^1(t)}{\psi_i(t)}\right) dt + \int_{-\infty}^{\infty} (1-2^{-\hat{k}})\varphi_i^0(t) \log\left(\frac{\varphi_i^0(t)}{\psi_i(t)}\right) dt,$$

with $\psi_i(t) \stackrel{\text{def}}{=} (1 - 2^{-\hat{k}})\varphi_i^0(t) + 2^{-\hat{k}-1}[\varphi_i^{-1}(t) + \varphi_i^1(t)]$ and $\varphi_i^\alpha(t)$ defined as in Attack 2.

3.3 An upper bound

One might wonder whether or not the bounds given in the previous subsection are sharp or not. It is clear that these lower bounds become negative when the number of pairs is large enough and that they are worthless in this case (since entropy is always nonnegative). However in all three cases it can be proved that as long the bound is non trivial it is quite sharp. We will prove this for the lower-bound (12). Similar techniques can be used for the other bounds but it would be too long to include them in this paper. To prove that (12) is sharp we will consider the case when

$$\sum_{i=1}^{n} \mathbf{Cap}(\sigma_i^2) \approx d$$

If the lower-bound is sharp, one might be tempted to say that the conditional entropy of $\mathbf{K}'$ given $\mathbf{Y}$ should be close to 0 which would mean that $\mathbf{K}'$ is determined from $\mathbf{Y}$ with probability close to 1. This is of course not always true, but it is the case *for most choices* of the coefficients κ_i^j. To give a precise meaning to this statement we will first consider what happens when the κ_i^j's are chosen *at random*.

Theorem 1. *Assume that the κ_i^j are chosen chosen uniformly at random and that $\sum_{i=1}^n \mathbf{Cap}(\sigma_i^2) \geq d + \delta n$ for some constant $\delta > 0$. Let P_{err} be the probability that the most likely value for $\mathbf{K}'$ given $\mathbf{Y}$ is not the right one. There exists a constant A such that*

$$P_{err} \leq \frac{A}{\delta^2 n} + 2^{-\delta n/2}.$$

The probability P_{err} is taken over $\mathbf{Y}$ but also over the choices of the κ_i^j's. It says nothing about a particular choice of the κ_i^j's. However it implies the aforementioned assertion about most choices of the κ_i's. Let us be more specific by bringing in $P_{\mathrm{err}}(\mathcal{C})$ which is the probability that the most likely key given $\mathbf{Y}$ is not the right one when the subspace of dimension d of the possible values for $\tilde{\mathbf{K}}$ is $\mathcal{C}$. A bound on P_{err} implies that for most choices of the κ_i's (and hence of $\mathcal{C}$) $P_{\mathrm{err}}(\mathcal{C})$ is small by using the following lemma

Lemma 2. *Assume that $P_{err} \leq \epsilon$. Then for any $t > 0$:*

$$\mathbf{Pr}_{\mathcal{C}}\left(P_{err}(\mathcal{C}) \geq t\epsilon\right) \leq \frac{1}{t}$$

Proof. Let us define $P \overset{\text{def}}{=} \mathbf{Pr}_{\mathcal{C}}\left(P_{\mathrm{err}}(\mathcal{C}) \geq t\epsilon\right)$. Then, we observe that $P_{\mathrm{err}} = \sum_{\mathcal{C}} P_{\mathrm{err}}(\mathcal{C})\mathbf{Pr}(\mathcal{C}) \geq Pt\epsilon$. This implies that $P \leq \frac{1}{t}$. ∎

Remark: The notation $\mathbf{Pr}_{\mathcal{C}}$ means here that the probability is taken over the choices for $\mathcal{C}$. It actually denotes the proportion of choices for $\mathcal{C}$ which lead to the specified event inside the probability.

3.4 Entropy vs. expected number of $\tilde{\mathbf{K}}$'s more likely than the right one

The aim of this subsection is to emphasize the fact that in a certain range of values of N (which is the number of plaintext-ciphertext pairs) the expected size $\mathcal{E}$ of the list of the $\tilde{\mathbf{K}}$'s which are more likely than the right one gives pessimistic estimates of the amount of plaintext-ciphertext pairs we need to mount an attack. Actually, the *gain* g of a type 1 attack defined in [11] relies on this statistic $\mathcal{E}$. Here, we compare this *gain* with the capacity defined in Subsection 3.2. In order to achieve top ranking for a d-bits key (that is the correct key is at the top of the list), the gain has to be equal to d and Theorem 1 shows that for $\sum_{i=1}^n \mathbf{Cap}(\sigma_i^2) \approx d$ the probability of top ranking is close to 1. The comparison shows that the estimate derived from $\mathcal{E}$ is twice bigger than the one derived from our entropy approach, as stated in Proposition 3.1. Proposition 3.1 holds for $N \cdot \epsilon_i^2 = o(1)$. This is often the case in multiple linear cryptanalysis where many approximations are used to drop the data complexity below the value required for a single approximation, that is $N = O(\epsilon^{-2})$.

Proposition 3.1 — *Suppose that N is in a range where $\forall i, N\epsilon_i^2 = o(1)$. Using our entropy approach, the estimate for the data complexity required to achieve top ranking on a d-bit key is*

$$N \approx \frac{d\ln(2)}{2\sum_{i=1}^n \epsilon_i^2}\left(1+o(1)\right).$$

The one obtained using the formula derived from the gain in [11] is

$$N \approx \frac{d\ln(2)}{\sum_{i=1}^n \epsilon_i^2}\left(1+o(1)\right).$$

Proof.
It can be found in [20, ex. 4.12] that $\sum_{i=1}^n \mathbf{Cap}(\sigma_i^2) = \frac{2N\sum_{i=1}^n \epsilon_i^2}{\ln(2)}(1+o(1))$, if for all i, $N \cdot \epsilon_i^2 = o(1)$. The corresponding estimate for N is $N \approx \frac{d\ln(2)}{2\sum_{i=1}^n \epsilon_i^2}\left(1+o(1)\right)$. The formula for the gain in [11] is:

$$g \approx -\log_2\left[2\cdot\Phi\left(-\sqrt{2N\cdot\sum_{i=1}^n \epsilon_i^2}\right)\right]. \tag{14}$$

The following estimate can be found in [21, p. 175]. For large x, $\ln(\Phi(-x)) = -x^2/2(1+o(1))$. We can apply this to (14) and find $N \approx \frac{d\ln(2)}{\sum_{i=1}^n \epsilon_i^2}\left(1+o(1)\right)$. ∎

4 Experimental Results

To corroborate the theoretical results presented in this paper, we performed some experiments. First, we confirm the tightness of the bound on entropy by comparing it to the empirical entropy computed for a toy example (namely a type 1 attack on the 8-round DES). Then, we performed a realistic type 1 attack on the full 16-round DES, ranked the subkeys with respect to their likelihoods and checked whether or not the rank of the right subkey is among the $2^{\mathcal{H}(\mathbf{K}'|\mathbf{Y})}$ most likely subkeys. The results we obtained confirmed that choosing lists of this size is indeed relevant. Finally, in order to emphasize the power of multiple linear cryptanalysis, we compare this type 1 attack using many approximations to Matsui's type 3 attack using the optimal ranking statistic suggested by Junod [15]. This is first time that such an attack is performed.

Accuracy of the bound on entropy
Concerning the bound on entropy given in 1, we checked our results on 8-round DES. For those simulations, we used a group of 76 linear approximations involving 13 key bits to perform a type 1 attack. The quality of the lower-bound (12) can be verified by estimating empirically the entropy. Figure 1 displays the empirical conditional entropy of $\mathbf{K}'$ given $\mathbf{Y}$ for these equations as a function of $\log_2(N)$, where N is the number of available plaintext-ciphertext pairs. There is an excellent agreement between the lower bound and the empirical entropies up to when we approach the critical value of N for which the lower bound is equal to zero. This kind of lower bound is really suited to the case when the amount of plaintext/ciphertext pairs is some order of magnitude below this critical value. This is typically the case when we want to decrease the amount of data needed at the expense of keeping a list of possible candidates for $\mathbf{K}'$.

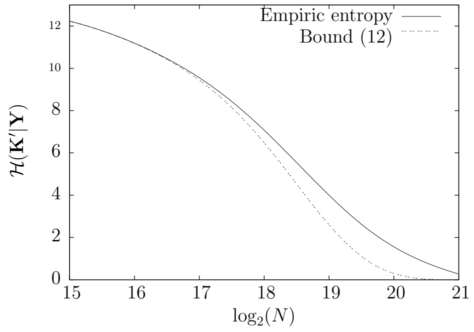

Fig. 1. comparison between lower bound and empirical value of entropy.

A realistic type 1 attack on the full 16-round DES
Our aim in this experiment was to confirm that most of the time the right value of $\mathbf{K}' = (< \kappa_i, \mathbf{K} >)_{1 \leq i \leq n}$ belongs to the list of $2^{\mathcal{H}(\mathbf{K}'|\mathbf{Y})}$ most likely candidates. We performed here the whole type 1 attack, with the exception of the search phase which is not relevant for our purpose. In [22], the analysis phase uses a soft decision decoding algorithm for Reed Muller codes over the Gaussian channel with erasures. This decoding algorithm can be efficiently performed using a fast Walsh-Hadamard trans-

form. Generating the list and sorting it are thus two operations with the same complexity $O(d2^d)$ where d is the dimension of the space spanned by κ's. This implies that to speed up the analysis phase, we have to use approximations that lead to a small d. In the case when the set of approximations does not have any structure, the analysis phase can be efficiently performed using a general decoding algorithm for random linear codes such as for instance the stochastic resonance decoding algorithm from Valembois [23]. There is still no proof of its complexity but it is quite simple to implement and actually efficient. The study of this decoding algorithm is out of the scope of this article but is a nice subject we wish to work on.

Using a Branch & Bound Algorithm, we found 74086 linear approximations on the 16-round DES with biases higher than $2^{-28.84}$ (the biases are obtained by using the piling-up lemma). The space spanned by these κ's turned out to be 56. This is too much to use directly the fast Walsh-Hadamard transform. We choose to consider a subset of these approximations (32968 out of 74086 spanning a vector space of dimension 42) which can be divided in 4 groups, each of them consisting of key masks κ_i spanning a vector space of small dimension d. We sum-up information about these groups in Figure 2.

Group	N	Nb. input masks	Nb. output masks	d
G1	12384	1500	82	19
G2	12384	82	1500	19
G3	4100	64	82	13
G4	4100	82	64	13

Fig. 2. characteristics of the groups of approximations.

The symmetry comes from the fact that enciphering or deciphering with the DES is the same algorithm (using subkeys in reverse order). We observe that the number of different masks in a group is much lower than the number of approximations. This will help us in speeding up the distillation phase using a trick similar to the one mentioned in [24]. Notice that some key bits are common to some groups. We performed a fast Walsh Hadamard transformation on each group separately and use a heuristic to combine the information for each group to compute the rank of the correct key inside the list of candidates sorted with respect to their likelihood. This is detailed in Appendix B.

The number N of available pairs was chosen to be small enough so that we can generate the data and perform the distillation phase in reasonable time. On the other hand, if we want our experiments to be relevant, we must get at least 1 bit of information about the key. These considerations lead us to choose $N = 2^{39}$ for which we get 2 bit of information out of 42 on the subkey.

We performed the attack 18 times. This attack recovers 42 bits of the key. For 2^{39} pairs, the information on the key is of 2 bits. The entropy on the key is thus 40 bits. Our theoretical work suggests to take a list of size $2^{\mathcal{H}(\mathbf{K}'|\mathbf{Y})} = 2^{40}$ to have a good success probability. Our experiments corroborate this. The worst rank over the 18 experiments is $2^{40.88}$ and the rank exceeded 2^{40} in only 3 experiments out of 18. The (ordered) list of ranks for the 18 experiments is:

$$2^{31.34}, 2^{33.39}, 2^{34.65}, 2^{35.24}, 2^{36.56}, 2^{37.32}, 2^{37.99}, 2^{38.11}, 2^{38.52}, 2^{38.97},$$

$$2^{39.04}, 2^{39.19}, 2^{39.27}, 2^{39.53}, 2^{39.85}, 2^{40.28}, 2^{40.82}, 2^{40.88}.$$

Comparison with Matsui's attack:
The attack from [2] uses two approximations on 14-round DES with biases $1.19.2^{-21}$ in a type 3 attack. This kind of attack uses approximations with much better biases than a type 1 attack because they involve only 14 rounds instead of the full 16 rounds.

Despite this fact, we show here that the gap between 14-round approximations and 16-round approximations can be filled by using many approximations in type 1 attack. We demonstrate here that for a rather large range of number of plaintext/ciphertext pairs N, a type 1 attack has a better complexity than the best version Matsui's type 3 attack [15]. This is first time that such a result is shown on the full DES.

Figure 3 gives the formulas used to compute the complexity of the two attacks. Due to space constraints, we do not detail how we obtained the distillation and analysis phase complexities but they are essentially a direct application of the tricks of [24] and the work of [19]. We denote by ν the XOR operation complexity and θ the DES enciphering complexity (including key schedule). From the same amount of data, our attack obtains

Attack	Distillation	Analysis	Search
Matsui's [15]	$N \cdot 2 \cdot 46 \cdot \nu$	$12 \cdot 2^{12} \cdot \nu$	$2^{\mathcal{H}(\mathbf{K}'|\mathbf{Y})} \cdot \theta$
Our	$N \cdot (82 + 82 + 64 + 64) \cdot 44 \cdot \nu$	$2^{26} \cdot \nu$	$2^{\mathcal{H}(\mathbf{K}'|\mathbf{Y})} \cdot \theta$

Fig. 3. complexities of the different steps.

more information on the key. It improves the final search complexity at the cost of increasing the distillation phase complexity. To measure the gain of using a type 1 attack, we have to estimate the ratio ν/θ. The lower this ratio is, the more we gain using multilinear type 1 attack. For a standard implementation, $600 \cdot \nu$ is a good estimate of θ. We computed the complexities of the two attacks in terms of DES evaluations (θ) and plotted it as functions of the number of pairs in Figure 4. We restrict the plot to the value of N where type 1 attack competes with type 3 and we can see that this attack is better for N less than 2^{42}. We also plotted the complexities of the type attack for $\theta = 400 \cdot \nu$ and $\theta = 200 \cdot \nu$ to show that type 1 attack still competes with type 3 whenever enciphering is very efficient. Notice that with these estimates of θ the complexity for Matsui's attack remains the same as long as N is less than $2^{42.5}$.

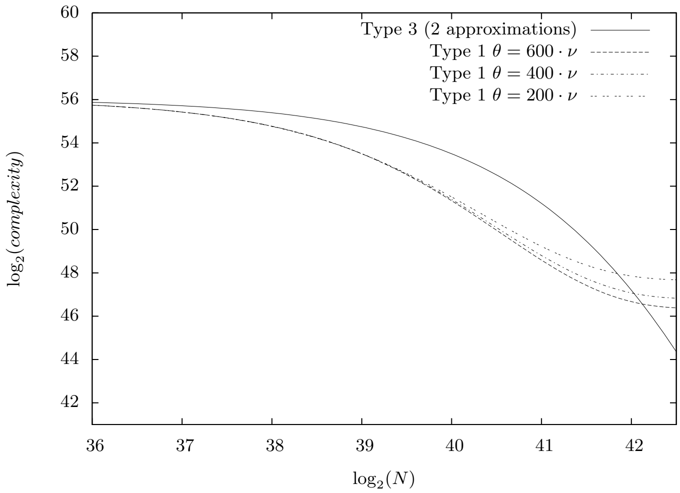

Fig. 4. complexities of Matsui's type 3 attack and our type 1 attack in terms of DES evaluations.

Remark on Matsui's attack complexity:

In [15], the author points out the fact that his Theorem 1 is pessimistic regarding the expected average rank of the good key. For 2^{43} pairs, the empirical complexity seems to be less than 2^{41} with high probability. Actually, the bound (12) suggests a complexity of precisely 2^{41} in this case (see Figure 4). This is a good illustration of the phenomenon mentioned in Section 3.4. The average rank of the good key is pessimistic because in some extremely rare cases the rank is sufficiently high to influence the mean. This observation, together with the complexity of computing multidimensional probability laws in a general case, may confirm the interest of the approach presented in this paper.

5 Conclusion and further work

We have presented here a rather general technique in Lemma 1 to derive a sharp lower bound on the entropy of a key given (independent) statistics.

We have applied it here to various linear cryptanalytic attacks, but the scope of this tool is much broader and it would be interesting to apply it for other classes of statistical attacks.

We performed a realistic type 1 attack on full 16-round DES using 32968 approximations and 2^{39} plaintext/ciphertext pairs that confirmed our theoretical results.

Moreover, theoretical results predict that for 2^{41} pairs, the DES can be broken with high probability with complexity close to 2^{48} while Matsui's attack 2 needs $2^{51.9}$ DES computations.

This work entails some further research interests.

It would be interesting to compare our theoretical results with some others [11, 25] for some particular type 3 attack.

Another interesting thing would be to perform a type 1 attack on another cipher (SERPENT for instance) to see if, for recent ciphers, type 1 attacks still can compete type 3 attacks.

A deep study of different decoding algorithms for the analysis phase is necessary as much as a precise complexity analysis of distillation phase complexity for type 1 attack (maybe using ideas from [19]).

References

1. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In: EUROCRYPT'93. Volume 765 of LNCS., Springer–Verlag (1993) 386–397
2. Matsui, M.: The First Experimental Cryptanalysis of the Data Encryption Standard. In: CRYPTO'94. Volume 839 of LNCS., Springer–Verlag (1994) 1–11
3. Tardy-Corfdir, A., Gilbert, H.: A Known Plaintext Attack of FEAL-4 and FEAL-6. In: CRYPTO'91. Volume 576 of LNCS., Springer–Verlag (1992) 172–181
4. Matsui, M., Yamagishi, A.: A New Method for Known Plaintext Attack of FEAL Cipher. In: EUROCRYPT'92. Volume 658 of LNCS., Springer–Verlag (1993) 81–91
5. Ohta, K., Aoki, K.: Linear Cryptanalysis of the Fast Data Encipherment Algorithm. In: CRYPTO'94. Volume 839 of LNCS., Springer–Verlag (1994) 12–16
6. Tokita, T., Sorimachi, T., Matsui, M.: Linear Cryptanalysis of LOKI and s2DES. In: ASIACRYPT'94. Volume 917 of LNCS., Springer–Verlag (1994) 293–303

7. Murphy, S., Piper, F., Walker, M., Wild, P.: Likelihood Estimation for Block Cipher Keys. Technical report, Information Security Group, University of London, England (1995)
8. Vaudenay, S.: An Experiment on DES Statistical Cryptanalysis. In: CCS 1996, ACM (1996) 139–147
9. Junod, P., Vaudenay, S.: Optimal key ranking procedures in a statistical cryptanalysis. In: FSE 2003. Volume 2887 of LNCS., Springer–Verlag (2003) 235–246
10. Kaliski, B.S., Robshaw, M.J.B.: Linear Cryptanalysis Using Multiple Approximations. In: CRYPTO'94. Volume 839 of LNCS., Springer–Verlag (1994) 26–39
11. Biryukov, A., Cannière, C.D., Quisquater, M.: On Multiple Linear Approximations. In: CRYPTO'04. Volume 3152 of LNCS., Springer–Verlag (2004) 1–22
12. Collard, B., Standaert, F.X., Quisquater, J.J.: Improved and Multiple Linear Cryptanalysis of Reduced Round Serpent. In: Inscrypt 2007. Volume 4990 of LNCS., Springer–Verlag (2007) 51–65
13. Collard, B., Standaert, F.X., Quisquater, J.J.: Experiments on the Multiple Linear Cryptanalysis of Reduced Round Serpent. In: FSE 2008. Volume 5086 of LNCS., Springer–Verlag (2008) 382–397
14. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In: ACISP 2008. Volume 5107 of LNCS., Springer–Verlag (2008) 203–215
15. Junod, P.: On the Complexity of Matsui's Attack. In: SAC 2001. Volume 2259 of LNCS., Springer–Verlag (2001)
16. Selçuk, A.: On Probability of Success in Linear and Differential Cryptanalysis. J. Cryptol. **21** (2008) 131–147
17. Murphy, S.: The Independence of Linear Approximations in Symmetric Cryptology. IEEE Transactions on Information Theory **52** (2006) 5510–5518
18. Cover, T., Thomas, J.: Information theory. Wiley series in communications. Wiley (1991)
19. Collard, B., Standaert, F.X., Quisquater, J.J.: Improving the Time Complexity of Matsui's Linear Cryptanalysis. In: ICISC 2007. Volume 4817 of LNCS., Springer–Verlag (2007) 77–88
20. Richardson, T., Urbanke, R.: Modern coding theory (2008)
21. Feller, W.: An introduction to probability theory and its applications. 3rd edn. Volume 1. John Wiley and Sons Inc., New York (1968)
22. Fourquet, R., Loidreau, P., Tavernier, C.: Finding Good Linear Approximations of Block Ciphers and its Application to Cryptanalysis of Reduced Round DES. In: WCC 2009. (2009) 501–515
23. Valembois, A.: Détection, Reconnaissance et Décodage des Codes Linéaires Binaires. PhD thesis, Université de Limoges (2000)
24. Biham, E., Dunkelman, O., Keller, N.: Linear Cryptanalysis of Reduced Round Serpent. In: FSE 2001. Volume 2355 of LNCS., Springer–Verlag (2001) 219–238
25. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Extension of Matsui's Algorithm 2. In: FSE 2009. LNCS, Springer–Verlag (2009)

A Proofs

A.1 Proof of Lemma 1

Let us use Equation (7) and write in two different ways the mutual information between $\mathbf{K}'$ and $\mathbf{Y}$: $\mathcal{I}(\mathbf{K}';\mathbf{Y}) = \mathcal{H}(\mathbf{K}') - \mathcal{H}(\mathbf{K}'|\mathbf{Y}) = \mathcal{H}(\mathbf{Y}) -$

$\mathcal{H}(\mathbf{Y}|\mathbf{K}')$. From this we deduce that

$$\begin{aligned}\mathcal{I}(\mathbf{K}';\mathbf{Y}) &= \mathcal{H}(\mathbf{Y}) - \mathcal{H}(\mathbf{Y}|\mathbf{K}') \\ &= \mathcal{H}(Y_1,\dots,Y_n) - \mathcal{H}(Y_1,\dots,Y_n|\mathbf{K}'). \qquad (15)\end{aligned}$$

Here Equation (15) is a consequence of the fact that the a priori distribution over $\mathbf{K}'$ is the uniform distribution and the entropy of a discrete random variable which is uniformly distributed is obviously nothing but the logarithm of the number of values it can take. Moreover (see [18, Theorem 2.6.6])

$$\mathcal{H}(Y_1,\dots,Y_n) \leq \mathcal{H}(Y_1) + \dots + \mathcal{H}(Y_n). \qquad (16)$$

On the other hand, by the chain rule for entropy [18, Theorem 2.5.1]:

$$\mathcal{H}(Y_1,\dots,Y_n|\mathbf{K}') = \mathcal{H}(Y_1|\mathbf{K}') + \mathcal{H}(Y_2|Y_1,\mathbf{K}') + \dots + \mathcal{H}(Y_n|Y_1,Y_2,\dots,Y_{n-1},\mathbf{K}'). \quad (17)$$

We notice now that $\mathcal{H}(Y_i|\mathbf{K}',Y_1\dots Y_{i-1})$ can be written as

$$\textstyle\sum_{\mathbf{k}} \int_{\mathbb{R}^{i-1}} \mathcal{H}(Y_i|\mathbf{K}'{=}\mathbf{k},Y_1{=}y_1,\dots,Y_{i-1}{=}y_{i-1}) f(y_1,\dots,y_{i-1}|\mathbf{K}'{=}\mathbf{k})\mathbf{Pr}(\mathbf{K}'{=}\mathbf{k})dy_1\dots dy_{i-1}, \qquad (18)$$

where the sum is taken over all possible values $\mathbf{k}$ of $\mathbf{K}'$ and $f(y_1,\dots,y_{i-1}|\mathbf{K}' = \mathbf{k})\mathbf{Pr}(\mathbf{K}' = \mathbf{k})$ is the density of the distribution of the vector $(Y_1,\dots,Y_{i-1})$ given the value $\mathbf{k}$ of $\mathbf{K}'$ at the point $(y_1,\dots,y_{i-1})$. From conditional independence assumption (9) we deduce that $\mathcal{H}(Y_i|\mathbf{K}' = \mathbf{k}, Y_1 = y_1,\dots,Y_{i-1} = y_{i-1}) = \mathcal{H}(Y_i|K'_i)$. By summing in Expression (18) over $y_1,\dots,y_{i-1}$ and all possible values of $K'_1,\dots,K'_{i-1},K'_{i+1},\dots,K'_n$ we obtain that

$$\mathcal{H}(Y_i|\mathbf{K}',Y_1,\dots,Y_{i-1}) = \frac{1}{2}\mathcal{H}(Y_i|K'_i = 0) + \frac{1}{2}\mathcal{H}(Y_i|K'_i = 1) = \mathcal{H}(Y_i|K'_i = k_i) \qquad (19)$$

Plugging in this last expression in Expression (17) we obtain that

$$\mathcal{H}(Y_1,\dots,Y_n|K'_1,\dots,K'_n) = \mathcal{H}(Y_1|K'_1) + \dots + \mathcal{H}(Y_n|K'_n). \qquad (20)$$

Using this last equation and Inequality (16) in (15) we finally deduce that

$$\begin{aligned}\mathcal{I}(\mathbf{K}';\mathbf{Y}) &\leq \mathcal{H}(Y_1) + \dots + \mathcal{H}(Y_n) - \mathcal{H}(Y_1|K'_1) - \dots - \mathcal{H}(Y_n|K'_n) \\ &\leq \sum_{i=1}^{n} \mathcal{H}(Y_i) - \mathcal{H}(Y_i|K'_i) \leq \sum_{i=1}^{n} \mathcal{I}(K'_i;Y_i). \qquad (21)\end{aligned}$$

The lower bound on the entropy follows from equality (7) that can be written as $\mathcal{H}(\mathbf{K}'|\mathbf{Y}) = \mathcal{H}(\mathbf{K}') - \mathcal{I}(\mathbf{K}';\mathbf{Y}) = k' - \mathcal{I}(\mathbf{K}';\mathbf{Y})$.

A.2 Proof of Theorem 1

The proof of this theorem follows closely standard proofs of the direct part of Shannon's channel capacity theorem [18], however most of the proofs given for this theorem are asymptotic in nature and are not suited to our case. There are proofs which are not asymptotic, but they are tailored for the case where all the σ_i's are equal and are rather involved. We prefer to follow a slightly different path here. The first argument we will use is an explicit form of the joint AEP (Asymptotic Equipartition Property) theorem.

For this purpose, we denote by $(\mathbf{X}, \mathbf{Y})$ a couple of random variables where $\mathbf{X} = (X_i)_{1 \leq i \leq n}$ is uniformly distributed over $\{0,1\}^n$ and $\mathbf{Y} = (Y_i)_{1 \leq i \leq n}$ is the output of the Gaussian channel described in Section 2 when $\mathbf{X}$ is sent through it. This means that

$$Y_i = (-1)^{X_i} + N_i, \tag{22}$$

where the N_i are independent centered normal variables of variance σ_i^2.

Let us first bring in the following definition.

Definition 1. *For $\epsilon > 0$, we define the set T_ϵ of ϵ-jointly typical sequences of $\{0,1\}^n \times \mathbb{R}^n$ by $T_\epsilon \overset{def}{=} \bigcup_{\mathbf{x} \in \{0,1\}^n} \{\mathbf{x}\} \times T_\epsilon(\mathbf{x})$ with*

$$\begin{aligned} T_\epsilon(\mathbf{x}) \overset{def}{=} \{\mathbf{y} \in \mathbb{R}^n : &|-\log_2(f(\mathbf{y})) - \mathcal{H}(\mathbf{Y})| < n\epsilon &(23)\\ &\left|-\log_2\left(f(\mathbf{y}|\mathbf{x})2^{-n}\right) - \mathcal{H}(\mathbf{X}, \mathbf{Y})\right| < n\epsilon\big\} &(24)\end{aligned}$$

where $f(\mathbf{y})$ is the density distribution of $\mathbf{Y}$ and $f(\mathbf{y}|\mathbf{x})$ is the density distribution of $\mathbf{Y}$ given that $\mathbf{X}$ is equal to $\mathbf{x}$.

The entropies of $\mathbf{Y}$ and $(\mathbf{X}, \mathbf{Y})$ are given by the following expressions

Lemma 3.

$$\mathcal{H}(\mathbf{Y}) = \sum_{i=1}^{n} \mathbf{Cap}(\sigma_i^2) + \frac{1}{2}\log_2(2\pi e \sigma_i^2)$$

$$\mathcal{H}(\mathbf{X}, \mathbf{Y}) = n + \sum_{i=1}^{n} \frac{1}{2}\log_2(2\pi e \sigma_i^2)$$

Proof. Notice that with our model the Y_i's are independent. Therefore $\mathcal{H}(\mathbf{Y}) = \sum_{i=1}^{n} \mathcal{H}(Y_i)$. Moreover, by the very definitions of entropy and mutual information: $\mathcal{H}(Y_i) = \mathcal{H}(Y_i|X_i) + \mathcal{I}(X_i; Y_i)$; X_i is uniformly distributed over $\{0,1\}$ and therefore by the definition of the capacity of a

Gaussian channel and the fact that the capacity attains its maximum for a binary input which is uniformly distributed we have $\mathcal{I}(X_i;Y_i) = \mathbf{Cap}(\sigma_i^2)$. On the other hand $\mathcal{H}(Y_i|X_i)$ is obviously the same as $\mathcal{H}(N_i)$. The calculation of this entropy is standard (see [18]) and gives

$$\mathcal{H}(N_i) = \frac{1}{2}\log_2(2\pi e\sigma_i^2) \tag{25}$$

By putting all these facts together we obtain the expression for $\mathcal{H}(\mathbf{Y})$. Concerning the other entropy, with similar arguments we obtain

$$\begin{aligned}\mathcal{H}(\mathbf{X},\mathbf{Y}) &= \mathcal{H}(\mathbf{X}) + \mathcal{H}(\mathbf{Y}|\mathbf{X}) \\ &= n + \sum_{\mathbf{x}\in\{0,1\}^n} \frac{1}{2^n}\mathcal{H}(\mathbf{Y}|\mathbf{X}=\mathbf{x}) \\ &= n + \sum_{\mathbf{x}\in\{0,1\}^n} \frac{1}{2^n}\mathcal{H}(N_1,\dots,N_n) \\ &= n + \sum_{i=1}^{n}\frac{1}{2}\log_2(2\pi e\sigma_i^2)\end{aligned}$$

∎

"T_ϵ" stands for "typical set" since it is highly unlikely that $(\mathbf{X},\mathbf{Y})$ does not belong to T_ϵ:

Lemma 4. *There exists a constant A such that*

$$\mathbf{Pr}\left((\mathbf{X},\mathbf{Y}) \notin T_\epsilon\right) \leq \frac{A}{\epsilon^2 n}.$$

Before giving the proof of this lemma we will first give an interpretation of entropy which provides an explanation of why the probability of falling outside the typical set becomes smaller as n increases.

Lemma 5. *Let $U_i \overset{def}{=} -\log_2 f_i(Y_i)$ where f_i is given by*

$$f_i(y) \overset{def}{=} \frac{1}{2\sqrt{2\pi\sigma_i^2}}\left(e^{-\frac{(y-1)^2}{2\sigma_i^2}} + e^{-\frac{(y+1)^2}{2\sigma_i^2}}\right).$$

We also denote by $V_i \overset{def}{=} -\log_2\left(\frac{g_i(Y_i-(-1)^{X_i})}{2}\right)$ *where* g_i *is the density distribution of a centered Gaussian variable of variance* σ_i^2*.*

$$-\log_2(f(\mathbf{Y})) - \mathcal{H}(\mathbf{Y}) = \sum_{i=1}^{n} U_i - \mathbb{E}\left(\sum_{i=1}^{n} U_i\right)$$
$$-\log_2(f(\mathbf{Y}|\mathbf{X})2^{-n}) - \mathcal{H}(\mathbf{X},\mathbf{Y}) = \sum_{i=1}^{n} V_i - \mathbb{E}\left(\sum_{i=1}^{n} V_i\right)$$

Proof. For the first equation we just have to notice that

$$-\log_2(f(\mathbf{Y})) = -\log_2\left(\mathit{\Pi}_{i=1}^{n} f_i(Y_i)\right) = -\sum_{i=1}^{n}\log_2(f_i(Y_i)) = \sum_{i=1}^{n} U_i$$

and that $\mathcal{H}(\mathbf{Y}) = \mathbb{E}(-\log_2 f(\mathbf{Y}))$, which follows directly from the definition of the entropy given in (5). The second equation can be obtained in a similar way. ▮

This implies that in order to estimate the probability that a point falls outside the typical set we have to estimate the probability that the deviation between a sum of n independent random variables and its expectation is at least of order ϵn. In our case, it can be proven that for fixed ϵ, this probability is exponentially small in n. However, we prefer to give a much weaker statement which is also easier to prove and which uses only Chebyschev's inequality, which we recall here

Lemma 6. *Consider a real random variable X of variance* $\mathrm{var}(X)$*. We have for any $t > 0$:*

$$\mathbf{Pr}\left(|X - \mathbb{E}(X)| \geq t\right) \leq \frac{\mathrm{var}(X)}{t^2}. \tag{26}$$

To use this inequality we have to estimate the variances of the U_i's and the V_i's. It can be checked that

Lemma 7. *There exists a constant A such that for any i we have*

$$\mathrm{var}(V_i) \leq A \quad \text{and} \quad \mathrm{var}(U_i) \leq A.$$

Proof. Let us prove the first statement. Recall that from (22), we have $N_i = Y_i - (-1)^{X_i}$.

$$\begin{aligned}\bar{V_i} \overset{\text{def}}{=} V_i - \mathbb{E}(V_i) &= -\log_2\left(\frac{g_i(N_i)}{2}\right) - \mathbb{E}\left(-\log_2\left(\frac{g_i(N_i)}{2}\right)\right)\\ &= -\log_2\left(g_i(N_i)\right) - \frac{1}{2}\log_2(2e\pi\sigma_i^2)\end{aligned}$$

where the last equation follows from Expression (25). Hence:

$$\bar{V_i} = \log_2(e)\frac{N_i^2}{2\sigma_i^2} + \frac{1}{2}\log_2(2\pi\sigma_i^2) - \frac{1}{2}\log_2(2e\pi\sigma_i^2) = \frac{\log_2(e)}{2}\left(\frac{N_i^2}{\sigma_i^2} - 1\right),$$

and therefore

$$\begin{aligned}\mathrm{var}(V_i) \overset{\mathrm{def}}{=} \mathbb{E}\left[\bar{V_i}^2\right] &= \frac{\log_2(e)^2}{4}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_i^2}}\left(\frac{u^2}{\sigma_i^2} - 1\right)^2 e^{-\frac{u^2}{2\sigma_i^2}}du \\ &= \frac{\log_2(e)^2}{4}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\left(v^2 - 1\right)^2 e^{-\frac{v^2}{2}}dv\end{aligned}$$

where the last equation follows by the change of variable $v = \frac{u}{\sigma_i}$ in the integral. This shows that the variance of V_i is constant. For the second statement we will make use of the following inequalities. For nonnegative u we have

$$e^{-\frac{(u-1)^2}{2\sigma_i^2}}/2\sqrt{2\pi\sigma_i^2} \le f_i(u) \le e^{-\frac{(u-1)^2}{2\sigma_i^2}}/\sqrt{2\pi\sigma_i^2}. \tag{27}$$

Recall that $\mathbb{E}(U_i) = \mathcal{H}(Y_i) = \mathcal{H}(Y_i|X_i) + \mathcal{I}(X_i;Y_i) = \mathcal{H}(N_i) + \mathcal{I}(X_i;Y_i)$. Note that $0 \le \inf(X_i;Y_i) \le \mathcal{H}(X_i) = 1$ by the properties of mutual information (see [18][chapter 2]). And since $\mathcal{H}(N_i) = \frac{1}{2}\log_2(2e\pi\sigma_i^2)$ we deduce that

$$\frac{1}{2}\log_2(2e\pi\sigma_i^2) \le \mathbb{E}(U_i) \le \frac{1}{2}\log_2(2e\pi\sigma_i^2) + 1. \tag{28}$$

To simplify the expressions below we let $u = Y_i$. Assume that U_i is greater that its expectation and that this expectation is nonnegative. This means that $-\log_2 f_i(u) \ge \mathbb{E}(U_i) \ge 0$. We notice that

$$\begin{aligned}\bar{U_i}^2 &= (U_i - \mathbb{E}(U_i))^2 \\ &= (-\log_2(f_i(u)) - \mathbb{E}(U_i))^2 \\ &\le \left(\log_2(e)\frac{(u-1)^2}{2\sigma_i^2} + 1 + \frac{1}{2}\log_2(2\pi\sigma_i^2) - \frac{1}{2}\log_2(2e\pi\sigma_i^2)\right)^2 \\ &= \left(\log_2(e)\frac{(u-1)^2}{2\sigma_i^2} - \frac{1}{2}\log_2(e/2)\right)^2 \end{aligned} \tag{29}$$

by using inequations (27) and (28). Let us now write

$$\begin{aligned}\mathrm{var}(U_i) = \mathbb{E}(\bar{U_i})^2 &= \int_{-\infty}^{\infty}\bar{U_i}^2 f_i(u)du \\ &= \int_{-\infty}^{0}\bar{U_i}^2 f_i(u)du + \int_{0}^{\mathbb{E}(U_i)}\bar{U_i}^2 f_i(u)du + \int_{\mathbb{E}(U_i)}^{\infty}\bar{U_i}^2 f_i(u)du\end{aligned}$$

From the previous upper-bound on $\bar{U_i}^2$ we deduce that

$$\begin{aligned}
\int_{\mathbb{E}(U_i)}^{\infty} \bar{U_i}^2 f_i(u)du &\leq \int_{\mathbb{E}(U_i)}^{\infty} \left(\log_2(e)\frac{(u-1)^2}{2\sigma_i^2} - \frac{1}{2}\log_2(e/2)\right)^2 f_i(u)du \\
&\leq \int_{\mathbb{E}(U_i)}^{\infty} \left(\log_2(e)\frac{(u-1)^2}{2\sigma_i^2} - \frac{1}{2}\log_2(e/2)\right)^2 \frac{e^{-\frac{(u-1)^2}{2\sigma_i^2}}}{\sqrt{2\pi\sigma_i^2}}du \qquad (30) \\
&= \int_{\frac{\mathbb{E}(U_i)-1}{\sigma_i}}^{\infty} \left(\log_2(e)\frac{v^2}{2} - \frac{1}{2}\log_2(e/2)\right)^2 \frac{e^{-\frac{v^2}{2}}}{\sqrt{2\pi}}dv \qquad (31) \\
&\leq \int_{-\infty}^{\infty} \left(\log_2(e)\frac{v^2}{2} - \frac{1}{2}\log_2(e/2)\right)^2 \frac{e^{-\frac{v^2}{2}}}{\sqrt{2\pi}}dv,
\end{aligned}$$

where Inequality (30) is a consequence of (27) and Equality (31) follows from the change of variable $v = \frac{u-1}{\sigma_i}$. The two other integrals in (29) can be treated similarly where instead of using (27) we use for negative values of u: $e^{-\frac{(u+1)^2}{2\sigma_i^2}}/2\sqrt{2\pi\sigma_i^2} \leq f_i(u) \leq e^{-\frac{(u+1)^2}{2\sigma_i^2}}/\sqrt{2\pi\sigma_i^2}$. This yields a constant upper-bound for all variances $\text{var}(U_i)$. ∎

We are ready now to prove Lemma 4:

Proof. We start the proof by writing

$$\begin{aligned}
\mathbf{Pr}((\mathbf{X},\mathbf{Y})\notin T_\epsilon) &= \mathbf{Pr}\big(\{|-\log_2(f(\mathbf{Y}))-\mathcal{H}(\mathbf{Y})|\geq n\epsilon\}\cup\big\{\big|-\log_2\big(f(\mathbf{Y}|\mathbf{X})2^{-n}\big)-\mathcal{H}(\mathbf{X},\mathbf{Y})\big|\geq n\epsilon\big\} \\
&\leq \mathbf{Pr}(|-\log_2(f(\mathbf{Y}))-\mathcal{H}(\mathbf{Y})|\geq n\epsilon)+\mathbf{Pr}\big(\big|-\log_2\big(f(\mathbf{Y}|\mathbf{X})2^{-n}\big)-\mathcal{H}(\mathbf{X},\mathbf{Y})\big|\geq n\epsilon\big) \\
&= \mathbf{Pr}(|U-\mathbb{E}(U)|\geq n\epsilon)+\mathbf{Pr}(|V-\mathbb{E}(V)|\geq n\epsilon)
\end{aligned}$$

with $U \overset{\text{def}}{=} \sum_{i=1}^n U_i$ and $V \overset{\text{def}}{=} \sum_{i=1}^n V_i$. We use now Chebyschev's inequality (Lemma 26) together with the upper-bounds $\text{var}(U) = \sum_{i=1}^n \text{var}(U_i) \leq nA$ and $\text{var}(V) = \sum_{i=1}^n \text{var}(V_i) \leq nA$ to obtain $\mathbf{Pr}\left((\mathbf{X},\mathbf{Y}) \notin T_\epsilon\right) \leq \frac{2A}{n\epsilon^2}$. ∎

Moreover, not only is it unlikely that $(\mathbf{X},\mathbf{Y})$ does not fall in T_ϵ, but the Euclidean volume (which we denote by "Vol") of this set is not too large:

Lemma 8.

$$\sum_{\mathbf{x}\in\{0,1\}^n} \text{Vol}(T_\epsilon(\mathbf{x})) \leq 2^{\mathcal{H}(\mathbf{X},\mathbf{Y})+\epsilon n}$$

Proof. Let us notice that

$$\begin{aligned}
1 = \sum_{\mathbf{x}\in\{0,1\}^n} \frac{1}{2^n}\int_{\mathbb{R}^n} f(\mathbf{y}|\mathbf{x})d\mathbf{y} &\geq \sum_{\mathbf{x}\in\{0,1\}^n} \frac{1}{2^n}\int_{T_\epsilon(\mathbf{x})} f(\mathbf{y}|\mathbf{x})d\mathbf{y} \\
&\geq \sum_{\mathbf{x}\in\{0,1\}^n} \text{Vol}(T_\epsilon(\mathbf{x}))2^{-\mathcal{H}(\mathbf{X},\mathbf{Y})-\epsilon n}
\end{aligned}$$

where the last inequality follows from (24) ∎

We will use this result to show that

Proposition 1. *If* $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$ *is a couple of independent random variables, where* $\tilde{\mathbf{X}}$ *is uniformly distributed and* $\tilde{\mathbf{Y}}$ *has the same distribution as* $\mathbf{Y}$*, then* $\mathbf{Pr}\left((\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) \in T_\epsilon\right) \leq 2^{-C+2n\epsilon}$ *with* $C \overset{def}{=} \sum_{i=1}^n \mathbf{Cap}(\sigma_i^2)$*.*

Proof. We evaluate $\mathbf{Pr}\left((\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) \in T_\epsilon\right)$ as follows

$$\mathbf{Pr}\left((\tilde{\mathbf{X}},\tilde{\mathbf{Y}}) \in T_\epsilon\right) = \sum_{\mathbf{x}\in\{0,1\}^n} \frac{1}{2^n} \int_{T_{\mathbf{x}}(\epsilon)} f(\mathbf{y}) \leq \sum_{\mathbf{x}\in\{0,1\}^n} \frac{1}{2^n} \operatorname{Vol}(T_{\mathbf{x}}(\epsilon)) 2^{-\mathcal{H}(\mathbf{Y})+\epsilon n}$$

The last inequality follows from (23) in the definition of the typical set. We use now Lemma 8 to obtain

$$\mathbf{Pr}\left((\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) \in T_\epsilon\right) \leq \frac{1}{2^n} 2^{\mathcal{H}(\mathbf{X},\mathbf{Y})+\epsilon n} 2^{-\mathcal{H}(\mathbf{Y})+\epsilon n} \leq 2^{-n+\mathcal{H}(\mathbf{X};\mathbf{Y})-\mathcal{H}(\mathbf{Y})+2\epsilon n}$$

By using the expressions for $\mathcal{H}(\mathbf{X}, \mathbf{Y})$ and $\mathcal{H}(\mathbf{Y})$ given in Lemma 3 we deduce $-n+\mathcal{H}(\mathbf{X}, \mathbf{Y})-\mathcal{H}(\mathbf{Y}) = -\sum_{i=1}^n \mathbf{Cap}(\sigma_i^2)$. This finishes the proof. ∎

These results can be used to analyze the following typical set decoder, which takes as inputs a vector $\mathbf{y}$ in $\mathbb{R}^n$ which is the output of the Gaussian channel described in Section 2 and a real parameter ϵ, and outputs either "Failure" or a possible key $\tilde{\mathbf{K}} \in \{0, 1\}^n$.

```
TYPICAL SET DECODER($\mathbf{y}, \epsilon$)
counter ← 0
for all possible values $\mathbf{k}$ of $\tilde{\mathbf{K}}$
    do if $\mathbf{y} \in T_{\mathbf{k}}(\epsilon)$
          then counter ← counter + 1
               result ← $\mathbf{k}$
if counter = 1
   then return result
   else return failure
```

This algorithm is therefore successful if and only if $\mathbf{y}$ is in the typical set of the right key and if there is no other value $\mathbf{k}$ for $\tilde{\mathbf{K}}$ for which $\mathbf{y}$ belongs to the typical set associated to $\mathbf{k}$. Let us now finish the proof of Theorem 1.

Proof. Let $\mathbf{k}$ be right value of $\tilde{\mathbf{K}}$ and let $\mathcal{C}$ be the set of possible values of $\tilde{\mathbf{K}}$. The probability P_{err} that the typical decoder fails is clearly upper-bounded by

$$P_{\text{err}} \leq \mathbf{Pr}_{\mathbf{y},\mathcal{C}}(\overline{T_{\mathbf{k}}(\epsilon)}) + \sum_{\mathbf{k}'\in\mathcal{C},\mathbf{k}'\neq\mathbf{k}} \mathbf{Pr}_{\mathbf{y},\mathcal{C}}\left(T_{\mathbf{k}'}(\epsilon)\right) \tag{32}$$

where $\overline{T_{\mathbf{k}}(\epsilon)}$ denotes the complementary set of $T_{\mathbf{k}}(\epsilon)$. On the one hand

$$\mathbf{Pr}_{\mathbf{y},\mathcal{C}}(\overline{T_{\mathbf{k}}(\epsilon)}) = \mathbf{Pr}((\mathbf{X},\mathbf{Y}) \notin T_\epsilon) \leq \frac{A}{\epsilon^2 n}.$$

by Lemma 4, and on the other hand for $\mathbf{k}' \neq \mathbf{k}$:

$$\sum_{\mathbf{k}'\in\mathcal{C},\mathbf{k}'\neq\mathbf{k}} \mathbf{Pr}_{\mathbf{y},\mathcal{C}}\left(T_{\mathbf{k}'}(\epsilon)\right) \leq \sum_{\mathbf{k}'\in\mathcal{C}} \mathbf{Pr}_{\mathbf{y},\mathcal{C}}\left(T_{\mathbf{k}'}(\epsilon)\right) = 2^r \mathbf{Pr}\left((\tilde{\mathbf{X}},\tilde{\mathbf{Y}}) \in T_\epsilon\right)$$
$$\leq 2^{r-\sum_{i=1}^n \mathbf{Cap}(\sigma_i^2)+2\epsilon n}$$

by Proposition 1. By plugging in these two upper bounds in the union bound (32) we obtain $P_{\text{err}} \leq \frac{A}{\epsilon^2 n} + 2^{r-\sum_{i=1}^n \mathbf{Cap}(\sigma_i^2)+2\epsilon n} \leq \frac{A}{\epsilon^2 n} + 2^{-\delta n+2\epsilon n}$. We finish the proof by choosing $\epsilon = \frac{\delta}{4}$. ∎

B How to combine information from the 4 groups

Let us recall the problematic. Four groups of approximations are used to recover 42 bits of the master key. The two first groups (namely G1 and G2) involve 19 bits of the key each and the two others (G3 and G4) involve only 13 key bits. Some bits are common to some groups what explains that the overall set of approximations only involves 42 bits. We denote by L_{G_i} $(1 \leq i \leq 4)$ the list sorted obtained after perfoming the Walsh Hadamard transformation on group G_i. Elements in L_{G_i} are of the form $(\mathbf{k}, l_i(\mathbf{k}))$ where $\mathbf{k} \in \mathbb{F}_2^{19/13}$ is the 13/19 bits candidate and $l_i(\mathbf{k})$ the log-likelihood of this candidate regarding the ith group of approximations. The question is how do we combine the information obtained by the 4 groups of approximations to efficiently recover the rank of the good subkey in the sorted list of candidates.

First of all, notice that the log-likelihood of a subkey is the sum of the log-likelihoods obtained with each group of approximations that is

$$l(\mathbf{k}) = l_1(\mathbf{k}) + l_2(\mathbf{k}) + l_3(\mathbf{k}) + l_4(\mathbf{k}).$$

The second thing to notice is that in our simulations, we know the value of the good key $\mathbf{k}^*$. Thus, we can compute $l(\mathbf{k}^*)$ from the L_i's.

We recall that the total dimension of the space spanned by κ's is 42 thus computing the log-likelihood of all subkeys and then sorting the list is not efficient enough.

We actually merge information from groups G1 and G3 (resp. G2 and G4) into 2 sorted lists L_1 and L_2. Since 6 bits are in common for each couple of groups, the size of these two lists is 2^{26}.

$$L_1 = \{(\mathbf{k}_1||\mathbf{k}_3, l_1(\mathbf{k}_1) + l_3(\mathbf{k}_3)), (\mathbf{k}_1, l(\mathbf{k}_1)) \in L_{G_1}, (\mathbf{k}_3, l(\mathbf{k}_3)) \in L_{G_3}\}$$

$$L_2 = \{(\mathbf{k}_2||\mathbf{k}_4, l_2(\mathbf{k}_2) + l_4(\mathbf{k}_4)), (\mathbf{k}_2, l(\mathbf{k}_2)) \in L_{G_2}, (\mathbf{k}_4, l(\mathbf{k}_4)) \in L_{G_4}\}$$

Finally, subkeys from L_1 and L_2 have 10 common bits thus we split L_2 into 2^{10} subkey cosets $L_{2,a}$ of size 2^{16} (with $L_{2,a}$ still sorted). Now, for each subkey $\mathbf{k} \in L_1$, we have a list $L_{2,\mathbf{k}_{10}}$ of all subkeys from L_2 with the same common 10 bits as $\mathbf{k}$. To compute the rank of the good subkey, we sum, for each subkey $\mathbf{k} \in L_1$, the number of subkeys in the subkey coset $L_{2,\mathbf{k}_{10}}$ that leads to a better global log-likelihood than the good subkey.

The complexity of the analysis phase is dominated by the merging operations of complexity 2^{26}.