Related-Key Linear Cryptanalysis on SIMON

Jung-Keun Lee, Bonwook Koo, and Woo-Hwan Kim

National Security Research Institute, Republic of Korea
jklee,bwkoo,whkim5@nsr.re.kr

Abstract. We present a general framework of the related-key linear attack that can be applied to a class of block ciphers including the key-alternating iterative ones with linear or affine key schedules. In contrast with the existing linear attacks that use the linear characteristics with linear correlations up to around $\pm 2^{-n/2}$, where n is the block length of the block cipher, the proposed attack can use linear characteristics with linear correlations up to around $\pm 2^{-k/2}$, where k is the key length, by exploiting related-key linear approximations introduced in this work. Thus the new attack can cover more rounds than the current linear attacks when the key length is much larger than the block length. The attack utilizes the key differences as the additional data source and it can use data of size larger than the codebook size. We also present a method of the related-key linear attack using multiple independent linear trails that provides a way to estimate the computational complexity, the data complexity, and the success probability of the attack in terms of the capacity of the system of the linear approximations. It is based on hypothesis testing with new decision rule and can be regarded as an alternative approach of the attack provided by A. Biryukov et al. at Crypto 2004 that is based on maximum likelihood approach and cannot estimate the success probability. By applying the framework to SIMON, we get various attack results, including the first full-round attacks on SIMON 64/128 and SIMON 128/256. Then we justify our claims by presenting experimental results that are close to expected ones with a small-scale variant of SIMON.

Keywords: related-key attack, linear cryptanalysis, linear key schedule, multiple linear attack, SIMON

1 Introduction

In the last decades many lightweight block ciphers have been proposed targeting resource-constrained platforms. They adopt simple key schedules to get competitive performance figures in terms of the resource requirements. In this regard not a few of them have linear or affine key schedules. (e.g. GIFT [2], SKINNY [5], MIDORI [1], SIMON [3], ZORRO [16], PRINCE [12], LED [17], PICCOLO [26], KATAN [13].) But no effective cryptanalytic methods have been

Table 1. Attack results on SIMON

cipher (# rounds)	# attacked rounds	computation	data	$\Pr_{\text{success}}$	Ref.
SIMON 32/64 (32)	27	$2^{61.34}$	2^{60}	0.5	This Work
	25	$2^{55.26}$	2^{55}	0.5	This Work
	23	$2^{56.3}$	$2^{31.19}$	0.28	[14]
SIMON 48/96 (36)	34	$2^{93.34}$	2^{94}	0.5	This Work
	30	$2^{80.18}$	$2^{80.1}$	0.5	This Work
	25	$2^{88.28}$	$2^{47.92}$	N/A	[14]
SIMON 64/128 (44)	44	$2^{125.34}$	2^{126}	0.5	This Work
	38	$2^{116.03}$	$2^{115.6}$	0.5	This Work
	31	2^{120}	$2^{63.53}$	N/A	[14]
SIMON 128/256 (72)	72	$2^{240.18}$	$2^{240.1}$	0.5	This Work
	53	$2^{248.01}$	$2^{127.6}$	N/A	[14]

discovered utilizing the linearity of the key schedule in the general framework as yet. Since the linear attack was made public by M. Matsui [22], there have been many extensions such as the attacks using the linear hulls [23] or the multiple linear approximations [19] [6]. Also there are lots of works regarding the related-key attacks against the block ciphers using differential characteristics. But there are not many works dealing with the related-key linear attacks, in striking contrast with the fact that the differential attack and the linear attack are regarded as the most important cryptanalytic methods to evaluate the security of block ciphers.

Our contributions

- We present a general framework of the related-key linear cryptanalysis that can be applied to the key-alternating iterative block ciphers with simple and close-to-affine key schedules. The framework enables to cover more rounds of such a block cipher than the existing linear attacks when the key length of the cipher is considerably larger than its block length.
- We present a method of the related-key linear attack using multiple independent trails. We provide an explicit formula based on hypothesis testing to estimate the complexity of the attack taking the success probability into account.
- We present related-key linear attacks on SIMON 32/64, SIMON 48/96, SIMON 64/128, and SIMON 128/256 whose results are summarized in Table 1. The presented attacks cover more rounds than the current best attacks while requiring larger data. In particular, the first attacks on the full SIMON 64/128 and SIMON 128/256 are presented.
- We present experimental results that show the validity of our framework including the appropriateness of the key hypotheses we presume.

Related Works

Related-key linear attacks. There are not so many works referring to the related-key linear attack in the literature. P. Vora et al. [27] presented an attack on a round-reduced DES claiming that using related keys one can amplify the weakness exploited in the single key recovery. M. Hermelin et al. [18] claim that using related keys, one can attack the full PRESENT-128. A. Bogdadov et al. [9] presented a key recovery attack using some related-key linear distinguishers with chosen key differences. C. Beierle et al. [5] argue that the SKINNY family of block ciphers are secure against the related-tweakey linear attack by presenting bounds on the correlations of linear trails as the number of rounds increases, taking into account the fact that the attacker may utilize the tweakey as the additional source of data. But the general framework of the linear attacks using the related-key linear approximations is introduced in this work for the first time.

Multiple linear attacks and the success probability. A linear attack using multiple trails was first introduced in [19]. But the method has severe limitations in applications as mentioned in [6], where the authors presented an attack using *independent* linear trails. They provided an explicit formula for the gain of the attack in terms of the capacity of the system of the linear approximations and the size of the data based on a maximum likelihood approach, but they were not able to estimate the success probability of the attack. Our multiple linear attack is based on hypothesis testing and makes use of a new decision rule that leads to an explicit formula for attack complexities together with the success probability. On the other hand, extensions of Matsui's Algorithm 2 using linear hulls were introduced and their computational complexity and the success probability were estimated with standard key hypotheses [23] [25]. Recently, the legitimacy of the standard key hypotheses for the linear attacks has been questioned and some adjusted key hypotheses have been considered [8] [7] [11] [10]. But the complexities of our related-key attacks can be estimated accurately with the standard key hypotheses.

Organization of the paper In Sect. 2 we introduce the terminology and notations used in the paper. In Sect. 3 we describe the general framework of the related-key linear cryptanalysis with emphasis on its application to the block ciphers with the affine key schedules. In Sect. 4 we present attack results on SIMON obtained from the framework presented in Sect. 3 together with some dedicated analysis. In Sect. 5 we provide experimental results on a small-scale variant of SIMON that corroborate the claims of the paper. In Sect. 6 we discuss the validity of the attacks in more detail. We conclude in Sect. 7.

2 Terminology and Notations

$\mathbb{F}_2$ denotes the field with 2 elements 0 and 1. A word is a bit string of the length w=12,16,24,32, or 64. For integers i, j with $i \leq j$, $[i..j]$ denotes the set of integers x such that $i \leq x \leq j$. The inner product of a mask γ and a value x of the same length is denoted by $\langle \gamma, x \rangle$. For a Boolean function $G : \mathbb{F}_2^l \to \mathbb{F}_2$, the correlation

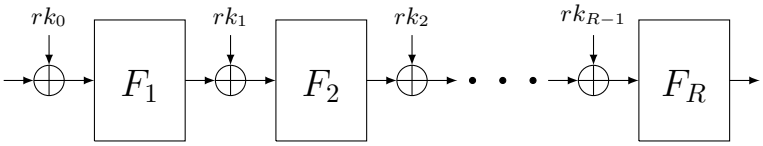

Fig. 1. a long-key cipher

of G is defined to be the imbalance $(|\{x : G(x) = 0\}| - |\{x : G(x) = 1\}|)/2^l$. For a vectorial Boolean function $F : \mathbb{F}_2^l \to \mathbb{F}_2^m$, an l-bit mask γ, and an m-bit mask λ, the (linear) correlation of F with respect to the mask pair (γ, λ) is defined to be the correlation of the Boolean function G given by $G(x) = \langle \gamma, x \rangle \oplus \langle \lambda, F(x) \rangle$ and is denoted by $\varepsilon_F(\gamma, \lambda)$. R denotes the the number of the rounds of the iterative block cipher in consideration. The intermediate state of such a block cipher just before xoring with the $(i+1)$-th round key is denoted by X_i $(i = 0, \ldots, R-1)$. E_i^j is the subcipher of the block cipher from the $(i+1)$-th round to the $(j+1)$-th round. The first mask and the last mask of a linear trail are called the initial (intermediate) mask and the final (intermediate) mask, respectively. The intermediate state of a block cipher coupled with the initial mask and the final mask are called the initial intermediate value and the final intermediate value, respectively. k and n denote the key length and the block length of the block cipher, respectively. K^* denotes the unknown fixed key to be recovered. rk_i^* denotes the round key for the $(i + 1)$-th round derived from K^*. The difference of the round keys for the $(i + 1)$-th round derived from K^* and $K^* \oplus \Delta K$ is denoted by δrk_i. The LSB (least significant bit) of a word is indexed as 0 and is located at the rightmost position. The $(i + 1)$-th rightmost bit of a word x is denoted by $x[i]$ so that $x[0]$ denotes the LSB of x. For a bit string X with even length, X_{L} and X_{R} denote the left half and right half of X, respectively. For a w-bit word x, $x[i]$ with $i \notin [0..(w - 1)]$ means $x[i \mod w]$. $x \lll a$ and $x \ggg a$ denote the circular shift of a word x to the left and right by a bits, respectively. The Hamming weight of a word x, denoted by $\mathrm{wt}(x)$, is the number of the nonzero bits of x. The support of a w-bit word x is defined to be the set of indices $\{i \in [0..(w - 1)] : x[i] \neq 0\}$ and is denoted by $\mathtt{supp}(x)$. $\|$ denotes the concatenation of bit strings. Bit strings such as words and masks are expressed in the hexadecimal representations. For example, `c201` represents the bit string $1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1$. For real numbers μ and $\sigma > 0$, $\mathrm{N}(\mu, \sigma^2)$ denotes the normal distribution with the mean μ and the standard deviation σ. Φ denotes the cumulative distribution function of the standard normal distribution.

3 The Related-Key Linear Cryptanalysis

3.1 The General Framework

Our related-key linear attacks incorporate both of Matsui's Algorithm 1 and Algorithm 2. But unlike the single-key linear attacks, they make use of the newly

introduced related-key linear approximations that involve the key differences together with the intermediate values and the round keys. Let R, r, and s be integers with $0 \le s \le s+r \le R$ and let $E : \mathbb{F}_2^k \times \mathbb{F}_2^n \to \mathbb{F}_2^n$ be an R-round key-alternating iterative block cipher with k-bit keys and n-bit blocks. We will describe an attack on E that adds rounds to an r-round related-key linear approximation that is obtained from an r-round linear trail of the long-key cipher and approximations between the key differences and the round keys. Let $\tilde{E}$ be the long-key cipher corresponding to E and φ be the key scheduling function. That is, $\tilde{E}$ is a function $\mathbb{F}_2^{Rn} \times \mathbb{F}_2^n \to \mathbb{F}_2^n$ defined by

$$\tilde{E}(rk_0 \| rk_1 \| \cdots \| rk_{R-1}, x) = F_R(rk_{R-1} \oplus \cdots F_2(rk_1 \oplus F_1(rk_0 \oplus x)) \cdots)$$

as in Fig. 1, where each F_i is a fixed n-bit permutation, φ is a function $\mathbb{F}_2^k \to \mathbb{F}_2^{Rn}$, and $E(K, x) = \tilde{E}(\varphi(K), x)$ for $(K, x) \in \mathbb{F}_2^k \times \mathbb{F}_2^n$.

Getting a related-key linear approximation Suppose that we have an r-round linear trail $[\gamma_s, \gamma_{s+1}, \ldots, \gamma_{s+r}]$ for $\tilde{E}$ such that the correlation $\varepsilon_{F_{i+1}}(\gamma_i, \gamma_{i+1})$ for the $(i+1)$-th round is ϵ_i for each $i \in [s..(s+r-1)]$. It is well-known that the average of the empirical correlations is $\tilde{\epsilon} = \epsilon_s \cdots \epsilon_{s+r-1}$. That is,

$$\Pr_{x,\mathbf{rk}}(\langle \gamma_s, x \rangle \oplus \langle \gamma_{s+r} \tilde{E}_s^{s+r-1}(\mathbf{rk}, x) \oplus \bigoplus\nolimits_{i=0}^{r-1} \langle \gamma_{s+i}, rk_{s+i} \rangle = 0) = \frac{1+\tilde{\epsilon}}{2}, \quad (1)$$

where $\mathbf{rk} = rk_0 \| rk_1 \| \cdots \| rk_{R-1}$ and $\tilde{E}_i^j$ is the subcipher of $\tilde{E}$ from the $(i+1)$-th round to the $(j+1)$-th round. (See e.g. [23].) Let K^* be the fixed unknown key to be recovered and let $\varphi(K^*) = \mathbf{rk}^* = rk_0^* \| rk_1^* \| \cdots \| rk_{R-1}^*$. Suppose that φ admits a linear approximation

$$\langle \bar{\gamma}_{s+i}, \Delta K \rangle \oplus \langle \gamma_{s+i}, \delta rk_{s+i} \rangle = 0 \quad (2)$$

with the correlation $\bar{\epsilon}_{s+i}$ for each i, where $\delta\mathbf{rk} := \varphi(K^* \oplus \Delta K) \oplus \varphi(K^*)$ is $\delta rk_0 \| \delta rk_1 \| \cdots \| \delta rk_{R-1}$. Such linear approximations between the key differences and the round key differences are trivially obtained if φ is affine, and in this case $\bar{\epsilon}_{s+i} = 1$ for each i. If not, we can try to get such appromations by combining two "same" linear approximations between the keys and the round keys. Writing $\mathbf{rk}$ as $\mathbf{rk}^* \oplus \delta\mathbf{rk}$ in (1) and using (2), we get a linear approximation

$$\langle \gamma_s, x \rangle \oplus \langle \gamma_{s+r}, E_s^{s+r-1}(K^* \oplus \Delta K, x) \rangle \oplus \bigoplus\nolimits_{i=0}^{r-1} (\langle \bar{\gamma}_{s+i}, \Delta K \rangle \oplus \langle \gamma_{s+i}, rk_{s+i}^* \rangle) = 0 \quad (3)$$

with the correlation $\epsilon = \tilde{\epsilon} \prod_{i=0}^{r-1} \bar{\epsilon}_{s+i}$ assuming that the linear approximations are all independent. Here the correlation of (3) is meant to be the imbalance of the approximation as $(x, \Delta K)$ takes all the values in $\mathbb{F}_2^{n+k}$ and we call (3) a *related-key linear approximation*.

The attack using a single linear approximation Since we have fixed the unknown key K^*, $\bigoplus_{i=0}^{r-1} \langle \gamma_{s+i}, rk_{s+i}^* \rangle$ is a constant. Thus (3) leads to the linear

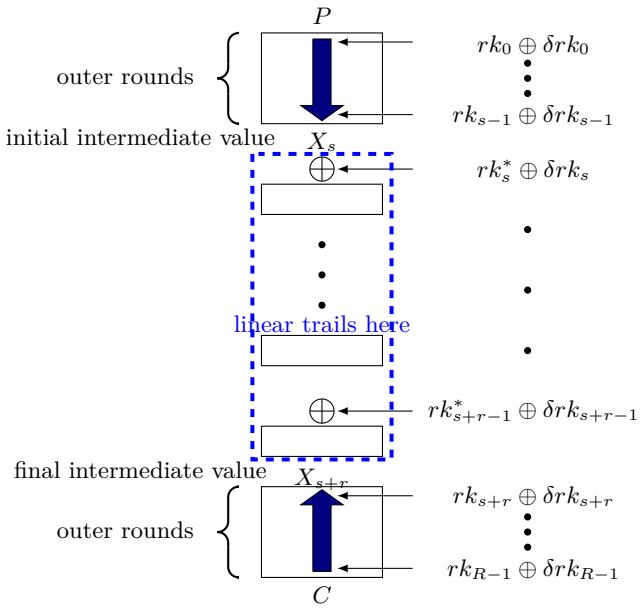

Fig. 2. The outline of the linear attacks using related-key linear approximations

approximation

$$\langle \gamma_s, X_s \rangle \oplus \langle \gamma_{s+r}, X_{s+r} \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \bar{\gamma}_{s+i}, \Delta K \rangle = 0, \tag{4}$$

with the correlation $\pm\epsilon$ that involves the key differences as well as the initial and the final intermediate values. We perform an attack using (4) as the distinguisher: To check whether a candidate key K is correct, for many triples $(P, C, \Delta K)$ with $C = E(K^* \oplus \Delta K, P)$, we compute $\langle \gamma_s, X_s \rangle$, $\langle \gamma_{s+r}, X_{s+r} \rangle$, and $\bigoplus_{i=0}^{r-1} \langle \bar{\gamma}_{s+i}, \Delta K \rangle$, where $X_s = E_0^{s-1}(K \oplus \Delta K, P)$ and $E_{s+r}^{R-1}(K \oplus \Delta K, X_{s+r}) = C$ and then compute

$$\langle \gamma_s, X_s \rangle \oplus \langle \gamma_{s+r}, X_{s+r} \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \bar{\gamma}_{s+i}, \Delta K \rangle \tag{5}$$

(See Fig. 2.) If K satisfies certain conditions that K^* does, the imbalance of (5) is likely to be close to $\pm\epsilon$. If not, the imbalance is likely to be close to 0. We assume that we have the related-key linear approximation (4) with the correlation $\pm\epsilon$. We also assume that we have data $\{(P_i, C_i, \Delta K_i) : 1 \le i \le N\}$ of size N available for the attack, where $C_i = E(K^* \oplus \Delta K_i, P_i)$ for each i. For a linear attack, we usually have to perform the data compression first to reduce the computational complexity. The data compression in our related-key setting needs to handle the key differences unlike that in the single-key linear attacks. So the "compression function" $H_c : \mathbb{F}_2^{2n+k} \to \mathbb{F}_2^d$ with $2^d \ll N$ we need to get for the data compression is one such that the computation of (5) using some of the "outer"

Alg. 1 Related-key linear attack using a single trail

1. Perform the data compression to get the compressed set of size 2^d and set $\tau(z)$ for each $z \in \mathcal{Z}$.
2. For each entry (v, n_v) in the compressed data,
   - For each $z \in \mathcal{Z}$, compute $h(v, z)$ and increment or decrement $\tau(z)$ by n_v depending on whether $h(v, z)$ is 0 or 1.
3. For each z for which $|\tau(z)| \geq tN|\epsilon|$, try to recover the whole key bits using z.

round key bits and $(P, C, \Delta K)$ can be carried out using the same round key bits and $H_c(P, C, \Delta K)$. More specifically, $\langle \gamma_s, X_s \rangle \oplus \langle \gamma_{s+r}, X_{s+r} \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \bar{\gamma}_{s+i}, \Delta K \rangle$ can be computed as $h(H_c(P, C, \Delta K), rk^O) \oplus g(rk^O)$ for some functions h and g, where rk^O is the concatenation of the outer round key bits. Note that the pair $(h,\ g)$ is not unique but we just need to choose one. Then the set of the concatenations of outer round key bits can be partitioned into outer key classes[1] and the second input of h is actually an outer key class. Then we apply the compression function to the data set to get the compressed data set $\{(v, n_v) \in \mathbb{F}_2^d \times \mathbb{Z} : n_v = |\{i : H_c(P_i, C_i, \Delta K_i) = v\}|\}$. We let $\mathcal{Z}$ be the set of the outer key classes used in the computation of (5) and let $k_O = \log_2 |\mathcal{Z}|$.

Hypothesis testing and the decision rule. Suppose that we mount a related-key linear attack using a single linear approximation (4) with available data of size N. For each $z \in \mathcal{Z}$, let $\tau(z) := |\{i : h(H_c(P_i, C_i, \Delta K_i), z) = 0\}| - |\{i : h(H_c(P_i, C_i, \Delta K_i), z) = 1\}|$. Let z^* be the correct outer key class. Then the right key hypothesis and the wrong key hypothesis we presume are that $\tau(z)/N$ can be regarded as a random variable such that

- $\tau(z^*)/N$ follows $\mathrm{N}(\pm\epsilon, (1 - \epsilon^2)/N) \approx \mathrm{N}(\pm\epsilon, 1/N)$ and
- $\tau(z)/N$ follows $\mathrm{N}(0, 1/N)$ for $z \neq z^*$.

We determine whether z is correct or not by the decision rule $|\tau(z)/N| \geq t\epsilon$. Here $t > 0$ is the threshold parameter that enables to get a tradeoff between the computational complexity and the success probability.

The attack algorithm. The attack proceeds as in Alg. 1. Consider the list of z's in $\mathcal{Z}$ for which $|\tau(z)| \geq tN|\epsilon|$. The attack is successful if z^* is in the list, and the list may also contain many wrong entries that are called the false alarms. We let $c_d = \sqrt{N}|\epsilon|$. By the following Lemmas, the success probability of the attack is

$$\Phi\left((1 - t)c_d\right) + \Phi\left((-1 - t)c_d\right)$$

and the false alarm probability is $2\Phi(-tc_d)$, which implies that the list is expected to contain about $2\Phi(-tc_d)|\mathcal{Z}| = 2^{k_O+1}\Phi(-tc_d)$ wrong entries.

[1] Two concatenations rk and rk' of outer round key bits are in the same outer key class if and only if $h(v, rk) = h(v, rk')$ for all v. In many cases, the outer key classes can be represented by the concatenation of some outer round key bits involved in the computation.

Lemma 1. *Let μ, t, σ be positive real numbers and let Y be a random variable with $Y \sim \mathrm{N}(0, \sigma^2)$. Then*

$$\Pr\left(|Y| \geq t\mu\right) = 2\Phi\left(-t\mu/\sigma\right).$$

Lemma 2. *Let μ, t, σ be positive real numbers and let Y be a random variable with $Y \sim \mathrm{N}(\pm\mu, \sigma^2)$. Then*

$$\Pr\left(|Y| \geq t\mu\right) = \Phi\left((1-t)\mu/\sigma\right) + \Phi\left((-1-t)\mu/\sigma\right).$$

To compare the computational complexity of the attack with that of the exhaustive key search, we say that the complexity of 1 encryption (including the key schedule) is 1. Let c_p be the complexity of 1 computation of H_c and c_o be the complexity of 1 computation of h using an entry in the compressed data and a candidate outer key class. Then we have

Theorem 1. *The computational complexity of the attack using Alg. 1 is*

$$c_p N + c_o 2^{d+k_O} + 2^{k+1}\Phi(-tc_d)$$

with the success probability $\Phi\left((1-t)c_d\right) + \Phi\left((-1-t)c_d\right)$.

Here we have assumed that the restored outer key class z reveals simple independent relations between the bits of K^*, meaning that using the k_O-bit information that z reveals about K^*, we can recover the whole k bits of K^* using other simple $(k - k_O)$ relations between the bits of K^*. (This is mostly the case if, for example, the key schedule is affine.) Letting $t = 1$, the success probability of the attack is slightly larger than 0.5, so the attack can be regarded to be more effective than the exhaustive key search when the computational complexity of the attack is less than 2^{k-1}. So if, for example, the absolute value of the correlation ϵ of the related-key linear approximation is larger than $2^{-k/2+1}$, the trail admits data compression with $d + k_O \ll k - 1$, and $c_p, c_o \leq 1$, then we have an effective attack with data complexity $N = \epsilon^{-2}$. We will also present an attack method that can halve the computational complexity in some cases without much loss in the success probability later.

The related-key linear attack using multiple linear approximations We can also consider the related-key linear attack that uses multiple linear trails to reduce the attack complexity as in the single-key setting. We present an attack method based on hypothesis testing using a newly introduced decision rule. Suppose that we have a system of $m \geq 1$ independent related-key linear approximations

$$\langle \gamma_s^{(j)}, X_s \rangle \oplus \langle \gamma_{s+r}^{(j)}, X_{s+r} \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \bar{\gamma}_{s+i}^{(j)}, \Delta K \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \gamma_{s+i}^{(j)}, rk_{s+i}^* \rangle = 0 \quad (6)$$

$(1 \leq j \leq m)$ and the correlation of the j-th approximation is $\epsilon_{(j)}$. We let $\epsilon = \left(\sum_{j=1}^{m} \epsilon_{(j)}^2\right)^{1/2}$, the square of which is called the capacity of the system in

the literature [6]. Suppose that we have related-key data $\{(P_i, C_i, \Delta K_i)\}$ of size N. Suppose also that we have a function $H_c : \mathbb{F}_2^{2n+k} \to \mathbb{F}_2^d$ from which we get a compressed data set $\{(v, n_v)\}$ such that the computation of the m values

$$\langle \gamma_s^{(j)}, X_s \rangle \oplus \langle \gamma_{s+r}^{(j)}, X_{s+r} \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \bar{\gamma}_{s+i}^{(j)}, \Delta K \rangle \ (j = 1, \ldots, m) \tag{7}$$

using some of the outer round key bits and $(P, C, \Delta K)$ can be carried out using the same round key bits and $H_c(P, C, \Delta K)$. That is, for each j, there are functions $h^{(j)}$ and $g^{(j)}$ such that (7) can be computed as $h^{(j)}(H_c(P, C, \Delta K), rk^O) \oplus g^{(j)}(rk^O)$, where rk^O is the concatenation of the outer round key bits. For each j, let $\mathcal{Z}_O^j$ be the set of outer key classes determined by $h^{(j)}$. Let $\mathcal{Z}_O$ be the set of outer key classes determined by all the $h^{(j)}$'s. That is, two concatenations rk and rk' of outer round key bits are in the same outer key class in $\mathcal{Z}_O$ if and only if $h^{(j)}(v, rk) = h^{(j)}(v, rk')$ for all j and v. Every z_O in $\mathcal{Z}_O$ can be represented as $z_O^{(1)} \cap \ldots \cap z_O^{(m)}$, where $z_O^{(j)} \in \mathcal{Z}_O^j$ for each $j \in [1..m]$. So for each j and each $z_O \in \mathcal{Z}_O$, there is a unique $z_O^{(j)} \in \mathcal{Z}_O^j$ such that $z_O \subset z_O^{(j)}$. We denote the correct outer key class by $z_O^* \in \mathcal{Z}_O$. Let $e_j^* = \bigoplus_{i=0}^{r-1} \langle \gamma_{s+i}^{(j)}, rk_{s+i}^* \rangle \oplus g^{(j)}(rk^{O*})$ and $\theta_j^* = (-1)^{e_j^*}$ for each $j \in [1..m]$. The m "parity bits" e_1^*, ..., e_m^* also have to be guessed in the multiple linear attack. We let z_I^* be the m-bit value such that $z_I^*[j] = e_{j+1}^*$ for each $j \in [0..m-1]$. z_I will denote a guess for z_I^* and $\mathcal{Z}_I$ will denote the set of all z_I's. Let $k_I = m$ so that $|\mathcal{Z}_I| = 2^{k_I}$. We let $c_d = \sqrt{N}|\epsilon|$.

Hypothesis testing and the decision rule. Suppose that we mount a (related-key) multiple linear attack using m linear approximations (6) with the correlations $\epsilon_{(j)}$ $(j = 1, \ldots, m)$ with available data of size N. For each $j \in [1..m]$ and $z_O \in \mathcal{Z}_O$, let $\tau_j(z_O) := |\{i : h^{(j)}(H_c(P_i, C_i, \Delta K_i), z_O) = 0\}| - |\{i : h^{(j)}(H_c(P_i, C_i, \Delta K_i), z_O) = 1\}|$. Then the wrong key hypothesis and the right key hypothesis we presume are that

- $\tau_j(z_O)/N$ is a random variable following $\mathrm{N}(0, 1/N)$ if $z_O^{(j)}$ containing z_O is wrong,
- $\tau_j(z_O)/N$ is a random variable following $\mathrm{N}(\theta_j^* \epsilon_{(j)}, (1 - \epsilon_{(j)}^2)/N) \approx \mathrm{N}(\theta_j^* \epsilon_{(j)}, 1/N)$ if $z_O^{(j)}$ containing z_O is correct, and
- $\tau_1(z_O)/N, \ldots, \tau_m(z_O)/N$ are independent for each z_O.

Under the right key hypothesis,

$$\Pr\Big(\sum_{j=1}^{m} \theta_j^* \epsilon_{(j)} \tau_j(z_O^*)/N \geq t\epsilon^2 \Big) = \Phi((1-t)c_d) \tag{8}$$

by Lemma 3 that seems to be well-known.

Lemma 3. *Let $a_1, \ldots, a_m, b$ be real numbers with at least one a_j nonzero and let $\mu_1, \ldots, \mu_m, \sigma_1, \ldots, \sigma_m$ be real numbers with each $\sigma_j > 0$. Let $Y_1, \ldots, Y_m$ be independent random variables with $Y_j \sim \mathrm{N}(\mu_j, \sigma_j^2)$ for each j. Then*

$$\Pr\Big(\sum_{j=1}^{m} a_j Y_j \geq b \Big) = \Phi\left(\Big(\sum_{j=1}^{m} a_j \mu_j - b \Big) \Big/ \Big(\sum_{j=1}^{m} a_j^2 \sigma_j^2 \Big)^{1/2} \right).$$

Alg. 2 Related-key linear attack using multiple linear trails

1. Perform the data compression to get the compressed set of size 2^d. Set $\tau^{(j)}(z_O)$=0 for each j and z_O.
2. For each entry (v, n_v) in the compressed data,
   - For each $z_O \in \mathcal{Z}_O$ and each $j \geq 1$, compute $h^{(j)}(v, z_O)$ and increment $\tau^{(j)}(z_O)$ by n_v if $h^{(j)}(v, z_O)$ is 0. Otherwise, decrement $\tau^{(j)}(z_O)$ by n_v.
3. For each (z_O, z_I) for which

$$\sum\nolimits_{j=1}^{m} \epsilon_{(j)}(-1)^{z_I[j-1]}\tau^{(j)}(z_O) \geq tN\epsilon^2,$$

try to recover the whole key bits from (z_O, z_I).

The attack algorithm. We mount the related-key multiple linear attack as in Alg. 2. (Note that the decision rule used in Step 3 is a new one introduced in this work for the first time.) Let $k_O = \log_2 |\mathcal{Z}_O|$. Since (z_O^*, z_I^*) with $z_I^*[j] = e_{j+1}^*$ for each j is the correct one we are looking for in the attack algorithm and

$$\begin{aligned} & \Pr\Big(\sum\nolimits_{j=1}^{m} \epsilon_{(j)}(-1)^{z_I^*[j-1]}\tau^{(j)}(z_O^*) \geq tN\epsilon^2\Big) \\ = \quad & \Pr\Big(\sum\nolimits_{j=1}^{m} \theta_j^*\epsilon_{(j)}\tau_j(z_O^*)/N \geq t\epsilon^2\Big), \end{aligned}$$

the success probability of the attack is $\Phi\left((1-t)c_d\right)$ by (8). Now we consider the false alarm probability. Let $z_O \in \mathcal{Z}_O$ and $z_I \in \mathcal{Z}_I$ be given. Let $\theta_j = (-1)^{z_I[j-1]}$ for $j \in [1..m]$. Let $S_O = S_O(z_O^*, z_O)$ and $S_I = S_I(z_I^*, z_I)$ be subsets of $[1..m]$ such that $z_O^{(j)} \in \mathcal{Z}_O^j$ containing z_O is correct exactly for $j \in S_O$ and $\theta_j = \theta_j^*$ exactly for $j \in S_I$. Then

$$\begin{aligned} & \sum\nolimits_{j=1}^{m} \epsilon_{(j)}(-1)^{z_I[j-1]}\tau^{(j)}(z_O) \geq tN\epsilon^2 \\ \Leftrightarrow \quad & \sum\nolimits_{j=1}^{m} \theta_j\epsilon_{(j)}\tau_j(z_O) \geq tN\epsilon^2 \\ \Leftrightarrow \quad & \sum\nolimits_{j=1}^{m} \theta_j^*u_j\epsilon_{(j)}\tau_j(z_O) \geq tN\epsilon^2 \end{aligned}$$

where $u_j = \theta_j^*\theta_j$ for each j. Thus,

$$\begin{aligned} & \Pr\Big(\sum\nolimits_{j=1}^{m} \epsilon_{(j)}(-1)^{z_I[j-1]}\tau^{(j)}(z_O) \geq tN\epsilon^2\Big) \\ = \quad & \Phi\big(-tc_d + q(S_O, S_I)\big), \end{aligned}$$

where $q(S_O, S_I) = N^{1/2}(\sum_{j\in S_O\cap S_I} \epsilon_{(j)}^2 - \sum_{j\in S_O\setminus S_I} \epsilon_{(j)}^2)/\epsilon$ by Lemma 3. Thus the false alarm probability $p_{\mathrm{fa}}(t)$ for the threshold parameter t is

$$\sum_{(S_O,S_I)\neq([1..m],[1..m])} \frac{n(S_O, S_I)}{2^{k_I+k_O}}\Big(\Phi\big(-tc_d + q(S_O, S_I)\big)\Big), \tag{9}$$

where $n(S_O, S_I)$ be the number of the pairs (z_O, z_I) such that $z_O \in \mathcal{Z}_O$ is contained in the correct $z_O^{(j)} \in \mathcal{Z}_O^j$ exactly for $j \in S_O$ and $\theta_j = \theta_j^*$ exactly for $j \in S_I$ for each $S_O, S_I \subset [1..m]$. Let c_p be the complexity of 1 computation of H_c, $c_o^{(j)}$ be the complexity of the computation of $h^{(j)}$ from an entry in the compressed data and z_O, and c_a be the complexity of 1 integer addition. Then we have

Theorem 2. *The computational complexity of the attack using Alg. 2 is*

$$c_p N + 2^{k_O}\left(\sum_{j=1}^{m} c_o^{(j)} 2^d + m c_a 2^{k_I}\right) + 2^k p_{fa}(t)$$

with the success probability $\Phi((1-t)c_d)$.

In Theorem 2 we have assumed further that the restored (z_O^*, z_I^*) reveal simple independent relations between the bits of K^*. (This is mostly the case if the key schedule is affine.) Though (9) may vary according to z_O^*, we can check whether it is the same regardless of z_O^*. If, for example, each $n(S_O, S_I)$ is the same regardless of z_O^*, then so is (9). In addition, if $|\{z_O \in \mathcal{Z}_O : z_O \subset z_O^j\}| = |\{z_O \in \mathcal{Z}_O : z_O \subset z_O'^j\}|$ for any $z_O^j, z_O'^j \in \mathcal{Z}_O^j$, then $p_{\mathrm{fa}}(t)$ is bounded by $\sum_{j=1}^{m} 2^{-k_O^{(j)}} \Phi((1-t)c_d) + \Phi(-tc_d)$, where $k_O^{(j)} = \log_2 |\mathcal{Z}_O^j|$ for each j: Separate the summands in (9) into two parts using the condition $S_O \neq \emptyset$ or $S_O = \emptyset$.

- The number of z_O's for which $z_O^{(j)}$ containing z_O is correct for some j is bounded by $\sum_{j=1}^{m} 2^{k_O - k_O^{(j)}} = 2^{k_O}\left(\sum_{j=1}^{m} 2^{-k_O^{(j)}}\right)$. Note that $q(S_O, S_I) \leq c_d$ for each (S_O, S_I).
- If $S_O = \emptyset$, then $q(S_O, S_I) = \Phi(-tc_d)$ for each S_I.

So if, for example, the capacity ϵ^2 of the system of the related-key linear approximations is larger than 2^{-k+2} and it admits data compression with $2^{k_O}\left(\sum_{j=1}^{m} c_o^{(j)} 2^d + m c_a 2^{k_I}\right) \ll 2^{k-1}$ and $\sum_{j=1}^{m} 2^{-k_O^{(j)}} \ll 1$, we have an effective attack with the data complexity ϵ^{-2}.

An improved attack using a single linear trail When we use a single trail, by guessing the additional parity bit $\bigoplus_{i=0}^{r-1} \langle \gamma_{s+i}, rk_{s+i}^* \rangle \oplus g(rk^{O*})$, we can get an attack with the reduced computational complexity in some cases. The attack proceeds as in Alg. 2 with $k_I = 1$. Then we have

Corollary 1. *The computational complexity of the attack with a single linear trail using Alg. 2 is*

$$c_p N + c_o 2^{d+k_O} + 2^k p_{fa}(t)$$

with the success probability $\Phi((1-t)c_d)$*, where* $p_{fa}(t) = (2^{k_O} - 1)\Phi(-tc_d)/2^{k_O} + \Phi((-1-t)c_d)/2^{k_O+1} \approx \Phi(-tc_d)$.

So as long as $c_p N + c_o 2^{d+k_O} \ll 2^k \Phi(-tc_d)$, we have an attack whose computational complexity is about half of the previous attack performed by Alg. 1 while maintaining almost the same success probability.

3.2 Related-Key Linear Attack on the Block Ciphers with the Affine Key Schedules

The attacks described in Sect. 3.1 are particularly powerful when $k \gg n$ and the key schedule is affine, i.e., if each round key rk_i is expressed as $L_i(K) \oplus b_i$ for a linear function L_i and a constant b_i. In this case, we have the following advantages compared with other types of key schedules:

- The difference of round keys for each round can be computed directly from the key difference with probability 1. That is, $\delta rk_i = L_i(\Delta K)$. So we have a linear approximation

$$\langle \gamma_s, X_s \rangle \oplus \langle \gamma_{s+r}, X_{s+r} \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \gamma_{s+i}, L_{s+i}(\Delta K) \rangle \oplus \bigoplus_{i=0}^{r-1} \langle \gamma_{s+i}, L_{s+i}(K^*) \rangle = 0$$

  without any decrease in the absolute value of the correlation compared with the single-key setting since no additional probabilities are introduced from the key schedule.
- The data compression for adding rounds before or/and after the related-key linear approximation(s) can be performed efficiently: Each $rk_i \oplus \delta rk_i$ can be computed as the xor of rk_i and a term determined only by ΔK.
- The restored (z_O^*, z_I^*) is very likely to reveal simple independent relations between the whole key bits in Alg. 2.

Thus if a block cipher has an affine key schedule and its key length is much larger than its block length, then it is very likely to be subject to a related-key linear attack that covers more rounds than the best single-key linear attacks.

4 Related-Key Linear Attacks on SIMON

The NSA published two families of lightweight block ciphers SIMON and SPECK [3]. They have remarkable performance figures on most software and hardware platforms and SIMON is the more hardware-oriented of the two. They have been the subject of intensive security analysis since their publication. It seems that the designers of SIMON have not recognized the type of the related-key attacks presented in this work since they expect SIMON to be secure against the related-key attacks [4].

4.1 The Simon Family of Block Ciphers

SIMON n/k is a block cipher of the classical Feistel structure with k-bit keys and n-bit blocks. Its round function f sends an $n/2$-bit input x onto $((x \lll 8) \& (x \lll 1)) \oplus (x \lll 2)$.(See Fig. 3.) It has an affine key schedule. We focus on the following ciphers whose key lengths are double the block lengths: SIMON 32/64, SIMON 48/96, SIMON 64/128, and SIMON 128/256. The details of this section can be applied equally well to the variant of SIMON to be used in Sect. 5.

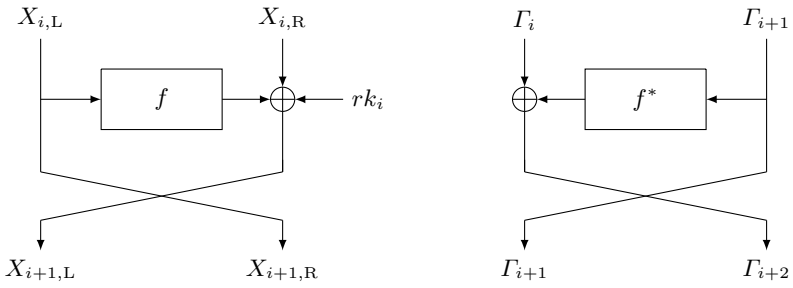

Fig. 3. A round of SIMON and a 1-round linear trail

4.2 Related-Key Linear Approximations of SIMON

Since SIMON n/k has the classical Feistel structure, an r-round linear trail can be represented as a sequence of $(r+2)$ $n/2$-bit masks: $\Gamma_s.\Gamma_{s+1}.\cdots.\Gamma_{s+r+1}$ represents a linear trail such that at the $(i+1)$-th round, the input and output masks are $\Gamma_i\|\Gamma_{i+1}$ and $\Gamma_{i+1}\|\Gamma_{i+2}$, respectively, for each $i \in [s..(s+r-1)]$. (See Fig. 3) Such a linear trail leads to the related-key linear approximation

$$\langle \Gamma_s, X_{s,\mathrm{L}}\rangle \oplus \langle \Gamma_{s+1}, X_{s,\mathrm{R}}\rangle \oplus \langle \Gamma_{s+r}, X_{s+r,\mathrm{L}}\rangle \oplus \langle \Gamma_{s+r+1}, X_{s+r,\mathrm{R}}\rangle \\ \oplus\langle \Lambda, \Delta K\rangle \oplus \bigoplus_{i=0}^{r-1}\langle \Gamma_{s+i+1}, rk^*_{s+i}\rangle = 0 \quad (10)$$

between $X_s, \Delta K, X_{s+r}$, where Λ is a mask with $\langle \Lambda, \Delta K\rangle = \bigoplus_{i=0}^{r-1}\langle \Gamma_{s+i+1}, \delta rk_{s+i}\rangle$. Such a mask exists and is easily obtained since the key schedule is affine.

4.3 Adding Outer Rounds

One of the pivotal processes of the related-key attacks as described in Alg. 1 and Alg. 2 is to get effective data compression with the related-key linear linear approximation(s). For SIMON, we can get effective data compression for prepending and appending many rounds when both the initial mask and the final mask have small Hamming weights. We will describe the way how to append outer rounds to a single linear trail from which how to do so to multiple linear trails can be easily deduced. In this subsection, we will explain in detail how to add 2+2 rounds. How to add $s+s$ rounds for $s = 3, 4, 5$ will be explained in Appendix 4.3, from which how to add $s+s'$ rounds for $s \neq s'$ and $2 \leq s, s' \leq 5$ will be obvious. For simplicity $a+b$ and ab (or $a \bullet b$) denote the XOR and AND of $a, b \in \mathbb{F}_2$ in this section, respectively.

2-round computation. Let rk_0, rk_1 be the round keys derived from the candidate key K for the first 2 rounds. For a plaintext $P = P_{\mathrm{L}}\|P_{\mathrm{R}}$ and a key difference ΔK, let $\delta rk_0, \delta rk_1$ be the derived round key differences for the first 2 rounds. Note that $\delta rk_0, \delta rk_1$ can be computed directly from ΔK since the key schedule is affine.

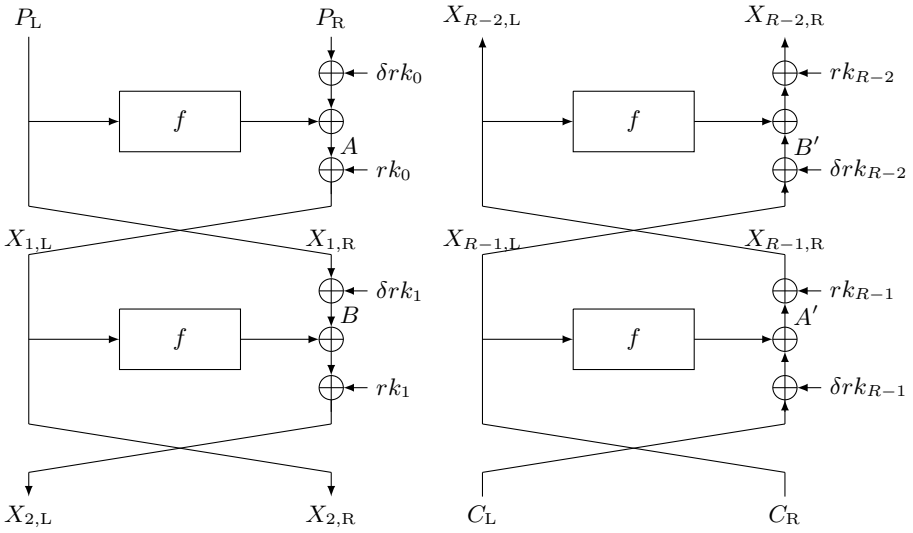

Fig. 4. 2-round computations of SIMON

We want to express each bit of $X_2 = X_{2,\mathrm{L}} \| X_{2,\mathrm{R}} = E_0^1(K \oplus \Delta K, P)$ in terms of P, rk_0, rk_1, δrk_0, and δrk_1. Let $A = f(P_{\mathrm{L}}) \oplus P_{\mathrm{R}} \oplus \delta rk_0$, and $B = P_{\mathrm{L}} \oplus \delta rk_1$. (See Fig. 4.) Since $X_{2,\mathrm{R}} = X_{1,\mathrm{L}} = f(P_{\mathrm{L}}) \oplus P_{\mathrm{R}} \oplus \delta rk_0 \oplus rk_0 = rk_0 \oplus A$ and $X_{2,\mathrm{L}} = f(X_{1,\mathrm{L}}) \oplus P_{\mathrm{L}} \oplus \delta rk_1 \oplus rk_1 = f(rk_0 \oplus A) \oplus B \oplus rk_1$,

$$\begin{aligned} X_{2,\mathrm{L}}[i] &= rk_0[i-8]rk_0[i-1] + rk_0[i-8]A[i-1] + A[i-8]rk_0[i-1] \\ &\quad + A[i-8]A[i-1] + \underline{rk_0[i-2]} + A[i-2] + B[i] + \underline{rk_1[i]}, \\ X_{2,\mathrm{R}}[i] &= \underline{rk_0[i]} + A[i] \end{aligned}$$

Here

- $X_{2,\mathrm{L}}[i]$ can be easily computed in terms of $rk_0[i-8], rk_0[i-1], A[i-8], A[i-1]$, up to $A[i-2] + B[i]$ xored with a constant determined only by rk_0, rk_1.
- The underlined terms $rk_0[i-2]$, $rk_1[i]$, and $rk_0[i]$ depending only on rk_0, rk_1 are absorbed into $g(rk^O)$ and thus they will be ignored in the case of using a single trail since they will not affect the absolute value of $\tau(z)$ for each guessed value of rk_0, rk_1. They will only affect the sign of $\tau(z)$. But they are relevant to the multiple linear attack and the improved linear attack..

By symmetry of the cipher structure, we get similar expressions for bits of $X_{R-2,\mathrm{R}}$ and $X_{R-2,\mathrm{L}}$ in terms of $A' = f(C_{\mathrm{R}}) \oplus C_{\mathrm{L}} \oplus \delta rk_{R-1}$, $B' = C_{\mathrm{R}} \oplus \delta rk_{R-2}$, rk_{R-2}, and rk_{R-1}. (See Fig. 4.)

The data compression. The above arguments tell us how to compress the data when adding 2+2 rounds. Suppose that we want to make use of the related-key linear approximation (10) with $s = 2$ and $s + r + 2 = R$. Let $w = n/2$ be the

word size. Let $\mathcal{I}_{\mathrm{L}} = \mathsf{supp}(\Gamma_s) = \{i \in [0..(w-1)] : \Gamma_s[i] \neq 0\}$, $\mathcal{I}_{\mathrm{R}} = \mathsf{supp}(\Gamma_{s+1})$, $\mathcal{I}'_{\mathrm{L}} = \mathsf{supp}(\Gamma_{R-2})$, and $\mathcal{I}'_{\mathrm{R}} = \mathsf{supp}(\Gamma_{R-1})$. The compression function extracts the following values from each data entry $(P, C, \Delta K)$:

- $A[i]$ for i such that $(i+8) \bmod w \in \mathcal{I}_{\mathrm{L}}$ or $(i+1) \bmod w \in \mathcal{I}_{\mathrm{L}}$
- $A'[i]$ for i such that $(i+8) \bmod w \in \mathcal{I}'_{\mathrm{R}}$ or $(i+1) \bmod w \in \mathcal{I}'_{\mathrm{R}}$
- $\bigoplus_{i \in \mathcal{I}_{\mathrm{L}}} (A[i-2] \oplus B[i]) \oplus \bigoplus_{i \in \mathcal{I}_{\mathrm{R}}} A[i] \oplus \bigoplus_{i \in \mathcal{I}'_{\mathrm{R}}} (A'[i-2] \oplus B'[i]) \oplus \bigoplus_{i \in \mathcal{I}'_{\mathrm{L}}} A'[i] \oplus \langle \Lambda, \Delta K \rangle$

In the attacks the outer round key bits we need to guess are as follows:

- $rk_0[i]$ for i such that $(i+8) \bmod w \in \mathcal{I}_{\mathrm{L}}$ or $(i+1) \bmod w \in \mathcal{I}_{\mathrm{L}}$
- $rk_{R-1}[i]$ for i such that $(i+8) \bmod w \in \mathcal{I}'_{\mathrm{R}}$ or $(i+1) \bmod w \in \mathcal{I}'_{\mathrm{R}}$

Note that

- the number k_O of guessed round key bits for outer rounds is at most $2\mathrm{wt}(\Gamma_s) + 2\mathrm{wt}(\Gamma_{s+r+1})$ and
- d, $\log_2$ of the size of the compressed data set, is $k_O + 1$.

4.4 The Attacks on SIMON

Now we are ready to present the attacks on Simon whose results are summarized in Table 1. Each attack is the "improved" related-key linear one using a single linear trail and we can argue that $c_o \leq R_{\mathrm{add}}/R$, where R is the number of rounds for the reduced Simon and R_{add} is the number of the added outer rounds. In each attack, the outer key classes are represented by the concatenations of the round key bits in some fixed positions that we will describe. We use the linear trails presented in Sect. C of the Appendix and set the threshold parameter t for each attack to 1 so that the success probability of each attack is $\Phi(0) = 0.5$ by Corollary 1. The attack method presented in this work can be applied to other Simon n/k with $k > n$ in a straightforward manner.

Attacks on SIMON 32/64 We have a 21-round linear trail with the correlation 2^{-30} whose initial and final mask are `00010044` and `44401010`, respectively. We can add 3+3 rounds to this linear trail:

- $k_O = 26$: Each outer key class is represented by a concatenation of 26 round key bits. The guessed round key bits are $rk_0[0, 1, 5, 6, 7, 10, 13, 14]$, $rk_1[8, 15]$, $rk_{25}[3, 4, 11, 12]$, $rk_{26}[1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14]$.
- $d = k_O + 1 = 27$.

Since $c_p \ll 1$ and $k_O + d \ll 61$, if we let $N = 2^{60}$, then $c_d = 1$ and the complexity of the improved attack on the 27-round reduced Simon 32/64 is about $2^{61.34}$. Likewise, using a 19-round trail with the correlation 2^{-26} whose initial and final mask are `00010044` and `44000100`, respectively, we get an attack on the 25-round reduced Simon 32/64 with data complexity 2^{55} and computational complexity $2^{55.26}$.

Attacks on SIMON 48/96 We have a 27-round linear trail with the correlation 2^{-47} whose initial and final mask are `000001000044` and `440040010000`, respectively. We can add 4+3 rounds to this linear trail:

- $k_O = 37$: The guessed outer round key bits are $rk_0[0, 3, 4, 6, 7, 10, 12, 13, 14, 16, 17, 19, 20, 21, 23]$, $rk_1[1, 5, 8, 14, 15, 18, 21, 22]$, $rk_2[16, 23]$, $rk_{32}[8, 15]$, and $rk_{33}[0, 5, 6, 7, 10, 13, 14, 17, 21, 22]$.
- $d = k_O + 1 = 38$.

Since $c_p \ll 1$ and $k_O + d \ll 93$, if we let $N = 2^{94}$, then $c_d = 1$ and the complexity of the improved attack on the 34-round reduced Simon is about $2^{93.34}$. Similarly, using a 23-round trail with the correlation 2^{-38} whose initial and final mask are `000001000044` and `000044000001`, respectively, we get an attack on the 30-round reduced Simon 48/96 with data complexity $2^{80.1}$ and computational complexity $2^{80.18}$.

Attacks on SIMON 64/128 We have a 36-round linear trail with the correlation 2^{-63} whose initial mask is `4000000400000001` and final mask is `0000000400000001`. We can add 4+4 rounds to this linear trail:

- $k_O = 56$: The guessed round key bits are $rk_0[1, 6, 10, 12, 13, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 29, 30, 31]$, $rk_1[0, 14, 18, 20, 21, 24, 25, 27, 28, 31]$, $rk_2[1, 22, 26, 29]$, $rk_{41}[24, 31]$, $rk_{42}[1, 6, 22, 23, 26, 29, 30]$, and $rk_{43}[0, 8, 14, 15, 18, 20, 21, 22, 24, 25, 27, 28, 29, 31]$.
- $d = k_O + 1 = 57$.

Since $c_p \ll 1$ and $k_O + d \ll 125$, if we let $N = 2^{126}$, then $c_d = 1$ and the complexity of the improved attack on the full Simon 64/128 is about $2^{125.34}$. Also, using a 31-round trail with the correlation 2^{-56} whose initial and final mask are `0000000100004044` and `0000404400000001`, respectively, we get an attack on the 38-round reduced Simon 64/128 with data complexity $2^{115.6}$ and computational complexity $2^{116.03}$.

An attack on the SIMON 128/256 We have a 62-round linear trail with the correlation 2^{-118} whose initial mask is `00000000000000040000000000000000` and final mask is `40000000000000040000000000000001`. We can add 5+5 rounds to this linear trail with $k_O \leq 39 + 2 \cdot 18 + 39 = 114$ and $d = k_O + 1 = 115$. Since $c_p \ll 1$ and $k_O + d \ll 240$, if we let $N = 2^{240.1}$, then $c_d = 2^{2.05}$ and the complexity of the improved attack on the full Simon 128/256 is about $2^{240.18}$.

5 Experiments

We carry out related-key linear attacks on an 18-round cipher that is a small-scale variant of Simon with 48-bit keys and 24-bit blocks. The round function and the key schedule of the cipher are defined exactly in the same way as Simon 32/64. The 31-bit constant used in the key schedule is also the same. We perform the following 3 experiments with the attack algorithm presented in Sect. 4 with $n = 24$:

1. A key recovery attack on the cipher using Alg. 1 by adding 2+2 rounds to a 14-round linear trail with the correlation 2^{-13}.
2. An improved key recovery attack on the 14-round reduced cipher by adding 2+2 rounds to a 10-round linear trail with the correlation 2^{-9}.
3. A related-key multiple linear attack on the 14-round reduced cipher using Alg. 2 by adding 2+2 rounds to 2 10-round linear trails with the correlations 2^{-9}.

In each experiment, the related-key data is chosen such that the number of data entries per related key is 2^{12}. Note that the first experiment is an attack using a linear trail with the correlation whose absolute value is less than $2^{-n/2}$ that alone is not so useful in the single-key linear attacks.

5.1 A Key Recovery Using a Single Linear Approximation

We perform a related-key linear attack on an 18-round toy cipher that can be regarded as SIMON 24/48. The attack proceeds as in Alg. 1 by adding 2+2 rounds to the following 14-round linear trail with the correlation 2^{-13} :

$$\texttt{001.000.001.410.001.000.001.410.001.000.001.410.001.000.001.400}$$

So the guessed outer ound key bits are $rk_0[4, 11], rk_{17}[2, 9]$, the number k_O of the guessed outer round key bits is 4, and the size of the compressed data set is 2^5. In the experiment:

- For each fixed K^* we do not recover the whole key bits. Instead we stop when we obtain the list of z_O's with $\tau(z_O) \geq tN|\epsilon|$. The attack is successful if the correct round key bits are obtained and the experimental false alarm probability is measured as the ratio of the number of the wrong entries in the list to 2^{k_O}.
- We repeat the experiment using 1,000 different keys K^* so that the experimental success probability is measured as the number of the successful attempts divided by 1,000, and the average false alarm probability is obtained as the average of the experimental false alarm probabilities.

The theoretical success probability p_S is $\Phi\left((1-t)c_d\right) + \Phi\left((-1-t)c_d\right)$ and the false alarm probability p_{fa} is $2\Phi(-tc_d)$ where $c_d = \sqrt{N}|\epsilon|$ by Theorem 1. The result is shown in Table 2 from which we can see that the experimental probabilities are close to the theoretical ones.

5.2 An Improved Key Recovery Using a Single Approximation

We perform an improved related-key linear attack on the cipher described in Sect. 5.1 reduced to 14-rounds by adding 2+2 rounds to a 10-round trail. The linear trail is

$$\texttt{001.000.001.410.001.000.001.410.001.000.001.400} \tag{11}$$

Table 2. Experimental result - key recovery using a single linear trail

	$\log_2 N$	25		26		27		28		29	
t		T	E	T	E	T	E	T	E	T	E
0.5	p_S	0.783	0.776	0.758	0.776	0.777	0.807	0.843	0.864	0.921	0.930
	p_{fa}	0.724	0.676	0.617	0.580	0.480	0.449	0.317	0.300	0.157	0.146
1.0	p_S	0.579	0.572	0.523	0.523	0.502	0.538	0.500	0.528	0.500	0.541
	p_{fa}	0.480	0.445	0.317	0.297	0.157	0.148	0.046	0.043	0.005	0.004
1.5	p_S	0.400	0.390	0.315	0.300	0.240	0.255	0.159	0.177	0.079	0.085
	p_{fa}	0.289	0.272	0.134	0.123	0.034	0.032	0.003	0.003	$<.001$	$<.001$

'T' and 'E' stand for "theoretical" and "experimental", resp.

Table 3. Experimental result - improved key recovery using a single linear trail

	$\log_2 N$	17		18		19		20		21	
t		T	E	T	E	T	E	T	E	T	E
0.5	p_S	0.638	0.640	0.691	0.706	0.760	0.759	0.841	0.835	0.921	0.924
	p_{fa}	0.344	0.345	0.291	0.293	0.225	0.228	0.149	0.151	0.074	0.077
1.0	p_S	0.500	0.502	0.500	0.529	0.500	0.512	0.500	0.500	0.500	0.497
	p_{fa}	0.227	0.230	0.149	0.152	0.074	0.075	0.021	0.022	0.002	0.002
1.5	p_S	0.401	0.422	0.362	0.365	0.309	0.320	0.240	0.257	0.159	0.162
	p_{fa}	0.137	0.138	0.063	0.063	0.016	0.016	0.001	0.001	$<.001$	$<.001$

with the correlation $\epsilon = 2^{-9}$. So the guessed outer ound key bits are $rk_0[4, 11]$ and $rk_{13}[2, 9]$, the number k_O of the guessed outer round key bits is 4, and the size of the compressed data set is 2^5. The total number of guessed key bits is $k_O + 1 = 5$ with 1 additional parity bit. Proceeding the experiment similarly as in Sect. 5.1, we get the result as shown in Table 3 from which we can see that the theoretical probabilities obtained by Corollary 1 are close to the experimental ones in this case, too.

5.3 A Key Recovery Using Multiple Approximations

We also perform a related-key multiple linear attack on the same 14-round cipher using 2 independent linear trails with the correlations 2^{-9}. One of the linear trails is (11) presented in the preceding experiment and the other is

004.000.004.041.004.000.004.041.004.000.004.001

that is obtained by rotating all the intermediate masks in (11) by 2 bits to the left. We try to restore the following 10 bits:

- 8 outer round key bits: $rk_0[1, 4, 6, 11], rk_{13}[2, 4, 9, 11]$
- 2 parity bits

The result of the experiment is shown in Table 4. The theoretical probabilities obtained by Theorem 2 are close to the experimental ones again.

Table 4. Experimental result - key recovery using 2 linear trails

	$\log_2 N$	16		17		18		19		20	
t		T	E	T	E	T	E	T	E	T	E
0.5	p_S	0.638	0.638	0.691	0.676	0.760	0.753	0.841	0.853	0.921	0.929
	p_{fa}	0.362	0.362	0.310	0.310	0.245	0.246	0.170	0.170	0.099	0.099
1.0	p_S	0.500	0.522	0.500	0.508	0.500	0.502	0.500	0.490	0.500	0.470
	p_{fa}	0.241	0.240	0.162	0.160	0.084	0.084	0.029	0.029	0.007	0.007
1.5	p_S	0.362	0.377	0.309	0.302	0.240	0.236	0.159	0.163	0.079	0.079
	p_{fa}	0.146	0.145	0.069	0.068	0.020	0.019	0.003	0.003	$<.001$	$<.001$

6 Discussions

6.1 Linear Trails versus Linear Hulls

The formulations in Sect. 3.1 indicate that in our related-key linear attacks it seems natural to use linear trails rather than linear hulls since they incorporate Matsui's Algorithm 1 and the use of the the key differences in the attacks seems to make it hard or inherently impossible to utilize the linear hulls. Though the use of a linear trail is known to lead to inaccurate analysis for single-key linear attacks [23], especially when the linear trail is not dominant, the experimental results in Sect. 5 show that our analyses of the complexities of the related-key linear attacks are accurate even when we use linear trails that are not dominant when many related keys are used. All the linear trails in the experiments in Sect. 5 are not dominant. For example, we have at least three 10-round linear trails with the correlation 2^{-9} whose initial and final mask are 001000 and 001400, respectively, other than the one used in Sect. 5.2: 001.000.001.410.001.000.001.c10.001.000.001.400, 001.000.001.c10.001.000.001.410.001.000.001.400, and 001.000.001.c10.001.000.001.c10.001.000.001.400.

6.2 Key Hypotheses

The standard key hypotheses for single-key linear attacks are regarded as rather misleading and some adjusted key hypotheses are adopted in recent works [8] [7] [11] [10]. But we claim that the standard key hypotheses are adequate for our attacks in the typical related-key scenario where the attacker has randomly chosen related-key data, or related-key data that is not concentrated at a small number of related keys. The standard *right* key hypothesis is applicable in our related-key setting since many keys are involved in the attack and the average of the empirical correlations of the linear trail as the key varies is much more uniform than when just a single key is involved. The standard *wrong* key hypotheses are also appropriate in the typical related-key scenario. In the extreme nontypical case when we have only one related key, our attack becomes one in the single-key setting and we may have to consider some adjusted wrong key hypotheses especially when the data size is large. But our hypotheses are appropriate in the typical related-key scenario. The reason is that the distribution of observed

correlations for wrong keys in the single-key setting has deviation $O(2^{-n})$ from the Daemen-Rijmen's analysis [15] on the distribution of the correlations of linear approximations for n-bit permutations. But in our related-key setting, we just need to consider $(k+n)$-bit-to-n-bit functions where the additional k-bit comes from key differences so the deviation of the distribution might be $O(2^{-(k+n)})$ that is negligible compared to $1/N$ with the data size N being much smaller than 2^{k+n}. The validity of the standard wrong key hypothesis in Matsui's Algorithm 2 in the single-key linear attack is questionable especially when the capacity is close to 2^{-n}. But the experimental result in Sect. 5.1 using a linear trail with correlation whose absolute value is less than $2^{-n/2}$ provides an evidence of our claims on the wrong key hypotheses as well as the right key hypotheses in our related-key attacks.

7 Conclusions

We have introduced a general framework of the related-key linear cryptanalysis on the block ciphers. The attack is most effective when the key schedule of the cipher is affine and in this case it is very likely to cover more rounds than the existing linear attacks if the key length of the cipher is much larger than its block length. We also have provided a method of the multiple related-key linear attack with an explicit formula for the attack complexities taking both the success probability and the false alarm probability into account. Using the framework, we are able to get effective related-key linear attacks on Simon that cover considerably longer rounds than the previous attacks. Then we have performed experiments with a small-scale variant of Simon, which corroborated the validity of the attack together with the suitability of the key hypotheses. Though the attack may be hard to apply in practice due to the requirement of large related-key data, it is certainly one that should be taken into account in the design of a block cipher with a small block length and a simple key schedule. For example, a block cipher with a small block length and an affine key schedule that is marginally secure against the single-key linear attacks may undergo a practical break by the attack method presented in this paper. A remarkable feature of our attack is that the complexity of the attack does not depend much on the detail of the key schedule once the key schedule is affine. As possible extensions of the current work, the followings seem worth investigation:

- Applying the attack method presented in this work to various block ciphers with affine key schedules other than Simon.
- Extending the current method to various cryptographic constructions based on block ciphers.

References

1. Banik, S., Bogdanov, A., Isobe, T., Shibutani, K., Hiwatari, H., Akishita, T., Regazzoni, F.: Midori: A block cipher for low energy. In: Iwata, T., Cheon, J.H.

(eds.) ASIACRYPT 2015, Part II. LNCS, vol. 9453, pp. 411–436. Springer (2015), https://doi.org/10.1007/978-3-662-48800-3_17
2. Banik, S., Pandey, S.K., Peyrin, T., Sasaki, Y., Sim, S.M., Todo, Y.: GIFT: A small present - towards reaching the limit of lightweight encryption. In: Fischer, W., Homma, N. (eds.) CHES 2017. LNCS, vol. 10529, pp. 321–345. Springer (2017), https://doi.org/10.1007/978-3-319-66787-4_16
3. Beaulieu, R., Shors, D., Smith, J., Treatman-Clark, S., Weeks, B., Wingers, L.: The simon and speck families of lightweight block ciphers. Cryptology ePrint Archive, Report 2013/404 (2013), http://eprint.iacr.org/2013/404
4. Beaulieu, R., Shors, D., Smith, J., Treatman-Clark, S., Weeks, B., Wingers, L.: Notes on the design and analysis of simon and speck. Cryptology ePrint Archive, Report 2017/560 (2017), http://eprint.iacr.org/2017/560
5. Beierle, C., Jean, J., Kölbl, S., Leander, G., Moradi, A., Peyrin, T., Sasaki, Y., Sasdrich, P., Sim, S.M.: The SKINNY family of block ciphers and its low-latency variant MANTIS. In: Robshaw, M., Katz, J. (eds.) CRYPTO 2016, Part II. LNCS, vol. 9815, pp. 123–153. Springer (2016), https://doi.org/10.1007/978-3-662-53008-5_5
6. Biryukov, A., Cannière, C.D., Quisquater, M.: On multiple linear approximations. In: Franklin, M.K. (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 1–22. Springer (2004), https://doi.org/10.1007/978-3-540-28628-8_1
7. Blondeau, C., Nyberg, K.: Improved parameter estimates for correlation and capacity deviates in linear cryptanalysis. IACR Trans. Symmetric Cryptol. 2016(2), 162–191 (2016), https://doi.org/10.13154/tosc.v2016.i2.162-191
8. Blondeau, C., Nyberg, K.: Joint data and key distribution of simple, multiple, and multidimensional linear cryptanalysis test statistic and its impact to data complexity. Des. Codes Cryptography 82(1-2), 319–349 (2017), https://doi.org/10.1007/s10623-016-0268-6
9. Bogdanov, A., Boura, C., Rijmen, V., Wang, M., Wen, L., Zhao, J.: Key difference invariant bias in block ciphers. In: Sako, K., Sarkar, P. (eds.) ASIACRYPT 2013, Part I. LNCS, vol. 8269, pp. 357–376. Springer (2013), https://doi.org/10.1007/978-3-642-42033-7_19
10. Bogdanov, A., Tischhauser, E.: On the wrong key randomisation and key equivalence hypotheses in matsui's algorithm 2. In: Moriai, S. (ed.) FSE 2013, Revised Selected Papers. LNCS, vol. 8424, pp. 19–38. Springer (2013), https://doi.org/10.1007/978-3-662-43933-3_2
11. Bogdanov, A., Vejre, P.S.: Linear cryptanalysis of DES with asymmetries. In: Takagi, T., Peyrin, T. (eds.) ASIACRYPT 2017, Part I. LNCS, vol. 10624, pp. 187–216. Springer (2017), https://doi.org/10.1007/978-3-319-70694-8_7
12. Borghoff, J., Canteaut, A., Güneysu, T., Kavun, E.B., Knezevic, M., Knudsen, L.R., Leander, G., Nikov, V., Paar, C., Rechberger, C., Rombouts, P., Thomsen, S.S., Yalçin, T.: PRINCE - A low-latency block cipher for pervasive computing applications - extended abstract. In: Wang, X., Sako, K. (eds.) ASIACRYPT 2012. LNCS, vol. 7658, pp. 208–225. Springer (2012), https://doi.org/10.1007/978-3-642-34961-4_14
13. Cannière, C.D., Dunkelman, O., Knezevic, M.: KATAN and KTANTAN - A family of small and efficient hardware-oriented block ciphers. In: Clavier, C., Gaj, K. (eds.) CHES 2009. LNCS, vol. 5747, pp. 272–288. Springer (2009), https://doi.org/10.1007/978-3-642-04138-9_20
14. Chen, H., Wang, X.: Improved linear hull attack on round-reduced simon with dynamic key-guessing techniques. In: Peyrin, T. (ed.) FSE 2016, Revised Selected

Papers. LNCS, vol. 9783, pp. 428–449. Springer (2016), https://doi.org/10.1007/978-3-662-52993-5_22
15. Daemen, J., Rijmen, V.: Probability distributions of correlation and differentials in block ciphers. Cryptology ePrint Archive, Report 2005/212 (2005), http://eprint.iacr.org/2005/212
16. Gérard, B., Grosso, V., Naya-Plasencia, M., Standaert, F.: Block ciphers that are easier to mask: How far can we go? In: Bertoni, G., Coron, J. (eds.) CHES 2013. LNCS, vol. 8086, pp. 383–399. Springer (2013), https://doi.org/10.1007/978-3-642-40349-1_22
17. Guo, J., Peyrin, T., Poschmann, A., Robshaw, M.J.B.: The LED block cipher. In: Preneel and Takagi [24], pp. 326–341, https://doi.org/10.1007/978-3-642-23951-9_22
18. Hermelin, M., Nyberg, K.: Linear cryptanalysis using multiple linear approximations. IACR Cryptology ePrint Archive 2011, 093 (2011), http://eprint.iacr.org/2011/093
19. Kaliski, B.S., Robshaw, M.J.B.: Linear cryptanalysis using multiple approximations. In: Desmedt, Y. (ed.) CRYPTO '94. LNCS, vol. 839, pp. 26–39. Springer (1994), https://doi.org/10.1007/3-540-48658-5_4
20. Kölbl, S., Leander, G., Tiessen, T.: Observations on the SIMON block cipher family. In: Gennaro, R., Robshaw, M. (eds.) CRYPTO 2015, Part I. LNCS, vol. 9215, pp. 161–185. Springer (2015), https://doi.org/10.1007/978-3-662-47989-6_8
21. Liu, Z., Li, Y., Wang, M.: The security of simon-like ciphers against linear cryptanalysis. Cryptology ePrint Archive, Report 2017/576 (2017), http://eprint.iacr.org/2017/576
22. Matsui, M.: Linear cryptanalysis method for DES cipher. In: Helleseth, T. (ed.) EUROCRYPT '93. LNCS, vol. 765, pp. 386–397. Springer (1993), https://doi.org/10.1007/3-540-48285-7_33
23. Nyberg, K.: Linear approximation of block ciphers. In: Santis, A.D. (ed.) EUROCRYPT '94. LNCS, vol. 950, pp. 439–444. Springer (1994), https://doi.org/10.1007/BFb0053460
24. Preneel, B., Takagi, T. (eds.): CHES 2011, LNCS, vol. 6917. Springer (2011), https://doi.org/10.1007/978-3-642-23951-9
25. Selçuk, A.A.: On probability of success in linear and differential cryptanalysis. J. Cryptology 21(1), 131–147 (2008), https://doi.org/10.1007/s00145-007-9013-7
26. Shibutani, K., Isobe, T., Hiwatari, H., Mitsuda, A., Akishita, T., Shirai, T.: Piccolo: An ultra-lightweight blockcipher. In: Preneel and Takagi [24], pp. 342–357, https://doi.org/10.1007/978-3-642-23951-9_23
27. Vora, P.L., Mir, D.J.: Related-key linear cryptanalysis. ISIT '06: Proceedings of the 2006 IEEE International Symposium of Information Theory 2006, 1609–1613 (July 2006)

A Proof of Lemma 3

Lemma 3 follows easily from Lemma 4.

Lemma 4. *Let $a_1, \ldots, a_m, b$ be real numbers with at least one a_j nonzero and let $Z_1, \ldots, Z_m$ be independent random variables with $Z_j \sim \mathrm{N}(0,1)$ for each j. Then*

$$\Pr\Big(\sum_{j=1}^{m} a_j Z_j \le b\Big) = \Pr\Big(\sum_{j=1}^{m} a_j Z_j \ge -b\Big) = \Phi\left(b/\Big(\sum_{j=1}^{m} a_j^2\Big)^{1/2}\right)$$

Proof. Replacing Z_j by $-Z_j$ if necessary, we may assume that $a_j \geq 0$ for each j. Let p be the density function defined by $p(x_1, \ldots, x_m) = (\frac{1}{\sqrt{2\pi}})^m exp\left(-\sum_{j=1}^{m} x_j^2/2\right)$ and let $\mathcal{D}$ be the half space $\{(x_1, \ldots, x_m) \in \mathbb{R}^m : \sum_{j=1}^{m} a_j x_j \leq b\}$ in $\mathbb{R}^m$. Let ρ be a rigid motion in $\mathbb{R}^m$ having 0 as a fixed point that maps $\mathcal{D}$ to the half space $\{(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m : x_1 \leq b/(\sum_{j=1}^{m} a_j^2)^{1/2}\}$. Since p is invariant under the rigid motions having 0 as a fixed point,

$$\begin{aligned}\Pr\left(\sum_{j=1}^{m} a_j Z_j \leq b\right) &= \int_{\mathcal{D}} p\, dx \\ &= \int_{\rho(\mathcal{D})} p\, dx = \Phi\left(b/\left(\sum_{j=1}^{m} a_j^2\right)^{1/2}\right).\end{aligned}$$

□

Now Lemma 3 can be proved as follows:

Proof. Let $Z_j = (Y_j - \mu_j)/\sigma_j$ for each j. Then each Z_j follows $\mathrm{N}(0, 1)$ and $\sum_{j=1}^{m} a_j Y_j \geq b$ if and only if $\sum_{j=1}^{m} a_j \sigma_j Z_j \geq b - \sum_{j=1}^{m} a_j \mu_j$. So the lemma follows by Lemma 4. □

B A searching method for the linear trails of SIMON

Linear trails with the correlations whose absolute values are less than $2^{-n/2}$ for Simon n/k are not found in the previous works. To find good linear trails of Simon n/k whose correlations are around $\pm 2^{-n}$, we apply an efficient search algorithm that we have designed by slightly modifying the searching method described in [21] that is based on M. Matsui's branch-and-bound algorithm. As noted before, an r-round linear trail of Simonn/k can be represented as a sequence of $(r+2)$ $n/2$-bit masks: $\Gamma_s.\Gamma_{s+1}.\cdots.\Gamma_{s+r+1}$ represents a linear trail such that at the $(i+1)$-th round, the input and output masks are $\Gamma_i \| \Gamma_{i+1}$ and $\Gamma_{i+1} \| \Gamma_{i+2}$ for each $i \in [s..(s+r-1)]$. Such a linear trail leads to the related-key linear approximation (10). The linear correlation of the round function of Simon with respect to various input-output mask pairs can be easily computed as explained in [20] and [21]. Throughout this section, f denotes the round function of Simon n/k that sends an $n/2$-bit input x onto $((x \lll 8)\&(x \lll 1)) \oplus (x \lll 2)$. The following is a basic fact about the linear trails of Simon noted in the previous works:

- If $\Gamma_s.\cdots.\Gamma_{s+r+1}$ is a linear trail with the correlation ϵ, so are the reversed trail $\Gamma_{s+r+1}.\cdots.\Gamma_s$ and the rotated trail $(\Gamma_s \lll l).\cdots.(\Gamma_{s+r+1} \lll l)$ for each l.

For each $l \geq 0$, let $\mathcal{L}_l$ be the list of mask pairs (α, β) such that $\varepsilon_f(\alpha, \beta) = 2^{-l}$. We denote $-\log_2(\varepsilon_f(\beta \ggg 2, \beta))$ by $\mathrm{lac}(\beta)$. Then the mask pairs in $\mathcal{L}_l$ are exactly those (α, β)'s for which $\mathrm{lac}(\beta) = l$ and $\varepsilon_f(\alpha, \beta) \neq 0$. A mask β is called a

Table 5. $-\log_2|\text{correlation}|$ of the best linear trails found

block length | # rounds	21	22	23	24	27	28	29	30	35	36	37	38	39
32	28	31	33	37	-	-	-	-	-	-	-	-	-
48	34	36	37	39	45	48	49	52	-	-	-	-	-
64	34	36	37	39	45	48	49	52	62	63	64	66	67

rotational representative if $(\beta \lll j) \geq \beta$ for each j. For each $l \geq 0$, let $\mathcal{L}_l^{\text{red}}$ be the set of mask pairs (α, β) in $\mathcal{L}_l$ such that β is a rotational representative. For an r-round trail $T = \Gamma_0.\Gamma_1.\cdots.\Gamma_{r+1}$ and $l \leq r+1$, T_l denotes Γ_l and $\mathrm{lac}(T)$ denotes $\sum_{l=1}^{r} \mathrm{lac}(\Gamma_l)$. Also for $l \leq r$, $T|_l$ denotes the l-round subtrail $\Gamma_0.\Gamma_1.\cdots.\Gamma_{l+1}$. When we search for an r-round trail T, we impose the following restrictions on the masks in the trail:

- $\mathrm{lac}(T_i) \leq 4$ for $i = 0, \ldots, r$.
- If $T_1 = 0$, T_0 is a rotational representative. Otherwise, T_1 is a rotational representative.

It turns out that these restrictions let us quickly find out the trails suitable for our purposes. For each r, let B_r be the minimum of $\mathrm{lac}(T)$ when T runs among all the r-round trails such that each mask in the trail except the last one has lac ≤ 4. It is easy to see that $B_1 = 0$, $B_2 = 1$, and $B_3 = 2$. In the search algorithm, for each $r \geq 4$, we get B_r and an r-round linear trail T with $\mathrm{lac}(T) = B_r$ assuming that we have already computed B_1, ..., B_{r-1}. The search algorithm is presented as Alg. 3. Before running it, we prepare two lists of $n/2$-bit masks in advance for acceleration: One with β's such that $\mathrm{lac}(\beta) \leq 4$ and the other with β's such that $\mathrm{lac}(\beta) \leq 4$ *and* β is a rotational representative. B_r's we have obtained are the same as those presented in [21] for Simon n/k with n=32, 48, or 64 whenever $B_r \leq n$. We remark that by modifying Alg. 3, we can also find many linear trails with large correlations and various constraints on the intermediate masks.

C Linear trails of SIMON

We get linear trails for Simon with large correlations as in Table 5, using the algorithm described in the preceding section. But we also get trails with large correlations that have small Hamming weights for the initial and final masks. The linear trails we have used in our attacks on Simon in Sect. 4.4 are as follows:

- a 21-round trail for Simon 32/64 with the correlation 2^{-30}:
  0001.0044.0010.0040.0000.0040.0010.0044.0001.4044.1010.4440.0100.4400.1000.4000.0000.4000.1000.4400.0100.4440.1010
- a 27-round trail for Simon 48/96 with the correlation 2^{-47}:
  000001.000044.000010.000040.000000.000040.000010.000044.000001.400044.100010.440040.610000.4c0040.300010.400044.000001.000044.000010.000040.000000.000040.000010.000044.000001.400044.100010.440040.010000

Alg. 3 The search algorithm

Set $\bar{B} = B_{r-1} - 1$, and $found = 0$.
repeat
  $\bar{B}$++
  ProcessR1( )
until $found == 1$
Output T, $\bar{B}$

function ProcessR1( )
  ProcessR2A( )
  for $l \leftarrow 1$ **to** $\min(4, \bar{B} - B_{r-1})$ **do**
    for $(\alpha, \beta) \in \mathcal{L}_l^{\text{red}}$ **do**
      ProcessR2(α, β)
    end for
  end for
  return
end function

function ProcessR2A( )
  for $l \leftarrow 1$ **to** $\min(4, \bar{B} - B_{r-2})$ **do**
    for $(\alpha, \beta) \in \mathcal{L}_l^{\text{red}}$ **do**
      Set $T_0 = \beta$, $T_1 = 0$, $T_2 = \beta$, $T_3 = \alpha$.
      ProcessR(3)
    end for
  end for
  return
end function

function ProcessR2(α,β)
  Set $c = \text{lac}(\beta)$.
  for $l \leftarrow 0$ **to** $\min(4, \bar{B} - c - B_{r-2})$ **do**
    for $(\alpha_1, \beta_1) \in \mathcal{L}_l$ for which $\text{lac}(\alpha \oplus \beta_1) \leq 4$ **do**
      Set $T_0 = \alpha \oplus \beta_1$, $T_1 = \beta$, $T_2 = \beta_1$, $T_3 = \beta \oplus \alpha_1$.
      ProcessR(3)
    end for
  end for
  return
end function

- a 31-round trail for Simon 64/128 with the correlation 2^{-56}:
  00000001.00004044.00001010.00004440.00000100.00004400.00001000.00004000.00000000.00004000.00001000.00004400.00000100.00004440.00001010.00004044.00000061.0000404c.00001030.00004440.00000100.00004400.00001000.00004000.00000000.00004000.00001000.00004400.00000100.00004440.00001010.00004044.00000001.
- a 36-round trail for Simon 64/128 with the correlation 2^{-63}:
  40000004.00000001.00000004.00000000.00000004.00000001.400000

Alg. 3 The search algorithm (continued)

```
function ProcessR(m)
    Set β_m = T_m, c = lac(β_m).
    if m < r then
        if (lac(T|_{m-1}) + c + B_{r-m} > B̄) or (c > 4) then
            return
        else
            for α_m for which ε_f(α_m, β_m) ≠ 0 do
                Set T_{m+1} = α_m ⊕ T_{m-1}.
                ProcessR(m + 1)
            end for
        end if
    else
        if lac(T|_{m-1}) + c == B̄ then
            Choose an α_m for which ε_f(α_m, β_m) ≠ 0.
            Set T_{m+1} = α_m ⊕ T_{m-1}.
            Set found = 1, and exit all the functions.
        else
            return
        end if
    end if
end function
```

04.10000000.44000004.01000001.04400004.06100000.04c00004.03000001.44000004.10000000.40000004.00000001.00000004.00000000.00000004.00000001.40000004.10000000.44000004.01000001.04400004.06100000.04c00004.03000001.44000004.10000000.40000004.00000001.00000004.00000000.00000004.00000001

– a 62-round trail for Simon 128/256 with the correlation 2^{-118}:
0000000000000004.0000000000000000.0000000000000004.0000000000000001.$\cdots$.0000000000000004.0000000000000000.0000000000000004.0000000000000001.$\cdots$.1000000000000000.4000000000000004.0000000000000001

Linear trails for Simon 32/64, Simon 48/96, Simon 64/128 were found by the search algorithm described in the preceding section. The linear trail for Simon 128/256 was constructed from an iterative trail in [21].

D Adding More Rounds to Related-Key Linear Approximations of Simon

In this section, we will explain to add $r_{\text{pre}} + r_{\text{post}}$ rounds for $(r_{\text{pre}}, r_{\text{post}}) = (3,3)$, (4,4), or (5,5). For simplicity $a + b$ and ab (or $a \bullet b$) denote the XOR and AND of $a, b \in \mathbb{F}_2$ in this section, respectively. Let $w = n/2$ be the word size as before.

Adding 3+3 rounds

3-round computation. Let rk_0, rk_1, rk_2 be the round keys derived from the candidate key K for the first 3 rounds. For a plaintext $P = P_{\mathrm{L}} \| P_{\mathrm{R}}$ and a key difference ΔK, let $\delta rk_0, \delta rk_1, \delta rk_2$ be the derived round key differences for the first 3 rounds. Let $A = f(P_{\mathrm{L}}) \oplus P_{\mathrm{R}} \oplus \delta rk_0$, and $B = P_{\mathrm{L}} \oplus \delta rk_1$ as before. Then using the relations $X_{3,\mathrm{R}} = X_{2,\mathrm{L}}$ and $X_{3,\mathrm{L}} = f(X_{2,\mathrm{L}}) \oplus X_{2,\mathrm{R}} \oplus \delta rk_2 \oplus rk_2$, we can compute each bit of $X_3 = X_{3,\mathrm{L}} \| X_{3,\mathrm{R}} = E_0^2(K \oplus \Delta K, P)$ in terms of bits of $(A, B, rk_0, rk_1, rk_2, \delta rk_0, \delta rk_1, \delta rk_2)$ as follows:

$$
\begin{aligned}
X_{3,\mathrm{L}}[i] = \big(&rk_0[i-9]rk_0[i-2] + rk_0[i-9]A[i-2] + A[i-9]rk_0[i-2] \\
&+A[i-9]A[i-2] + rk_0[i-3] + A[i-3] + B[i-1] + rk_1[i-1]\big) \bullet \\
&\big(rk_0[i-16]rk_0[i-9] + rk_0[i-16]A[i-9] + A[i-16]rk_0[i-9] \\
&+A[i-16]A[i-9] + rk_0[i-10] + A[i-10] + B[i-8] + rk_1[i-8]\big) \\
&+rk_0[i-10]rk_0[i-3] + rk_0[i-10]A[i-3] + A[i-10]rk_0[i-3] \\
&+A[i-10]A[i-3] + \underline{rk_0[i-4]} + A[i-4] + B[i-2] + \underline{rk_1[i-2]} \\
&+\underline{rk_0[i]} + A[i] + \underline{rk_2[i]} + \delta rk_2[i], \\
X_{3,\mathrm{R}}[i] = X_{2,\mathrm{L}}[i].&
\end{aligned}
$$

Thus

- $X_{3,\mathrm{L}}[i]$ can be easily computed in terms of $rk_0[i-9]$, $rk_0[i-2]$, $rk_0[i-3]$, $rk_0[i-16]$, $rk_0[i-10]$, $rk_1[i-8]$, $rk_1[i-1]$, $A[i-9]$, $A[i-2]$, $A[i-3]$, $A[i-16]$, $A[i-10]$, $B[i-1]$, $B[i-8]$, xored with $A[i-4] + B[i-2] + A[i] + \delta rk_2[i]$ and a constant determined only by rk_0, rk_1, rk_2.
- $X_{3,\mathrm{R}}[i]$ can be easily computed in terms of $rk_0[i-8], rk_0[i-1], A[i-8], A[i-1]$, xored with $A[i-2] + B[i]$ and a constant determined only by rk_0, rk_1, rk_2.
- The underlined terms that depend only on the outer round keys rk_0, rk_1, rk_2 will be ignored in the attack using a single trail as in the 2-round computations.

By symmetry of the cipher structure, we get similar expressions for bits of $X_{R-3,\mathrm{R}}$ and $X_{R-3,\mathrm{L}}$ in terms of $A' = f(C_{\mathrm{R}}) \oplus C_{\mathrm{L}} \oplus \delta rk_{R-1}$, $B' = C_{\mathrm{R}} \oplus \delta rk_{R-2}$, δrk_{R-3}, rk_{R-1}, rk_{R-2}, and rk_{R-3}.

The data compression. Suppose that we want to make use of the related-key linear approximation (10) with $s = 3$ and $s + r + 3 = R$. Let $\mathcal{I}_{\mathrm{L}} = \mathsf{supp}(\Gamma_s)$, $\mathcal{I}_{\mathrm{R}} = \mathsf{supp}(\Gamma_{s+1})$, $\mathcal{I}'_{\mathrm{L}} = \mathsf{supp}(\Gamma_{R-3})$, and $\mathcal{I}'_{\mathrm{R}} = \mathsf{supp}(\Gamma_{R-2})$. The compression function extracts the following values from each data entry $(P, C, \Delta K)$:

- $A[i]$ for i such that one of $i+9, i+2, i+3, i+16, i+10 \bmod w$ is in $\mathcal{I}_{\mathrm{L}}$
- $B[i]$ for i such that one of $i+1, i+8 \bmod w$ is in $\mathcal{I}_{\mathrm{R}}$
- $A'[i]$ for i such that one of $i+9, i+2, i+3, i+16, i+10 \bmod w$ is in $\mathcal{I}'_{\mathrm{R}}$
- $B'[i]$ for i such that one of $i+1, i+8 \bmod w$ is in $\mathcal{I}'_{\mathrm{L}}$
- $\bigoplus_{i \in \mathcal{I}_{\mathrm{L}}} (A[i-4] + B[i-2] + A[i] + \delta rk_2[i]) \oplus \bigoplus_{i \in \mathcal{I}_{\mathrm{R}}} (A[i-2] + B[i]) \oplus \bigoplus_{i \in \mathcal{I}'_{\mathrm{R}}} (A'[i-4] + B'[i-2] + A'[i] + \delta rk_{R-3}[i]) \oplus \bigoplus_{i \in \mathcal{I}'_{\mathrm{L}}} (A'[i-2] + B'[i]) \oplus \langle \Lambda, \Delta K \rangle$

In the attacks the outer round key bits we need to guess are as follows:

- $rk_0[i]$ for i such that one of $i+9, i+2, i+3, i+16, i+10 \bmod w$ is in $\mathcal{I}_{\mathrm{L}}$
- $rk_1[i]$ for i such that one of $i+1, i+8 \bmod w$ is in $\mathcal{I}_{\mathrm{R}}$
- $rk_{R-1}[i]$ for i such that one of $i+9, i+2, i+3, i+16, i+10 \bmod w$ is in $\mathcal{I}'_{\mathrm{R}}$
- $rk_{R-2}[i]$ for i such that one of $i+1, i+8 \bmod w$ is in $\mathcal{I}'_{\mathrm{L}}$

Note that

- the number k_O of guessed round key bits for outer rounds is at most $k_O \leq 7\mathrm{wt}(\Gamma_s) + 2\mathrm{wt}(\Gamma_{s+1}) + 2\mathrm{wt}(\Gamma_{s+r}) + 7\mathrm{wt}(\Gamma_{s+r+1})$ and
- $d = k_O + 1$.

Adding 4+4 rounds Let $A = f(P_{\mathrm{L}}) \oplus P_{\mathrm{R}} \oplus \delta rk_0$, and $B = P_{\mathrm{L}} \oplus \delta rk_1$ as before. Using the relations $X_{4,\mathrm{R}} = X_{3,\mathrm{L}}$ and $X_{4,\mathrm{L}} = f(X_{3,\mathrm{L}}) \oplus X_{3,\mathrm{R}} \oplus \delta rk_3 \oplus rk_3$, we see that $X_{4,\mathrm{L}}[i]$ (up to a constant determined only by rk_0, rk_1, rk_2, rk_3) is a function of the following values and round key bits

- $rk_0[i-1, i-3, i-4, i-5, i-8, i-10, i-11, i-12, i-17, i-18, i-24]$, $A[i-1, i-3, i-4, i-5, i-8, i-10, i-11, i-12, i-17, i-18, i-24]$
- $rk_1[i-2, i-3, i-9, i-10, i-16]$, $B[i-2, i-3, i-9, i-10, i-16]$
- $rk_2[i-1, i-8]$, $\delta rk_2[i-1, i-8]$

xored by $A[i-6] + B[i] + B[i-4] + \delta rk_2[i-2] + \delta rk_3[i]$. Note also that $X_{4,\mathrm{R}}[i] = X_{3,\mathrm{L}}[i]$ and the backward computations can be carried out similarly. So when we use a single trail, we have a compression with

- $k_O \leq 18\mathrm{wt}(\Gamma_s) + 7\mathrm{wt}(\Gamma_{s+1}) + 7\mathrm{wt}(\Gamma_{s+r}) + 18\mathrm{wt}(\Gamma_{s+r+1})$ and
- $d = k_O + 1$.

Adding 5+5 rounds Let $A = f(P_{\mathrm{L}}) \oplus P_{\mathrm{R}} \oplus \delta rk_0$, and $B = P_{\mathrm{L}} \oplus \delta rk_1$. $X_{4,\mathrm{L}}[i]$ (up to a constant determined only by $rk_0, rk_1, rk_2, rk_3, rk_4$) is a function of the following values and round key bits

- $rk_0[i-2, i-3, i-4, i-5, i-6, i-7, i-9, i-10, i-11, i-12, i-13, i-14, i-16, i-18, i-19, i-20, i-25, i-26, i-32]$, $A[i-2, i-3, i-4, i-5, i-6, i-7, i-9, i-10, i-11, i-12, i-13, i-14, i-16, i-18, i-19, i-20, i-25, i-26, i-32]$
- $rk_1[i-1, i-3, i-4, i-5, i-8, i-10, i-11, i-12, i-17, i-18, i-24]$, $B[i-1, i-3, i-4, i-5, i-8, i-10, i-11, i-12, i-17, i-18, i-24]$
- $rk_2[i-1, i-2, i-3, i-8, i-9, i-10, i-16]$, $\delta rk_2[i-1, i-2, i-3, i-8, i-9, i-10, i-16]$
- $rk_3[i-1, i-8]$, $\delta rk_3[i-1, i-8]$

xored by $A[i] + A[i-4] + A[i-8] + B[i-6] + \delta rk_2[i] + \delta rk_2[i-4] + \delta rk_3[i-2] + \delta rk_4[i]$. We also have $X_{5,\mathrm{R}}[i] = X_{4,\mathrm{L}}[i]$ and the backward computations can be carried out similarly. So when we use a single trail, we have a compression with

- $k_O \leq 39\mathrm{wt}(\Gamma_s) + 18\mathrm{wt}(\Gamma_{s+1}) + 18\mathrm{wt}(\Gamma_{s+r}) + 39\mathrm{wt}(\Gamma_{s+r+1})$ and
- $d = k_O + 1$.