

Revisiting the Wrong-Key-Randomization Hypothesis

Tomer Ashur, Tim Beyne, and Vincent Rijmen

ESAT/COSIC, KU Leuven and iMinds, Leuven, Belgium
[tomer.ashur,vincent.rijmen] @ esat.kuleuven.be
[tim.beyne] @ student.kuleuven.be

Abstract. Linear cryptanalysis can be considered to be one of the strongest techniques in the cryptanalyst’s arsenal. In most cases, Matsui’s Algorithm 2 is used for the key recovery part of the attack. The success rate analysis of this algorithm is based on an assumption regarding the bias of a linear approximation for a wrong key, known as the wrong-key-randomization hypothesis. This hypothesis was refined by Bogdanov and Tischhauser to take into account the stochastic nature of the bias for a wrong key. We provide further refinements to the analysis of Matsui’s algorithm 2 by considering the more natural setting of sampling without replacement.

This paper derives the distribution for the observed bias for wrong keys when sampling is done without replacement and shows that less data is required when duplicate pairs are discarded. It also develops formulas for the success probability and the required data complexity when this approach is taken. The formulas predict that the success probability may reach a peak, then decrease as more pairs are considered. We provide a new explanation for this behavior and derive the conditions for encountering it. We empirically verify our results and compare them to previous work.

Keywords: linear cryptanalysis, wrong-key-randomization hypothesis, success probability, data complexity

1 Introduction

Linear cryptanalysis can be considered to be one of the most powerful cryptanalytic techniques. Due to its enormous importance, it is standard practice for algorithm designers to show that their new schemes are resistant to it. This is usually done by providing upper bounds for the cryptographic properties exploited by a linear attack, which is used to argue that the success probability of an adversary is negligible.

From the cryptanalyst’s side, much effort was invested in order to improve the attack, either by better understanding it, or by developing possible extensions. Among these extension we can name the zero correlation attack [4,5], extensions using more than a single approximation such as multiple linear cryptanalysis [1] and multidimensional linear cryptanalysis [10], the partition attack [9], and

more. On the front of better understanding the attack we can find research on the linear hull effect [14], various papers suggesting statistical models for analyzing the attack [16,6,2,11], and attempts to quantify the success probability of the various algorithms underlying the attack and the amount of data required.

In his original paper, Matsui estimated that the data complexity should be approximately the squared inverse of the bias of the linear approximation employed in the attack. Selçuk improved this analysis based on the work of Junod and Vaudenay [11] and suggested a complete statistical model, yielding closed formulas for the required data complexity and the success probability. Being able to estimate the success probability of an attack is of great importance to cipher designers as, without it, larger security margins need to be used, in contradiction to the growing trend of using lightweight cryptography.

Estimates of the success probability of linear attacks have traditionally used a simplifying assumption about the behavior of wrong keys. This assumption, commonly known as the *wrong-key-randomization hypothesis*, says that when the adversary tries to determine the right key among a set including some wrong keys, the key-dependent bias is much larger for the right key than for wrong ones. The wrong-key-randomization hypothesis was commonly understood to mean that the bias for a wrong key is exactly zero. However, Bogdanov and Tischhauser noted in [6] that the bias for a wrong key is a random variable with a normal distribution rather than a fixed value. They proposed a corrected wrong key hypothesis, taking this distribution into account, and developed a model for the distribution of the empirical bias. Using this model, they have extended Selçuk’s formula [16] for the success probability of a linear attack.

An interesting consequence of the new formula is that it reaches a maximum for certain parameters. This result was described as counter-intuitive as it implies that increasing the data complexity may sometimes lead to a reduced success probability [6].

1.1 Our Contributions

In this paper we point out the importance of the sampling strategy employed for obtaining plaintext/ciphertext pairs in a linear attack. We argue that the more common case is that plaintext/ciphertext pairs are not used more than once, and thus, the theoretical model should be based on sampling without replacement.

Under this assumption, we redevelop a model for the distribution of the empirical bias. Our result shows that, for sampling without replacement, the correction introduced by Bogdanov and Tischhauser is not necessary and is even undesired.

As a second contribution, we derive a formula for the success probability and the data complexity of a linear attack for sampling without replacement. These formulas confirm the intuitive notion that, for sampling without replacement, the empirical bias converges faster to its real value, which means that the data complexity can be reduced.

We also confirm that the success probability can show non-monotonous behavior as was observed by Tischhauser and Bogdanov [6]. However, their explana-

tion does not explain all of our observations. Hence, we explain the phenomenon anew and derive necessary conditions for its occurrence. The average-case condition, given in Theorem 1, represents a prerequisite for the applicability of Matsui’s algorithm 2.

The paper is organized as follows: Section 2 briefly recalls a few basic notions from probability theory and discusses previous work on the wrong-key-randomization hypothesis. In Section 3 we discuss the behavior of the empirical bias. The influence of the sampling strategy is clarified, and the distribution of the sample bias is derived. Section 4 deals with the non-monotonicity of the success probability. The phenomenon is explained, and the conditions for its occurrence are developed. A discussion of the success probability and the data complexity is provided in section 5. Finally, we verify our results experimentally in Section 6. Section 7 concludes the paper.

2 Preliminaries and Related Work

A random variable is denoted by bold letters *e.g.*, \mathbf{X} , \mathbf{Y} . The expected value of a random variable \mathbf{X} is denoted by $\mathbf{E}[\mathbf{X}]$, and its variance by $\mathbf{Var}[\mathbf{X}]$. By writing $\mathbf{X} \sim \mathcal{N}(\mu, \sigma^2)$, we mean that \mathbf{X} follows a normal distribution with mean μ and variance σ^2 . Similarly, $\mathbf{X} \sim \text{Hypergeometric}(N, M, R)$ means that \mathbf{X} follows a hypergeometric distribution, *i.e.*, \mathbf{X} is a random variable counting the number of occurrences of an item of type I in N draws from a population of size M known to include R such items, where the draws are performed without replacement. The conditional random variable \mathbf{X} given \mathbf{Y} is denoted by $\mathbf{X} | \mathbf{Y}$. This notation carries over to conditional expectations and variances.

During our analysis, it will frequently be convenient to approximate the hypergeometric distribution. Several accurate χ^2 and normal approximations exist, see for example [13]. For our purposes, the following result given by Feller [8] shall suffice. The interested reader may find two proofs of the result by Pinsky in [15].

Lemma 1 ([8]). *Let $\mathbf{X} \sim \text{Hypergeometric}(N, M, pM)$. If $N, M \rightarrow \infty$ in such manner that $N/M \rightarrow t \in (0, 1)$, then \mathbf{X} has asymptotic distribution*

$$\mathbf{X} \sim \mathcal{N}(pN, N(1-t)p(1-p)).$$

2.1 Linear Cryptanalysis

We now describe linear cryptanalysis [12]. Let $f : \mathbb{F}_2^n \rightarrow \mathbb{F}_2^n$ be an n -bit permutation. A linear approximation for f is a pair of masks $(\alpha, \beta) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ such that $\alpha^t f(\mathbf{Y}) = \beta^t \mathbf{Y}$ holds with probability p for \mathbf{Y} uniform over \mathbb{F}_2^n . The bias of the approximation is defined as $\epsilon = p - 1/2$. In the following, the bit-length of the round key of the cipher under attack will be denoted by m .

Without loss of generality, in a key-recovery attack using Matsui’s Algorithm 2, the approximation (α, β) covers $R - 1$ rounds of an R -round cipher used to encrypt N plaintexts. The resulting N ciphertexts are then 1-round decrypted

using 2^m different key guesses, and the linear approximation is evaluated against the resulting pairs. For each key k_i , the adversary keeps a counter \mathbf{T}_i counting the number of times a pair satisfies the linear approximation. Once enough data have been observed, the adversary calculates the empirical bias $\hat{\epsilon}_i = \mathbf{T}_i/N - 1/2$. The biases are sorted in descending order according to their magnitude, and their respective keys are tried in this order. If the bias corresponding to the right key is ranked among the highest 2^{m-a} biases, the attacker is said to obtain an advantage a over brute-force. Throughout our paper, the subscript zero refers to the right key. In particular, the counter and bias for the correct key are denoted respectively by \mathbf{T}_0 and ϵ_0 .

2.2 The Wrong-Key-Randomization Hypothesis

The success rate analysis performed by Selçuk [16] uses order statistics to investigate the probability that the right key is among the 2^{m-a} top ranked keys. The main underlying assumption of this analysis is that the real bias for a wrong key is zero, and thus, that the sample bias would have a normal distribution centered around zero. This assumption may be summarized in the following hypothesis:

Hypothesis 1 (Simple wrong-key-randomization hypothesis). *The bias for a wrong key equals zero:*

$$\epsilon_w = 0.$$

If Hypothesis 1 is true, we have the following lemma.

Lemma 2. *Let $\hat{\epsilon}_w$ be the empirical bias obtained from a counter associated with a wrong key using N pairs of plaintexts and ciphertexts. Assuming Hypothesis 1 is true and sampling is performed with replacement, we have approximately*

$$\hat{\epsilon}_w \sim \mathcal{N}\left(0, \frac{1}{4N}\right).$$

However, Bogdanov and Tischhauser noted in [6] that, in accordance with Daemen and Rijmen [7], the underlying bias of a random linear approximation is not necessarily zero but a random variable. This resulted in the following extension of Hypothesis 1:

Hypothesis 2 (Bogdanov and Tischhauser [6]). *The bias ϵ_w for a wrong key is distributed as for a random permutation, i.e.*

$$\epsilon_w \sim \mathcal{N}\left(0, 2^{-n-2}\right).$$

This hypothesis requires the usage of a compound model for the empirical bias, which takes into account the distribution of the wrong bias. This leads to the following statement about the distribution of the sample bias for wrong keys.

Lemma 3 ([6], Lemma 1). *Let $\hat{\epsilon}_w$ be defined as before, then assuming the validity of Hypothesis 2, we have approximately*

$$\hat{\epsilon}_w \sim \mathcal{N}\left(0, \frac{1}{4} \cdot \left(\frac{1}{N} - \frac{1}{2^n}\right)\right).$$

Selçuk gives the success probability of a linear attack as

$$P_S(N) = \Phi\left(2\sqrt{N}|\epsilon_0| - \Phi^{-1}(1 - 2^{-a-1})\right), \quad (1)$$

which holds under Hypothesis 1. However, as was noted by Bogdanov and Tishchauer, the bias for wrong keys has a normal distribution centered around zero, in accordance with Hypothesis 2. Using the distribution of Lemma 3, they extend Selçuk's formula as follows:

$$P_S(N) = \Phi\left(2\sqrt{N}|\epsilon_0| - \sqrt{1 + \frac{N}{2^n}}\Phi^{-1}(1 - 2^{-a-1})\right). \quad (2)$$

An experimental verification of the above formula is provided in [3].

3 Sample Bias for Wrong Keys

This section deals with the distribution of the sample bias for wrong keys. The details of this distribution are relevant for the construction and the analysis of statistical procedures that attempt to distinguish the right key from wrong keys.

As mentioned in the related work section, the distribution of $\hat{\epsilon}_w$ must be described by a compound model whenever Hypothesis 2 is used. This is clarified in Section 3.1.

Generally speaking, the sample bias can be fully described given the distribution of the bias for the wrong keys and a *sampling strategy*. The former is completely determined by the choice of Hypothesis 1 or 2 and requires no further discussion. The latter will be discussed in Section 3.2.

Finally, the main result of this section is presented in Lemma 4, which approximates the distribution of the sample bias for a random wrong key in the case of sampling without replacement. The resulting distribution turns out to be the same as the one given in Lemma 2.

3.1 Compound model

It is important to distinguish between two different random variables: the sample bias for a specific wrong key, and the sample bias for a uniformly selected random wrong key. We shall refer to the former by $\hat{\epsilon}_w | \epsilon_w$ and the latter will be written as $\hat{\epsilon}_w$. The idea of a compound model is that a parameter in the distribution of a random variable is randomized, *i.e.* is itself a random variable. The probability density of the compound variable $\hat{\epsilon}_w$ is given by the probability density of $\hat{\epsilon}_w$ given $\epsilon_w = \varepsilon$, weighted by the probability that $\epsilon_w = \varepsilon$, for any possible ε that ϵ_w can take. Formally, if $f_{\hat{\epsilon}_w | \epsilon_w}$ is the probability density function of $\hat{\epsilon}_w | \epsilon_w$, and f_{ϵ_w} likewise for ϵ_w , then we may write

$$f_{\hat{\epsilon}_w}(\epsilon_w) = \int_{\varepsilon} f_{\hat{\epsilon}_w | \epsilon_w}(\epsilon_w, \varepsilon) f_{\epsilon_w}(\varepsilon) d\varepsilon,$$

for the density of $\hat{\epsilon}_w$. This is depicted in Figure 1.

The behavior of the random variable $\hat{\epsilon}_w \mid \epsilon_w$ is completely determined by the sampling strategy. For example, for sampling without replacement, $\mathbf{T}_w \mid \epsilon_w$ follows a hypergeometric distribution centered around $N(\frac{1}{2} + \epsilon_w)$.

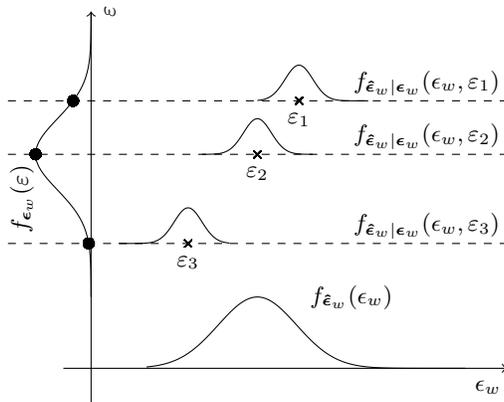


Fig. 1: The curve along the vertical axis represents the density function of ϵ_w . The probability density function of $\hat{\epsilon}_w$ is shown at the bottom. $\hat{\epsilon}_w$ has a compound distribution obtained by weighed integration over the smaller curves which represent the sample biases for specific keys.

3.2 Sampling strategies

The way plaintext/ciphertext pairs are obtained in linear cryptanalysis corresponds to sampling from a population of size 2^n . For each sample, a trial is conducted: it is checked whether or not a fixed linear approximation holds. Recall that the sum of the outcomes (zero or one) of these trials is stored in a counter \mathbf{T}_i .

One could conceive of many strategies for sampling, but here we will limit the discussion to two common cases:

1. Sampling with replacement, trials are independent and $N > 2^n$ is possible.
2. Sampling without replacement. Trials are not independent and $N \leq 2^n$.

The existing analyses that we are aware of [2,6,16] implicitly start from the assumption that the first strategy is used. This leads to a binomial distribution for the counters.

We now argue that the second strategy is preferable. This will lead to a hypergeometric distribution for the counters. Since for a given key, a specific plaintext

always results in the same ciphertext, duplicates provide no new information for the estimation of the real bias. Moreover, increasing the data complexity beyond what is needed for the attack reduces its efficiency, and may render it worse than exhaustive search in some cases. Hence, whenever possible, an adversary would prefer sampling without replacement.

3.3 Sampling without Replacement

Given that duplicate draws provide no additional information to the cryptanalyst, we argue that sampling without replacement is a more practical strategy.

Assume then that N distinct plaintext/ciphertext pairs are sampled at random from the total population of 2^n pairs. The counter for a specific wrong key follows a hypergeometric distribution

$$\mathbf{T}_w \mid \mathbf{R} \sim \text{Hypergeometric}(N, 2^n, R),$$

where $R = 2^n(\epsilon_w + 1/2)$ equals the amount of plaintext/ciphertext pairs in the population for which the linear approximation holds. Given this starting point, the proof of the next lemma derives the distribution of the sample bias for a random wrong key.

Lemma 4 (Lemma 2, *stet.*). *Under Hypothesis 2, and for sampling without replacement, we have for the sample bias $\hat{\epsilon}_w$ of a random wrong key*

$$\hat{\epsilon}_w \sim \mathcal{N}\left(0, \frac{1}{4N}\right),$$

approximately.

Proof. By Lemma 1, we have asymptotically

$$\mathbf{T}_w \mid \epsilon_w \sim \mathcal{N}\left(N\left(\frac{1}{2} + \epsilon_w\right), N\left(1 - \frac{N}{2^n}\right)\left(\frac{1}{4} - \epsilon_w^2\right)\right).$$

Since ϵ_w^2 is small, we have approximately

$$\mathbf{T}_w \mid \epsilon_w \sim \mathcal{N}\left(N\left(\frac{1}{2} + \epsilon_w\right), \frac{N}{4}\left(1 - \frac{N}{2^n}\right)\right).$$

It follows that for the sample bias

$$\hat{\epsilon}_w \mid \epsilon_w \sim \mathcal{N}\left(\epsilon_w, \frac{1}{4N}\left(1 - \frac{N}{2^n}\right)\right).$$

A compound normal distribution with normally distributed mean is again normal. That is, if $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma_1^2)$ with $\boldsymbol{\mu} \sim \mathcal{N}(\mu, \sigma_2^2)$, then $\mathbf{X} \sim \mathcal{N}(\mu, \sigma_1^2 + \sigma_2^2)$.¹ In this particular case we obtain

$$\begin{aligned} \mathbf{E}[\hat{\epsilon}_w] &= 0 \\ \mathbf{Var}[\hat{\epsilon}_w] &= \frac{1}{4N}\left(1 - \frac{N}{2^n}\right) + \frac{1}{2^{n+2}} = \frac{1}{4N}. \end{aligned}$$

□

¹ For a proof, note that $\mathbf{X} - \boldsymbol{\mu} \sim \mathcal{N}(0, \sigma_1^2)$ is independent of $\boldsymbol{\mu}$.

Observation. The preceding lemma shows that $\hat{\epsilon}_w$ for sampling without replacement has approximately the same distribution as given by Lemma 2. In other words, the correction provided in Lemma 3 should *not* be taken into account for sampling without replacement. Note, however, that the result is based on Hypothesis 2 rather than the simpler Hypothesis 1.

The mean and variance of \mathbf{T}_w can also be computed directly, without using any of the approximations used in Lemma 4. The first two central moments of \mathbf{R} are given by

$$\begin{aligned}\mathbf{E}[\mathbf{R}] &= 2^{n-1} \\ \mathbf{Var}[\mathbf{R}] &= 2^{2n} \mathbf{Var}[\hat{\epsilon}_w + 1/2] = 2^{n-2}.\end{aligned}$$

Hence, the expected value of \mathbf{T}_w is given by

$$\mathbf{E}[\mathbf{T}_w] = \mathbf{E}[\mathbf{E}[\mathbf{T}_w \mid \mathbf{R}]] = \mathbf{E}[N\mathbf{R}2^{-n}] = N/2.$$

For the variance of T_w we have

$$\begin{aligned}\mathbf{Var}[\mathbf{T}_w] &= \mathbf{E}[\mathbf{Var}[\mathbf{T}_w \mid \mathbf{R}]] + \mathbf{Var}[\mathbf{E}[\mathbf{T}_w \mid \mathbf{R}]] \\ &= \mathbf{E}\left[N\frac{\mathbf{R}}{2^n} \frac{2^n - \mathbf{R}}{2^n} \frac{2^n - N}{2^n - 1}\right] + \mathbf{Var}\left[N\frac{\mathbf{R}}{2^n}\right] \\ &= \frac{N(2^n - N)}{2^{2n}(2^n - 1)} \mathbf{E}[\mathbf{R}(2^n - \mathbf{R})] + \mathbf{Var}\left[N\frac{\mathbf{R}}{2^n}\right] \\ &= \frac{N(2^n - N)}{2^{2n}(2^n - 1)} (2^n \mathbf{E}[\mathbf{R}] - \mathbf{E}[\mathbf{R}^2]) + \frac{N^2}{2^{2n}} \mathbf{Var}[R] \\ &= \frac{N(2^n - N)}{2^{2n}(2^n - 1)} (2^n \mathbf{E}[\mathbf{R}] - \mathbf{Var}[\mathbf{R}] - \mathbf{E}[\mathbf{R}]^2) + \frac{N^2}{2^{2n}} \mathbf{Var}[\mathbf{R}] \\ &= \frac{N(2^n - N)}{2^{2n}(2^n - 1)} (2^{2n-2} - 2^{n-2}) + \frac{N^2}{2^{n+2}}.\end{aligned}$$

If n is sufficiently large, it is reasonable to assume that $2^{2n-2} - 2^{n-2} \approx 2^{2n-2}$. This gives

$$\mathbf{Var}[\mathbf{T}_w] \approx \frac{N(2^n - N)}{2^{n+2} - 1/4} + \frac{N^2}{2^{n+2}} \approx \frac{N}{4}.$$

Assuming that the distribution of \mathbf{T}_w can be approximated using a normal distribution, we also obtain Lemma 4.

4 Non-monotonicity of the Success Probability

For sampling with replacement, Selçuk estimates the success probability as given by (1). Bogdanov and Tischhauser have extended this result to (2). Due to the fact that the biases for the wrong keys are selected at random from $\mathcal{N}(0, 2^{-n-2})$,

the success probability $P_S(N)$ is in fact a random variable. Hence, the probability given by (2) is an estimate for the average success probability. Despite this complication, we shall continue to use the notation $P_S(N)$ in the sense that $P_S(N) = \mathbf{E}[\mathbf{P}_S(N)]$.

We give two results related to two aspects of this random variable:

- According to (2), $P_S(N)$ can be non-monotonous. This effect was also observed in [6]. We provide an explanation for this phenomenon, as well as necessary conditions for its occurrence.
- In Section 5, we derive a formula for the average success probability in the case of sampling without replacement.

4.1 Explanation of Non-monotonicity

Bogdanov and Tischhauser have observed that, in some cases, the success probability shows a maximum [6]. They attribute this effect to the fact that, as N increases, the number of duplicate samples increases which amplifies the “noise” due to the random distribution of the biases for the wrong keys. In the present section we propose a different explanation for the non-monotonicity and show that the phenomenon is not counter-intuitive. Note that a maximum can also be observed for sampling without replacement, hence without duplicates, which is difficult to reconcile with the explanation given by [6]. The discussion below and the conditions given in the next section are independent of the sampling strategy.

When the bias ϵ_0 of the right key is close to zero, there is a non-negligible probability that some of the wrong keys have a higher absolute bias than $|\epsilon_0|$. This is depicted in Figure 2. In this case, the correct key should *not* be expected to be ranked higher than (some of the) wrong keys. As N increases, the accuracy of the ranking increases because the variances of all sample biases decrease. It follows that, if there are wrong keys with absolute bias higher than $|\epsilon_0|$, then for large N those will be ranked higher than the right key. If there are more such keys than the attacker advantage allows, *i.e.* more than 2^{m-a} , then this causes a decreasing success probability. In this case, if $N \rightarrow \infty$ (or $N = 2^n$ without replacement) then also $\mathbf{P}_S(N) \rightarrow 0$, almost surely. In other words: given all possible information, the attack will always fail.

From the discussion above, we may conclude that non-monotonous behavior indicates that the attack can not be conducted using Matsui’s algorithm 2. In fact, a correct identification of the right key amounts to a false positive. This is formalized in the next section, by giving a bound on the required bias for given values of 2^n and a . This bound hence also expresses a prerequisite for Matsui’s algorithm 2.

4.2 Conditions for Non-monotonicity

This subsection derives necessary conditions for non-monotonous behavior of the success probability. These conditions are necessarily probabilistic, since they are

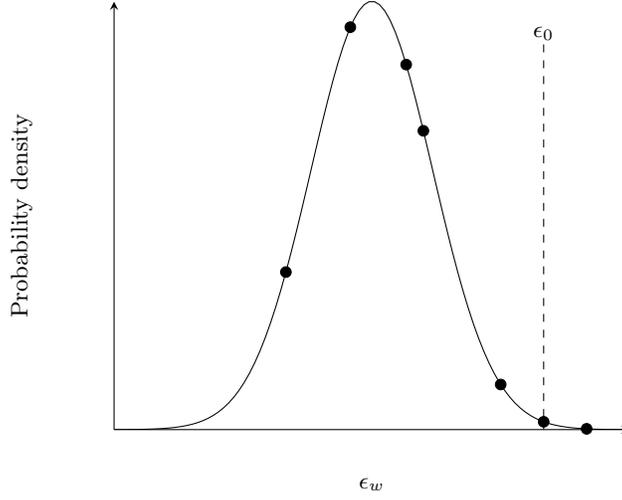


Fig. 2: The biases for a few keys are indicated by dots, the dashed line represents the bias for the right key. One of the wrong keys has a larger bias than the right key.

determined by the biases of individual wrong keys. Hence, it can be expected that for some values of a, n, m and ϵ_0 , $\mathbf{P}_S(N)$ is non-monotonous only for *some* keys. The main result is Theorem 1, which gives a necessary condition for average non-monotonicity and hence for the applicability of Matsui's algorithm 2.

The following lemma gives a first result on the probability of monotonicity. It is independent of the sampling strategy.

Lemma 5. *The probability that $\mathbf{P}_S(N)$ is a monotonous function is given by*

$$\Pr[\mathbf{P}_S(N) \text{ is monotonous}] = \Phi\left(\frac{(2^{-a} - 2^{-m-1} - p) 2^{m/2}}{\sqrt{p(1-p)}}\right),$$

where

$$p = 2\left(1 - \Phi\left(|\epsilon_0| 2^{n/2+1}\right)\right).$$

Proof. The probability that a random permutation has absolute bias larger than $|\epsilon_0|$ can be computed as

$$\Pr[|\epsilon_w| \geq |\epsilon_0|] = 2(1 - \Pr[\epsilon_w < |\epsilon_0|]) = 2\left(1 - \Phi\left(|\epsilon_0| 2^{n/2+1}\right)\right),$$

since $\epsilon_w \sim \mathcal{N}(0, 2^{-n-2})$. For a decreasing success probability, we require at least 2^{m-a} keys with bias larger than $|\epsilon_0|$. Let \mathbf{C} be the random variable describing the amount of such keys, then \mathbf{C} is approximately binomially distributed. Furthermore, if the amount of keys 2^m is sufficiently large, \mathbf{C} can be approximated

with a normal distribution:

$$\mathbf{C} \sim \mathcal{N}(p2^m, p(1-p)2^m) \text{ where } p = 2 \left(1 - \Phi \left(|\epsilon_0| 2^{n/2+1} \right) \right).$$

The probability that $P_S(N)$ is monotonous for some $|\epsilon_0|$, m and a can hence be computed as

$$\Pr[\mathbf{P}_S(N) \text{ is monotonous}] = \Pr[\mathbf{C} < 2^{m-a}].$$

Using the normal approximation of \mathbf{C} , we get:

$$\begin{aligned} \Pr[\mathbf{P}_S(N) \text{ is monotonous}] &= \Phi \left(\frac{2^{m-a} - 2^{-1} - p2^m}{\sqrt{p(1-p)2^m}} \right) \\ &= \Phi \left(\frac{(2^{-a} - 2^{-m-1} - p)2^{m/2}}{\sqrt{p(1-p)}} \right). \end{aligned}$$

□

Observe that, as the bias ϵ_0 approaches 0, the success probability is almost surely non-monotonous. If $\mathbf{E}[\mathbf{C}] \geq 2^{m-a}$, then $\mathbf{P}_S(N)$ is non-monotonous on average, *i.e.* over all keys and all *distinct* attacks. This condition can be used to derive the following theorem.

Theorem 1 (Prerequisite of Matsui's algorithm 2). *The success probability is monotonous on average if and only if*

$$|\epsilon_0| > 2^{-n/2-1} \Phi^{-1}(1 - 2^{-a-1}).$$

Proof. The condition $\mathbf{E}[\mathbf{C}] \geq 2^{m-a}$ corresponds to the inequality

$$2 \left(1 - \Phi \left(|\epsilon_0| 2^{n/2+1} \right) \right) 2^m \geq 2^{m-a},$$

which can be rewritten as

$$\begin{aligned} 1 - \Phi \left(|\epsilon_0| 2^{n/2+1} \right) &\geq 2^{-a-1} \\ \iff |\epsilon_0| &\leq 2^{-n/2-1} \Phi^{-1}(1 - 2^{-a-1}). \end{aligned}$$

□

Theorem 1 expresses a necessary condition for a nonzero success probability in the average case as $N \rightarrow \infty$ (or $N = 2^n$ without replacement). Hence, when the condition does not hold, the advantage a can only be obtained as a false-positive during key recovery with Matsui's algorithm 2.

It can be verified that the condition of Theorem 1 ensures that the maximum of $P_S(N)$ as defined by equation (2) corresponds to a positive value of N .

Bogdanov and Tischhauser have observed non-monotonous behavior with $|\epsilon_0| = 2^{-10}$, $a = 12$ and $n = 20$ [6]. Theorem 1 gives:

$$p = 2 \left(1 - \Phi(2^{-10} \cdot 2^{11}) \right) \approx 0.0455.$$

and

$$|\epsilon_0| \leq 2^{-11} \Phi^{-1}(1 - 2^{-13}) \approx 2^{-9.125}.$$

Hence, with these parameters, the average attack setup will lead to non-monotonous behavior. By Lemma 5, the probability of monotonicity is $\Phi(-218.78) \approx 0$.

5 Average Success Probability and Data Complexity

To compute the average success probability, we will make the approximation that the non-identically distributed sample biases for wrong keys can be replaced by an equal amount of independent and identically distributed random variables with distribution given by Lemma 4. A similar approximation was also implicitly made in [6] and greatly simplifies the distribution of the order statistics.

The derivation of $P_S(N)$ is similar to that of Selçuk [16], with the important difference that the counter for the right key is distributed as

$$\mathbf{T}_0 \sim \text{Hypergeometric} \left(N, 2^n, \left(\epsilon_0 + \frac{1}{2} \right) 2^n \right). \quad (3)$$

The corresponding distribution function of $\hat{\epsilon}_0$ will be denoted by $F_{\hat{\epsilon}_0}$ and can be written in terms of the distribution function $F_{\mathbf{T}_0}$ of \mathbf{T}_0 :

$$F_{\hat{\epsilon}_0}(\varepsilon) = F_{\mathbf{T}_0} \left(\frac{N}{2} + N\varepsilon \right).$$

Following Selçuk, without loss of generality, we only consider the case $\epsilon_0 \geq 0$. The discussion for $\epsilon_0 < 0$ is completely analogous. It will be assumed that the distribution of an order statistic of the sample biases $\hat{\epsilon}_w$ for wrong keys is approximately degenerate relative to that of the right key — Selçuk makes the same approximation in his discussion. The mean of the $(2^m - 2^{m-a})$ th order statistic ζ is approximately given by $\mathbf{E}[\zeta] = \Phi^{-1}(1 - 2^{-a-1}) / (2\sqrt{N})$ [16]. Noting that $\Pr[\hat{\epsilon}_0 < 0] \approx 0$, we have for the average success probability

$$\begin{aligned} P_S(N) &= \Pr[\hat{\epsilon}_0 - \zeta > 0] \\ &\approx \Pr[\hat{\epsilon}_0 > \mathbf{E}[\zeta]] \\ &= 1 - F_{\hat{\epsilon}_0} \left(\frac{\Phi^{-1}(1 - 2^{-a-1})}{2\sqrt{N}} \right) \\ &= 1 - F_{\mathbf{T}_0} \left(\underbrace{\frac{N}{2} + \frac{\sqrt{N}}{2} \Phi^{-1}(1 - 2^{-a-1})}_{k(N)} \right). \end{aligned}$$

An accurate approximation of $F_{\mathbf{T}_0}$ can be obtained by using a normal approximation with respect to N . Indeed, by applying Lemma 1 to \mathbf{T}_0 , one obtains the approximation (assuming $\epsilon_0^2 \approx 0$)

$$\mathbf{T}_0 \sim \mathcal{N} \left(N \left(\frac{1}{2} + |\epsilon_0| \right), \left(1 - \frac{N}{2^n} \right) \frac{N}{4} \right),$$

which is accurate if N and 2^n are sufficiently large. It can be verified that the above expression also holds for $\epsilon_0 < 0$. In terms of the standard normal distribution, we have

$$\begin{aligned} F_{\mathbf{T}_0}(k(N)) &\approx \Phi\left(\frac{k(N) - N\left(\frac{1}{2} + |\epsilon_0|\right)}{\sqrt{\frac{N}{4}\left(1 - \frac{N}{2^n}\right)}}\right) \\ &= \Phi\left(\frac{\sqrt{N}\Phi^{-1}(1 - 2^{-a-1})/2 - N|\epsilon_0|}{\sqrt{\frac{N}{4}\left(1 - \frac{N}{2^n}\right)}}\right) \\ &= \Phi\left(\frac{\Phi^{-1}(1 - 2^{-a-1}) - 2\sqrt{N}|\epsilon_0|}{\sqrt{1 - \frac{N}{2^n}}}\right). \end{aligned}$$

By symmetry, we then obtain the simple result of the theorem below.

Theorem 2. *Assume Hypothesis 2 holds. Let $P_S(N)$ denote the average success probability of a linear attack on a block cipher given N distinct known plaintext/ciphertext pairs. If the linear approximation has bias ϵ_0 for the right key, and the desired advantage is a , then we have*

$$P_S(N) = \Phi\left(\frac{2\sqrt{N}|\epsilon_0| - \Phi^{-1}(1 - 2^{-a-1})}{\sqrt{1 - \frac{N}{2^n}}}\right),$$

for sampling without replacement.

Theorem 2 directly leads to an expression for the data complexity, which is given below. Note that when $P_S(N)$ is non-monotonous, a value of the success probability P_S will in general correspond to two data complexities N . For simplicity, and because non-monotonicity is a prerequisite for practical usage of Matsui's algorithm 2, we only deal with the monotonous case.

Corollary 1. *Under the same conditions as Theorem 2, and when the condition given by Theorem 1 is satisfied, the number of plaintext/ciphertext pairs required to obtain an average success probability P_S is*

$$N = \left(\frac{2|\epsilon_0|\alpha \pm \sqrt{(2\epsilon_0\alpha)^2 - (\alpha^2 - \beta^2)(2^{-n}\beta^2 + 4\epsilon_0^2)}}{4\epsilon_0^2 + 2^{-n}\beta^2}\right)^2,$$

where $\alpha = \Phi^{-1}(1 - 2^{-a-1})$ and $\beta = \Phi^{-1}(P_S)$. The plus sign applies whenever $P_S \geq 1/2$, otherwise the minus sign applies. For $|\epsilon_0| \gg 2^{-n/2-1}\Phi^{-1}(P_S)$, this simplifies to

$$N = \left(\frac{\Phi^{-1}(1 - 2^{-a-1}) + \Phi^{-1}(P_S)}{2\epsilon_0}\right)^2.$$

Proof. The result is obtained by solving the equation

$$\Phi^{-1}(P_S)\sqrt{1 - \frac{N}{2^n}} = 2\sqrt{N}|\epsilon_0| - \Phi^{-1}(1 - 2^{-a-1}).$$

A trite calculation (*cf.* Appendix A) yields

$$(2^{-n}\beta^2 + 4\epsilon_0^2)N - 4|\epsilon_0|\alpha\sqrt{N} + \alpha^2 - \beta^2 = 0,$$

which can be solved to obtain the result. \square

Note that the approximation for large $|\epsilon_0|$ gives the same data complexity as Selçuk [16]. This is due to the fact that large biases require fewer plain-text/ciphertext pairs, and for very small N the difference between sampling with and without replacement is negligible.

In general, the data complexity for sampling without replacement is lower. This is a consequence of the fact that no duplicates are used. Bogdanov and Tischhauser provide an algorithm to compute the data complexity for a given success probability [6].² Here, the following equivalent closed-form formula for the monotonous case will be used instead:

$$N = \left(\frac{2|\epsilon_0|\beta + \sqrt{(2\epsilon_0\beta)^2 - (\alpha^2 - \beta^2)(2^{-n}\alpha^2 - 4\epsilon_0^2)}}{4\epsilon_0^2 - 2^{-n}\alpha^2} \right)^2,$$

where $\alpha = \Phi^{-1}(1 - 2^{-a-1})$ and $\beta = \Phi^{-1}(P_S)$.

Figure 3 shows the data complexity for a large bias and for a small bias close to the bound of Theorem 1. For $|\epsilon_0| = 2^{-14}$ the difference between the data complexities is relatively small. For instance, at a success probability of 95%, the data complexity is about 14% higher for sampling with replacement. The difference with sampling without replacement is much more significant for small values of the bias. In this case, for a success probability of 95%, the data complexity is 69% higher for sampling with replacement. Note that, due to duplicates, the data complexity for sampling with replacement can exceed the size of the codebook, but not that of the key.

² In the non-monotonous case, their algorithm returns the lowest data complexity corresponding to the given success probability.

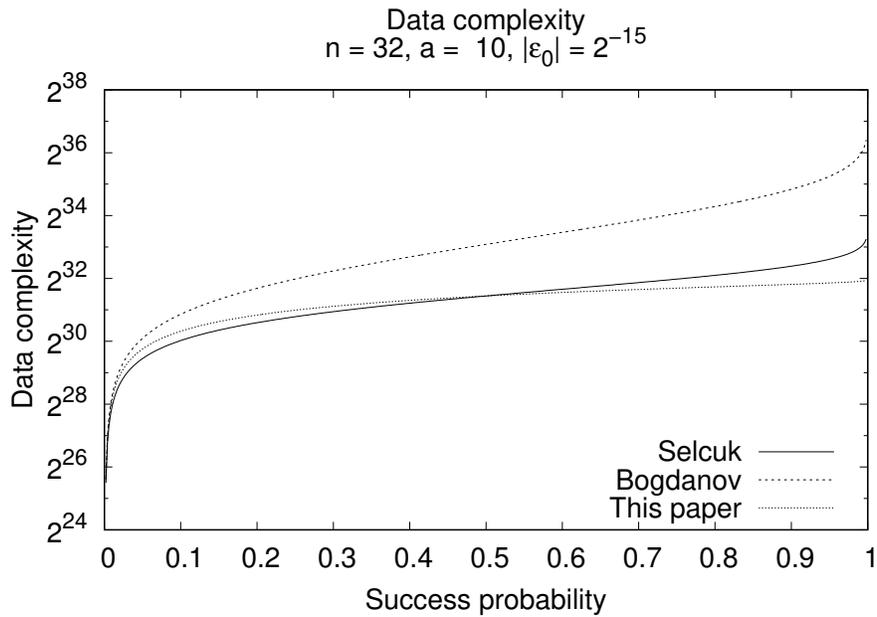
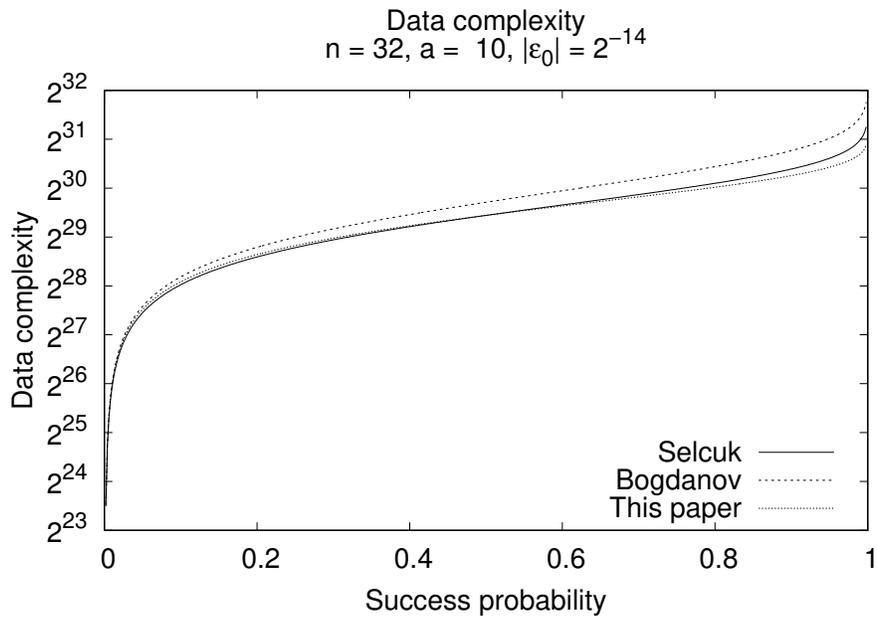


Fig. 3: The theoretical data complexity for a given success probability. The top figure on corresponds to a relatively large bias compared to the bias in the bottom figure.

For completeness and comparison with [6], we also compute the maximum of $P_S(N)$.

Corollary 2. *Under the conditions of Theorem 2 and Theorem 1, the success probability attains a maximum at*

$$\arg \max_N P_S(N) = \left(\frac{2^{n+1}|\epsilon_0|}{\Phi^{-1}(1 - 2^{-a-1})} \right)^2.$$

Proof. Maximizing $P_S(N)$ amounts to solving

$$\frac{d}{dN} \left(\frac{2\sqrt{N}|\epsilon_0| - \Phi^{-1}(1 - 2^{-a-1})}{\sqrt{1 - \frac{N}{2^n}}} \right) = 0.$$

A trite calculation (*cf.* Appendix B) shows that the solution is

$$N = \left(\frac{2^{n+1}|\epsilon_0|}{\Phi^{-1}(1 - 2^{-a-1})} \right)^2.$$

Finally, note that the condition $N \leq 2^n$ corresponds exactly to the condition given by Theorem 1. \square

6 Experimental Verification

Bogdanov *et al.* have conducted a series of large-scale experiments to verify their model for the success probability in the monotonous case [3]. Hence, Hypothesis 2 has already been verified implicitly. We use this to simplify the verification of our theoretical results by using simulations that are based on Hypothesis 2.

Since our model for the success probability relies heavily on the validity of the distribution of the sample bias $\hat{\epsilon}_w$ for wrong keys as given by Lemma 4, we verify the accuracy of this claim in Section 6.1

In Section 6.2, we provide simulation data for the success probability and compare the results with Theorem 2.

Finally, our explanation for the non-monotonicity of the success probability given in Section 5 suggests that the probabilistic nature of the phenomenon becomes relevant when the bias is close to the bound of Theorem 1. We verify this for parameters such that the success probability is monotonous with probability 1/2.

6.1 Sample bias

In this section we verify the proposition that, for sampling without replacement, the distribution of the sample bias is given by Lemma 4. Our experiments do not attempt to evaluate the validity of Hypothesis 2. We therefore propose the following simulation procedure:

1. Sample 2^{13} biases from $\mathcal{N}(0, 2^{-n-2})$ and, for each such bias ϵ_i , compute the corresponding number of plaintext/ciphertext pairs for which the approximation holds, *i.e.* $R = 2^n(1/2 + \epsilon_i)$.
2. For each ϵ_i , sample N values from the corresponding population without replacement. Keep a counter T_i for the number of successful approximations. For this purpose, the simulation performs a series of dependent Bernoulli trials, *i.e.* after N samples, increasing the counter has probability $(R - T_i)/(2^n - N)$.

With respect to Lemma 4, two aspects must be verified: Firstly, the normality of $\hat{\epsilon}_w$ and secondly the variance, which we claim is $1/(4N)$ such that the additional term 2^{-n-2} from Lemma 2 only applies to sampling with replacement. The normal probability plots in Figure 4 allow for a quick graphical check of both aspects.

From Figure 4, we see that the empirical quantiles of the observed sample biases are a linear function of the quantiles of the standard normal distribution. This indicates that the distribution of the sample bias $\hat{\epsilon}_w$ for wrong keys is approximately normal. For both plots, the slope of a least-squares fit is close to $1/(2\sqrt{N})$, the standard deviation of $\hat{\epsilon}_w$. In Figure 4, we have drawn another straight line with slope $\frac{1}{2}\sqrt{1/N + 1/2^n}$ to clarify the difference between Lemma 2 and Lemma 4.

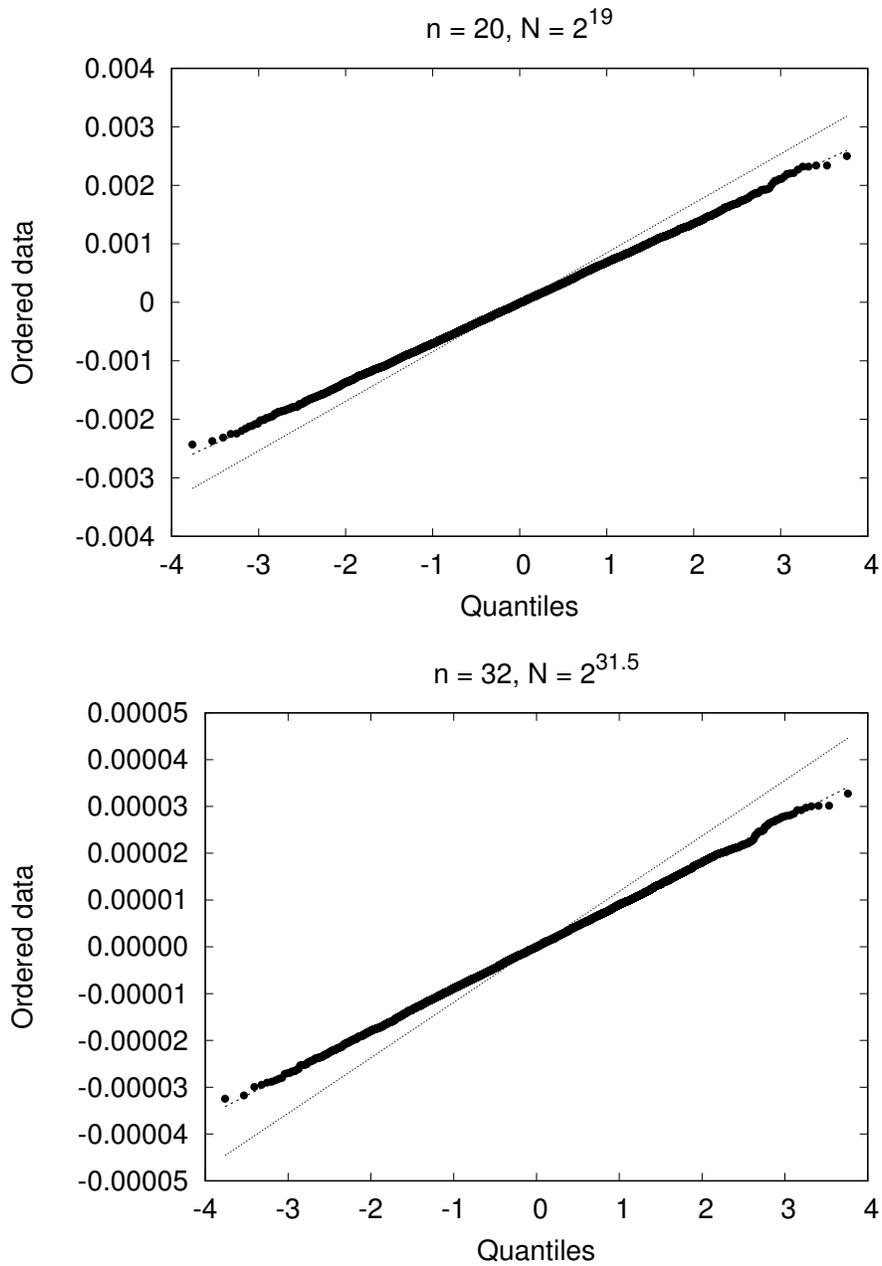


Fig. 4: Standard normal probability plots of the sample bias for sampling without replacement. Each plot contains 2^{13} data points. The straight line matching the data corresponds to the distribution for the sample bias as given by Lemma 4. The other straight line represents the result of Lemma 3.

6.2 Success probability

In this subsection, we verify the formula for the success probability given by Theorem 2. We simulate the attack and compare our formula with the result of the simulation. This allows a simple verification of Theorem 2 independent of the accuracy of the assumptions on the sample bias. An attack is simulated as follows:

1. Generate a list of $2^m - 1$ zero-initialized counters, each corresponding to a key with associated bias sampled from $\mathcal{N}(0, 2^{-n-2})$ in accordance with Hypothesis 2. An additional counter is kept for the right key, which has bias ϵ_0 .
2. For several values of N , repeat the following steps 2^8 times to estimate the proportion of successes:
 - (a) The value of each counter with corresponding bias ϵ is sampled directly from

$$\mathcal{N}\left(N\left(\epsilon + \frac{1}{2}\right), \frac{N^2}{4}\left(1 - \frac{N}{2^n}\right)\right),$$

- (b) Check whether the absolute bias corresponding to the right key is among the first 2^{m-a} entries of the list of absolute biases.

To capture the stochastic nature of the success probability, we repeat the above procedure 400 times.

The second step of the procedure can be justified by the results of the previous section, where it was shown that the distribution of the sample bias provided by Lemma 4 is sufficiently accurate. That is, the distribution from which the counters are sampled was the starting point for the proof of Lemma 4. The correspondence between the simulation and our theoretical results is shown in Figure 5.

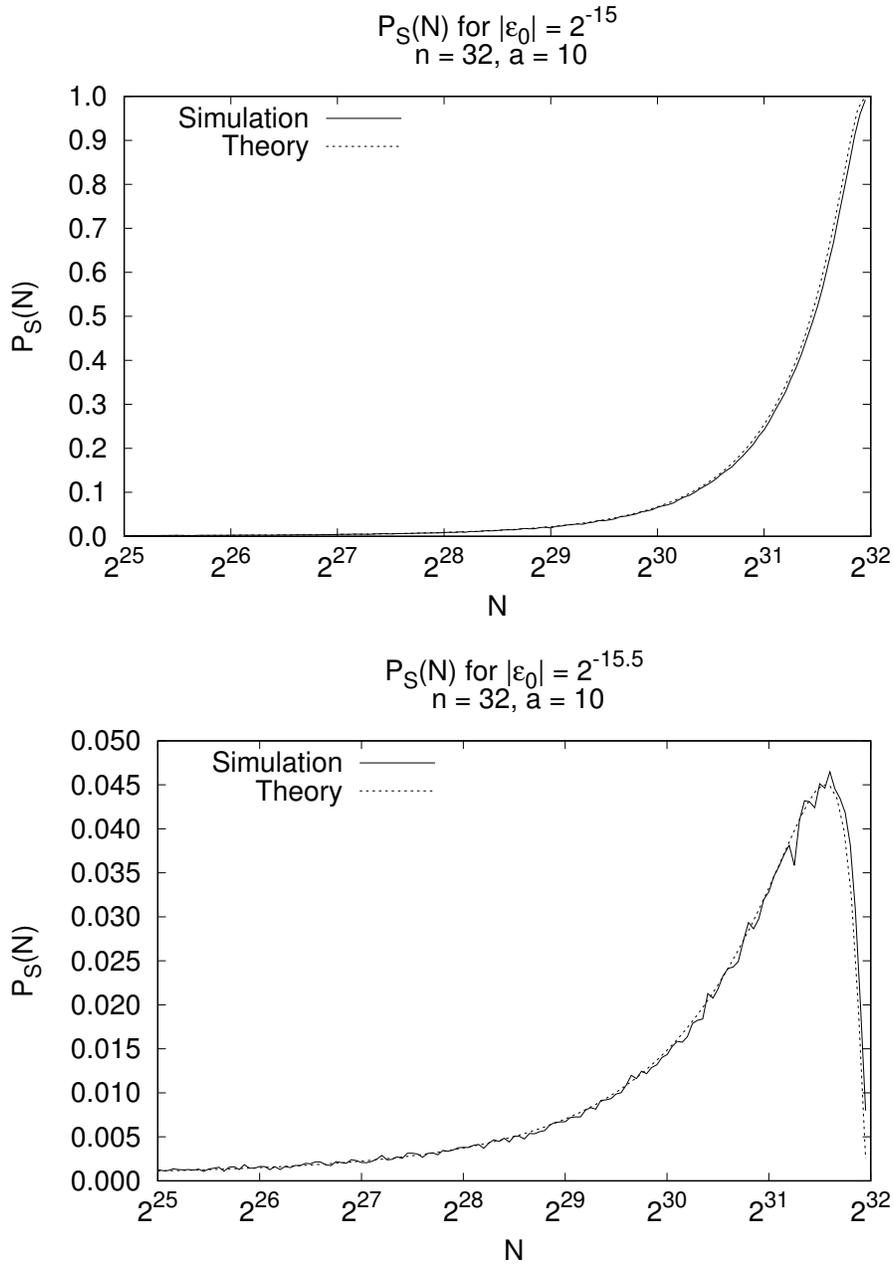


Fig. 5: Comparison of the theoretical success probability with a simulation of the success probability. The top figure on corresponds to the monotonous case, *i.e.* an attack using a bias larger than the bound of Theorem 1. The other figure shows the non-monotonous behavior.

6.3 Non-monotonicity

To test the stochastic nature of the non-monotonicity of the success probability, we choose parameters $n, |\epsilon_0|$ and a such that Lemma 5 predicts monotonicity with some fixed probability. For $n = 32, m = 14$ and $a = 10$, Lemma 5 shows that $\mathbf{P}_S(N)$ is monotonous with probability $1/2$ for

$$|\epsilon_0| = 2^{-n/2-1} \Phi^{-1} (1 - 2^{-a-1} + 2^{-m-2}) \approx 2^{-15.27486986}.$$

Note that since m is relatively small³, the above value for $|\epsilon_0|$ is not exactly equal to the bound of Theorem 1 contrary to what one might expect due to symmetry.

To detect whether an observation of the success probability is monotonous, it suffices to check that it reaches one for $N = 2^n$. In a total of 400 simulations, we have observed monotonous behavior in 204 cases (51%). Figure 6 shows the average curves $\mathbf{E}[\mathbf{P}_S(N) \mid \mathbf{P}_S(2^n) = 0]$ and $\mathbf{E}[\mathbf{P}_S(N) \mid \mathbf{P}_S(2^n) = 1]$, corresponding respectively to the monotonous and non-monotonous case.

From Corollary 2, we see that if there is a maximum success probability, it should occur for N very close to 2^n . Hence, for relatively small values of N , the average success probability should behave the same in the monotonous and non-monotonous case:

$$\mathbf{E}[\mathbf{P}_S(N) \mid \mathbf{P}_S(2^n) = 0] \approx \mathbf{E}[\mathbf{P}_S(N) \mid \mathbf{P}_S(2^n) = 1],$$

for small N . This is reflected by Figure 6.

³ More precisely, the continuity correction in Lemma 5 is non-negligible.

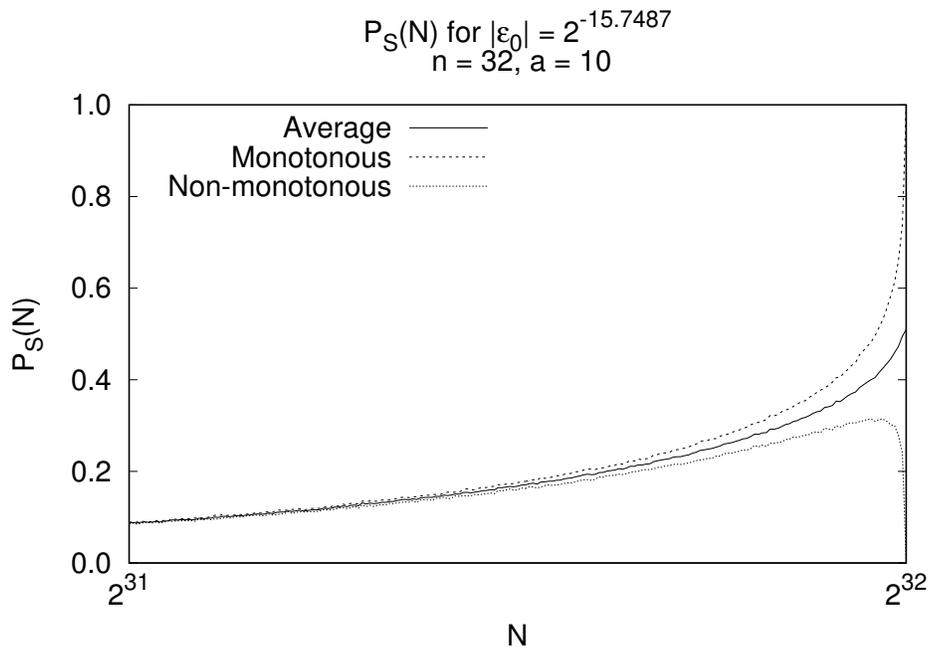


Fig. 6: Averages of the experimental success probability. The middle curve is the average over all experiments. The upper curve was computed by averaging over the experiments with monotonous behavior. Finally, the lower curve corresponds to the average over all non-monotonous experiments. Note that the scale for the horizontal axis is logarithmic and does not start from zero.

7 Conclusion

In this paper we revisited the behavior of the empirical bias for wrong keys. We have pointed out that previous works implicitly assume that plaintext/ciphertext pairs are sampled with replacement, which results in larger data complexities than necessary. We have redeveloped the theory under the assumption that the adversary can discard duplicate plaintext/ciphertext pairs, and have presented formulas for the success probability and the data complexity. The previously observed non-monotonous behavior of the success probability, which was characterized as counter-intuitive, has been explained and the conditions for its occurrence have been derived.

All the results in this paper have been verified through simulations. We conclude that when an adversary using Matsui's Algorithm 2 attempts to increase his advantage beyond certain bounds, what is witnessed to be an increase in the

success probability is in fact a false positive. As the accuracy of the estimation increases, the attack is doomed to fail, which is evidenced by a decreasing success probability as N increases beyond a certain point.

Our simulations also show that, independent of the success probability, the attack converges faster to its final result when sampling without replacement is preferred over sampling with replacement. This results in a reduced data complexity. Since the overall time complexity of an attack includes the time it takes to generate the required data (*i.e.*, the data complexity), this reduction may decrease the overall complexity of attacks believed to be more expensive than exhaustive search.

We believe that alternatives for or extensions to Matsui's algorithm 2 could extend the reach of linear cryptanalysis to absolute biases below the bound given by Theorem 1, provided that they take into account the origins of this prerequisite.

A Data Complexity

This section provides the calculations in the proof of Corollary 1. The objective is to solve the equation

$$\Phi^{-1}(P_S)\sqrt{1 - \frac{N}{2^n}} = 2\sqrt{N}|\epsilon_0| - \alpha.$$

Letting $\alpha = \Phi^{-1}(1 - 2^{-a-1})$ and $\beta = \Phi^{-1}(P_S)$, and squaring yields

$$\beta^2 (1 - 2^{-n}N) = 4N|\epsilon_0|^2 - 4\sqrt{N}|\epsilon_0|\alpha + \alpha^2.$$

Grouping terms appropriately, we obtain

$$(4|\epsilon_0|^2 + 2^{-n}\beta^2)N - 4\sqrt{N}|\epsilon_0|\alpha + \alpha^2 - \beta^2 = 0.$$

This equation is quadratic in \sqrt{N} and has the solutions

$$\sqrt{N} = \frac{2|\epsilon_0|\alpha \pm \sqrt{(2\epsilon_0\alpha)^2 - (\alpha^2 - \beta^2)(2^{-n}\beta^2 + 4|\epsilon_0|^2)}}{4|\epsilon_0|^2 + 2^{-n}\beta^2}.$$

B Maximum of $P_S(N)$

In the proof of Corollary 2, it is mentioned that the maximum is obtained by solving

$$\frac{d}{dN} \left(\frac{2\sqrt{N}|\epsilon_0| - \Phi^{-1}(1 - 2^{-a-1})}{\sqrt{1 - \frac{N}{2^n}}} \right) = 0.$$

Note that

$$\frac{d}{dN} \left(\frac{1}{\sqrt{1 - \frac{N}{2^n}}} \right) = \frac{1}{2^{n+1}\sqrt{(1 - \frac{N}{2^n})^3}},$$

such that we obtain the equivalent equation

$$\frac{|\epsilon_0|}{\sqrt{N} \left(1 - \frac{N}{2^n}\right)} = \frac{\Phi^{-1}(1 - 2^{-a-1}) - 2\sqrt{N}|\epsilon_0|}{2^{n+1}\sqrt{(1 - \frac{N}{2^n})^3}}.$$

This is readily simplified to

$$|\epsilon_0| \left(1 - \frac{N}{2^n}\right) = \sqrt{N}2^{-n-1}\Phi^{-1}(1 - 2^{-a-1}) - 2^{-n}N|\epsilon_0|,$$

and further

$$|\epsilon_0| = \sqrt{N}2^{-n-1}\Phi^{-1}(1 - 2^{-a-1}).$$

Finally, we obtain the result:

$$N = \left(\frac{|\epsilon_0|2^{n+1}}{\Phi^{-1}(1 - 2^{-a-1})} \right)^2.$$

References

1. Biryukov, A., Cannière, C.D., Quisquater, M.: On Multiple Linear Approximations. In: Franklin, M.K. (ed.) *Advances in Cryptology - CRYPTO 2004*, 24th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 2004, Proceedings. *Lecture Notes in Computer Science*, vol. 3152, pp. 1–22. Springer (2004), http://dx.doi.org/10.1007/978-3-540-28628-8_1
2. Blondeau, C., Gérard, B., Tillich, J.: Accurate estimates of the data complexity and success probability for various cryptanalyses. *Des. Codes Cryptography* 59(1-3), 3–34 (2011), <http://dx.doi.org/10.1007/s10623-010-9452-2>
3. Bogdanov, A., Kavun, E.B., Tischhauser, E., Yalçın, T.: Large-scale high-resolution computational validation of novel complexity models in linear cryptanalysis. *J. Computational Applied Mathematics* 259, 592–598 (2014), <http://dx.doi.org/10.1016/j.cam.2013.10.020>
4. Bogdanov, A., Rijmen, V.: Zero-Correlation Linear Cryptanalysis of Block Ciphers. *IACR Cryptology ePrint Archive* 2011, 123 (2011), <http://eprint.iacr.org/2011/123>
5. Bogdanov, A., Rijmen, V.: Linear hulls with correlation zero and linear cryptanalysis of block ciphers. *Des. Codes Cryptography* 70(3), 369–383 (2014), <http://dx.doi.org/10.1007/s10623-012-9697-z>
6. Bogdanov, A., Tischhauser, E.: On the Wrong Key Randomisation and Key Equivalence Hypotheses in Matsui’s Algorithm 2. In: Moriai, S. (ed.) *Fast Software Encryption - 20th International Workshop, FSE 2013*, Singapore, March 11-13, 2013. *Revised Selected Papers. Lecture Notes in Computer Science*, vol. 8424, pp. 19–38. Springer (2013), http://dx.doi.org/10.1007/978-3-662-43933-3_2
7. Daemen, J., Rijmen, V.: Probability distributions of correlation and differentials in block ciphers. *J. Mathematical Cryptology* 1(3), 221–242 (2007), <http://dx.doi.org/10.1515/JMC.2007.011>
8. Feller, W.: *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons (1967), exercise 10
9. Harpes, C., Massey, J.L.: Partitioning Cryptanalysis. In: Biham, E. (ed.) *Fast Software Encryption, 4th International Workshop, FSE '97*, Haifa, Israel, January 20-22, 1997, Proceedings. *Lecture Notes in Computer Science*, vol. 1267, pp. 13–27. Springer (1997), <http://dx.doi.org/10.1007/BFb0052331>
10. Hermelin, M., Cho, J.Y., Nyberg, K.: Multidimensional Linear Cryptanalysis of Reduced Round Serpent. In: Mu, Y., Susilo, W., Seberry, J. (eds.) *Information Security and Privacy, 13th Australasian Conference, ACISP 2008*, Wollongong, Australia, July 7-9, 2008, Proceedings. *Lecture Notes in Computer Science*, vol. 5107, pp. 203–215. Springer (2008), http://dx.doi.org/10.1007/978-3-540-70500-0_15
11. Junod, P., Vaudenay, S.: Optimal Key Ranking Procedures in a Statistical Cryptanalysis. In: Johansson, T. (ed.) *Fast Software Encryption, 10th International Workshop, FSE 2003*, Lund, Sweden, February 24-26, 2003, Revised Papers. *Lecture Notes in Computer Science*, vol. 2887, pp. 235–246. Springer (2003), http://dx.doi.org/10.1007/978-3-540-39887-5_18
12. Matsui, M.: Linear Cryptanalysis Method for DES Cipher. In: Helleseth, T. (ed.) *Advances in Cryptology - EUROCRYPT '93, Workshop on the Theory and Application of Cryptographic Techniques, Lofthus, Norway, May 23-27, 1993*, Proceedings. *Lecture Notes in Computer Science*, vol. 765, pp. 386–397. Springer (1993), http://dx.doi.org/10.1007/3-540-48285-7_33

13. Molenaar, W.: Approximations to the Poisson, Binomial and Hypergeometric Distribution Functions. Ph.D. thesis, Mathematisch Centrum Amsterdam (1970)
14. Nyberg, K.: Linear Approximation of Block Ciphers. In: Santis, A.D. (ed.) Advances in Cryptology - EUROCRYPT '94, Workshop on the Theory and Application of Cryptographic Techniques, Perugia, Italy, May 9-12, 1994, Proceedings. Lecture Notes in Computer Science, vol. 950, pp. 439–444. Springer (1994), <http://dx.doi.org/10.1007/BFb0053460>
15. Pinsky, M.A.: The normal approximation to the hypergeometric distribution. Unpublished manuscript, https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/pinsky-hypergeometric.pdf
16. Selçuk, A.A.: On Probability of Success in Linear and Differential Cryptanalysis. *J. Cryptology* 21(1), 131–147 (2008), <http://dx.doi.org/10.1007/s00145-007-9013-7>