Improving Recent Side-Channel Attacks Against the DES Key Schedule

March 1, 2021

Andreas Wiemers and Johannes Mittmann

Bundesamt für Sicherheit in der Informationstechnik (BSI), Germany
firstname.lastname@bsi.bund.de

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 controler. 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. [HZZW16, WHZZ16, WH17b, WH17a, WH18] attempt side-channel attacks against the key schedule of the Data Encryption Standard (DES), which are further investigated in [HMU+20]. 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 [WH17a], 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 [FP85].

We apply our attack to side-channel measurements provided by the authors of [HMU+20]. 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.

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.

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 $\boldsymbol{I}_n \in \mathbb{R}^{n \times n}$ the identity matrix, and by $\mathbf{0}_{m,n} \in \mathbb{R}^{m \times n}$ the zero matrix. The Euclidean norm of a vector $\boldsymbol{v} \in \mathbb{R}^n$ is denoted by $\|\boldsymbol{v}\|$.

Let $\boldsymbol{a} = a_1 \cdots a_n \in \{0,1\}^n$ be a bit-string. Depending on the context, we identify $\boldsymbol{a}$ 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 $\boldsymbol{a}, \boldsymbol{b} \in \{0,1\}^n$ is denoted by $\boldsymbol{a} \oplus \boldsymbol{b}$. The bit-wise complement of $\boldsymbol{a} \in \{0,1\}^n$ is denoted by $\overline{\boldsymbol{a}} := \boldsymbol{a} \oplus \mathbf{1}_n$, the cyclic left-shift (rotation) of $\boldsymbol{a}$ by $k \in \mathbb{Z}$ positions is denoted by $\boldsymbol{a} \lll k := a_{1+k} \cdots a_{n+k}$ (where indices are to be interpreted modulo n with representatives in $[n]$), and the Hamming weight of $\boldsymbol{a}$ is denoted by $\mathrm{wt}(\boldsymbol{a}) := \sum_{i=1}^n a_i$.

.

2.2 DES key schedule

The Data Encryption Standard (DES) is defined in [Nat99]. 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 $\boldsymbol{k} = (\boldsymbol{c}, \boldsymbol{d}) \in \{0,1\}^{56}$, where $\boldsymbol{c}, \boldsymbol{d} \in \{0,1\}^{28}$ 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. $\boldsymbol{c}, \boldsymbol{d}$ correspond to $\mathsf{C}_0, \mathsf{D}_0$ in the notation of [Nat99]).

The DES round keys $\boldsymbol{k}_1, \dots, \boldsymbol{k}_{16} \in \{0,1\}^{48}$ are derived from $\boldsymbol{k} = (\boldsymbol{c}, \boldsymbol{d})$ as follows. We write $\boldsymbol{c} = c_1 \cdots c_{28}$ and $\boldsymbol{d} = d_1 \cdots d_{28}$. 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 $2 \le i \le 16$. The values of the C- and D-registers in round i are therefore given by

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

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 $(\boldsymbol{c}_i, \boldsymbol{d}_i)$ to obtain the round key $\boldsymbol{k}_i$. The map PC-2 is defined as

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

where $\sigma\colon [24] \to [28] \setminus \{9, 18, 22, 25\}$ and $\tau\colon [24] \to [28] \setminus \{7, 10, 15, 26\}$ 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 σ and τ, respectively. The round keys are finally defined as $\boldsymbol{k}_i := \text{PC-2}(\boldsymbol{c}_i, \boldsymbol{d}_i)$ for $i \in [16]$. Written in terms of the original key $\boldsymbol{k} = (\boldsymbol{c}, \boldsymbol{d})$, we have

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

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 [HZZW16, WHZZ16, WH17b, WH17a, WH18, HMU+20]. The models assume that the key-dependent leakage originates from updates $(\boldsymbol{c}_{i+1}, \boldsymbol{d}_{i+1}) \leftarrow (\boldsymbol{c}_i, \boldsymbol{d}_i)$ of the C- and D-registers and/or updates $\boldsymbol{k}_{i+1} \leftarrow \boldsymbol{k}_i$ 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 $\boldsymbol{a} \in \{0,1\}^{28}$ be one half of a DES key $(\boldsymbol{c}, \boldsymbol{d}) \in \{0,1\}^{56}$ in the C- or D-register. By (2), (3), and (4), the bit transitions occuring 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 $\boldsymbol{a} \in \{0,1\}^{28}$, we define the vector

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

Written differently, we have

$$\Delta_1(\boldsymbol{a}) = \big((-1)^{a_1 \oplus a_2}, (-1)^{a_2 \oplus a_3}, \ldots, (-1)^{a_{27} \oplus a_{28}}, (-1)^{a_{28} \oplus a_1}\big)^\top \quad \text{and}$$
$$\Delta_2(\boldsymbol{a}) = \big((-1)^{a_1 \oplus a_3}, (-1)^{a_2 \oplus a_4}, \ldots, (-1)^{a_{27} \oplus a_1},\ (-1)^{a_{28} \oplus a_2}\big)^\top.$$

Furthermore, we define the stacked vectors

$$\Delta(\boldsymbol{a}) := \begin{pmatrix} \Delta_1(\boldsymbol{a}) \\ \Delta_2(\boldsymbol{a}) \end{pmatrix} \in \{\pm 1\}^{56} \quad \text{and} \quad \Delta(\boldsymbol{c}, \boldsymbol{d}) := \begin{pmatrix} \Delta(\boldsymbol{c}) \\ \Delta(\boldsymbol{d}) \end{pmatrix} \in \{\pm 1\}^{112} \tag{6}$$

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

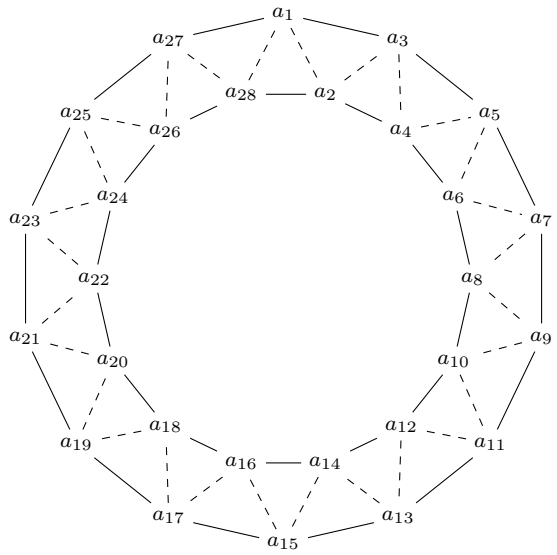

Figure 1: The nodes in this graph represent the bits of a key half $\boldsymbol{a} \in \{0,1\}^{28}$ in the C- or D-register. The dashed edges $\{a_i, a_{i+1}\}$ correspond to the components $(-1)^{a_i \oplus a_{i+1}}$ of $\Delta_1(\boldsymbol{a})$ and the solid edges $\{a_i, a_{i+2}\}$ to the components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

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

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 $(\boldsymbol{c},\boldsymbol{d})$ is given by an $\mathbb{R}$-linear function of $\Delta(\boldsymbol{c},\boldsymbol{d})$ and a key-independent error term.

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

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

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

We refer to realizations $\boldsymbol{y} \in \mathbb{R}^m$ of $\boldsymbol{Y}$ as observations, to realizations of $\boldsymbol{\varepsilon}$ 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* $\boldsymbol{W}_1, \boldsymbol{W}_2 \in \mathbb{R}^{m \times 56}$ *such that* $\boldsymbol{W} = (\boldsymbol{W}_1, \boldsymbol{W}_2)$.

(a) We have $\boldsymbol{W}\Delta(\boldsymbol{C},\boldsymbol{D}) = \boldsymbol{W}_1\Delta(\boldsymbol{C}) + \boldsymbol{W}_2\Delta(\boldsymbol{D})$.

(b) We have $\mathrm{E}(\Delta(\boldsymbol{C})) = \mathrm{E}(\Delta(\boldsymbol{D})) = \boldsymbol{0}_{56}$ *and* $\mathrm{Cov}(\Delta(\boldsymbol{C})) = \mathrm{Cov}(\Delta(\boldsymbol{D})) = \boldsymbol{I}_{56}$.

(c) We have $\mathrm{E}(\Delta(\boldsymbol{C},\boldsymbol{D})) = \boldsymbol{0}_{112}$ *and* $\mathrm{Cov}(\Delta(\boldsymbol{C},\boldsymbol{D})) = \boldsymbol{I}_{112}$.

(d) We have $\mathrm{E}(\boldsymbol{Y}) = \boldsymbol{0}_m$ *and* $\mathrm{Cov}(\boldsymbol{Y}) = \boldsymbol{W}\boldsymbol{W}^\top + \mathrm{Cov}(\boldsymbol{\varepsilon}) = \boldsymbol{W}_1\boldsymbol{W}_1^\top + \boldsymbol{W}_2\boldsymbol{W}_2^\top + \mathrm{Cov}(\boldsymbol{\varepsilon})$.

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

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

$$\mathrm{E}(X_iX_j) = \mathrm{E}\big((-1)^{C_p\oplus C_q}(-1)^{C_r\oplus C_s}\big) = \mathrm{E}\big((-1)^{C_p}(-1)^{C_q}(-1)^{C_r}(-1)^{C_s}\big)\,.$$

  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 $\mathrm{E}(X_iX_j) = \mathrm{E}\big((-1)^{C_p}\big)\,\mathrm{E}\big((-1)^{C_q}(-1)^{C_r}(-1)^{C_s}\big) = 0$.

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

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 [WH17a, Section 5].

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

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

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 $\boldsymbol{W} \in \{-\frac{1}{2},0\}^{15\times 112}$, and error $\boldsymbol{\varepsilon} = \boldsymbol{0}_{15}$. The weight matrix $\boldsymbol{W}$ is completely determined by the model assumptions and will be derived in Subsection 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 $\boldsymbol{K} = (\boldsymbol{C},\boldsymbol{D})$ and $\boldsymbol{Y}$ be the random variables as defined in Leakage Model 2. We want to determine a weight matrix $\boldsymbol{W} \in \{-\frac{1}{2},0\}^{15\times 112}$ such that $\boldsymbol{Y} = \boldsymbol{W}\Delta(\boldsymbol{C},\boldsymbol{D})$. Let $i \in [15]$.

Then

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

The images of σ and τ 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. Subsection 2.2). Changing the summation order, we obtain the representation

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

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 $\boldsymbol{W} \in \{-\frac{1}{2}, 0\}^{15 \times 112}$ can be easily read off, see Figure 2.

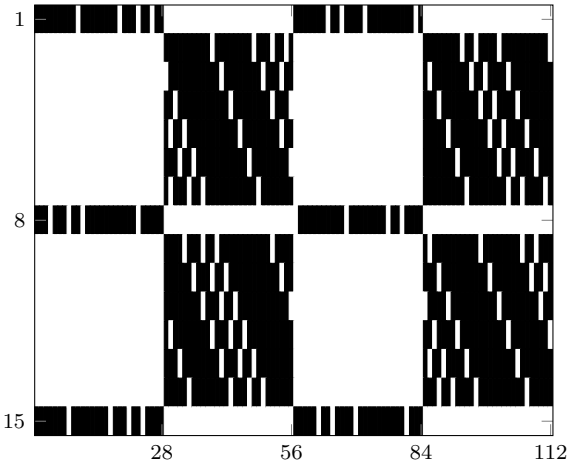

Figure 2: Matrix plot of the weight matrix $\boldsymbol{W}$ with element values $-\frac{1}{2}$ and 0 depicted in black and white, respectively.

Next we want to determine the covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{Y})$. By Lemma 1 (d), we have $\boldsymbol{\Sigma} = \boldsymbol{W}\boldsymbol{W}^\top$. 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

$$\sigma_{i,j} = \frac{1}{4} \Big(56 - \#\big((\rho(i) + M_\sigma) \cup (\rho(j) + M_\sigma)\big) - \#\big((\rho(i) + M_\tau) \cup (\rho(j) + M_\tau)\big) \Big) .$$

More concretely, we get

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

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

3.2 Key ranking and key enumeration

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

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

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

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

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

At first glance, enumerating the set $\mathcal{C}(\boldsymbol{y})$ looks like a 56-bit (or 54-bit) problem. However, we can apply a meet-in-the-middle approach (cf. [WH17a, Section 5]). By Lemma 1 (a), we have the decomposition

$$\boldsymbol{W}\Delta(\boldsymbol{c}, \boldsymbol{d}) = \boldsymbol{W}_1\Delta(\boldsymbol{c}) + \boldsymbol{W}_2\Delta(\boldsymbol{d}) \quad \text{for all } (\boldsymbol{c}, \boldsymbol{d}) \in \{0,1\}^{56}\,, \tag{12}$$

where $\boldsymbol{W}_1, \boldsymbol{W}_2 \in \mathbb{R}^{15\times 56}$ such that $\boldsymbol{W} = (\boldsymbol{W}_1, \boldsymbol{W}_2)$. This leads to the following simple enumeration procedure.

Algorithm 1.

Input: A vector $\boldsymbol{y} \in ([-24, 24] \cap \mathbb{Z})^{15}$.
Output: The set of key candidates $\mathcal{C}(\boldsymbol{y})$.

1. Compute the lists

$$\mathcal{L}_1 \leftarrow \big\{(\boldsymbol{c}, \boldsymbol{y} - \boldsymbol{W}_1\Delta(\boldsymbol{c})) \mid \boldsymbol{c} = 0, \ldots, 2^{27} - 1\big\}\,,$$
$$\mathcal{L}_2 \leftarrow \big\{(\boldsymbol{d}, \boldsymbol{W}_2\Delta(\boldsymbol{d})) \mid \boldsymbol{d} = 0, \ldots, 2^{27} - 1\big\}$$

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

2. Set $\mathcal{C} \leftarrow \varnothing$. For all $(\boldsymbol{c}, \boldsymbol{y}_1) \in \mathcal{L}_1$ and $(\boldsymbol{d}, \boldsymbol{y}_2) \in \mathcal{L}_2$ with $\boldsymbol{y}_1 = \boldsymbol{y}_2$, set

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

(Since $\mathcal{L}_1$ and $\mathcal{L}_1$ 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 [BEKS17] and conducted 1000 trials. For each trial we chose a random DES key and calculated the associated observation vector $\boldsymbol{y}$. Then we enumerated all candidates $(\boldsymbol{c}, \boldsymbol{d})$ such that $\boldsymbol{y} = \boldsymbol{W}\Delta(\boldsymbol{c}, \boldsymbol{d})$. 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.

Table 1: Empirical distribution of the logarithmic key rank based on 1000 trials with random keys. Q1 and Q3 denote the first and third quartile, respectively.

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

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.

3.4 Theoretical estimation of the remaining entropy

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

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

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

3.4.1 A lower bound for the remaining entropy

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

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

$$\mathrm{H}(\boldsymbol{Y}) \le \frac{1}{2} \log_2\bigl(\det(2\pi e \boldsymbol{\Sigma})\bigr) + m \log_2\left(\frac{1 + e^{-2\pi^2 \lambda}}{1 - e^{-2\pi^2 \lambda}}\right), \tag{13}$$

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

Proof. Let $\mathcal{Y} \subseteq \mathbb{Z}^m$ be the support of $\boldsymbol{Y}$. By Gibbs' inequality (cf. [CT06, Theorem 2.6.3]), we have

$$\mathrm{H}(\boldsymbol{Y}) = -\sum_{\boldsymbol{y} \in \mathcal{Y}} \Pr(\boldsymbol{Y} = \boldsymbol{y}) \log_2\bigl(\Pr(\boldsymbol{Y} = \boldsymbol{y})\bigr) \le -\sum_{\boldsymbol{y} \in \mathcal{Y}} \Pr(\boldsymbol{Y} = \boldsymbol{y}) \log_2\bigl(p(\boldsymbol{y})\bigr)$$

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

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

we obtain

$$\begin{aligned}
\mathrm{H}(\boldsymbol{Y}) &\leq -\log_2(\mu) + \frac{\log_2(e)}{2} \sum_{\boldsymbol{y}\in\mathcal{Y}} \Pr(\boldsymbol{Y}=\boldsymbol{y}) \boldsymbol{y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{y} \\
&= -\log_2(\mu) + \frac{\log_2(e)}{2} \mathrm{E}(\boldsymbol{Y}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{Y}) \\
&= -\log_2(\mu) + \frac{\log_2(e)}{2} \operatorname{tr}\big(\boldsymbol{\Sigma}^{-1} \mathrm{E}(\boldsymbol{Y}\boldsymbol{Y}^\top)\big) \qquad \text{(by the “trace trick”)} \\
&= -\log_2(\mu) + \frac{m\log_2(e)}{2} .
\end{aligned}$$

Using the Poisson summation formula (cf. [Ban93, Lemma (1.1) (i)]), we get

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

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

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

finishing the proof. □

Remark 3. We note that [Ban93, Lemma (1.5) (i)] implies a better bound for the term $m\log_2(\ldots)$ of equation (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 $\mathrm{H}(\boldsymbol{Y}) \leq 41.73$ by numerical methods. We also note that the term

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

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

$$\mathrm{H}(\boldsymbol{C},\boldsymbol{D} \mid \boldsymbol{Y}) = 56 - \mathrm{H}(\boldsymbol{Y}) \geq 14.27 \tag{15}$$

for the remaining entropy. The experiments reported in Subsection 3.3 (cf. Table 1) suggest that the remaining entropy is close to this lower bound. In Subsubsection 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 Subsection 3.3, we propose the heuristic formula

$$\mathrm{H}(\boldsymbol{C}, \boldsymbol{D} \mid \boldsymbol{Y}) \approx 56 - \frac{1}{2}\log_2\bigl(\det(2\pi e \boldsymbol{\Sigma})\bigr) \approx 14.27 \tag{16}$$

for the remaining entropy.

Each component Y_i of $\boldsymbol{Y}$ 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 $\boldsymbol{Y}$ by a 15-dimensional normal distribution?

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

- If the support of $\boldsymbol{AX}$ is contained in a subset $S \subseteq \mathbb{R}^m$, then clearly $\mathrm{H}(\boldsymbol{AX}) \le \log_2\bigl(\#(S \cap \mathbb{Z}^m)\bigr)$. In addition, we can expect that $\log_2\bigl(\#(S \cap \mathbb{Z}^m)\bigr) \approx \log_2\bigl(\mathrm{vol}(S)\bigr)$ for "natural" sets S.

- Let $\boldsymbol{A} = \boldsymbol{UDV}^\top$ be the singular value decomposition of $\boldsymbol{A}$, where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ are orthogonal matrices and $\boldsymbol{D} \in \mathbb{R}^{m\times n}$ is a rectangular diagonal matrix with non-negative elements on the diagonal. This representation of $\boldsymbol{A}$ easily implies that $\boldsymbol{AX}$ takes values in an m-dimensional ellipsoid with semiaxes equal to the non-zeros elements of $\boldsymbol{D}$ times $\sqrt{n}$. The volume of this ellipsoid is

$$V_m(1) n^{m/2} \sqrt{\det\bigl(\boldsymbol{DD}^\top\bigr)} = V_m(1) n^{m/2} \sqrt{\det\bigl(\boldsymbol{AA}^\top\bigr)}\,,$$

  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 $\boldsymbol{AX}$.
  (ii) The components of $\boldsymbol{V}^\top\boldsymbol{X}$ have expectation 0 and variance 1. Therefore, the bulk of the support of $\boldsymbol{AX}$ lies in a smaller ellipsoid with semiaxes equal to the non-zero elements of $\boldsymbol{D}$ times $\sqrt{m}$. The volume of this smaller ellipsoid is

$$V_m(1) m^{m/2} \sqrt{\det\bigl(\boldsymbol{AA}^\top\bigr)}\,.$$

  (iii) Furthermore, if (i) is valid, we expect that the discrete distribution of $\boldsymbol{AX}$ is "similar" to the continuous distribution $\boldsymbol{AZ}$, where $\boldsymbol{Z}$ is normally distributed with covariance matrix $\boldsymbol{I}_n$. $\boldsymbol{AZ}$ is therefore normally distributed with covariance matrix $\boldsymbol{AA}^\top$. The entropy of $\boldsymbol{AZ}$ is well known and given by the formula

$$\frac{1}{2}\log_2\bigl(\det(2\pi e \boldsymbol{AA}^\top)\bigr)\,.$$

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

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

3.4.3 Distribution of the remaining entropy

Why has the remaining entropy in the experiments in Subsection 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 ± 24, we have a large amount of information about the key $(\boldsymbol{c}, \boldsymbol{d})$. On the other

hand, for $y_i = 0$ there are many candidates for $(\boldsymbol{c}, \boldsymbol{d})$. As argued in Subsubsection 3.4.2, we expect that in our case

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

This leads to the following approximation of the remaining entropy $\mathrm{H}(\boldsymbol{C}, \boldsymbol{D} \mid \boldsymbol{Y} = \boldsymbol{y})$ for fixed $\boldsymbol{y}$:

$$\begin{aligned}\mathrm{H}(\boldsymbol{C}, \boldsymbol{D} \mid \boldsymbol{Y} = \boldsymbol{y}) &= \log_2\bigl(\#\{(\boldsymbol{c}, \boldsymbol{d}) \in \{0,1\}^{56} \mid \boldsymbol{W}\Delta(\boldsymbol{c}, \boldsymbol{d}) = \boldsymbol{y}\}\bigr) \\ &= 56 + \log_2\bigl(\Pr(\boldsymbol{Y} = \boldsymbol{y})\bigr) \\ &\approx \max\Bigl\{2\,,\ 56 - \tfrac{1}{2}\log_2\bigl(\det(2\pi\boldsymbol{\Sigma})\bigr) - \tfrac{1}{2}\log_2(e)\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}\|^2\Bigr\} \\ &\approx \max\Bigl\{2\,,\ 25.09 - 0.72 \cdot \|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}\|^2\Bigr\}\,.\end{aligned} \tag{17}$$

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

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

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

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

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

In order to test the heuristic (17), we conducted some further experiments. For every $a \in \{0.6, 0.8, \ldots, 7.4\}$, we generated 10 random observations $\boldsymbol{y} = \boldsymbol{W}\Delta(\boldsymbol{k})$ such that $\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}\| \in [a, a + 0.2]$ by rejection sampling and computed $\#\mathcal{C}(\boldsymbol{y})$ using Algorithm 1. Figure 3 shows a plot of the remaining entropy estimate (17) as a function of $\|\boldsymbol{\Sigma}^{-1/2}\boldsymbol{y}\|$ 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. Subsection 3.3). The experiments confirm that the approximation (17) is good.

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

$$\begin{aligned}\mathrm{E}\Bigl(25.09 - \tfrac{1}{2}\boldsymbol{u}^\top \boldsymbol{u} \log_2(e)\Bigr) &= 25.09 - \frac{1}{2}15\log_2(e) \approx 14.27\,, \\ \mathrm{Var}\Bigl(25.09 - \tfrac{1}{2}\boldsymbol{u}^\top \boldsymbol{u} \log_2(e)\Bigr) &= 30 \cdot \Bigl(\frac{1}{2}\log_2(e)\Bigr)^2 \approx 15.61\,.\end{aligned}$$

Note that the expectation fits to the value $\mathrm{H}(\boldsymbol{C}, \boldsymbol{D} \mid \boldsymbol{Y}) = 56 - \mathrm{H}(\boldsymbol{Y}) \approx 56 - 41.7 \approx 14.3$ 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 Subsection 3.3.

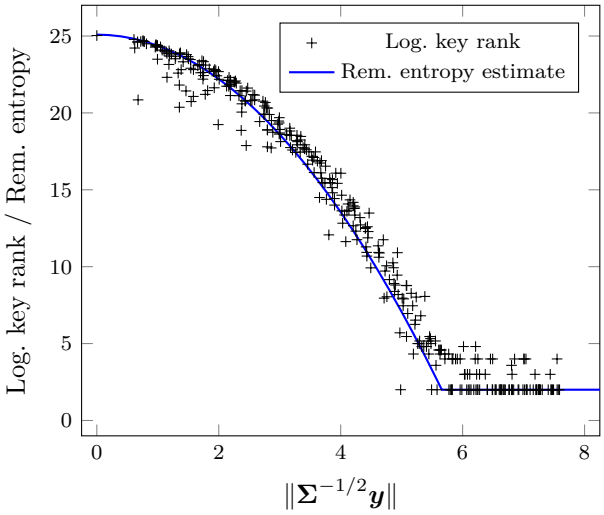

Figure 3: Graph of the remaining entropy estimate (17) as a function of $\|\Sigma^{-1/2}\boldsymbol{y}\|$ together with the logarithmic key ranks computed in the experiments.

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 [HMU+20], 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 $\boldsymbol{c}$ 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 $\boldsymbol{d}$ in the D-register. Find a set $\mathcal{D}$ of likely candidates for the D-register.

3. Check all combinations $(\boldsymbol{c}, \boldsymbol{d}) \in \mathcal{C} \times \mathcal{D}$.

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 $\boldsymbol{W}_2\Delta(\boldsymbol{D})$ takes on certain values in a 15-dimensional space and its entropy $\mathrm{H}(\boldsymbol{W}_2\Delta(\boldsymbol{D}))$ is clearly bounded by 27. By Lemma 2, we have

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

where we have neglected the second term in (13). Let $\boldsymbol{U}_2$ 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

$$\mathrm{H}(\boldsymbol{W}_2\boldsymbol{U}_2) \approx 34.05\,.$$

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

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

Applying Lemma 2 to $\boldsymbol{Y}_1$, we get

$$\begin{aligned}\mathrm{I}(\boldsymbol{Y}_1;\boldsymbol{C}) &= \mathrm{I}(\boldsymbol{W}_1\Delta(\boldsymbol{C}) + \boldsymbol{W}_2\boldsymbol{U}_2;\boldsymbol{C}) = \mathrm{H}(\boldsymbol{Y}_1) - \mathrm{H}(\boldsymbol{W}_2\boldsymbol{U}_2)\\ &\approx \frac{1}{2}\log_2\bigl(\det(2\pi e\boldsymbol{W}\boldsymbol{W}^\top)\bigr) - \frac{1}{2}\log_2\bigl(\det(2\pi e\boldsymbol{W}_2\boldsymbol{W}_2^\top)\bigr)\\ &= \frac{1}{2}\log_2\bigl(\det(\boldsymbol{W}\boldsymbol{W}^\top)\bigr) - \frac{1}{2}\log_2\bigl(\det(\boldsymbol{W}_2\boldsymbol{W}_2^\top)\bigr)\,.\end{aligned}$$

Using this approximation, we obtain

$$\mathrm{I}(\boldsymbol{Y}_1;\boldsymbol{C}) \approx 7.68 \quad \text{and} \quad \mathrm{I}(\boldsymbol{Y}_2;\boldsymbol{D}) \approx 8.68\,,$$

where $\boldsymbol{Y}_2 := \boldsymbol{W}_1\boldsymbol{U}_1 + \boldsymbol{W}_2\Delta(\boldsymbol{D})$ is defined analoguously to $\boldsymbol{Y}_1$ with $\boldsymbol{U}_1$ uniformly distributed on $\{\pm1\}^{56}$.

Remark 4. These small concrete values do not come as a complete surprise. By construction of $\boldsymbol{W}$ we know that $\boldsymbol{W}_1\boldsymbol{W}_1^\top \approx \boldsymbol{W}_2\boldsymbol{W}_2^\top$, so that we can expect roughly

$$\begin{aligned}\mathrm{I}(\boldsymbol{Y}_1;\boldsymbol{C}) &\approx \frac{1}{2}\log_2\bigl(\det(\boldsymbol{W}\boldsymbol{W}^\top)\bigr) - \frac{1}{2}\log_2\bigl(\det(\boldsymbol{W}_2\boldsymbol{W}_2^\top)\bigr)\\ &\approx \frac{1}{2}\log_2\bigl(\det(2\boldsymbol{W}_2\boldsymbol{W}_2^\top)\bigr) - \frac{1}{2}\log_2\bigl(\det(\boldsymbol{W}_2\boldsymbol{W}_2^\top)\bigr) = \frac{1}{2}\log_2(2^{15}) = \frac{15}{2}\,.\end{aligned}$$

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

$$\boldsymbol{Y} = \boldsymbol{W}\Delta(\boldsymbol{C},\boldsymbol{D}) + \boldsymbol{\varepsilon}\,, \tag{18}$$

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

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 Subsubsection 4.3.1.

4.1 Key ranking and key enumeration

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

We define the evaluation function

$$\eta_{\boldsymbol{W},\boldsymbol{y}}\colon \{0,1\}^{56} \to \mathbb{R}_{\geq 0}\,, \quad (\boldsymbol{c},\boldsymbol{d}) \mapsto \bigl\|\boldsymbol{y} - \boldsymbol{W}\Delta(\boldsymbol{c},\boldsymbol{d})\bigr\|^2\,. \tag{19}$$

We denote by

$$\mathcal{C}_{\boldsymbol{W}}(\boldsymbol{y},B) := \bigl\{(\boldsymbol{c},\boldsymbol{d}) \in \{0,1\}^{56} \mid \eta_{\boldsymbol{W},\boldsymbol{y}}(\boldsymbol{c},\boldsymbol{d}) \leq B\bigr\} \tag{20}$$

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

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

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

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

$$\mathcal{R}_{\boldsymbol{W},\boldsymbol{y}}(\boldsymbol{k}^*) \approx \frac{2^{56}}{N} \cdot \#\big\{j \in [N] \,\big|\, \eta_{\boldsymbol{W},\boldsymbol{y}}(\boldsymbol{c}_j, \boldsymbol{d}_j) \leq \eta_{\boldsymbol{W},\boldsymbol{y}}(\boldsymbol{c}^*, \boldsymbol{d}^*)\big\}\,. \tag{22}$$

Next we develop an algorithm to enumerate the set $\mathcal{C}_{\boldsymbol{W}}(\boldsymbol{y}, B)$ based on the Fincke–Pohst lattice point enumeration algorithm [FP85]. 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 $\boldsymbol{W} \in \mathbb{R}^{m \times 112}$ can be replaced by an upper triangular matrix $\boldsymbol{R} \in \mathbb{R}^{112 \times 112}$. At the same time, the observation vector $\boldsymbol{y}$ is projected onto the range of $\boldsymbol{W}$.

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

$$\big\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\big\|^2 = \big\|\boldsymbol{R}(\boldsymbol{t} - \boldsymbol{x})\big\|^2 + \big\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{t}\big\|^2 \quad \textit{for all } \boldsymbol{x} \in \mathbb{R}^{112}\,,$$

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

Proof. Since $\boldsymbol{W}$ has full column rank, the matrix $\boldsymbol{W}^\top \boldsymbol{W}$ is symmetric positive-definite and, in particular, non-singular. The existence and uniqueness of $\boldsymbol{R}$ follow from the Cholesky factorization of $\boldsymbol{W}^\top \boldsymbol{W}$ (cf. [GL96, Theorem 4.2.5]). Since $\boldsymbol{W}(\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top$ is the orthogonal projection onto the range of $\boldsymbol{W}$, we have $\langle \boldsymbol{W}\boldsymbol{x}, \boldsymbol{y} - \boldsymbol{W}\boldsymbol{t} \rangle = 0$ for all $\boldsymbol{x} \in \mathbb{R}^{112}$, hence $\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{x}\|^2 = \|\boldsymbol{W}(\boldsymbol{t} - \boldsymbol{x}) + \boldsymbol{y} - \boldsymbol{W}\boldsymbol{t}\|^2 = \|\boldsymbol{W}(\boldsymbol{t} - \boldsymbol{x})\|^2 + \|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{t}\|^2$ for all $\boldsymbol{x} \in \mathbb{R}^{112}$. Finally, we have $\|\boldsymbol{W}(\boldsymbol{t} - \boldsymbol{x})\|^2 = (\boldsymbol{t} - \boldsymbol{x})^\top \boldsymbol{W}^\top \boldsymbol{W}(\boldsymbol{t} - \boldsymbol{x}) = (\boldsymbol{t} - \boldsymbol{x})^\top \boldsymbol{R}^\top \boldsymbol{R}(\boldsymbol{t} - \boldsymbol{x}) = \|\boldsymbol{R}(\boldsymbol{t} - \boldsymbol{x})\|^2$ for all $\boldsymbol{x} \in \mathbb{R}^{112}$. □

Remark 6. Consider the situation of Lemma 3. By the thin/reduced QR factorization of $\boldsymbol{W}$, there exists a unique matrix $\boldsymbol{Q} \in \mathbb{R}^{m \times 112}$ with orthonormal columns and a unique upper triangular matrix $\boldsymbol{R} \in \mathbb{R}^{112 \times 112}$ with positive diagonal elements such that $\boldsymbol{W} = \boldsymbol{Q}\boldsymbol{R}$ (cf. [GL96, Theorem 5.2.2]). We have $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{R}^\top \boldsymbol{R}$ (in particular, $\boldsymbol{R}$ is the Cholesky factor of $\boldsymbol{W}^\top \boldsymbol{W}$) and $\boldsymbol{t} = (\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{y} = \boldsymbol{R}^{-1} \boldsymbol{Q}^\top \boldsymbol{y}$.

By Lemma 3, we have $\mathcal{C}_{\boldsymbol{W}}(\boldsymbol{y}, B) = \mathcal{C}_{\boldsymbol{R}}(\boldsymbol{R}\boldsymbol{t}, B - \|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{t}\|^2)$. Let $(\boldsymbol{c}, \boldsymbol{d}) \in \{0,1\}^{56}$ and $\boldsymbol{x} := \Delta(\boldsymbol{c}, \boldsymbol{d}) \in \{\pm 1\}^{112}$. Then $(\boldsymbol{c}, \boldsymbol{d}) \in \mathcal{C}_{\boldsymbol{W}}(\boldsymbol{y}, B)$ if and only if

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

In principle, we could enumerate all vectors $\boldsymbol{x} \in \{\pm 1\}^{112}$ satisfying (23) using backtracking (cf. [Knu19, Section 7.2.2, Algorithm B]). If the elements of $\boldsymbol{x}$ are traversed in the order $x_{112}, x_{111}, \ldots, x_1$, then partial assignments $x_s \cdots x_{112}$ violating the condition

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

can be rejected immediately (without trying further values for $x_1 \cdots x_{s-1}$). For each vector $\boldsymbol{x} \in \{\pm 1\}^{112}$ satisfying (23) we could then compute the preimage under Δ 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 $\boldsymbol{k} = (\boldsymbol{c}, \boldsymbol{d}) \in \{0,1\}^{56}$ directly. In order to make this approach work, we have to reorder the components of $\boldsymbol{x} = \Delta(\boldsymbol{k}) = \Delta(\boldsymbol{c}, \boldsymbol{d})$ and the columns of $\boldsymbol{W}$ have to be reordered accordingly (before applying Lemma 3).

Let $\boldsymbol{k} = k_1 \cdots k_{56} = (\boldsymbol{c}, \boldsymbol{d}) = c_1 \cdots c_{28}\, d_1 \cdots d_{28} \in \{0,1\}^{56}$. First we choose a permutation $\pi \colon [56] \to [56]$ that determines the order $k_{\pi(1)}, k_{\pi(2)}, \ldots, k_{\pi(56)}$ in which we want to traverse the bits of $\boldsymbol{k}$ in the enumeration procedure. By Remark 1 we can keep one bit of $\boldsymbol{c}$ and one bit of $\boldsymbol{d}$ 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 $\Delta(\boldsymbol{k})$ that are determined by the current choices. In other words, we pick nodes in the graph of Figure 1 such that the number of edges between the chosen nodes is maximized. For example, we may choose

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

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

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

and we may choose $\boldsymbol{P}$ such that

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

Applying Lemma 3 to the column-permuted matrix $\boldsymbol{W}\boldsymbol{P}$ and the observation $\boldsymbol{y} \in \mathbb{R}^m$, we obtain an upper triangular matrix $\boldsymbol{R} \in \mathbb{R}^{112 \times 112}$ and $\boldsymbol{t} \in \mathbb{R}^{112}$ such that

$$\left\|\boldsymbol{y} - \boldsymbol{W}\Delta(\boldsymbol{k})\right\|^2 = \left\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{P}\boldsymbol{P}^\top\Delta(\boldsymbol{k})\right\|^2 = \left\|\boldsymbol{R}(\boldsymbol{t} - \boldsymbol{x})\right\|^2 + \left\|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{P}\boldsymbol{t}\right\|^2.$$

Therefore, we have $\boldsymbol{k} \in \mathcal{C}_{\boldsymbol{W}}(\boldsymbol{y}, B)$ if and only if

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

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

$\rho_2 = 0$ and

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

Using the backtracking scheme described in [Knu19, Section 7.2.2, Algorithm B], we obtain the following algorithm.

Algorithm 2.

Input: A matrix $\boldsymbol{W} \in \mathbb{R}^{m\times 112}$ of full column rank, a vector $\boldsymbol{y} \in \mathbb{R}^m$, and a bound $B \in \mathbb{R}_{\ge 0}$.
Output: The set of key candidates $\mathcal{C}_{\boldsymbol{W}}(\boldsymbol{y}, B)$.

1. [Initialize.] Set $\mathcal{C} \leftarrow \varnothing$, $\ell \leftarrow 3$, $\boldsymbol{k} \leftarrow \boldsymbol{0}_{56}$, $\boldsymbol{x} \leftarrow \boldsymbol{1}_{112}$, and $\boldsymbol{\rho} \leftarrow \boldsymbol{0}_{56}$.

2. [Preprocess.] Set $\boldsymbol{W} \leftarrow \boldsymbol{WP}$. Compute an upper triangular matrix $\boldsymbol{R} \in \mathbb{R}^{112\times 112}$ such that $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{R}^\top \boldsymbol{R}$. Set $\boldsymbol{t} \leftarrow (\boldsymbol{W}^\top \boldsymbol{W})^{-1} \boldsymbol{W}^\top \boldsymbol{y}$ and $B \leftarrow B - \|\boldsymbol{y} - \boldsymbol{Wt}\|^2$. If $B < 0$, return $\mathcal{C}$ and stop.

3. [Enter level ℓ.] If $\ell = 57$, set $(\boldsymbol{c}, \boldsymbol{d}) \leftarrow \boldsymbol{k}$, set $\mathcal{C} \leftarrow \mathcal{C} \cup \{(\boldsymbol{c}, \boldsymbol{d}), (\overline{\boldsymbol{c}}, \boldsymbol{d}), (\boldsymbol{c}, \overline{\boldsymbol{d}}), (\overline{\boldsymbol{c}}, \overline{\boldsymbol{d}})\}$, and go to step **6**. Otherwise set $k_{\pi(\ell)} \leftarrow 0$.

4. [Try $k_{\pi(\ell)}$.] Set $x_i \leftarrow (\boldsymbol{P}^\top \Delta(\boldsymbol{k}))_i$ for $i = s_\ell, \ldots, s_{\ell-1} - 1$ 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 $\rho_\ell \le B$, 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 $\ell \ge 3$, go to step **5**. Otherwise return $\mathcal{C}$ and stop.

Remark 7. We note some possible variations and optimizations of Algorithm 2.

(a) To avoid repeated computations, Algorithm 2 can be modified as follows (cf. [GNR10, Appendix B] and [LN13, 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 $3 \le \ell \le 56$. By using these recurrence relations and by reusing values of $\sigma_{i,h}$ that are still valid during the enumeration, the partial squared norms ρ_ℓ can be computed with fewer operations. For details, see Algorithm 3 in Appendix A.

(b) Using pruning [SE94, SH95], we can heuristically reject partial assignments $k_{\pi(1)} \cdots k_{\pi(\ell)}$ during the enumeration if the partial squared norm ρ_ℓ is already so large that $\rho_{56} \le B$ becomes unlikely for any choice of $k_{\pi(\ell+1)} \cdots k_{\pi(56)}$. This can be done by replacing the if-condition "$\rho_\ell \le B$" in step **4** by "$\rho_\ell \le B_\ell$" for suitable bounds $0 \le B_3 \le \cdots \le B_{56} = B$. In our experiments we used the bounds

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

Note that we cannot use extreme pruning [GNR10], since we want to find all (or almost all) key candidates in $\mathcal{C}_{\boldsymbol{W}}(\boldsymbol{y}, B)$.

(c) For a given weight matrix $\boldsymbol{W}$, the running time of Algorithm 2 may be optimized by choosing different permutations π and $\boldsymbol{P}$ (in compliance with the conditions outlined above). Preprocessing $\boldsymbol{W}$ with general unimodular transformations (e.g. lattice basis reduction, cf. [FP85, (2.12)]) seems not possible in our setting.

(d) To find N best key candidates for the evaluation function $\eta_{\boldsymbol{W},\boldsymbol{y}}$, 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 $\eta_{\boldsymbol{W},\boldsymbol{y}}$ 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 $\boldsymbol{W} \in \mathbb{R}^{m\times 112}$ 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 $\boldsymbol{y}_{\text{prf},j} \in \mathbb{R}^m$ of Leakage Model 3 for known and randomly chosen keys $\boldsymbol{k}_{\text{prf},j} = (\boldsymbol{c}_{\text{prf},j}, \boldsymbol{d}_{\text{prf},j}) \in \{0,1\}^{56}$ for $j \in [N_{\text{prf}}]$. We denote by $\boldsymbol{X}_{\text{prf}} \in \{\pm 1\}^{112\times N_{\text{prf}}}$ the matrix with columns $\boldsymbol{x}_{\text{prf},j} := \Delta(\boldsymbol{c}_{\text{prf},j}, \boldsymbol{d}_{\text{prf},j})$ and by $\boldsymbol{Y}_{\text{prf}} \in \mathbb{R}^{m\times N_{\text{prf}}}$ the matrix with columns $\boldsymbol{y}_{\text{prf},j}$ for $j \in [N_{\text{prf}}]$.

We want to find an approximation $\widetilde{\boldsymbol{W}} \in \mathbb{R}^{m\times 112}$ of $\boldsymbol{W}$ such that $\boldsymbol{Y}_{\text{prf}} \approx \widetilde{\boldsymbol{W}}\boldsymbol{X}_{\text{prf}}$. Since the error vector in Leakage Model 3 has independent components, we may estimate the rows of $\boldsymbol{W}$ independently. Let $i \in [m]$ and let $\boldsymbol{y}_{\text{prf},i} \in \mathbb{R}^{1\times N_{\text{prf}}}$ denote the i-th row of $\boldsymbol{Y}_{\text{prf}}$. We approximate the i-th row of $\boldsymbol{W}$ by a least squares estimate, i.e. by a vector $\boldsymbol{w} \in \mathbb{R}^{1\times 112}$ minimizing

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

Since $N_{\text{prf}} \gg 112$, we may assume that $\boldsymbol{X}_{\text{prf}}$ has full row rank and $\boldsymbol{X}_{\text{prf}}\boldsymbol{X}_{\text{prf}}^\top \in \mathbb{R}^{112\times 112}$ is non-singular. This implies that (24) is minimized by the (unique) vector $\widetilde{\boldsymbol{w}}_i := \boldsymbol{y}_{\text{prf},i}\boldsymbol{X}_{\text{prf}}^\top(\boldsymbol{X}_{\text{prf}}\boldsymbol{X}_{\text{prf}}^\top)^{-1} \in \mathbb{R}^{1\times 112}$ (cf. [GL96, Section 5.3.1]). Combining the estimated rows of the weight matrix, we obtain the matrix

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

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

4.3 Experiments

We performed experiments using real and simulated measurements.

4.3.1 Real measurements

The authors of [HMU+20] 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 $\boldsymbol{k}_{\text{prf},j} = (\boldsymbol{c}_{\text{prf},j}, \boldsymbol{d}_{\text{prf},j}) \in \{0,1\}^{56}$ for $j \in [N_{\text{prf}}]$. We denote by $\boldsymbol{Y}_{\text{prf}} \in \mathbb{R}^{m\times N_{\text{prf}}}$ the matrix of measurements (arranged in columns) and by $\boldsymbol{X}_{\text{prf}} \in \{\pm 1\}^{112\times N_{\text{prf}}}$ the matrix with columns $\Delta(\boldsymbol{c}_{\text{prf},j}, \boldsymbol{d}_{\text{prf},j})$.

The attack set comprises $N_{\text{att}} = 247088$ measurements of DES operations with random keys $\boldsymbol{k}_{\text{att},j} \in \{0,1\}^{56}$ for $j \in [288]$, where $N_{\text{att},j} \in \{761,\dots,927\}$ measurements have been performed with key $\boldsymbol{k}_{\text{att},j}$. (The authors of [HMU+20] also carried out measurements

for so-called *weak keys*, but we do not consider them in this article.) We denote by $\boldsymbol{Y}_{\mathrm{att},j} \in \mathbb{R}^{m \times N_{\mathrm{att},j}}$ the matrix of measurements with key $\boldsymbol{k}_{\mathrm{att},j}$ (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

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

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

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

Using the empirical covariance matrix

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

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

$$\boldsymbol{Y}_{\mathrm{prf}} \leftarrow \widetilde{\boldsymbol{\Sigma}}^{-1/2} \boldsymbol{Y}_{\mathrm{prf}} \quad \text{and} \quad \boldsymbol{Y}_{\mathrm{att},j} \leftarrow \widetilde{\boldsymbol{\Sigma}}^{-1/2} \boldsymbol{Y}_{\mathrm{att},j} \quad \text{for } j \in [288] \,.$$

Finally, we computed the averaged measurements

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

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

$$(\overline{\boldsymbol{y}}_1, \ldots, \overline{\boldsymbol{y}}_{288}) \leftarrow (\overline{\boldsymbol{y}}_1, \ldots, \overline{\boldsymbol{y}}_{288})(\boldsymbol{I}_{288} - 288^{-1} \, \mathbf{1}_{288} \, \mathbf{1}_{288}^\top) \,.$$

Using the preprocessed data of the profiling set, we computed an estimate $\widetilde{\boldsymbol{W}} \in \mathbb{R}^{m \times 112}$ of the weight matrix according to Leakage Model 3 as in (25). A matrix plot of $\widetilde{\boldsymbol{W}}$ is shown in Figure 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. Figure 2). This visual structure of $\widetilde{\boldsymbol{W}}$ is a first indication that Leakage Model 3 is adequate for the measurements.

We implemented Algorithm 2 in the Julia programming language [BEKS17] with the optimizations of Remark 7 (a) and (b). We computed the key ranks

$$\mathcal{R}_i := \mathcal{R}_{\widetilde{\boldsymbol{W}}, \overline{\boldsymbol{y}}_i}(\boldsymbol{k}_{\mathrm{att},i})$$

for 287 of the 288 averaged measurements $\overline{\boldsymbol{y}}_i$ 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. A log-log plot of the running times and key ranks is shown in Figure 5.

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 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.

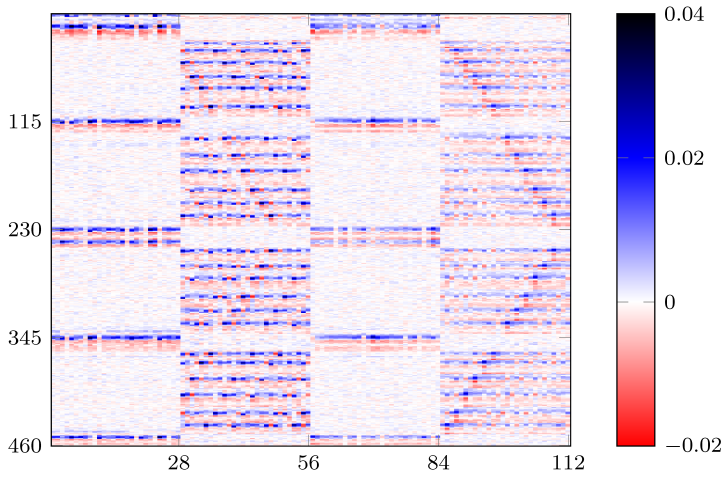

Figure 4: Matrix plot of the estimated weight matrix $\widetilde{\boldsymbol{W}}$ with element values depicted according to the color bar.

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

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 $\widetilde{\boldsymbol{W}} \in \mathbb{R}^{m \times 112}$ with $m = 460$ estimated from the real measurements as described in Subsubsection 4.3.1 (cf. Figure 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 $\sigma^2 \boldsymbol{I}_m$. In contrast to Subsubsection 4.3.1, we used the observations directly without averaging over several observations.

For each $\sigma \in \{0.02, 0.03, \ldots, 0.07\}$, we generated 100 observations and computed the corresponding key ranks explicitly using our implementation of Algorithm 2 with the optimizations of Remark 7 (a) and (b). The distributions of the computed ranks and the single-core running times are described in Table 3.

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

For larger values of σ, we have resorted to the Monte-Carlo heuristic (22). For each $\sigma \in \{0.2, 0.3, \ldots, 0.7\}$, 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.

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

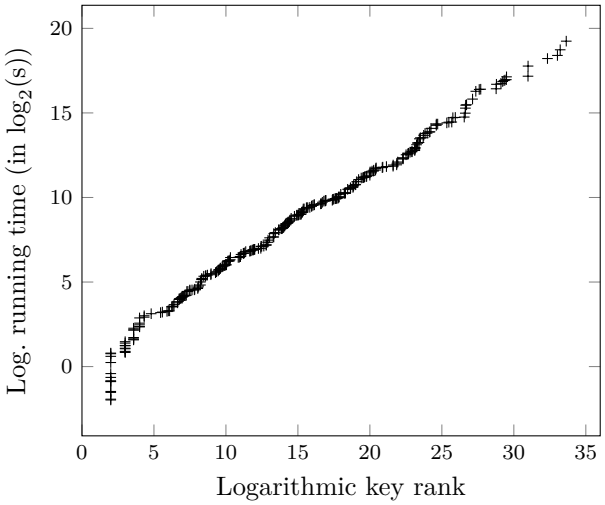

Figure 5: Log-log plot of the key enumeration running times and key ranks for the real measurements.

Table 3: Empirical distribution of the logarithmic key ranks and key enumeration running times for 100 simulated measurements per value of σ. The key ranks were computed explicitly using Algorithm 2 with the optimizations of Remark 7 (a) and (b).

Noise	**Logarithmic key rank**					**Running time**		
σ	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

random variables should be treated with care, since some of its properties are different compared to the discrete case (cf. [CT06]). The following lemma provides an upper bound for the mutual information of $\boldsymbol{Y}$ and $(\boldsymbol{C}, \boldsymbol{D})$ in Leakage Model 3.

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

$$\mathrm{I}(\boldsymbol{Y}; \boldsymbol{C}, \boldsymbol{D}) \leq \frac{1}{2} \log_2\bigl(\det(\sigma^{-2} \boldsymbol{W}^\top \boldsymbol{W} + \boldsymbol{I}_{112})\bigr) .$$

Proof. We have $\mathrm{I}(\boldsymbol{Y}; \boldsymbol{C}, \boldsymbol{D}) = H(\boldsymbol{Y}) - H(\boldsymbol{\varepsilon})$. Let $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{Y})$. By Lemma 1 (d), we have $\boldsymbol{\Sigma} = \boldsymbol{W}\boldsymbol{W}^\top + \sigma^2 \boldsymbol{I}_m$. By [CT06, Theorem 8.6.5], we obtain $H(\boldsymbol{Y}) \leq \frac{1}{2} \log_2\bigl(\det(2\pi e \boldsymbol{\Sigma})\bigr)$ and $H(\boldsymbol{\varepsilon}) = \frac{1}{2} \log_2\bigl(\det(2\pi e \sigma^2 \boldsymbol{I}_m)\bigr)$. This implies

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

Since $\boldsymbol{W}\boldsymbol{W}^\top$ and $\boldsymbol{W}^\top\boldsymbol{W}$ have identical non-zero eigenvalues, we get

$$\det(\sigma^{-2} \boldsymbol{W}\boldsymbol{W}^\top + \boldsymbol{I}_m) = \det(\sigma^{-2} \boldsymbol{W}^\top \boldsymbol{W} + \boldsymbol{I}_{112})$$

Table 4: Empirical distribution of the estimated logarithmic key ranks for 100 simulated measurements per value of σ. The key ranks were estimated using the Monte-Carlo heuristic (22) with N random keys.

Noise	**Logarithmic key rank**					**# Random keys**
σ	Min	Q1	Median	Q3	Max	N
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}

and the assertion follows. □

Based on the experiments in Subsubsection 4.3.2, we propose the heuristic formula

$$\mathrm{H}(\boldsymbol{C}, \boldsymbol{D} \mid \boldsymbol{Y}) \approx \max\Big\{2\,,\ 56 - \tfrac{1}{2}\log_2\big(\det(\sigma^{-2}\boldsymbol{W}^\top\boldsymbol{W} + \boldsymbol{I}_{112})\big)\Big\} \tag{26}$$

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

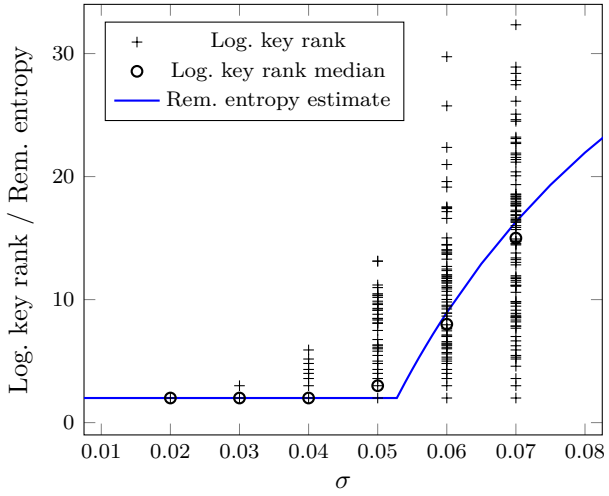

Figure 6: Graph of the remaining entropy estimate (26) as a function of σ together with the logarithmic key ranks computed in the experiments reported in Table 3.

4.5 Isolated consideration of the C- and D-register

In Subsection 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 $\boldsymbol{c}$ in the C-register. Find a set $\mathcal{C}$ of likely candidates for the C-register.

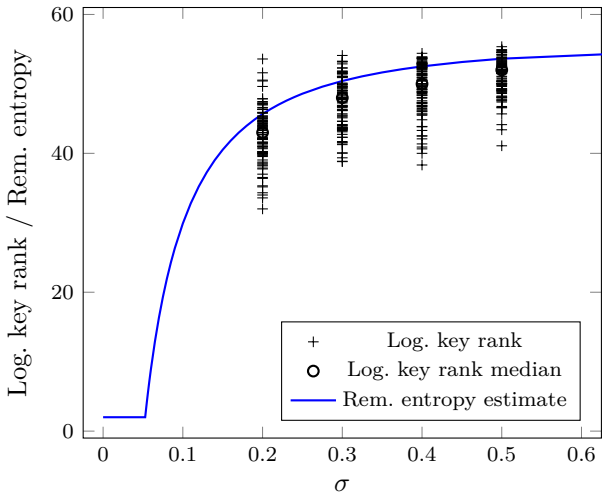

Figure 7: Graph of the remaining entropy estimate (26) as a function of σ together with the logarithmic key ranks estimated in the experiments reported in Table 4.

2. Define an appropriate evaluation function that depends only on the key part $\boldsymbol{d}$ in the D-register. Find a set $\mathcal{D}$ of likely candidates for the D-register.

3. Check all combinations $(\boldsymbol{c}, \boldsymbol{d}) \in \mathcal{C} \times \mathcal{D}$.

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 $\Delta(\boldsymbol{D})$ by a normal distributed random variable $\boldsymbol{N}_2$ with mean $\mathbf{0}_m$ and covariance matrix $\boldsymbol{I}_m$ in $\boldsymbol{Y}$, i.e. we set

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

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

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

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

$$\eta_{\boldsymbol{W},\boldsymbol{y},1} \colon \{0,1\}^{28} \to \mathbb{R}_{\geq 0}\,, \quad \boldsymbol{c} \mapsto \left\| \boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{y} - \boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{W}_1 \Delta(\boldsymbol{c}) \right\|^2. \tag{27}$$

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

$$\begin{aligned} \mathrm{H}(\boldsymbol{C} \mid \boldsymbol{Y}'_1) &\approx \max\Big\{1\,,\ 28 - \tfrac{1}{2} \log_2\big(\det(\boldsymbol{\Sigma}_2^{-1/2} \boldsymbol{W}_1 \boldsymbol{W}_1^\top \boldsymbol{\Sigma}_2^{-1/2} + \boldsymbol{I}_m)\big)\Big\} \\ &= \max\Big\{1\,,\ 28 - \tfrac{1}{2} \log_2\big(\det(\sigma^{-2} \boldsymbol{W} \boldsymbol{W}^\top + \boldsymbol{I}_m) / \det(\sigma^{-2} \boldsymbol{W}_2 \boldsymbol{W}_2^\top + \boldsymbol{I}_m)\big)\Big\}\,. \end{aligned} \tag{28}$$

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

Applying heuristic (28) to the weight matrix $\widetilde{\boldsymbol{W}}$ estimated in Subsubsection 4.3.1 for the real measurements and $\sigma = 0.07$ as estimated in Subsubsection 4.3.2, we obtain $\mathrm{H}(\boldsymbol{C} \mid \boldsymbol{Y}'_1) \approx 6.55$ and $\mathrm{H}(\boldsymbol{D} \mid \boldsymbol{Y}'_2) \approx 16.64$. We computed the ranks of the key parts in the

C- and D-register with respect to $\eta_{\widetilde{\boldsymbol{W}},\overline{\boldsymbol{y}}_i,1}$ and the analogously defined evaluation function $\eta_{\widetilde{\boldsymbol{W}},\overline{\boldsymbol{y}}_i,2}$, 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).

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. Subsubsection 4.3.1).

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

Remark 8. 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 $2^{\mathrm{H}(\boldsymbol{C}|\boldsymbol{Y}'_1)}$ and $2^{\mathrm{H}(\boldsymbol{D}|\boldsymbol{Y}'_2)}$ 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 Subsection 4.1 does not have this drawback.

Acknowledgements

We thank the authors of [HMU+20] for providing us with their measurement data.

References

[Ban93] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296(4):625–636, 1993.

[BEKS17] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of information theory (2. ed.)*. Wiley, 2006.

[FP85] U. Fincke and M. Pohst. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Mathematics of Computation*, 44(170):463–471, 1985.

[GL96] Gene H. Golub and Charles F. Van Loan. *Matrix Computations, Third Edition*. Johns Hopkins University Press, 1996.

[GNR10] Nicolas Gama, Phong Q. Nguyen, and Oded Regev. Lattice enumeration using extreme pruning. In Henri Gilbert, editor, *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*, volume 6110 of *Lecture Notes in Computer Science*, pages 257–278. Springer, 2010.

[HMU+20] Johann Heyszl, Katja Miller, Florian Unterstein, Marc Schink, Alexander Wagner, Horst A. Gieser, Sven Freud, Tobias Damm, Dominik Klein, and Dennis Kügler. Investigating profiled side-channel attacks against the DES key schedule. *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, 2020(3):22–72, 2020.

[HZZW16] Yongbo Hu, Chen Zhang, Yeyang Zheng, and Mathias Wagner. Ciphertext and plaintext leakage reveals the entire TDES key. *IACR Cryptology ePrint Archive*, 2016:1143, 2016.

[Knu19] Donald E. Knuth. *The Art of Computer Programming, Volume 4, Fascicle 5: Mathematical Preliminaries Redux; Introduction to Backtracking; Dancing Links.* Addison-Wesley Professional, 2019.

[LN13] Mingjie Liu and Phong Q. Nguyen. Solving BDD by enumeration: An update. In Ed Dawson, editor, *Topics in Cryptology - CT-RSA 2013 - The Cryptographers' Track at the RSA Conference 2013, San Francisco,CA, USA, February 25-March 1, 2013. Proceedings*, volume 7779 of *Lecture Notes in Computer Science*, pages 293–309. Springer, 2013.

[Nat99] National Institute of Standards and Technology. FIPS Publication 46-3: Data encryption standard (DES), 1999.

[SE94] Claus-Peter Schnorr and M. Euchner. Lattice basis reduction: Improved practical algorithms and solving subset sum problems. *Math. Program.*, 66:181–199, 1994.

[SH95] Claus-Peter Schnorr and Horst Helmut Hörner. Attacking the Chor-Rivest cryptosystem by improved lattice reduction. In Louis C. Guillou and Jean-Jacques Quisquater, editors, *Advances in Cryptology - EUROCRYPT '95, International Conference on the Theory and Application of Cryptographic Techniques, Saint-Malo, France, May 21-25, 1995, Proceeding*, volume 921 of *Lecture Notes in Computer Science*, pages 1–12. Springer, 1995.

[WH17a] Mathias Wagner and Stefan Heyse. Brute-force search strategies for single-trace and few-traces template attacks on the DES round keys of a recent smart card. *IACR Cryptology ePrint Archive*, 2017:614, 2017.

[WH17b] Mathias Wagner and Stefan Heyse. Single-trace template attack on the DES round keys of a recent smart card. *IACR Cryptology ePrint Archive*, 2017:57, 2017.

[WH18] Mathias Wagner and Stefan Heyse. Improved brute-force search strategies for single-trace and few-traces template attacks on the DES round keys. *IACR Cryptology ePrint Archive*, 2018:937, 2018.

[WHZZ16] Mathias Wagner, Yongbo Hu, Chen Zhang, and Yeyang Zheng. Comparative study of various approximations to the covariance matrix in template attacks. *IACR Cryptology ePrint Archive*, 2016:1155, 2016.

A An optimized version of Algorithm 2

The following algorithm is a variation of Algorithm 2 with the optimization described in Remark 7 (a).

Algorithm 3.

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

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

1. [Initialize.] Set $\mathcal{C} \leftarrow \varnothing$, $\ell \leftarrow 3$, $\boldsymbol{k} \leftarrow \mathbf{0}_{56}$, $\boldsymbol{x} \leftarrow \mathbf{1}_{112}$, $\boldsymbol{\rho} \leftarrow \mathbf{0}_{56}$, $\boldsymbol{\sigma} \leftarrow \mathbf{0}_{112,113}$, and $\boldsymbol{v} \leftarrow 112 \cdot \mathbf{1}_{56}$.

2. [Preprocess.] Set $\boldsymbol{W} \leftarrow \boldsymbol{W}\boldsymbol{P}$. Compute an upper triangular matrix $\boldsymbol{R} \in \mathbb{R}^{112\times 112}$ such that $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{R}^\top \boldsymbol{R}$. Set $\boldsymbol{t} \leftarrow (\boldsymbol{W}^\top \boldsymbol{W})^{-1}\boldsymbol{W}^\top \boldsymbol{y}$ and $B \leftarrow B - \|\boldsymbol{y} - \boldsymbol{W}\boldsymbol{t}\|^2$. If $B < 0$, return $\mathcal{C}$ and stop.

3. [Enter level ℓ.] If $\ell = 57$, set $(\boldsymbol{c}, \boldsymbol{d}) \leftarrow \boldsymbol{k}$, set $\mathcal{C} \leftarrow \mathcal{C} \cup \{(\boldsymbol{c}, \boldsymbol{d}), (\overline{\boldsymbol{c}}, \boldsymbol{d}), (\boldsymbol{c}, \overline{\boldsymbol{d}}), (\overline{\boldsymbol{c}}, \overline{\boldsymbol{d}})\}$, 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 $i \leftarrow s_{\ell-1} - 1, s_{\ell-1} - 2, \ldots, s_\ell$, do the following:

a. Set $x_i \leftarrow (\boldsymbol{P}^\top \Delta(\boldsymbol{k}))_i$.

b. For $j \leftarrow v_\ell, v_\ell - 1, \ldots, i$, 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 $\rho_\ell \le B$, 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 $\ell \ge 3$, set $v_\ell \leftarrow s_{\ell-1} - 1$ and go to step **5**. Otherwise return $\mathcal{C}$ and stop.