



# Improving recent side-channel attacks against the DES key schedule

Andreas Wiemers<sup>1</sup> · Johannes Mittmann<sup>1</sup>

Received: 2 March 2021 / Accepted: 26 October 2021 / Published online: 4 December 2021  
© The Author(s) 2021

## Abstract

Recent publications consider side-channel attacks against the key schedule of the Data Encryption Standard (DES). These publications identify a leakage model depending on the XOR of register values in the DES key schedule. Building on this leakage model, we first revisit a discrete model which assumes that the Hamming distances between subsequent round keys leak without error. We analyze this model formally and provide theoretical explanations for observations made in previous works. Next we examine a continuous model which considers more points of interest and also takes noise into account. The model gives rise to an evaluation function for key candidates and an associated notion of key ranking. We develop an algorithm for enumerating key candidates up to a desired rank which is based on the Fincke–Pohst lattice point enumeration algorithm. We derive information-theoretic bounds and estimates for the remaining entropy and compare them with our experimental results. We apply our attack to side-channel measurements of a security controller. Using our enumeration algorithm we are able to significantly improve the results reported previously for the same measurement data.

**Keywords** Side-channel analysis · Data Encryption Standard (DES) · Key schedule · XOR leakage · Linear regression · Fincke–Pohst enumeration · Mutual information

## 1 Introduction

Publications by Wagner et al. [8, 14–17] attempt side-channel attacks against the key schedule of the Data Encryption Standard (DES), which are further investigated in [7]. They conduct template attacks against several microcontrollers and demonstrate that the entropy of the 56-bit DES keys can be reduced to 48 bits on average in their experimental setting.

In this article we consider the leakage model identified in aforementioned works. The model assumes that information about the XOR of register values in the DES key schedule leaks.

First we revisit a discrete model examined in [14], which assumes that the Hamming distances between subsequent round keys leak without error. We analyze this model formally and provide theoretical explanations for observations made in previous works. Next we examine a continuous model which considers more points of interest (POI) and

also takes noise into account. The parameters of this model can be learned in a profiling phase using linear regression. The model gives rise to an evaluation function for key candidates and an associated notion of key ranking. We develop an algorithm for enumerating key candidates up to a desired rank which is based on the Fincke–Pohst lattice point enumeration algorithm [4].

We apply our attack to side-channel measurements provided by the authors of [7]. The measurements are obtained from an implementation without countermeasures. In the profiling phase we use nearly 900,000 measurements to learn the parameters of our model and in the attack phase we use averages of several hundred measurements. Using our enumeration algorithm we are able to explicitly compute the ranks of the correct keys and find that the entropy of the DES keys is reduced to 15 bits on average and below 21 bits in 75% of the considered cases. Furthermore, we conduct a series of experiments on simulated measurements in different noise regimes.

We derive information-theoretic bounds and estimates for the remaining entropy and compare them with our experimental results. Our bounds and heuristics may be used by evaluators as theoretical tools for assessing side-channel leakage of DES implementations. The considered leakage model is quite general and should be adaptable to typical cryptographic implementations on security controllers.

✉ Andreas Wiemers  
andreas.wiemers@bsi.bund.de

Johannes Mittmann  
johannes.mittmann@bsi.bund.de

<sup>1</sup> Bundesamt für Sicherheit in der Informationstechnik (BSI),  
Bonn, Germany

Moreover, the information-theoretic bounds are independent of the DES algorithm and depend only on the leakage model. Therefore, we believe that our theoretical results could be useful for assessing side-channel leakage of other cryptographic implementations.

Fortunately, our attack becomes infeasible in the presence of large noise. Therefore it is possible to design effective countermeasures against this attack based on randomization (e.g. masking) and/or limited key usage.

The original DES cipher can already be broken by exhaustive key search, but the use of Triple DES is still considered practically secure for some applications. Although more secure block ciphers such as AES are recommended in general, the use of Triple DES remains quite widespread (e.g. in electronic payment systems). Therefore, many security controllers still offer hardware support for (Triple) DES today and assessing the security of these implementations remains an important topic. We note that our attack extends to Triple DES by applying it to the three DES invocations individually and by using the classic meet-in-the-middle approach.

## 2 Preliminaries

### 2.1 Notation

Let  $[n] := \{1, \dots, n\}$ .

We denote by  $\mathbf{0}_n = (0, \dots, 0)^\top \in \mathbb{R}^n$  the all-zero-vector, by  $\mathbf{1}_n = (1, \dots, 1)^\top \in \mathbb{R}^n$  the all-one-vector, by  $\mathbf{I}_n \in \mathbb{R}^{n \times n}$  the identity matrix, and by  $\mathbf{0}_{m,n} \in \mathbb{R}^{m \times n}$  the zero matrix. The Euclidean norm of a vector  $\mathbf{v} \in \mathbb{R}^n$  is denoted by  $\|\mathbf{v}\|$ .

Let  $\mathbf{a} = a_1 \dots a_n \in \{0, 1\}^n$  be a bit-string. Depending on the context, we identify  $\mathbf{a}$  with the (column) vector  $(a_1, \dots, a_n)^\top \in \mathbb{R}^n$  or the (big-endian represented) integer  $\sum_{i=1}^n a_i 2^{n-i} \in \{0, 1, \dots, 2^n - 1\}$ . The bit-wise XOR of  $\mathbf{a}, \mathbf{b} \in \{0, 1\}^n$  is denoted by  $\mathbf{a} \oplus \mathbf{b}$ . The bit-wise complement of  $\mathbf{a} \in \{0, 1\}^n$  is denoted by  $\bar{\mathbf{a}} := \mathbf{a} \oplus \mathbf{1}_n$ , the cyclic left-shift (rotation) of  $\mathbf{a}$  by  $k \in \mathbb{Z}$  positions is denoted by  $\mathbf{a} \lll k := a_{1+k} \dots a_{n+k}$  (where indices are to be interpreted modulo  $n$  with representatives in  $[n]$ ), and the Hamming weight of  $\mathbf{a}$  is denoted by  $\text{wt}(\mathbf{a}) := \sum_{i=1}^n a_i$ .

### 2.2 DES key schedule

The Data Encryption Standard (DES) is defined in [11]. In this article we are only concerned with the DES key schedule, which we describe below.

For simplicity and without loss of generality, we assume that DES keys are represented by  $\mathbf{k} = (\mathbf{c}, \mathbf{d}) \in \{0, 1\}^{56}$ , where  $\mathbf{c}, \mathbf{d} \in \{0, 1\}^{28}$  denote the contents of the C- and D-register after the map PC-1 (permuted choice 1) has been applied to the actual DES master key KEY (i.e.  $\mathbf{c}, \mathbf{d}$  correspond to  $C_0, D_0$  in the notation of [11]).

The DES round keys  $\mathbf{k}_1, \dots, \mathbf{k}_{16} \in \{0, 1\}^{48}$  are derived from  $\mathbf{k} = (\mathbf{c}, \mathbf{d})$  as follows. We write  $\mathbf{c} = c_1 \dots c_{28}$  and  $\mathbf{d} = d_1 \dots d_{28}$ . In each round  $i \in [16]$ , the values of the C- and D-registers are cyclically shifted (i.e. rotated) by 1 or 2 positions to the left. The number  $\delta(i)$  of shifts in round  $i$  is given by

$$\delta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \end{pmatrix}. \tag{1}$$

The accumulated number  $\rho(i)$  of shifts (modulo 28) up to round  $i$  is given by

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 \\ 1 & 2 & 4 & 6 & 8 & 10 & 12 & 14 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 0 \end{pmatrix}, \tag{2}$$

i.e. we have  $\rho(1) = \delta(1)$  and  $\rho(i) = \rho(i - 1) + \delta(i) \pmod{28}$  for  $2 \leq i \leq 16$ . The values of the C- and D-registers in round  $i$  are therefore given by

$$(\mathbf{c}_i, \mathbf{d}_i) := (c_{\rho(i)+1} \dots c_{\rho(i)+28}, d_{\rho(i)+1} \dots d_{\rho(i)+28}) \text{ for } i \in [16], \tag{3}$$

where the indices are to be interpreted modulo 28 (with representatives in [28]).

In each round  $i \in [16]$ , the map PC-2 (permuted choice 2) is applied to  $(\mathbf{c}_i, \mathbf{d}_i)$  to obtain the round key  $\mathbf{k}_i$ . The map PC-2 is defined as

$$\text{PC-2: } \{0, 1\}^{56} \rightarrow \{0, 1\}^{48}, \quad (c_1 \dots c_{28}, d_1 \dots d_{28}) \\ \mapsto (c_{\sigma(1)} \dots c_{\sigma(24)}, d_{\tau(1)} \dots d_{\tau(24)}),$$

where  $\sigma: [24] \rightarrow [28] \setminus \{9, 18, 22, 25\}$  and  $\tau: [24] \rightarrow [28] \setminus \{7, 10, 15, 26\}$  are the bijections defined by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\ 14 & 17 & 11 & 24 & 1 & 5 & 3 & 28 & 15 & 6 & 21 & 10 & 23 & 19 & 12 & 4 & 26 & 8 & 16 & 7 & 27 & 20 & 13 & 2 \end{pmatrix},$$

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 & 23 & 24 \\ 13 & 24 & 3 & 9 & 19 & 27 & 2 & 12 & 23 & 17 & 5 & 20 & 16 & 21 & 11 & 28 & 6 & 25 & 18 & 14 & 22 & 8 & 1 & 4 \end{pmatrix}.$$

We denote by  $M_\sigma := \{9, 18, 22, 25\}$  and  $M_\tau := \{7, 10, 15, 26\}$  the sets of elements in [28] which are “missing” from the images of  $\sigma$  and  $\tau$ , respectively. The round keys are finally defined as  $k_i := \text{PC-2}(c_i, d_i)$  for  $i \in [16]$ . Written in terms of the original key  $k = (c, d)$ , we have

$$k_i = (c_{\rho(i)+\sigma(1)} \cdots c_{\rho(i)+\sigma(24)}, d_{\rho(i)+\tau(1)} \cdots d_{\rho(i)+\tau(24)}) \text{ for } i \in [16], \tag{4}$$

where the indices are again to be interpreted modulo 28.

### 2.3 Leakage models

We consider variations of the leakage models identified in previous works [7,8,14–17]. The models assume that the key-dependent leakage originates from updates  $(c_{i+1}, d_{i+1}) \leftarrow (c_i, d_i)$  of the C- and D-registers and/or updates  $k_{i+1} \leftarrow k_i$  of the round-key register for  $i \in [15]$ . Moreover, it is assumed that the leakage stemming from a bit transition  $b \leftarrow a$  in those register updates depends only on  $a \oplus b$  (XOR leakage) for  $a, b \in \{0, 1\}$ .

Let  $a \in \{0, 1\}^{28}$  be one half of a DES key  $(c, d) \in \{0, 1\}^{56}$  in the C- or D-register. By (2), (3), and (4), the bit transitions occurring in the DES key schedule are of the form  $a_{i+1} \leftarrow a_i$  (shift-1 transitions) or  $a_{i+2} \leftarrow a_i$  (shift-2 transitions) for some  $i \in [28]$  and with indices interpreted modulo 28. Hence we may assume that the leakage depends only on  $a_i \oplus a_{i+1}$  and  $a_i \oplus a_{i+2}$  or, equivalently, only on  $(-1)^{a_i \oplus a_{i+1}}$  and  $(-1)^{a_i \oplus a_{i+2}}$  for all  $i \in [28]$ . Since shift-1 transitions appear in 3 rounds and shift-2 transitions in 12 rounds of the key schedule (cf. (1)), it is conceivable that shift-2 transitions will have a higher impact on the total leakage.

Based on this discussion, we introduce explanatory variables for the leakage models as follows. For a shift  $k \in \{1, 2\}$  and a key half  $a \in \{0, 1\}^{28}$ , we define the vector

$$\Delta_k(a) := \mathbf{1}_{28} - 2 \cdot (a \oplus (a \lll k)) \in \{\pm 1\}^{28}. \tag{5}$$

Written differently, we have

$$\Delta_1(a) = ((-1)^{a_1 \oplus a_2}, (-1)^{a_2 \oplus a_3}, \dots, (-1)^{a_{27} \oplus a_{28}}, (-1)^{a_{28} \oplus a_1})^\top \text{ and}$$

$$\Delta_2(a) = ((-1)^{a_1 \oplus a_3}, (-1)^{a_2 \oplus a_4}, \dots, (-1)^{a_{27} \oplus a_1}, (-1)^{a_{28} \oplus a_2})^\top.$$

Furthermore, we define the stacked vectors

$$\Delta(a) := \begin{pmatrix} \Delta_1(a) \\ \Delta_2(a) \end{pmatrix} \in \{\pm 1\}^{56} \text{ and}$$

$$\Delta(c, d) := \begin{pmatrix} \Delta(c) \\ \Delta(d) \end{pmatrix} \in \{\pm 1\}^{112} \tag{6}$$

for all key halves  $a \in \{0, 1\}^{28}$  and full keys  $(c, d) \in \{0, 1\}^{56}$ . The components of  $\Delta(a)$  are illustrated in Fig. 1. The vector  $\Delta(c, d)$  captures all possible bit transitions in the key schedule of  $(c, d)$  and will serve as explanatory variable for the leakage models.

**Remark 1** Let  $k \in \{1, 2\}$ . The map  $\Delta_k$  is a group homomorphism from  $(\{0, 1\}^{28}, \oplus)$  to  $(\{\pm 1\}^{28}, \odot)$ , where  $\odot$  denotes componentwise multiplication. A vector  $x \in \{\pm 1\}^{28}$  is in the image of  $\Delta_k$  iff  $x$  has an even number of positive components iff  $x$  has an even number of negative components iff  $\sum_{i=1}^{28} x_i = 0 \pmod{4}$ . The kernel of  $\Delta_k$  is the cyclic group generated by  $\mathbf{1}_{28}$ . In particular, we have  $\Delta_k(a) = \Delta_k(\bar{a})$  and  $\Delta(a) = \Delta(\bar{a})$  for all  $a \in \{0, 1\}^{28}$ , as well as  $\Delta(c, d) = \Delta(\bar{c}, d) = \Delta(c, \bar{d}) = \Delta(\bar{c}, \bar{d})$  for all  $c, d \in \{0, 1\}^{28}$ .

Now we can define the general form of the leakage models under consideration. We restrict ourselves to one of the simplest conceivable settings in which the leakage for a key  $(c, d)$  is given by an  $\mathbb{R}$ -linear function of  $\Delta(c, d)$  and a key-independent error term.

**Leakage Model 1 (General model)** Let  $m \geq 1$ , let  $W \in \mathbb{R}^{m \times 112}$  be a fixed weight matrix, and let  $K = (C, D)$  be a uniformly distributed random variable on  $\{0, 1\}^{56}$ . We define the random variable  $Y$  on  $\mathbb{R}^m$  by

$$Y = W \Delta(C, D) + \epsilon, \tag{7}$$

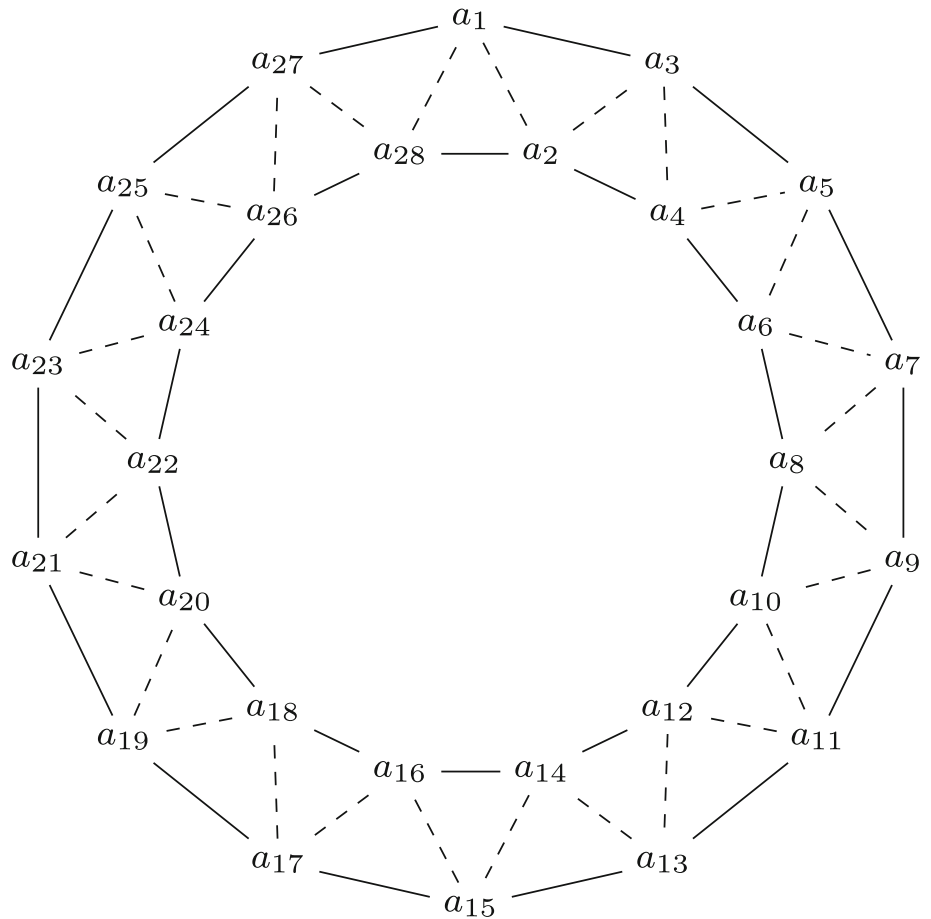
where  $\epsilon$  is a zero-mean random variable on  $\mathbb{R}^m$  which is independent of  $K$ .

We refer to realizations  $y \in \mathbb{R}^m$  of  $Y$  as observations, to realizations of  $\epsilon$  in  $\mathbb{R}^m$  as errors or noise, and to  $m$  as the number of points of interest (POIs). The following lemma collects some general properties of the random variables in Leakage Model 1.

**Lemma 1** Consider the situation of Leakage Model 1 and let  $W_1, W_2 \in \mathbb{R}^{m \times 56}$  such that  $W = (W_1, W_2)$ .

(a) We have  $W \Delta(C, D) = W_1 \Delta(C) + W_2 \Delta(D)$ .

**Fig. 1** The nodes in this graph represent the bits of a key half  $\mathbf{a} \in \{0, 1\}^{28}$  in the C- or D-register. The dashed edges  $\{a_i, a_{i+1}\}$  correspond to the components  $(-1)^{a_i \oplus a_{i+1}}$  of  $\Delta_1(\mathbf{a})$  and the solid edges  $\{a_i, a_{i+2}\}$  to the components  $(-1)^{a_i \oplus a_{i+2}}$  of  $\Delta_2(\mathbf{a})$ . The union of all edges corresponds to the components of  $\Delta(\mathbf{a})$ . For a full key  $(\mathbf{c}, \mathbf{d}) \in \{0, 1\}^{56}$ , the components of  $\Delta(\mathbf{c}, \mathbf{d})$  can be described by two disjoint copies of this graph with nodes labelled by the bits of  $\mathbf{c}$  and  $\mathbf{d}$ , respectively



- (b) We have  $E(\Delta(\mathbf{C})) = E(\Delta(\mathbf{D})) = \mathbf{0}_{56}$  and  $\text{Cov}(\Delta(\mathbf{C})) = \text{Cov}(\Delta(\mathbf{D})) = \mathbf{I}_{56}$ .
- (c) We have  $E(\Delta(\mathbf{C}, \mathbf{D})) = \mathbf{0}_{112}$  and  $\text{Cov}(\Delta(\mathbf{C}, \mathbf{D})) = \mathbf{I}_{112}$ .
- (d) We have  $E(\mathbf{Y}) = \mathbf{0}_m$  and  $\text{Cov}(\mathbf{Y}) = \mathbf{W}\mathbf{W}^\top + \text{Cov}(\boldsymbol{\epsilon}) = \mathbf{W}_1\mathbf{W}_1^\top + \mathbf{W}_2\mathbf{W}_2^\top + \text{Cov}(\boldsymbol{\epsilon})$ .

**Proof** Assertion (a) is obvious. To show (b), denote  $\mathbf{X} := \Delta(\mathbf{C})$ . Clearly  $E(\mathbf{X}) = \mathbf{0}_{56}$ , hence  $\text{Cov}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^\top) = (E(X_i X_j))_{i,j \in [56]}$ . Let  $i, j \in [56]$ . We distinguish two cases:

- If  $i = j$ , then  $E(X_i X_j) = E(X_i^2) = 1$ .
- If  $i \neq j$ , then there are  $p, q, r, s \in [28]$  such that  $\{p, q\} \neq \{r, s\}$ ,  $q - p \pmod{28} \in \{1, 2\}$ ,  $s - r \pmod{28} \in \{1, 2\}$ , and

$$\begin{aligned} E(X_i X_j) &= E((-1)^{C_p \oplus C_q} (-1)^{C_r \oplus C_s}) \\ &= E((-1)^{C_p} (-1)^{C_q} (-1)^{C_r} (-1)^{C_s}). \end{aligned}$$

Since  $p \neq q$ ,  $r \neq s$ , and  $\{p, q\} \neq \{r, s\}$ , we have  $p \notin \{q, r, s\}$  or  $q \notin \{p, r, s\}$ . Let us assume  $p \notin \{q, r, s\}$  (the case  $q \notin \{p, r, s\}$  can be handled analogously). Then  $(-1)^{C_p}$  and  $(-1)^{C_q} (-1)^{C_r} (-1)^{C_s}$

are independent, therefore  $E(X_i X_j) = E((-1)^{C_p}) E((-1)^{C_q} (-1)^{C_r} (-1)^{C_s}) = 0$ .

We have shown that  $\text{Cov}(\mathbf{X}) = \mathbf{I}_{56}$ . The remaining assertions of (b) follow analogously. Since  $\Delta(\mathbf{C})$  and  $\Delta(\mathbf{D})$  are independent, (b) implies (c). Assertion (d) follows from (a), (b), and (c) by linearity of expectation and independence of  $\mathbf{C}$ ,  $\mathbf{D}$ , and  $\boldsymbol{\epsilon}$ .  $\square$

### 3 Hamming weight model

In this section we consider a discrete leakage model, whose observations consist of the (centered) Hamming distances between subsequent round keys and are error-free. This model was already examined in [14, Section 5].

**Leakage Model 2 (Hamming weight model)** Let  $\mathbf{K} = (\mathbf{C}, \mathbf{D})$  be a uniformly distributed random variable on  $\{0, 1\}^{56}$  and let  $\mathbf{K}_1, \dots, \mathbf{K}_{16}$  be the random variables on  $\{0, 1\}^{48}$  derived from  $\mathbf{K}$  as defined by Eq. (4). We define the random variable  $\mathbf{Y}$  on  $\mathbb{Z}^{15}$  by

$$Y_i := \text{wt}(\mathbf{K}_i \oplus \mathbf{K}_{i+1}) - 24, \quad i \in [15]. \tag{8}$$

The components  $Y_i$  take values in  $[-24, 24] \cap \mathbb{Z}$ .

This leakage model is a special instance of Leakage Model 1 with  $m = 15$ , a weight matrix  $\mathbf{W} \in \{-\frac{1}{2}, 0\}^{15 \times 112}$ , and error  $\boldsymbol{\varepsilon} = \mathbf{0}_{15}$ . The weight matrix  $\mathbf{W}$  is completely determined by the model assumptions and will be derived in Sect. 3.1.

**Remark 2** In the case of noisy measurements, the error can be reduced by averaging repeated measurements for a fixed key. If the maximum norm of the error vector is less than  $\frac{1}{2}$ , an exact observation as in (8) can be recovered from the noisy version by rounding each component to the nearest integer.

### 3.1 Determination of the weight and covariance matrix

Let  $\mathbf{K} = (\mathbf{C}, \mathbf{D})$  and  $\mathbf{Y}$  be the random variables as defined in Leakage Model 2. We want to determine a weight matrix  $\mathbf{W} \in \{-\frac{1}{2}, 0\}^{15 \times 112}$  such that  $\mathbf{Y} = \mathbf{W}\Delta(\mathbf{C}, \mathbf{D})$ . Let  $i \in [15]$ . Then

$$\begin{aligned} Y_i &= \text{wt}(\mathbf{K}_i \oplus \mathbf{K}_{i+1}) - 24 \\ &= \sum_{j=1}^{24} (C_{\rho(i)+\sigma(j)} \oplus C_{\rho(i+1)+\sigma(j)}) \\ &\quad + \sum_{j=1}^{24} (D_{\rho(i)+\tau(j)} \oplus D_{\rho(i+1)+\tau(j)}) - 24 \\ &= -\frac{1}{2} \sum_{j=1}^{24} (-1)^{C_{\rho(i)+\sigma(j)} \oplus C_{\rho(i+1)+\sigma(j)}} \\ &\quad - \frac{1}{2} \sum_{j=1}^{24} (-1)^{D_{\rho(i)+\tau(j)} \oplus D_{\rho(i+1)+\tau(j)}}. \end{aligned}$$

The images of  $\sigma$  and  $\tau$  are  $[28] \setminus M_\sigma$  and  $[28] \setminus M_\tau$ , respectively, where  $M_\sigma = \{9, 18, 22, 25\}$  and  $M_\tau = \{7, 10, 15, 26\}$  (cf. Sect. 2.2). Changing the summation order, we obtain the representation

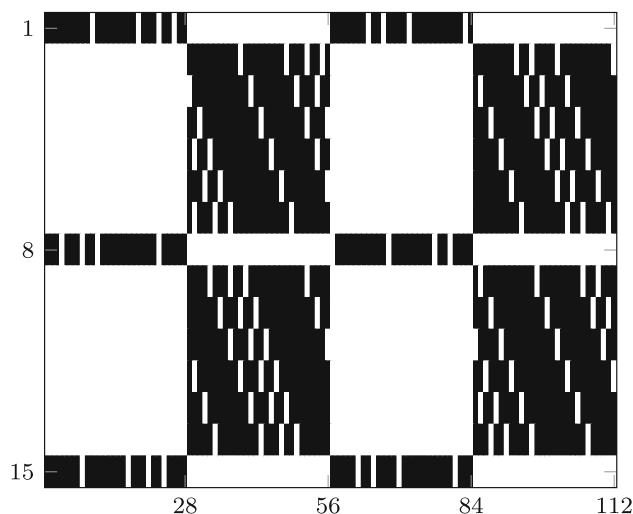


Fig. 2 Matrix plot of the weight matrix  $\mathbf{W}$  with element values  $-\frac{1}{2}$  and 0 depicted in black and white, respectively

$$\begin{aligned} Y_i &= -\frac{1}{2} \sum_{j \in [28] \setminus (\rho(i) + M_\sigma)} (-1)^{C_j \oplus C_{j+\delta(i+1)}} \\ &\quad - \frac{1}{2} \sum_{j \in [28] \setminus (\rho(i) + M_\tau)} (-1)^{D_j \oplus D_{j+\delta(i+1)}}, \end{aligned} \tag{9}$$

where the elements in the shifted sets  $\rho(i) + M_\sigma$  and  $\rho(i) + M_\tau$  are to be interpreted modulo 28 (with representatives in  $[28]$ ). From (9) the weight matrix  $\mathbf{W} \in \{-\frac{1}{2}, 0\}^{15 \times 112}$  can be easily read off, see Fig. 2.

Next we want to determine the covariance matrix  $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{Y})$ . By Lemma 1(d), we have  $\boldsymbol{\Sigma} = \mathbf{W}\mathbf{W}^\top$ . Let  $i, j \in [15]$ . If  $\delta(i+1) \neq \delta(j+1)$  (rounds  $i+1$  and  $j+1$  have different shifts), then  $\sigma_{i,j} = 0$ . If  $\delta(i+1) = \delta(j+1)$  (rounds  $i+1$  and  $j+1$  have the same shift), then

$$\begin{aligned} \sigma_{i,j} &= \frac{1}{4} \left( 56 - \#((\rho(i) + M_\sigma) \cup (\rho(j) + M_\sigma)) \right. \\ &\quad \left. - \#((\rho(i) + M_\tau) \cup (\rho(j) + M_\tau)) \right). \end{aligned}$$

More concretely, we get

$$\Sigma = \frac{1}{4} \begin{pmatrix} 48 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 41 & 0 & 0 & 0 & 0 & 0 & 40 \\ 0 & 48 & 40 & 41 & 40 & 41 & 40 & 0 & 41 & 41 & 41 & 42 & 41 & 41 & 0 \\ 0 & 40 & 48 & 40 & 41 & 40 & 41 & 0 & 41 & 41 & 41 & 41 & 42 & 41 & 0 \\ 0 & 41 & 40 & 48 & 40 & 41 & 40 & 0 & 42 & 41 & 41 & 41 & 41 & 42 & 0 \\ 0 & 40 & 41 & 40 & 48 & 40 & 41 & 0 & 41 & 42 & 41 & 41 & 41 & 41 & 0 \\ 0 & 41 & 40 & 41 & 40 & 48 & 40 & 0 & 41 & 41 & 42 & 41 & 41 & 41 & 0 \\ 0 & 40 & 41 & 40 & 41 & 40 & 48 & 0 & 42 & 41 & 41 & 42 & 41 & 41 & 0 \\ 41 & 0 & 0 & 0 & 0 & 0 & 0 & 48 & 0 & 0 & 0 & 0 & 0 & 0 & 41 \\ 0 & 41 & 41 & 42 & 41 & 41 & 42 & 0 & 48 & 40 & 41 & 40 & 41 & 40 & 0 \\ 0 & 41 & 41 & 41 & 42 & 41 & 41 & 0 & 40 & 48 & 40 & 41 & 40 & 41 & 0 \\ 0 & 41 & 41 & 41 & 41 & 42 & 41 & 0 & 41 & 40 & 48 & 40 & 41 & 40 & 0 \\ 0 & 42 & 41 & 41 & 41 & 41 & 42 & 0 & 40 & 41 & 40 & 48 & 40 & 41 & 0 \\ 0 & 41 & 42 & 41 & 41 & 41 & 41 & 0 & 41 & 40 & 41 & 40 & 48 & 40 & 0 \\ 0 & 41 & 41 & 42 & 41 & 41 & 41 & 0 & 40 & 41 & 40 & 41 & 40 & 48 & 0 \\ 40 & 0 & 0 & 0 & 0 & 0 & 0 & 41 & 0 & 0 & 0 & 0 & 0 & 0 & 48 \end{pmatrix}.$$

We note that  $\det(\Sigma) = 4^{-15} \cdot 4650233960271024$ .

### 3.2 Key ranking and key enumeration

Let  $y \in \mathbb{Z}^{15}$  be an observation under Leakage Model 2 corresponding to an unknown key  $k^* = (c^*, d^*) \in \{0, 1\}^{56}$ , i.e. we have  $y = W\Delta(c^*, d^*)$ . We denote by

$$\mathcal{C}(y) := \{(c, d) \in \{0, 1\}^{56} \mid y = W\Delta(c, d)\} \tag{10}$$

the set of key candidates for observation  $y$ . The rank of  $k^*$  is defined as

$$\mathcal{R}(k^*) := \#\mathcal{C}(y). \tag{11}$$

Note that  $\mathcal{R}(k^*)$  is a multiple of 4 (cf. Remark 1). We call  $\log_2 \mathcal{R}(k^*)$  the logarithmic key rank of  $k^*$ .

At first glance, enumerating the set  $\mathcal{C}(y)$  looks like a 56-bit (or 54-bit) problem. However, we can apply a meet-in-the-middle approach (cf. [14, Section 5]). By Lemma 1(a), we have the decomposition

$$W\Delta(c, d) = W_1\Delta(c) + W_2\Delta(d) \quad \text{for all } (c, d) \in \{0, 1\}^{56}, \tag{12}$$

where  $W_1, W_2 \in \mathbb{R}^{15 \times 56}$  such that  $W = (W_1, W_2)$ . This leads to the following simple enumeration procedure.

#### Algorithm 1

**Input:** A vector  $y \in ([-24, 24] \cap \mathbb{Z})^{15}$ .

**Output:** The set of key candidates  $\mathcal{C}(y)$ .

##### 1. Compute the lists

$$\mathcal{L}_1 \leftarrow \{(c, y - W_1\Delta(c)) \mid c = 0, \dots, 2^{27} - 1\},$$

$$\mathcal{L}_2 \leftarrow \{(d, W_2\Delta(d)) \mid d = 0, \dots, 2^{27} - 1\}$$

and sort them by the second component of their elements (e.g. using the lexicographical order on  $\mathbb{Z}^{15}$ ).

##### 2. Set $\mathcal{C} \leftarrow \emptyset$ . For all $(c, y_1) \in \mathcal{L}_1$ and $(d, y_2) \in \mathcal{L}_2$ with $y_1 = y_2$ , set

$$\mathcal{C} \leftarrow \mathcal{C} \cup \{(c, d), (\bar{c}, d), (c, \bar{d}), (\bar{c}, \bar{d})\}.$$

(Since  $\mathcal{L}_1$  and  $\mathcal{L}_2$  are ordered by the second component of their elements, the lists only have to be traversed once in order to find all collisions in the second component.) Return  $\mathcal{C}$  and stop.

### 3.3 Experiments

In order to estimate the expected logarithmic key rank, we implemented Algorithm 1 in the Julia programming language [2] and conducted 1000 trials. For each trial we chose a random DES key and calculated the associated observation vector  $y$ . Then we enumerated all candidates  $(c, d)$  such that  $y = W\Delta(c, d)$ . Each attempt took approximately 140 seconds of single-core computing time on a standard computer. The results of the experiments are given in Table 1.

We observe that in one half of all cases the logarithmic key rank is less than 16. With such low logarithmic key ranks we note that the classic meet-in-the-middle approach against 3-key Triple DES has very moderate running time. For average keys we can expect a running time of roughly  $2^{32}$  DES encryptions/decryptions.

**Table 1** Empirical distribution of the logarithmic key rank based on 1000 trials with random keys

Logarithmic key rank					Running time
Min	Q1	Median	Q3	Max	
2	13	16	18	23	140s

Q1 and Q3 denote the first and third quartile, respectively

### 3.4 Theoretical estimation of the remaining entropy

The conditional entropy  $H(\mathbf{C}, \mathbf{D} \mid \mathbf{Y})$  is an information-theoretic measure for the expected logarithmic key rank, which we call remaining entropy. We have

$$H(\mathbf{C}, \mathbf{D} \mid \mathbf{Y}) = H(\mathbf{C}, \mathbf{D}) - I(\mathbf{Y}; \mathbf{C}, \mathbf{D}) = 56 - I(\mathbf{Y}; \mathbf{C}, \mathbf{D}),$$

where  $I(\mathbf{Y}; \mathbf{C}, \mathbf{D})$  is the mutual information of  $\mathbf{Y}$  and  $(\mathbf{C}, \mathbf{D})$ . Since  $\mathbf{Y}$  is a deterministic function of  $(\mathbf{C}, \mathbf{D})$ , we have  $I(\mathbf{Y}; \mathbf{C}, \mathbf{D}) = H(\mathbf{Y})$ , hence  $H(\mathbf{C}, \mathbf{D} \mid \mathbf{Y}) = 56 - H(\mathbf{Y})$ .

#### 3.4.1 A lower bound for the remaining entropy

The following lemma provides an upper bound for  $H(\mathbf{Y})$ .

**Lemma 2** *Let  $\mathbf{Y}$  be a random variable on  $\mathbb{Z}^m$  with  $E(\mathbf{Y}) = \mathbf{0}_m$  and positive-definite covariance matrix  $\Sigma := \text{Cov}(\mathbf{Y}) \in \mathbb{R}^{m \times m}$ . Then*

$$H(\mathbf{Y}) \leq \frac{1}{2} \log_2(\det(2\pi e \Sigma)) + m \log_2 \left( \frac{1 + e^{-2\pi^2 \lambda}}{1 - e^{-2\pi^2 \lambda}} \right), \quad (13)$$

where  $\lambda > 0$  is the smallest eigenvalue of  $\Sigma$ .

**Proof** Let  $\mathcal{Y} \subseteq \mathbb{Z}^m$  be the support of  $\mathbf{Y}$ . By Gibbs' inequality (cf. [3, Theorem 2.6.3]), we have

$$\begin{aligned} H(\mathbf{Y}) &= - \sum_{\mathbf{y} \in \mathcal{Y}} \Pr(\mathbf{Y} = \mathbf{y}) \log_2(\Pr(\mathbf{Y} = \mathbf{y})) \\ &\leq - \sum_{\mathbf{y} \in \mathcal{Y}} \Pr(\mathbf{Y} = \mathbf{y}) \log_2(p(\mathbf{y})) \end{aligned}$$

for any probability distribution  $p: \mathcal{Y} \rightarrow [0, 1]$  with support  $\mathcal{Y}$ . Setting

$$p(\mathbf{y}) := \mu e^{-\frac{1}{2} \mathbf{y}^\top \Sigma^{-1} \mathbf{y}} \quad \text{with} \quad \mu := \left( \sum_{\mathbf{y} \in \mathcal{Y}} e^{-\frac{1}{2} \mathbf{y}^\top \Sigma^{-1} \mathbf{y}} \right)^{-1},$$

we obtain

$$\begin{aligned} H(\mathbf{Y}) &\leq -\log_2(\mu) + \frac{\log_2(e)}{2} \sum_{\mathbf{y} \in \mathcal{Y}} \Pr(\mathbf{Y} = \mathbf{y}) \mathbf{y}^\top \Sigma^{-1} \mathbf{y} \\ &= -\log_2(\mu) + \frac{\log_2(e)}{2} E(\mathbf{Y}^\top \Sigma^{-1} \mathbf{Y}) \\ &= -\log_2(\mu) + \frac{\log_2(e)}{2} \text{tr}(\Sigma^{-1} E(\mathbf{Y} \mathbf{Y}^\top)) \\ &\quad \text{(by the "trace trick")} \\ &= -\log_2(\mu) + \frac{m \log_2(e)}{2}. \end{aligned}$$

Using the Poisson summation formula (cf. [1, Lemma (1.1) (i)]), we get

$$\begin{aligned} H(\mathbf{Y}) &\leq \frac{m \log_2(e)}{2} + \log_2 \left( \sum_{\mathbf{y} \in \mathcal{Y}} e^{-\frac{1}{2} \mathbf{y}^\top \Sigma^{-1} \mathbf{y}} \right) \\ &\leq \frac{m \log_2(e)}{2} + \log_2 \left( \sum_{\mathbf{y} \in \mathbb{Z}^m} e^{-\frac{1}{2} \mathbf{y}^\top \Sigma^{-1} \mathbf{y}} \right) \\ &= \frac{m \log_2(e)}{2} + \log_2 \left( \sqrt{(2\pi)^m \det(\Sigma)} \right) \\ &\quad + \log_2 \left( \sum_{\mathbf{y} \in \mathbb{Z}^m} e^{-2\pi^2 \mathbf{y}^\top \Sigma \mathbf{y}} \right) \\ &= \frac{1}{2} \log_2(\det(2\pi e \Sigma)) + \log_2 \left( \sum_{\mathbf{y} \in \mathbb{Z}^m} e^{-2\pi^2 \mathbf{y}^\top \Sigma \mathbf{y}} \right). \end{aligned}$$

Since  $\mathbf{y}^\top \Sigma \mathbf{y} \geq \lambda \|\mathbf{y}\|^2$  for all  $\mathbf{y} \in \mathbb{R}^m$ , we have

$$\begin{aligned} \log_2 \left( \sum_{\mathbf{y} \in \mathbb{Z}^m} e^{-2\pi^2 \mathbf{y}^\top \Sigma \mathbf{y}} \right) &\leq \log_2 \left( \sum_{\mathbf{y} \in \mathbb{Z}^m} e^{-2\pi^2 \lambda \|\mathbf{y}\|^2} \right) \\ &= m \log_2 \left( \sum_{z \in \mathbb{Z}} e^{-2\pi^2 \lambda z^2} \right) \\ &\leq m \log_2 \left( -1 + 2 \sum_{n \geq 0} e^{-2\pi^2 \lambda n} \right) \\ &= m \log_2 \left( \frac{1 + e^{-2\pi^2 \lambda}}{1 - e^{-2\pi^2 \lambda}} \right), \end{aligned}$$

finishing the proof. □

**Remark 3** We note that [1, Lemma (1.5) (i)] implies a better bound for the term  $m \log_2(\dots)$  of Eq. (13) in general. However, the bound of Lemma 2 is sufficient for our purposes.

Applying Lemma 2 to Leakage Model 2, we obtain  $\lambda \approx 0.65$  and  $H(\mathbf{Y}) \leq 41.73$  by numerical methods. We also note that the term

$$15 \log_2 \left( \frac{1 + e^{-2\pi^2 \lambda}}{1 - e^{-2\pi^2 \lambda}} \right) \approx 0.0001 \quad (14)$$

in (13) is negligible for this random variable. We obtain the lower bound

$$H(\mathbf{C}, \mathbf{D} \mid \mathbf{Y}) = 56 - H(\mathbf{Y}) \geq 14.27 \quad (15)$$

for the remaining entropy. The experiments reported in Sect. 3.3 (cf. Table 1) suggest that the remaining entropy

is close to this lower bound. In Sect. 3.4.2 we support this hypothesis by geometric considerations.

### 3.4.2 A heuristic for the remaining entropy

Based on the experiments reported in Sect. 3.3, we propose the heuristic formula

$$H(C, D | Y) \approx 56 - \frac{1}{2} \log_2(\det(2\pi e \Sigma)) \approx 14.27 \quad (16)$$

for the remaining entropy.

Each component  $Y_i$  of  $Y$  is a sum of independent random variables and we can certainly approximate the distribution of each  $Y_i$  by a continuous normal distribution. But is it a valid approximation, if we replace the distribution of  $Y$  by a 15-dimensional normal distribution?

Let  $m \leq n$ , let  $X$  be a uniformly distributed random variable on  $\{\pm 1\}^n$ , and let  $A \in \mathbb{R}^{m \times n}$  be a matrix of full row rank such that  $AX$  takes values in  $\mathbb{Z}^m$ . We have the following general properties:

- If the support of  $AX$  is contained in a subset  $S \subseteq \mathbb{R}^m$ , then clearly  $H(AX) \leq \log_2(\#(S \cap \mathbb{Z}^m))$ . In addition, we can expect that  $\log_2(\#(S \cap \mathbb{Z}^m)) \approx \log_2(\text{vol}(S))$  for “natural” sets  $S$ .
- Let  $A = UDV^T$  be the singular value decomposition of  $A$ , where  $U \in \mathbb{R}^{m \times m}$  and  $V \in \mathbb{R}^{n \times n}$  are orthogonal matrices and  $D \in \mathbb{R}^{m \times n}$  is a rectangular diagonal matrix with non-negative elements on the diagonal. This representation of  $A$  easily implies that  $AX$  takes values in an  $m$ -dimensional ellipsoid with semiaxes equal to the non-zeros elements of  $D$  times  $\sqrt{n}$ . The volume of this ellipsoid is

$$V_m(1)n^{m/2} \sqrt{\det(DD^T)} = V_m(1)n^{m/2} \sqrt{\det(AA^T)},$$

where  $V_m(1)$  denotes the volume of the  $m$ -dimensional ball with radius 1.

- The heuristic argumentation based on the singular value decomposition:
  - (i) If the volume of this ellipsoid is smaller than  $2^n$ , then we can expect that all integer points of this ellipsoid occur in the support of  $AX$ .
  - (ii) The components of  $V^T X$  have expectation 0 and variance 1. Therefore, the bulk of the support of  $AX$  lies in a smaller ellipsoid with semiaxes equal to the non-zero elements of  $D$  times  $\sqrt{m}$ . The volume of this smaller ellipsoid is

$$V_m(1)m^{m/2} \sqrt{\det(AA^T)}.$$

- (iii) Furthermore, if (i) is valid, we expect that the discrete distribution of  $AX$  is “similar” to the continuous distribution  $AZ$ , where  $Z$  is normally distributed with covariance matrix  $I_n$ .  $AZ$  is therefore normally distributed with covariance matrix  $AA^T$ . The entropy of  $AZ$  is well known and given by the formula

$$\frac{1}{2} \log_2(\det(2\pi e AA^T)).$$

Note that the approaches (ii) and (iii) lead to very similar approximations, since

$$\log_2(V_m(1)m^{m/2}) \approx \frac{m}{2} \log_2(2\pi e).$$

### 3.4.3 Distribution of the remaining entropy

Why has the remaining entropy in the experiments in Sect. 3.3 such a large variation?

Each  $y_i$  is a realization of a binomially distributed random variable. If  $y_i$  takes on extreme values near  $\pm 24$ , we have a large amount of information about the key  $(c, d)$ . On the other hand, for  $y_i = 0$  there are many candidates for  $(c, d)$ . As argued in Sect. 3.4.2, we expect that in our case

$$\begin{aligned} \Pr(Y = y) &= \frac{\#\{(c, d) \in \{0, 1\}^{56} \mid W\Delta(c, d) = y\}}{2^{56}} \\ &\approx \frac{1}{\sqrt{(2\pi)^{15} \det(\Sigma)}} \exp\left(-\frac{1}{2} y^T \Sigma^{-1} y\right). \end{aligned}$$

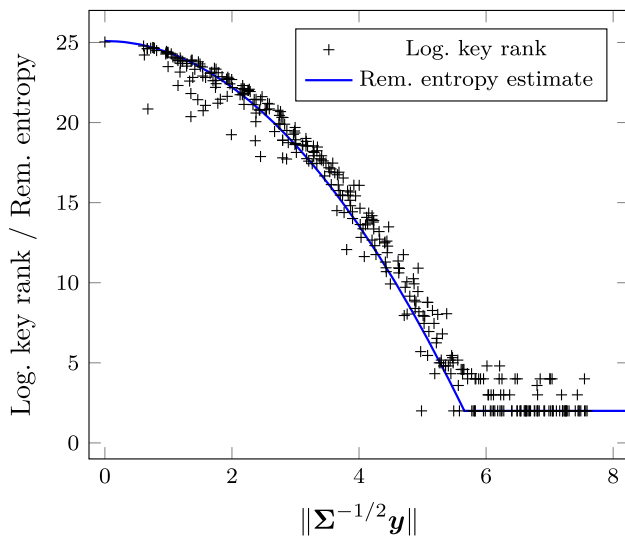
This leads to the following approximation of the remaining entropy  $H(C, D | Y = y)$  for fixed  $y$ :

$$\begin{aligned} H(C, D | Y = y) &= \log_2(\#\{(c, d) \in \{0, 1\}^{56} \\ &\quad \mid W\Delta(c, d) = y\}) \\ &= 56 + \log_2(\Pr(Y = y)) \\ &\approx \max\left\{2, 56 - \frac{1}{2} \log_2(\det(2\pi \Sigma)) \right. \\ &\quad \left. - \frac{1}{2} \log_2(e) \|\Sigma^{-1/2} y\|^2\right\} \\ &\approx \max\left\{2, 25.09 - 0.72 \cdot \|\Sigma^{-1/2} y\|^2\right\}. \end{aligned} \quad (17)$$

In approximation the remaining entropy depends only on  $\|\Sigma^{-1/2} y\|$ . If  $\|\Sigma^{-1/2} y\|$  is small, the remaining entropy  $H(C, D | Y = y)$  is large (“strong keys”). If  $\|\Sigma^{-1/2} y\|$  is large, the remaining entropy  $H(C, D | Y = y)$  is small (“weak keys”).

We expect that the largest remaining entropy should occur near  $y = \mathbf{0}_{15}$ . The largest number of candidates we found experimentally was indeed





**Fig. 3** Graph of the remaining entropy estimate (17) as a function of  $\|\Sigma^{-1/2}\mathbf{y}\|$  together with the logarithmic key ranks computed in the experiments

$$\#\mathcal{C}(\mathbf{0}_{15}) = 34296072,$$

exactly for the observation  $\mathbf{y} = \mathbf{0}_{15}$ . Since  $\log_2(34296072) \approx 25.03$ , this fits well to the heuristic argumentation above. At the other extreme, for the key  $\mathbf{k} = \mathbf{0}_{56}$  we have

$$\mathcal{R}(\mathbf{0}_{56}) = \#\mathcal{C}(-24 \cdot \mathbf{1}_{15}) = 4.$$

In order to test the heuristic (17), we conducted some further experiments. For every  $a \in \{0.6, 0.8, \dots, 7.4\}$ , we generated 10 random observations  $\mathbf{y} = \mathbf{W}\Delta(\mathbf{k})$  such that  $\|\Sigma^{-1/2}\mathbf{y}\| \in [a, a + 0.2]$  by rejection sampling and computed  $\#\mathcal{C}(\mathbf{y})$  using Algorithm 1. Figure 3 shows a plot of the remaining entropy estimate (17) as a function of  $\|\Sigma^{-1/2}\mathbf{y}\|$  together with the logarithmic key ranks computed in the experiments. Note that for unconditionally chosen random keys the observations would be concentrated around the middle region of the graph (cf. Sect. 3.3). The experiments confirm that the approximation (17) is good.

In the end, we can expect that the distribution of  $H(\mathbf{C}, \mathbf{D} | \mathbf{Y} = \mathbf{y})$  is near the continuous distribution of the random variable  $25.1 - \frac{1}{2}\mathbf{y}^\top \Sigma^{-1} \mathbf{y} \log_2(e)$ , where  $\mathbf{y}$  is normally distributed with expectation  $\mathbf{0}_{15}$  and covariance matrix  $\Sigma$ . Note that the probability density function of this continuous distribution is identical to the probability density function of the random variable  $25.1 - \frac{1}{2}\mathbf{u}^\top \mathbf{u} \log_2(e)$ , where  $\mathbf{u}$  is a standard normal random vector. Since  $\mathbf{u}^\top \mathbf{u}$  is  $\chi^2$ -distributed we know that

$$E\left(25.09 - \frac{1}{2}\mathbf{u}^\top \mathbf{u} \log_2(e)\right) = 25.09 - \frac{1}{2}15 \log_2(e) \approx 14.27,$$

$$\text{Var}\left(25.09 - \frac{1}{2}\mathbf{u}^\top \mathbf{u} \log_2(e)\right) = 30 \cdot \left(\frac{1}{2} \log_2(e)\right)^2 \approx 15.61.$$

Note that the expectation fits to the value  $H(\mathbf{C}, \mathbf{D} | \mathbf{Y}) = 56 - H(\mathbf{Y}) \approx 56 - 41.7 \approx 14.3$  as estimated above. In comparison, we obtained a mean value of 15.08 and empirical variance of 12.02 for the logarithmic key ranks in the experiment reported in Sect. 3.3.

### 3.5 Isolated consideration of the C- and D-register

As a natural approach one might try to find the values of the C- and D-register separately. In the publication [7], for instance, the authors discuss a template attack on even smaller parts of the C- and D-registers. Here we want to clarify that this reduces the amount of mutual information significantly. We consider the following general strategy:

1. Define an appropriate evaluation function that depends only on the key part  $\mathbf{c}$  in the C-register. Find a set  $\mathcal{C}$  of likely candidates for the C-register.
2. Define an appropriate evaluation function that depends only on the key part  $\mathbf{d}$  in the D-register. Find a set  $\mathcal{D}$  of likely candidates for the D-register.
3. Check all combinations  $(\mathbf{c}, \mathbf{d}) \in \mathcal{C} \times \mathcal{D}$ .

The work load of this approach is bounded by  $2^{27}$  in step 1 and 2, but in step 3 we have to check all combinations. We can now give an easy to compute indication of how successful such an approach could be. The random variable  $\mathbf{W}_2\Delta(\mathbf{D})$  takes on certain values in a 15-dimensional space and its entropy  $H(\mathbf{W}_2\Delta(\mathbf{D}))$  is clearly bounded by 27. By Lemma 2, we have

$$H(\mathbf{W}_2\Delta(\mathbf{D})) \lesssim \frac{1}{2} \log_2(\det(2\pi e \mathbf{W}_2 \mathbf{W}_2^\top)) \approx 34.05,$$

where we have neglected the second term in (13). Let  $\mathbf{U}_2$  be a uniformly distributed random variable on  $\{\pm 1\}^{56}$ . We assume that in our case the upper bound of Lemma 2 is in fact a good approximation

$$H(\mathbf{W}_2\mathbf{U}_2) \approx 34.05.$$

Now we use the following heuristic. The success of an evaluation function that depends only on the key part  $\mathbf{c}$  of the C-register should be restricted to the mutual information of  $\mathbf{C}$  and the random variable

$$\mathbf{Y}_1 := \mathbf{W}_1\Delta(\mathbf{C}) + \mathbf{W}_2\mathbf{U}_2.$$

Applying Lemma 2 to  $Y_1$ , we get

$$\begin{aligned} I(Y_1; C) &= I(W_1 \Delta(C) + W_2 U_2; C) = H(Y_1) - H(W_2 U_2) \\ &\approx \frac{1}{2} \log_2(\det(2\pi e W W^\top)) \\ &\quad - \frac{1}{2} \log_2(\det(2\pi e W_2 W_2^\top)) \\ &= \frac{1}{2} \log_2(\det(W W^\top)) - \frac{1}{2} \log_2(\det(W_2 W_2^\top)). \end{aligned}$$

Using this approximation, we obtain

$$I(Y_1; C) \approx 7.68 \quad \text{and} \quad I(Y_2; D) \approx 8.68,$$

where  $Y_2 := W_1 U_1 + W_2 \Delta(D)$  is defined analogously to  $Y_1$  with  $U_1$  uniformly distributed on  $\{\pm 1\}^{56}$ .

**Remark 4** These small concrete values do not come as a complete surprise. By construction of  $W$  we know that  $W_1 W_1^\top \approx W_2 W_2^\top$ , so that we can expect roughly

$$\begin{aligned} I(Y_1; C) &\approx \frac{1}{2} \log_2(\det(W W^\top)) - \frac{1}{2} \log_2(\det(W_2 W_2^\top)) \\ &\approx \frac{1}{2} \log_2(\det(2W_2 W_2^\top)) - \frac{1}{2} \log_2(\det(W_2 W_2^\top)) \\ &= \frac{1}{2} \log_2(2^{15}) = \frac{15}{2}. \end{aligned}$$

## 4 Linear regression model

In this section we consider a continuous leakage model, whose observations cover more points of interest but may contain errors. The weight matrix of this model is not derived by theoretical considerations, but must be learned in a profiling phase using linear regression.

**Leakage Model 3** (*Linear regression model*) Let  $m \geq 112$ , let  $W \in \mathbb{R}^{m \times 112}$  be a fixed weight matrix of full column rank, and let  $K = (C, D)$  be a uniformly distributed random variable on  $\{0, 1\}^{56}$ . We define the random variable  $Y$  on  $\mathbb{R}^m$  by

$$Y = W \Delta(C, D) + \epsilon, \quad (18)$$

where  $\epsilon$  is a zero-mean normal random variable on  $\mathbb{R}^m$  with covariance matrix  $\sigma^2 I_m$  which is independent of  $K$ .

**Remark 5** The leakage of a real implementation might not strictly follow this leakage model. For instance, there might be a non-linear key-dependent influence on the leakage and the error might follow a different distribution or be key-dependent as well. One may attempt to fit real measurements better to this model by recentering and decorrelating the measurements. Moreover, the error can be reduced by averaging over several observations for the same key, see Sect. 4.3.1.

**Remark 6** Leakage Model 3 is expressed in the setting of the DES algorithm. However, in principle it should be adaptable to other algorithms on security controllers. As a protection against side-channel attacks, an implementation should use many operations in parallel if key-dependent information is being processed. In our leakage model each observation is a large, weighted sum of certain key-dependent bits that is disturbed by noise.

### 4.1 Key ranking and key enumeration

Let  $y \in \mathbb{R}^m$  be an observation under Leakage Model 3 corresponding to an unknown key  $k^* = (c^*, d^*) \in \{0, 1\}^{56}$ , i.e. we have  $y = W \Delta(c^*, d^*) + \epsilon$  for some unknown noise vector  $\epsilon \in \mathbb{R}^m$ . We also assume for the moment that we know the weight matrix  $W \in \mathbb{R}^{m \times 112}$  (in Sect. 4.2 we describe how  $W$  can be estimated).

We define the evaluation function

$$\eta_{W,y}: \{0, 1\}^{56} \rightarrow \mathbb{R}_{\geq 0}, \quad (c, d) \mapsto \|y - W \Delta(c, d)\|^2. \quad (19)$$

We denote by

$$\mathcal{C}_W(y, B) := \{(c, d) \in \{0, 1\}^{56} \mid \eta_{W,y}(c, d) \leq B\} \quad (20)$$

the set of key candidates for observation  $y$  with error bound  $B \in \mathbb{R}_{\geq 0}$ . The rank of the correct key  $k^*$  with respect to  $W$  and  $y$  is defined as

$$\mathcal{R}_{W,y}(k^*) := \#\mathcal{C}_W(y, \eta_{W,y}(k^*)). \quad (21)$$

Note that  $\mathcal{R}_{W,y}(k^*)$  is a multiple of 4 (cf. Remark 1). We call  $\log_2 \mathcal{R}_{W,y}(k^*)$  the logarithmic key rank of  $k^*$ .

A quick check for assessing the quality of  $\eta_{W,y}$  can be obtained by testing the condition  $\eta_{W,y}(c_j, d_j) \leq \eta_{W,y}(c^*, d^*)$  for several random key candidates  $(c_j, d_j) \in \{0, 1\}^{56}$  with  $j = 1, \dots, N$ . If  $N$  is large and  $\mathcal{R}_{W,y}(k^*)$  is not too small, we can expect

$$\begin{aligned} \mathcal{R}_{W,y}(k^*) &\approx \frac{2^{56}}{N} \cdot \#\{j \in [N] \mid \eta_{W,y}(c_j, d_j) \\ &\leq \eta_{W,y}(c^*, d^*)\}. \end{aligned} \quad (22)$$

Next we develop an algorithm to enumerate the set  $\mathcal{C}_W(y, B)$  based on the Fincke–Pohst lattice point enumeration algorithm [4]. In principle, this algorithm explores the whole key space  $\{0, 1\}^{56}$  (or  $\{0, 1\}^{54}$ ), but in many instances the search tree can be pruned considerably.

The following preparatory lemma shows that the weight matrix  $W \in \mathbb{R}^{m \times 112}$  can be replaced by an upper triangular matrix  $R \in \mathbb{R}^{112 \times 112}$ . At the same time, the observation vector  $y$  is projected onto the range of  $W$ .

**Lemma 3** Let  $m \geq 112$ , let  $W \in \mathbb{R}^{m \times 112}$  be a matrix of full column rank, and let  $y \in \mathbb{R}^m$ . Then there exists a unique upper triangular matrix  $R \in \mathbb{R}^{112 \times 112}$  with positive diagonal elements such that  $W^T W = R^T R$ . We have

$$\|y - Wx\|^2 = \|R(t - x)\|^2 + \|y - Wt\|^2 \quad \text{for all } x \in \mathbb{R}^{112},$$

where  $t := (W^T W)^{-1} W^T y \in \mathbb{R}^{112}$ .

**Proof** Since  $W$  has full column rank, the matrix  $W^T W$  is symmetric positive-definite and, in particular, non-singular. The existence and uniqueness of  $R$  follow from the Cholesky factorization of  $W^T W$  (cf. [6, Theorem 4.2.5]). Since  $W(W^T W)^{-1} W^T$  is the orthogonal projection onto the range of  $W$ , we have  $\langle Wx, y - Wt \rangle = 0$  for all  $x \in \mathbb{R}^{112}$ , hence  $\|y - Wx\|^2 = \|W(t - x) + y - Wt\|^2 = \|W(t - x)\|^2 + \|y - Wt\|^2$  for all  $x \in \mathbb{R}^{112}$ . Finally, we have  $\|W(t - x)\|^2 = (t - x)^T W^T W(t - x) = (t - x)^T R^T R(t - x) = \|R(t - x)\|^2$  for all  $x \in \mathbb{R}^{112}$ .  $\square$

**Remark 7** Consider the situation of Lemma 3. By the thin/reduced QR factorization of  $W$ , there exists a unique matrix  $Q \in \mathbb{R}^{m \times 112}$  with orthonormal columns and a unique upper triangular matrix  $R \in \mathbb{R}^{112 \times 112}$  with positive diagonal elements such that  $W = QR$  (cf. [6, Theorem 5.2.2]). We have  $W^T W = R^T R$  (in particular,  $R$  is the Cholesky factor of  $W^T W$ ) and  $t = (W^T W)^{-1} W^T y = R^{-1} Q^T y$ .

By Lemma 3, we have  $C_W(y, B) = C_R(Rt, B - \|y - Wt\|^2)$ . Let  $(c, d) \in \{0, 1\}^{56}$  and  $x := \Delta(c, d) \in \{\pm 1\}^{112}$ . Then  $(c, d) \in C_W(y, B)$  if and only if

$$\|R(t - x)\|^2 = \sum_{i=1}^{112} \left( \sum_{j=i}^{112} r_{i,j}(t_j - x_j) \right)^2 \leq B - \|y - Wt\|^2. \tag{23}$$

In principle, we could enumerate all vectors  $x \in \{\pm 1\}^{112}$  satisfying (23) using backtracking (cf. [9, Section 7.2.2, Algorithm B]). If the elements of  $x$  are traversed in the order  $x_{112}, x_{111}, \dots, x_1$ , then partial assignments  $x_s \dots x_{112}$  violating the condition

$$\sum_{i=s}^{112} \left( \sum_{j=i}^{112} r_{i,j}(t_j - x_j) \right)^2 \leq B - \|y - Wt\|^2$$

can be rejected immediately (without trying further values for  $x_1 \dots x_{s-1}$ ). For each vector  $x \in \{\pm 1\}^{112}$  satisfying (23) we could then compute the preimage under  $\Delta$  to obtain the corresponding key candidates (cf. Remark 1). However, this approach is impractical due to the size of  $\{\pm 1\}^{112}$ . Therefore, we enumerate the keys  $k = (c, d) \in \{0, 1\}^{56}$  directly. In order to make this approach work, we have to reorder the

components of  $x = \Delta(k) = \Delta(c, d)$  and the columns of  $W$  have to be reordered accordingly (before applying Lemma 3).

Let  $k = k_1 \dots k_{56} = (c, d) = c_1 \dots c_{28} d_1 \dots d_{28} \in \{0, 1\}^{56}$ . First we choose a permutation  $\pi: [56] \rightarrow [56]$  that determines the order  $k_{\pi(1)}, k_{\pi(2)}, \dots, k_{\pi(56)}$  in which we want to traverse the bits of  $k$  in the enumeration procedure. By Remark 1 we can keep one bit of  $c$  and one bit of  $d$  fixed, so we move those bits to the front and do not change them during the enumeration. For example, we may choose  $\pi(1) = 1$  and  $\pi(2) = 29$ , hence  $k_{\pi(1)} = c_1$  and  $k_{\pi(2)} = d_1$ . We proceed by choosing bits as to maximize the number of components in  $\Delta(k)$  that are determined by the current choices. In other words, we pick nodes in the graph of Fig. 1 such that the number of edges between the chosen nodes is maximized. For example, we may choose

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & 28 & 29 & 30 & 31 & \dots & 55 & 56 \\ 1 & 29 & 2 & 3 & \dots & 27 & 28 & 30 & 31 & \dots & 55 & 56 \end{pmatrix}.$$

Next we determine a permutation matrix  $P \in \mathbb{R}^{112 \times 112}$  and integers  $s_1 = s_2 = 113 > s_3 > \dots > s_{56} = 1$  such that for all  $\ell \in [56]$  the components of  $x := P^T \Delta(k) \in \{\pm 1\}^{112}$ , which are determined by  $k_{\pi(1)}, \dots, k_{\pi(\ell)}$ , are the trailing components  $x_{s_\ell}, \dots, x_{112}$  of  $x$ . In our example, we have

$$\begin{aligned} s_3 &= 112, & s_{3+i} &= 112 - 2i & \text{for } i \in [24], \\ s_{28} &= 61, & s_{29} &= 57, \\ s_{30} &= 56, & s_{30+i} &= 56 - 2i & \text{for } i \in [24], \\ s_{55} &= 5, & s_{56} &= 1, \end{aligned}$$

and we may choose  $P$  such that

$$\begin{aligned} x_1, x_2, x_3, x_4 &= (-1)^{d_1 \oplus d_{28}}, (-1)^{d_2 \oplus d_{28}}, (-1)^{d_{26} \oplus d_{28}}, \\ & \quad (-1)^{d_{27} \oplus d_{28}}, \\ x_5, x_6, x_7 &= (-1)^{d_1 \oplus d_{27}}, (-1)^{d_{25} \oplus d_{27}}, (-1)^{d_{26} \oplus d_{27}}, \\ x_{56-2i}, x_{57-2i} &= (-1)^{d_i \oplus d_{i+2}}, (-1)^{d_{i+1} \oplus d_{i+2}} \quad \text{for } i \in [24], \\ x_{56} &= (-1)^{d_1 \oplus d_2}, \\ x_{57}, x_{58}, x_{59}, x_{60} &= (-1)^{c_1 \oplus c_{28}}, (-1)^{c_2 \oplus c_{28}}, (-1)^{c_{26} \oplus c_{28}}, \\ & \quad (-1)^{c_{27} \oplus c_{28}}, \\ x_{61}, x_{62}, x_{63} &= (-1)^{c_1 \oplus c_{27}}, (-1)^{c_{25} \oplus c_{27}}, (-1)^{c_{26} \oplus c_{27}}, \\ x_{112-2i}, x_{113-2i} &= (-1)^{c_i \oplus c_{i+2}}, (-1)^{c_{i+1} \oplus c_{i+2}} \quad \text{for } i \in [24], \\ x_{112} &= (-1)^{c_1 \oplus c_2}. \end{aligned}$$

Applying Lemma 3 to the column-permuted matrix  $WP$  and the observation  $y \in \mathbb{R}^m$ , we obtain an upper triangular

matrix  $\mathbf{R} \in \mathbb{R}^{112 \times 112}$  and  $\mathbf{t} \in \mathbb{R}^{112}$  such that

$$\begin{aligned} \|\mathbf{y} - \mathbf{W}\Delta(\mathbf{k})\|^2 &= \|\mathbf{y} - \mathbf{W}\mathbf{P}\mathbf{P}^\top\Delta(\mathbf{k})\|^2 \\ &= \|\mathbf{R}(\mathbf{t} - \mathbf{x})\|^2 + \|\mathbf{y} - \mathbf{W}\mathbf{P}\mathbf{t}\|^2. \end{aligned}$$

Therefore, we have  $\mathbf{k} \in \mathcal{C}_W(\mathbf{y}, B)$  if and only if

$$\begin{aligned} \rho_\ell &:= \sum_{i=s_\ell}^{112} \left( \sum_{j=i}^{112} r_{i,j}(t_j - x_j) \right)^2 \\ &\leq B - \|\mathbf{y} - \mathbf{W}\mathbf{P}\mathbf{t}\|^2 \text{ for all } \ell \in [56]. \end{aligned}$$

Note that  $\rho_\ell$  depends on the components  $x_{s_\ell}, \dots, x_{112}$  of  $\mathbf{x} = \mathbf{P}^\top\Delta(\mathbf{k})$  which are determined completely by  $k_{\pi(1)}, \dots, k_{\pi(\ell)}$ . Furthermore,  $\rho_\ell$  can be computed recursively, since  $\rho_1 = \rho_2 = 0$  and

$$\rho_\ell = \rho_{\ell-1} + \sum_{i=s_\ell}^{s_{\ell-1}-1} \left( \sum_{j=i}^{112} r_{i,j}(t_j - x_j) \right)^2 \text{ for } 3 \leq \ell \leq 56.$$

Using the backtracking scheme described in [9, Section 7.2.2, Algorithm B], we obtain the following algorithm.

**Algorithm 2**

**Input:** A matrix  $\mathbf{W} \in \mathbb{R}^{m \times 112}$  of full column rank, a vector  $\mathbf{y} \in \mathbb{R}^m$ , and a bound  $B \in \mathbb{R}_{\geq 0}$ .

**Output:** The set of key candidates  $\mathcal{C}_W(\mathbf{y}, B)$ .

1. [Initialize.] Set  $\mathcal{C} \leftarrow \emptyset$ ,  $\ell \leftarrow 3$ ,  $\mathbf{k} \leftarrow \mathbf{0}_{56}$ ,  $\mathbf{x} \leftarrow \mathbf{1}_{112}$ , and  $\rho \leftarrow \mathbf{0}_{56}$ .
2. [Preprocess.] Set  $\mathbf{W} \leftarrow \mathbf{W}\mathbf{P}$ . Compute an upper triangular matrix  $\mathbf{R} \in \mathbb{R}^{112 \times 112}$  such that  $\mathbf{W}^\top\mathbf{W} = \mathbf{R}^\top\mathbf{R}$ . Set  $\mathbf{t} \leftarrow (\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top\mathbf{y}$  and  $B \leftarrow B - \|\mathbf{y} - \mathbf{W}\mathbf{t}\|^2$ . If  $B < 0$ , return  $\mathcal{C}$  and stop.
3. [Enter level  $\ell$ .] If  $\ell = 57$ , set  $(\mathbf{c}, \mathbf{d}) \leftarrow \mathbf{k}$ , set  $\mathcal{C} \leftarrow \mathcal{C} \cup \{(\mathbf{c}, \mathbf{d}), (\bar{\mathbf{c}}, \mathbf{d}), (\mathbf{c}, \bar{\mathbf{d}}), (\bar{\mathbf{c}}, \bar{\mathbf{d}})\}$ , and go to step 6. Otherwise set  $k_{\pi(\ell)} \leftarrow 0$ .
4. [Try  $k_{\pi(\ell)}$ .] Set  $x_i \leftarrow (\mathbf{P}^\top\Delta(\mathbf{k}))_i$  for  $i = s_\ell, \dots, s_{\ell-1} - 1$  and set

$$\rho_\ell \leftarrow \rho_{\ell-1} + \sum_{i=s_\ell}^{s_{\ell-1}-1} \left( \sum_{j=i}^{112} r_{i,j}(t_j - x_j) \right)^2.$$

If  $\rho_\ell \leq B$ , set  $\ell \leftarrow \ell + 1$  and go to step 3.

5. [Try again.] If  $k_{\pi(\ell)} = 0$ , set  $k_{\pi(\ell)} \leftarrow 1$  and go to step 4.
6. [Backtrack.] Set  $\ell \leftarrow \ell - 1$ . If  $\ell \geq 3$ , go to step 5. Otherwise return  $\mathcal{C}$  and stop.

**Remark 8** We note some possible variations and optimizations of Algorithm 2.

- (a) To avoid repeated computations, Algorithm 2 can be modified as follows (cf. [5, Appendix B] and [10, Appendix A]). Let  $\sigma_{i,h} := \sum_{j=h}^{112} r_{i,j}(t_j - x_j)$  for  $i \in [112]$  and  $h \in [113]$ . The value  $\sigma_{i,h}$  can be computed recursively, since  $\sigma_{i,113} = 0$  and  $\sigma_{i,h} = \sigma_{i,h+1} + r_{i,h}(t_h - x_h)$  for all  $i, h \in [112]$ . Further, we have  $\rho_\ell = \rho_{\ell-1} + \sum_{i=s_\ell}^{s_{\ell-1}-1} \sigma_{i,i}^2$  for  $3 \leq \ell \leq 56$ . By using these recurrence relations and by reusing values of  $\sigma_{i,h}$  that are still valid during the enumeration, the partial squared norms  $\rho_\ell$  can be computed with fewer operations. For details, see Algorithm 3 in the appendix.
- (b) Using pruning [12,13], we can heuristically reject partial assignments  $k_{\pi(1)} \cdots k_{\pi(\ell)}$  during the enumeration if the partial squared norm  $\rho_\ell$  is already so large that  $\rho_{56} \leq B$  becomes unlikely for any choice of  $k_{\pi(\ell+1)} \cdots k_{\pi(56)}$ . This can be done by replacing the if-condition “ $\rho_\ell \leq B$ ” in step 4 by “ $\rho_\ell \leq B_\ell$ ” for suitable bounds  $0 \leq B_3 \leq \dots \leq B_{56} = B$ . In our experiments we used the bounds

$$B_\ell := \begin{cases} \frac{\ell+17}{54}B, & \text{if } 3 \leq \ell \leq 36, \\ B, & \text{if } 37 \leq \ell \leq 56. \end{cases}$$

Note that we cannot use extreme pruning [5], since we want to find all (or almost all) key candidates in  $\mathcal{C}_W(\mathbf{y}, B)$ .

- (c) For a given weight matrix  $\mathbf{W}$ , the running time of Algorithm 2 may be optimized by choosing different permutations  $\pi$  and  $\mathbf{P}$  (in compliance with the conditions outlined above). Preprocessing  $\mathbf{W}$  with general unimodular transformations (e.g. lattice basis reduction, cf. [4, (2.12)]) seems not possible in our setting.
- (d) To find  $N$  best key candidates for the evaluation function  $\eta_{W,\mathbf{y}}$ , Algorithm 2 can be modified as follows. We start the algorithm with  $B := \infty$ . The set  $\mathcal{C}$  is replaced by a list that is ordered by  $\eta_{W,\mathbf{y}}$  and keeps only the  $N$  best key candidates. Each time a key candidate gets evicted from  $\mathcal{C}$ , the bound  $B$  can be updated according to the currently worst key candidate in  $\mathcal{C}$ .

**4.2 Estimation of the weight matrix**

The weight matrix  $\mathbf{W} \in \mathbb{R}^{m \times 112}$  of Leakage Model 3 can be estimated in a profiling phase using linear regression if observations for several known keys are available.

Let  $N_{\text{prf}} \gg 112$ . Assume we are given observations  $\mathbf{y}_{\text{prf},j} \in \mathbb{R}^m$  of Leakage Model 3 for known and randomly chosen keys  $\mathbf{k}_{\text{prf},j} = (\mathbf{c}_{\text{prf},j}, \mathbf{d}_{\text{prf},j}) \in \{0, 1\}^{56}$  for  $j \in [N_{\text{prf}}]$ . We denote by  $\mathbf{X}_{\text{prf}} \in \{\pm 1\}^{112 \times N_{\text{prf}}}$  the matrix with columns  $\mathbf{x}_{\text{prf},j} := \Delta(\mathbf{c}_{\text{prf},j}, \mathbf{d}_{\text{prf},j})$  and by  $\mathbf{Y}_{\text{prf}} \in \mathbb{R}^{m \times N_{\text{prf}}}$  the matrix with columns  $\mathbf{y}_{\text{prf},j}$  for  $j \in [N_{\text{prf}}]$ .

We want to find an approximation  $\tilde{\mathbf{W}} \in \mathbb{R}^{m \times 112}$  of  $\mathbf{W}$  such that  $\mathbf{Y}_{\text{prf}} \approx \tilde{\mathbf{W}}\mathbf{X}_{\text{prf}}$ . Since the error vector in Leakage

Model 3 has independent components, we may estimate the rows of  $\mathbf{W}$  independently. Let  $i \in [m]$  and let  $\mathbf{y}_{\text{prf},i} \in \mathbb{R}^{1 \times N_{\text{prf}}}$  denote the  $i$ -th row of  $\mathbf{Y}_{\text{prf}}$ . We approximate the  $i$ -th row of  $\mathbf{W}$  by a least squares estimate, i.e. by a vector  $\mathbf{w} \in \mathbb{R}^{1 \times 112}$  minimizing

$$\|\mathbf{y}_{\text{prf},i} - \mathbf{w}\mathbf{X}_{\text{prf}}\|^2. \tag{24}$$

Since  $N_{\text{prf}} \gg 112$ , we may assume that  $\mathbf{X}_{\text{prf}}$  has full row rank and  $\mathbf{X}_{\text{prf}}\mathbf{X}_{\text{prf}}^\top \in \mathbb{R}^{112 \times 112}$  is non-singular. This implies that (24) is minimized by the (unique) vector  $\tilde{\mathbf{w}}_i := \mathbf{y}_{\text{prf},i}\mathbf{X}_{\text{prf}}^\top(\mathbf{X}_{\text{prf}}\mathbf{X}_{\text{prf}}^\top)^{-1} \in \mathbb{R}^{1 \times 112}$  (cf. [6, Section 5.3.1]). Combining the estimated rows of the weight matrix, we obtain the matrix

$$\tilde{\mathbf{W}} := \mathbf{Y}_{\text{prf}}\mathbf{X}_{\text{prf}}^\top(\mathbf{X}_{\text{prf}}\mathbf{X}_{\text{prf}}^\top)^{-1} \in \mathbb{R}^{m \times 112}. \tag{25}$$

Since the “true”, unknown weight matrix  $\mathbf{W}$  is assumed to have full column rank, we may hope that the same holds for the estimated weight matrix  $\tilde{\mathbf{W}}$ . We just mention that this was indeed the case in our experiments with real measurements reported in Sect. 4.3.1.

### 4.3 Experiments

We performed experiments using real and simulated measurements.

#### 4.3.1 Real measurements

The authors of [7] provided us with their measurement data. The provided data set consists of a profiling set and an attack set. The measurements are already aligned and trimmed to  $m = 460$  points of interest.

The profiling set comprises  $N_{\text{prf}} = 882547$  measurements of DES operations with random keys  $\mathbf{k}_{\text{prf},j} = (\mathbf{c}_{\text{prf},j}, \mathbf{d}_{\text{prf},j}) \in \{0, 1\}^{56}$  for  $j \in [N_{\text{prf}}]$ . We denote by  $\mathbf{Y}_{\text{prf}} \in \mathbb{R}^{m \times N_{\text{prf}}}$  the matrix of measurements (arranged in columns) and by  $\mathbf{X}_{\text{prf}} \in \{\pm 1\}^{112 \times N_{\text{prf}}}$  the matrix with columns  $\Delta(\mathbf{c}_{\text{prf},j}, \mathbf{d}_{\text{prf},j})$ .

The attack set comprises  $N_{\text{att}} = 247088$  measurements of DES operations with random keys  $\mathbf{k}_{\text{att},j} \in \{0, 1\}^{56}$  for  $j \in [288]$ , where  $N_{\text{att},j} \in \{761, \dots, 927\}$  measurements have been performed with key  $\mathbf{k}_{\text{att},j}$ . (The authors of [7] also carried out measurements for so-called *weak keys*, but we do not consider them in this article.) We denote by  $\mathbf{Y}_{\text{att},j} \in \mathbb{R}^{m \times N_{\text{att},j}}$  the matrix of measurements with key  $\mathbf{k}_{\text{att},j}$  (arranged in columns) for  $j \in [288]$ .

In order to fit the measurement data to Leakage Model 3, we preprocessed the data sets as follows. Using the mean

$$\tilde{\boldsymbol{\mu}}_{\text{prf}} := N_{\text{prf}}^{-1}\mathbf{Y}_{\text{prf}}\mathbf{1}_{N_{\text{prf}}} \in \mathbb{R}^m$$

of the measurements  $\mathbf{Y}_{\text{prf}}$  of the profiling set, we centered  $\mathbf{Y}_{\text{prf}}$  and  $\mathbf{Y}_{\text{att},j}$  by replacing

$$\begin{aligned} \mathbf{Y}_{\text{prf}} &\leftarrow \mathbf{Y}_{\text{prf}} - \tilde{\boldsymbol{\mu}}_{\text{prf}}\mathbf{1}_{N_{\text{prf}}}^\top \quad \text{and} \\ \mathbf{Y}_{\text{att},j} &\leftarrow \mathbf{Y}_{\text{att},j} - \tilde{\boldsymbol{\mu}}_{\text{prf}}\mathbf{1}_{N_{\text{att},j}}^\top \quad \text{for } j \in [288]. \end{aligned}$$

Using the empirical covariance matrix

$$\tilde{\boldsymbol{\Sigma}}_{\text{prf}} := N_{\text{prf}}^{-1}\mathbf{Y}_{\text{prf}}\mathbf{Y}_{\text{prf}}^\top \in \mathbb{R}^{m \times m}$$

of the (centered) measurements  $\mathbf{Y}_{\text{prf}}$  of the profiling set, we decorrelated  $\mathbf{Y}_{\text{prf}}$  and  $\mathbf{Y}_{\text{att},j}$  (Mahalanobis whitening) by replacing

$$\begin{aligned} \mathbf{Y}_{\text{prf}} &\leftarrow \tilde{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Y}_{\text{prf}} \quad \text{and} \\ \mathbf{Y}_{\text{att},j} &\leftarrow \tilde{\boldsymbol{\Sigma}}^{-1/2}\mathbf{Y}_{\text{att},j} \quad \text{for } j \in [288]. \end{aligned}$$

Finally, we computed the averaged measurements

$$\bar{\mathbf{y}}_j := N_{\text{att},j}^{-1}\mathbf{Y}_{\text{att},j}\mathbf{1}_{N_{\text{att},j}} \quad \text{for } j \in [288].$$

Due to a slight shift in the averaged measurements, we also recentered them amongst each other by replacing

$$(\bar{\mathbf{y}}_1, \dots, \bar{\mathbf{y}}_{288}) \leftarrow (\bar{\mathbf{y}}_1, \dots, \bar{\mathbf{y}}_{288})(\mathbf{I}_{288} - 288^{-1}\mathbf{1}_{288}\mathbf{1}_{288}^\top).$$

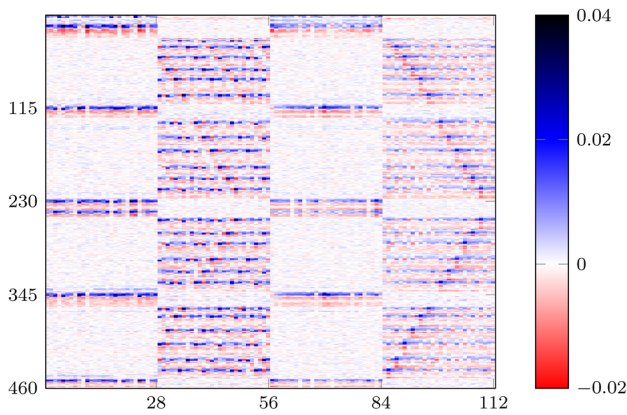
Using the preprocessed data of the profiling set, we computed an estimate  $\tilde{\mathbf{W}} \in \mathbb{R}^{m \times 112}$  of the weight matrix according to Leakage Model 3 as in (25). A matrix plot of  $\tilde{\mathbf{W}}$  is shown in Fig. 4. The plot illustrates the locations where updates of the C- and D-register (rotation by 1 resp. 2 positions) take place. In particular, it is visible that the measurement covers a DES-encryption followed by a full DES-decryption, presumably as a countermeasure against fault attacks. The upper half and the mirrored lower half of the plot bears some resemblance with the weight matrix of Leakage Model 2 (cf. Fig. 2). This visual structure of  $\tilde{\mathbf{W}}$  is a first indication that Leakage Model 3 is adequate for the measurements.

We implemented Algorithm 2 in the Julia programming language [2] with the optimizations of Remark 8(a) and (b). We computed the key ranks

$$\mathcal{R}_i := \mathcal{R}_{\tilde{\mathbf{W}}, \bar{\mathbf{y}}_i}(\mathbf{k}_{\text{att},i})$$

for 287 of the 288 averaged measurements  $\bar{\mathbf{y}}_i$  on a standard computer by explicit enumeration. The distribution of the computed ranks  $\mathcal{R}_i$  and the single-core running times is described in Table 2.

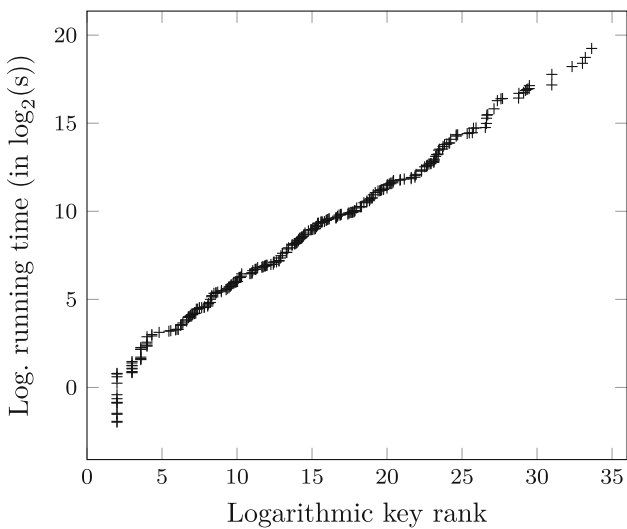
One half of the computed ranks are below  $2^{15}$  and 75% of them are below  $2^{21}$ . The key enumerations finished in under 7 minutes in one half of the cases using a single CPU-core.



**Fig. 4** Matrix plot of the estimated weight matrix  $\tilde{W}$  with element values depicted according to the color bar

**Table 2** Empirical distribution of the logarithmic key ranks and key enumeration running time for 287/288 averaged, real measurements

Logarithmic key rank					Running time		
Min	Q1	Median	Q3	Max	Min	Median	Max
2	9	15	21	34	1 s	7 min	8 d



**Fig. 5** Log-log plot of the key enumeration running times and key ranks for the real measurements

A log-log plot of the running times and key ranks is shown in Fig. 5. With such low key ranks we note that the classic meet-in-the-middle approach against 3-key Triple DES has very moderate running time. For average keys we can expect a running time of roughly  $2^{30}$  DES encryptions/decryptions.

The experiments demonstrate that Leakage Model 3 is adequate for the measurement data. Although the leakage model might only approximate the real leakage, the attack is successful. We note that, apart from model errors, there may be further obstacles to a successful attack. If the number of

**Table 3** Empirical distribution of the logarithmic key ranks and key enumeration running times for 100 simulated measurements per value of  $\sigma$

Noise $\sigma$	Logarithmic key rank					Running time		
	Min	Q1	Median	Q3	Max	Min	Median	Max
0.02	2	2	2	2	2	1 s	1 s	1 s
0.03	2	2	2	2	3	1 s	1 s	1 s
0.04	2	2	2	2	6	1 s	1 s	6 s
0.05	2	2	3	7	14	1 s	9 s	30 min
0.06	2	5	8	12	30	2 s	2 min	14 h
0.07	2	9	15	19	33	5 s	29 min	13 d

The key ranks were computed explicitly using Algorithm 2 with the optimizations of Remark 8(a) and (b)

measurements in the profiling phase is insufficient, the estimated weight matrix may differ significantly from the “true” weight matrix. If the number of measurements in the attack phase is insufficient, the noise of the averaged measurements may be too large.

### 4.3.2 Simulated measurements

In order to investigate the influence of the error distribution on the key rank, we performed a series of experiments with simulated measurements in different noise regimes.

For the simulated measurements, we used the weight matrix  $\tilde{W} \in \mathbb{R}^{m \times 112}$  with  $m = 460$  estimated from the real measurements as described in Sect. 4.3.1 (cf. Fig. 4) and generated the observations as samples from Leakage Model 3, where the keys were drawn uniformly from  $\{0, 1\}^{56}$  and the errors were drawn from a centered normal distribution on  $\mathbb{R}^m$  with covariance matrix  $\sigma^2 \mathbf{I}_m$ . In contrast to Sect. 4.3.1, we used the observations directly without averaging over several observations.

For each  $\sigma \in \{0.02, 0.03, \dots, 0.07\}$ , we generated 100 observations and computed the corresponding key ranks explicitly using our implementation of Algorithm 2 with the optimizations of Remark 8(a) and (b). The distributions of the computed ranks and the single-core running times are described in Table 3.

Comparing the distributions of Tables 2 and 3, we recognize that the averaged observations in Sect. 4.3.1 behave similarly to observations of Leakage Model 3 with  $\sigma \approx 0.07$ .

For larger values of  $\sigma$ , we have resorted to the Monte-Carlo heuristic (22). For each  $\sigma \in \{0.2, 0.3, \dots, 0.7\}$ , we generated 100 observations and estimated the corresponding key ranks using the Monte-Carlo heuristic (22) with an appropriate number  $N$  of random keys. The distribution of the estimated ranks is described in Table 4.

**Table 4** Empirical distribution of the estimated logarithmic key ranks for 100 simulated measurements per value of  $\sigma$

Noise $\sigma$	Logarithmic key rank					# Random keys $N$
	Min	Q1	Median	Q3	Max	
0.2	32	40	43	46	54	$2^{24}$
0.3	39	44	48	50	55	$2^{22}$
0.4	39	48	50	53	55	$2^{20}$
0.5	42	50	52	53	56	$2^{20}$

The key ranks were estimated using the Monte-Carlo heuristic (22) with  $N$  random keys

### 4.4 Theoretical estimation of the remaining entropy

Similar to the discrete case we use mutual information as a measure for the uncertainty about the key if an observation is given. However, mutual information of continuous random variables should be treated with care, since some of its properties are different compared to the discrete case (cf. [3]). The following lemma provides an upper bound for the mutual information of  $Y$  and  $(C, D)$  in Leakage Model 3.

**Lemma 4** Let  $K = (C, D)$  and  $Y = W\Delta(C, D) + \epsilon$  with  $\text{Cov}(\epsilon) = \sigma^2 I_m$  as in Leakage Model 3. Then

$$I(Y; C, D) \leq \frac{1}{2} \log_2(\det(\sigma^{-2}W^T W + I_{112})).$$

*Proof* We have  $I(Y; C, D) = H(Y) - H(\epsilon)$ . Let  $\Sigma = \text{Cov}(Y)$ . By Lemma 1(d), we have  $\Sigma = W W^T + \sigma^2 I_m$ . By [3, Theorem 8.6.5], we obtain  $H(Y) \leq \frac{1}{2} \log_2(\det(2\pi e \Sigma))$  and  $H(\epsilon) = \frac{1}{2} \log_2(\det(2\pi e \sigma^2 I_m))$ . This implies

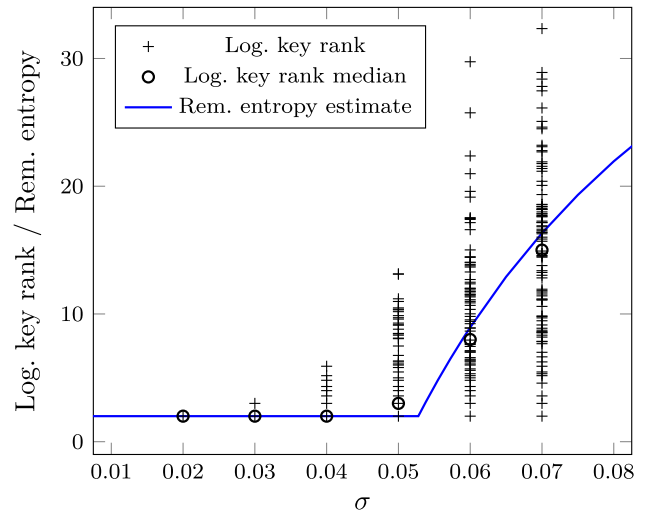
$$\begin{aligned} I(Y; C, D) &\leq \frac{1}{2} \log_2(\det(2\pi e \Sigma) / \det(2\pi e \sigma^2 I_m)) \\ &= \frac{1}{2} \log_2(\det(\sigma^{-2}W W^T + I_m)). \end{aligned}$$

Since  $W W^T$  and  $W^T W$  have identical non-zero eigenvalues, we get

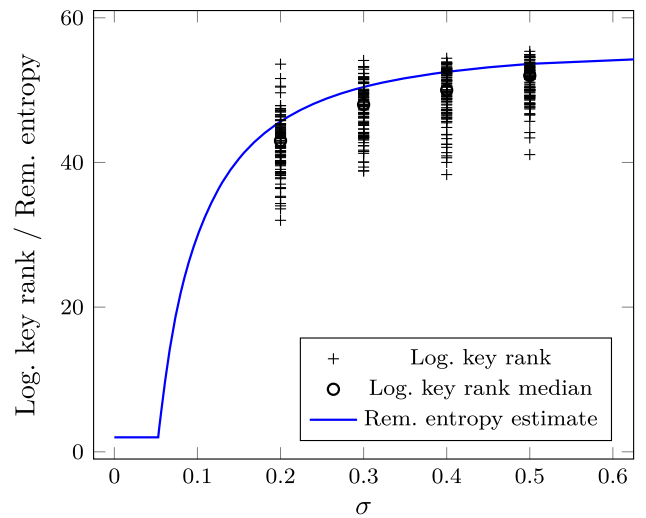
$$\det(\sigma^{-2}W W^T + I_m) = \det(\sigma^{-2}W^T W + I_{112})$$

and the assertion follows.  $\square$

**Remark 9** The validity of the upper bound in Lemma 4 only depends on the leakage model and the properties of the random vector  $\Delta(C, D)$  in Lemma 1. Although  $\Delta(C, D)$  is specific to the DES setting, a similar upper bound could be derived for other cryptographic implementations if a leakage model with the respective properties is applicable.



**Fig. 6** Graph of the remaining entropy estimate (26) as a function of  $\sigma$  together with the logarithmic key ranks computed in the experiments reported in Table 3



**Fig. 7** Graph of the remaining entropy estimate (26) as a function of  $\sigma$  together with the logarithmic key ranks estimated in the experiments reported in Table 4

Based on the experiments in Sect. 4.3.2, we propose the heuristic formula

$$\begin{aligned} H(C, D | Y) &\approx \max \left\{ 2, 56 - \frac{1}{2} \log_2(\det(\sigma^{-2}W^T W + I_{112})) \right\} \quad (26) \end{aligned}$$

for the remaining entropy. Figures 6 and 7 compare this heuristic with the results of the experiments with simulated measurements reported in Tables 3 and 4, respectively.

### 4.5 Isolated consideration of the C- and D-register

In Sect. 3.5 we looked at an approach that considers the C- and D-register separately. We argued that the mutual information is much lower in this setting. However, in the continuous case with error the situation is different. On the one hand, we have an additional error so that each observation gives less information compared to Leakage Model 2. On the other hand, we have much more POIs in Leakage Model 3. The general strategy is again as follows:

1. Define an appropriate evaluation function that depends only on the key part  $\mathbf{c}$  in the C-register. Find a set  $\mathcal{C}$  of likely candidates for the C-register.
2. Define an appropriate evaluation function that depends only on the key part  $\mathbf{d}$  in the D-register. Find a set  $\mathcal{D}$  of likely candidates for the D-register.
3. Check all combinations  $(\mathbf{c}, \mathbf{d}) \in \mathcal{C} \times \mathcal{D}$ .

The work load of this approach is again bounded by  $2^{27}$  in step 1 and 2, but in step 3 we have to check all combinations. Here we consider the following heuristic. We replace the random variable  $\Delta(\mathbf{D})$  by a normal distributed random variable  $N_2$  with mean  $\mathbf{0}_m$  and covariance matrix  $\mathbf{I}_m$  in  $\mathbf{Y}$ , i.e. we set

$$\mathbf{Y}_1 := \mathbf{W}_1 \Delta(\mathbf{C}) + \mathbf{W}_2 N_2 + \boldsymbol{\varepsilon}.$$

As an indication of the success of an evaluation function that depends only on the key part  $\mathbf{c}$  of the C-register, we compute the mutual information of  $\mathbf{C}$  and the observation  $\mathbf{Y}_1$ . First we normalize the new resulting error by setting

$$\mathbf{Y}'_1 := \boldsymbol{\Sigma}_2^{-1/2} \mathbf{Y}_1,$$

where  $\boldsymbol{\Sigma}_2 := \text{Cov}(\mathbf{W}_2 N_2 + \boldsymbol{\varepsilon}) = \mathbf{W}_2 \mathbf{W}_2^\top + \sigma^2 \mathbf{I}_m$ . Analogously to Sect. 4.1, we define the evaluation function for the key part  $\mathbf{c}$  in the C-register as

$$\begin{aligned} \eta_{\mathbf{W}, \mathbf{y}, 1} : \{0, 1\}^{28} &\rightarrow \mathbb{R}_{\geq 0}, \\ \mathbf{c} &\mapsto \|\boldsymbol{\Sigma}_2^{-1/2} \mathbf{y} - \boldsymbol{\Sigma}_2^{-1/2} \mathbf{W}_1 \Delta(\mathbf{c})\|^2. \end{aligned} \tag{27}$$

We assume that (26) can be applied analogously and get

$$\begin{aligned} H(\mathbf{C} | \mathbf{Y}'_1) &\approx \max \left\{ 1, 28 - \frac{1}{2} \log_2(\det(\boldsymbol{\Sigma}_2^{-1/2} \mathbf{W}_1 \mathbf{W}_1^\top \boldsymbol{\Sigma}_2^{-1/2} + \mathbf{I}_m)) \right\} \\ &= \max \left\{ 1, 28 - \frac{1}{2} \log_2(\det(\sigma^{-2} \mathbf{W} \mathbf{W}^\top + \mathbf{I}_m) / \det(\sigma^{-2} \mathbf{W}_2 \mathbf{W}_2^\top + \mathbf{I}_m)) \right\}. \end{aligned} \tag{28}$$

We expect that the work load of step 3 is roughly of size  $2^{H(\mathbf{C}|\mathbf{Y}'_1)+H(\mathbf{D}|\mathbf{Y}'_2)}$  in the algorithm above, where  $\mathbf{Y}'_2$  is defined analogously for the D-register.

**Table 5** Empirical distribution of the logarithmic key ranks of the key parts in the C- and D-register for 288 averaged, real measurements (cf. Sect. 4.3.1)

Register	Logarithmic key rank				
	Min	Q1	Median	Q3	Max
C	1	2	6	10	20
D	1	15	18	21	27

Applying heuristic (28) to the weight matrix  $\tilde{\mathbf{W}}$  estimated in Sect. 4.3.1 for the real measurements and  $\sigma = 0.07$  as estimated in Sect. 4.3.2, we obtain  $H(\mathbf{C} | \mathbf{Y}'_1) \approx 6.55$  and  $H(\mathbf{D} | \mathbf{Y}'_2) \approx 16.64$ . We computed the ranks of the key parts in the C- and D-register with respect to  $\eta_{\tilde{\mathbf{W}}, \bar{\mathbf{y}}_i, 1}$  and the analogously defined evaluation function  $\eta_{\tilde{\mathbf{W}}, \bar{\mathbf{y}}_i, 2}$ , respectively. The distribution of the computed ranks is described in Table 5. The average logarithmic key ranks were 6.55 and 17.26 for the C- and D-register in good agreement with heuristic (28).

**Remark 10** The key ranks in the experiments vary a lot. Therefore, in practice, the sets  $\mathcal{C}$  and  $\mathcal{D}$  of likely candidates have to be chosen larger than  $2^{H(\mathbf{C}|\mathbf{Y}'_1)}$  and  $2^{H(\mathbf{D}|\mathbf{Y}'_2)}$  in step 1 and 2, respectively, or one has to accept that the algorithm finds the correct combined key only with a certain probability. Algorithm 2 in Sect. 4.1 does not have this drawback.

**Acknowledgements** We thank the authors of [7] for providing us with their measurement data.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### Appendix: An optimized version of Algorithm 2

The following algorithm is a variation of Algorithm 2 with the optimization described in Remark 8(a).

**Algorithm 3**

**Input:** A matrix  $\mathbf{W} \in \mathbb{R}^{m \times 112}$  of full column rank, a vector  $\mathbf{y} \in \mathbb{R}^m$ , and a bound  $B \in \mathbb{R}_{\geq 0}$ .

**Output:** The set of key candidates  $\mathcal{C}_{\mathbf{W}}(\mathbf{y}, B)$ .



1. [Initialize.] Set  $\mathcal{C} \leftarrow \emptyset$ ,  $\ell \leftarrow 3$ ,  $\mathbf{k} \leftarrow \mathbf{0}_{56}$ ,  $\mathbf{x} \leftarrow \mathbf{1}_{112}$ ,  $\boldsymbol{\rho} \leftarrow \mathbf{0}_{56}$ ,  $\boldsymbol{\sigma} \leftarrow \mathbf{0}_{112,113}$ , and  $\mathbf{v} \leftarrow 112 \cdot \mathbf{1}_{56}$ .
2. [Preprocess.] Set  $\mathbf{W} \leftarrow \mathbf{W}\mathbf{P}$ . Compute an upper triangular matrix  $\mathbf{R} \in \mathbb{R}^{112 \times 112}$  such that  $\mathbf{W}^T \mathbf{W} = \mathbf{R}^T \mathbf{R}$ . Set  $\mathbf{t} \leftarrow (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}$  and  $B \leftarrow B - \|\mathbf{y} - \mathbf{W}\mathbf{t}\|^2$ . If  $B < 0$ , return  $\mathcal{C}$  and stop.
3. [Enter level  $\ell$ .] If  $\ell = 57$ , set  $(\mathbf{c}, \mathbf{d}) \leftarrow \mathbf{k}$ , set  $\mathcal{C} \leftarrow \mathcal{C} \cup \{(\mathbf{c}, \mathbf{d}), (\bar{\mathbf{c}}, \mathbf{d}), (\mathbf{c}, \bar{\mathbf{d}}), (\bar{\mathbf{c}}, \bar{\mathbf{d}})\}$ , and go to step 6. Otherwise set  $k_{\pi(\ell)} \leftarrow 0$  and  $v_\ell \leftarrow \max\{v_{\ell-1}, v_\ell\}$ .
4. [Try  $k_{\pi(\ell)}$ .] Set  $\rho_\ell \leftarrow \rho_{\ell-1}$ . For  $i \leftarrow s_{\ell-1} - 1, s_{\ell-1} - 2, \dots, s_\ell$ , do the following:
  - a. Set  $x_i \leftarrow (\mathbf{P}^T \Delta(\mathbf{k}))_i$ .
  - b. For  $j \leftarrow v_\ell, v_\ell - 1, \dots, i$ , set  $\sigma_{i,j} \leftarrow \sigma_{i,j+1} + r_{i,j}(t_j - x_j)$ .
  - c. Set  $\rho_\ell \leftarrow \rho_\ell + \sigma_{i,i}^2$ .
 If  $\rho_\ell \leq B$ , set  $\ell \leftarrow \ell + 1$  and go to step 3.
5. [Try again.] If  $k_{\pi(\ell)} = 0$ , set  $k_{\pi(\ell)} \leftarrow 1$  and go to step 4.
6. [Backtrack.] Set  $\ell \leftarrow \ell - 1$ . If  $\ell \geq 3$ , set  $v_\ell \leftarrow s_{\ell-1} - 1$  and go to step 5. Otherwise return  $\mathcal{C}$  and stop.

## References

1. Banaszczyk, W.: New bounds in some transference theorems in the geometry of numbers. *Math. Ann.* **296**(4), 625–636 (1993)
2. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**(1), 65–98 (2017)
3. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, 2nd edn. Wiley, London (2006)
4. Fincke, U., Pohst, M.: Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Math. Comput.* **44**(170), 463–471 (1985)
5. Gama, N., Nguyen, P.Q., Regev, O.: Lattice enumeration using extreme pruning. In: Gilbert, H. (ed.) *Advances in Cryptology—EUROCRYPT 2010*, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Monaco/French Riviera, May 30–June 3, 2010. Proceedings, vol. 6110. *Lecture Notes in Computer Science*, pp. 257–278. Springer, Berlin (2010)
6. Golub, G.H., Van Loan, C.F.: *Matrix Computations*, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
7. Heyszl, J., Miller, K., Unterstein, F., Schink, M., Wagner, A., Gieser, H.A., Freud, S., Damm, T., Klein, D., Kügler, D.: Investigating profiled side-channel attacks against the DES key schedule. *IACR Trans. Cryptogr. Hardw. Embed. Syst.* **2020**(3), 22–72 (2020)
8. Hu, Y., Zhang, C., Zheng, Y., Wagner, M.: Ciphertext and plaintext leakage reveals the entire TDES key. *IACR Cryptol. ePrint Arch.* **2016**, 1143 (2016)
9. Knuth, D.E.: *The Art of Computer Programming. Fascicle 5: Mathematical Preliminaries Redux; Introduction to Backtracking; Dancing Links*, vol. 4. Addison-Wesley Professional, Reading (2019)
10. Liu, M., Nguyen, P.Q.: Solving BDD by enumeration: an update. In: Dawson, E. (ed.) *Topics in Cryptology—CT-RSA 2013—The Cryptographers’ Track at the RSA Conference 2013*, San Francisco, CA, USA, February 25–March 1, 2013. Proceedings, vol. 7779. *Lecture Notes in Computer Science*, pp. 293–309. Springer, Berlin (2013)
11. National Institute of Standards and Technology. FIPS Publication 46-3: Data encryption standard (DES) (1999)
12. Schnorr, C.-P., Euchner, M.: Lattice basis reduction: improved practical algorithms and solving subset sum problems. *Math. Program.* **66**, 181–199 (1994)
13. Schnorr, C.-P., Hörner, H.H.: Attacking the Chor-Rivest cryptosystem by improved lattice reduction. In: Guillou, L.C., Quisquater, J.-J. (eds) *Advances in Cryptology—EUROCRYPT’95*, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21–25, 1995. Proceeding, vol. 921. *Lecture Notes in Computer Science*, pp. 1–12. Springer, Berlin (1995)
14. Wagner, M., Heyse, S.: Brute-force search strategies for single-trace and few-traces template attacks on the DES round keys of a recent smart card. *IACR Cryptol. ePrint Arch.* **2017**, 614 (2017)
15. Wagner, M., Heyse, S.: Single-trace template attack on the DES round keys of a recent smart card. *IACR Cryptol. ePrint Arch.* **2017**, 57 (2017)
16. Wagner, M., Heyse, S.: Improved brute-force search strategies for single-trace and few-traces template attacks on the DES round keys. *IACR Cryptol. ePrint Arch.* **2018**, 937 (2018)
17. Wagner, M., Yongbo, H., Zhang, C., Zheng, Y.: Comparative study of various approximations to the covariance matrix in template attacks. *IACR Cryptol. ePrint Arch.* **2016**, 1155 (2016)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.