# A Cryptanalysis of PRINTcipher: The Invariant Subspace Attack

Gregor Leander, Mohamed Ahmed Abdelraheem,
Hoda AlKhzaimi, and Erik Zenner

Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark
{G.Leander,M.A.Abdelraheem,H.Alkhzaimi,E.Zenner}@mat.dtu.dk

**Abstract.** At CHES 2010, the new block cipher PRINTcipher was presented as a light-weight encryption solution for printable circuits [15]. The best attack to date is a differential attack [1] that breaks less than half of the rounds. In this paper, we will present a new attack called *invariant subspace attack* that breaks the full cipher for a significant fraction of its keys. This attack can be seen as a weak-key variant of a statistical saturation attack. For such weak keys, a chosen plaintext distinguishing attack can be mounted in unit time. In addition to breaking PRINTcipher, the new attack also gives us new insights into other, more well-established attacks. We derive a truncated differential characteristic with a round-independent but highly key-dependent probability. In addition, we also show that for weak keys, strongly biased linear approximations exists for any number of rounds. In this sense, PRINTcipher behaves very differently to what is usually – often implicitly – assumed.

**Keywords:** Symmetric cryptography, block cipher, invariant subspace attack, truncated differentials, linear cryptanalysis, statistical saturation attack.

## 1 Introduction

Block ciphers are often said to be amongst the best understood subjects in the area of symmetric cryptography. Compared to – for example – stream ciphers and hash functions, the design of a secure block cipher is probably more straightforward. However, designing a secure block cipher that is at the same time very efficient is still challenging.

Incidentally, most recent block cipher proposals aim for the area of light-weight cryptography [3,5,13]. Light-weight cryptography provides building blocks for secure communication on extremely constrained devices. The constraints are mainly cost driven and result in highly limited computing power, chip area and/or power supply. It is an ongoing competition to design the most efficient block cipher for such devices. This competition resulted in more and more aggressive designs that often show two characteristics: (1) Innovative techniques are used to improve upon known ciphers, often leading to less standard and thus less well-understood designs. (2) The security margins that block ciphers are

P. Rogaway (Ed.): CRYPTO 2011, LNCS 6841, pp. 206–221, 2011.

traditionally equipped with are reduced as much as possible in order to optimize the cipher performance.

Unsurprisingly, this has led to a number of attacks against these newer designs [4,7,11]. In addition to constituting a break of the light-weight cipher in question, these attacks sometimes also have an additional quality: They improve our understanding of block ciphers in general. Note that an attack that breaks a light-weight cipher may be prevented by a conventional block cipher not by design, but by accident: Even though the attack was not even known by the time of designing the cipher, it may not pose a threat to the cipher simply because of the security margin.

In the following, we will present such a new attack called *invariant subspace attack* that breaks the block cipher PRINTcipher [15] proposed at CHES 2010. The best currently known analysis of PRINTcipher is a differential-style attack presented at FSE 2011 [1] that could be applied for less than half the rounds of PRINTcipher, making use of the full code book. Apart from breaking PRINTcipher and providing us with a new tool for attacking block ciphers, the invariant subspace attack also displays interesting relationships to other well-established attack techniques that increase our understanding of block cipher cryptanalysis in general.

### 1.1 Our Contribution

In this paper, detailed in Section 2.2, we present a new attack on PRINTcipher. In a nutshell, the attack is based on the observation that for PRINTcipher there exist cosets of subspaces of $\mathbb{F}_2^n$ that the round function maps to cosets of the same subspace. The exact coset is determined by the round key only. Now, if the round key is such that a coset gets mapped to itself, the fact that all round keys are identical in PRINTcipher (almost) immediately leads to the conclusion that for certain (weak) keys some affine subspaces are invariant under encryption. The round constants, mainly introduced to avoid slide attacks, do not prevent the attack as the round constants are included in the subspace. The principle of the attack is described in Section 2.1.

More particular, using this new attack technique, which we call (for obvious reasons) *invariant subspace attack*, we demonstrate the existence of $2^{52}$ weak keys (out of $2^{80}$) for PRINTcipher-48 and $2^{102}$ weak keys (out of $2^{160}$) for PRINTcipher-96. If a key is weak, our attack results in a distinguisher using less than 5 chosen plain- or ciphertexts. That is, even in the case of RFID-tags, where the amount of data available for a practical attack is strictly limited, our attacks apply. In a known plain- or ciphertext scenario the data complexity increases by a factor of $2^{16}$ (PRINTcipher-48) resp. $2^{32}$ (PRINTcipher-96).

Besides the low data complexity of the distinguisher, the attack technique has interesting relations to more established attacks which we like to highlight. Firstly, see Section 3, the invariant subspace attack implies a truncated differential attack, where the probability of the truncated differential characteristic is highly key-dependent. For a weak key, this probability is $2^{-16}$, independent of

the number of rounds – while for a non-weak key the probability is zero for any number of rounds greater or equal to two.

Secondly, the invariant subspace attack can be interpreted as a statistical saturation attack [7,8]. Here a weak key, together with a special choice of the fixed bits in a statistical saturation attack, leads to a maximal bias, independent of the number of rounds. Taking into account the close relation of statistical saturation attacks to multi-dimensional linear attacks, we show that the invariant subspace attack implies the existence of strongly biased linear approximations for weak keys, again independent of the number of rounds. Details can be found in Section 4.

It follows in particular that PRINTcipher is an example of a non-toy cipher where attacks do not behave as we usually expect them to. The probability of truncated differential characteristics, the bias for statistical saturation attacks, and the bias of linear hulls are extremely key-dependent. For a weak key, increasing the number of rounds up to the full number of rounds does not increase the security of the cipher with respect to these attacks.

### 1.2 Related Work

As already mentioned in the abstract, our attack can be seen as a weak key variant of statistical saturation attacks [7,8]. As the statistical saturation attack itself is a special case of partitioning cryptanalysis [12], so is our attack. Again, the main difference is that we make use of weak keys and for those keys the bias is maximal. More loosely our work is related to conditional cryptanalysis [2,14] in the sense that the truncated differential characteristic described in Section 3 is conditioned to certain key and message bits. Moreover, our attack can also be interpreted as an extreme case of a dynamic cube attack [11]. Here, the algebraic normal form of certain ciphertext bits becomes a constant when a weak key is used and certain message bits are fixed correctly.

## 2 The Invariant Subspace Attack

### 2.1 General Idea

Consider an $n$-bit block cipher with a round function $E_k$ consisting of a key addition and an SP-layer

$$E : \mathbb{F}_2^n \to \mathbb{F}_2^n,$$

that is $E_k$ is defined by $E_k(x) = E(x + k)$. Assume that the SP-layer $E$ is such that there exists a subspace $U \subseteq \mathbb{F}_2^n$ and two constants $c, d \in \mathbb{F}_2^n$ with the property:

$$E(U + c) = U + d.$$

Then, given a (round) key $k = u + c + d$ with $u \in U$, the following holds:

$$E_k(U + d) = E((U + d) + (u + c + d)) = E(U + c) = U + d,$$

i.e. the round function maps the affine subspace $U + d$ onto itself. If all round keys are in $k \in U + (c + d)$ (in particular if a constant round key is used), then this property is iterative over an arbitrary number of rounds. This yields a very efficient distinguisher for a fraction of the keys. $U$ should be as large as possible to increase this fraction. We call this new attack technique an *invariant subspace attack*. In the next section we show an example of how to apply it to the light-weight block cipher PRINTcipher.

### 2.2 Attack against PRINTcipher

**Description of PRINTcipher.** PRINTcipher is a block cipher proposed by Knudsen et al. at CHES 2010 [15]. It is a class of two SP-networks with a block size of $n = 48$ (resp. $n = 96$) bits, a key size of $l = 80$ (resp. $l = 160$) bit, and 48 (resp. 96) rounds. One round of PRINTcipher-48 is shown in Figure 1.

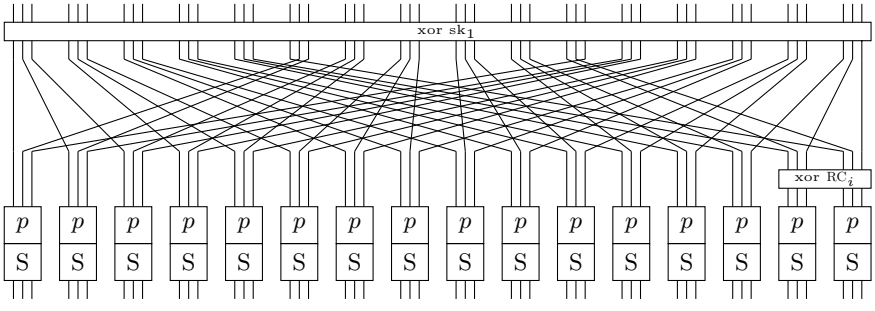


**Fig. 1.** One round of PRINTcipher-48 illustrating the bit-mapping between the 16 3-bit S-boxes from one round to the next. $sk_1$ denotes the xor key, $p$ the permutation key, and $RC_i$ the round counter.

PRINTcipher uses the same key for all rounds. It is split into two parts: The first $n$ bits are used as an xor key, the remaining $l - n$ bits control the permutations $p$. In order to introduce differences between the rounds, a round counter $RC_i$ is used which is generated by an LFSR (for details, see [15]). The other elements of the round function are defined as follows.

The **linear layer** consists of a bit permutation, where bit $i$ of the current state is moved to bit position $P(i)$ where

$$P(i) = \begin{cases} 3i \bmod n - 1 & \text{for } 0 \leq i \leq n-2, \\ n-1 & \text{for } i = n-1, \end{cases}$$

where $n \in \{48, 96\}$ is the block size.

Then the state bits are arranged in 16 (resp. 32) blocks of 3 bits each, which are permuted individually in the **permutation layer**. Out of 6 possible permutations on 3 bits, only four are valid permutations for PRINTcipher. Specifically,

the three input bits $c_2||c_1||c_0$ are permuted to give the following output bits according to two key bits $a_1||a_0$.

| nr. | $a_1\|\|a_0$ | $p$ |
|---|---|---|
| 0 | 00 | $c_2\|\|c_1\|\|c_0$ |
| 1 | 01 | $c_1\|\|c_2\|\|c_0$ |
| 2 | 10 | $c_2\|\|c_0\|\|c_1$ |
| 3 | 11 | $c_0\|\|c_1\|\|c_2$ |

Finally, in the **non-linear layer**, each 3-bit block is processed by the same s-box, which is shown in the following table.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $S[x]$ | 0 | 1 | 3 | 6 | 7 | 4 | 5 | 2 |

**An Attack on PRINTcipher.** One interesting property of the PRINTcipher s-box is that a one bit difference in the input causes a one bit difference in the same bit in the output with probability 2/8. That is, there exists exactly one pair for each one bit input difference resulting in a one bit output difference (at the same position). More precisely, denoting by $*$ an arbitrary value in $\mathbb{F}_2$, the following holds for the PRINTcipher s-box:

| | | |
|---|---|---|
| $\mathrm{S}(000) = 000$<br>$\mathrm{S}(001) = 001$ | $\Leftrightarrow$ | $\mathrm{S}(00*) = 00*$ |
| $\mathrm{S}(100) = 111$<br>$\mathrm{S}(110) = 101$ | $\Leftrightarrow$ | $\mathrm{S}(1*0) = 1*1$ |
| $\mathrm{S}(011) = 110$<br>$\mathrm{S}(111) = 010$ | $\Leftrightarrow$ | $\mathrm{S}(*11) = *10$ |

In addition, there exists a subset of s-boxes such that (1) two output bits of those s-boxes map onto two input bits of the same s-boxes in the next round and (2) the round-dependent $\mathrm{RC}_i$ is not involved (see Figure 2).

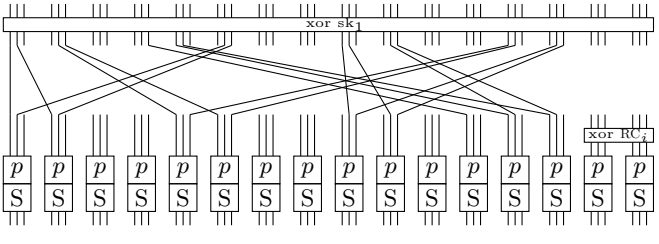


**Fig. 2.** A subset of PRINTcipher-48 s-boxes mapping onto itself

Now consider an xor-key $sk_1$ of the form

Xor key = 01* *11 *** *** 01* *11 *** *** 01* *11 *** *** 01* *11 *** ***,

and a permutation key with the following restrictions:

$$\text{Perm. key} = \texttt{0* 11 ** ** 10 01 ** ** 11 *0 ** ** *0 11 ** **},$$

where again $*$ denotes an arbitrary value in $\mathbb{F}_2$. For those keys the following structural *iterative* one round property holds:

| | |
|---|---|
| Start | 00* *10 *** *** 00* *10 *** *** 00* *10 *** *** 00* *10 *** *** |
| Key xoring | 01* *01 *** *** 01* *01 *** *** 01* *01 *** *** 01* *01 *** *** |
| Lin. layer | 00* 11* *** *** 0*0 1*1 *** *** *00 *11 *** *** 00* 11* *** *** |
| RC | 00* 11* *** *** 0*0 1*1 *** *** *00 *11 *** *** 00* 11* *** *** |
| Perm. layer | 00* *11 *** *** 00* *11 *** *** 00* *11 *** *** 00* *11 *** *** |
| S-box layer | 00* *10 *** *** 00* *10 *** *** 00* *10 *** *** 00* *10 *** *** |

This property holds with probability one if both keys are of the above form. The fraction of those keys is $(1/2)^{16}$ for the XOR key and $(1/2)^{13}$ for the permutation key, meaning that the property is met for a fraction of $(1/2)^{29}$ of all keys. In other words, there exist $2^{51}$ weak keys of this form.

Thus, one can very efficiently check if a key of the above form is used by encrypting a few texts of the above form and check if the ciphertext is again of the same form. Given that the probability for false positives is $\approx 2^{-16}$, trial encrypting just a handful of selected plaintexts will uniquely identify such a weak key. If such a key is found, we do of course immediately have a distinguisher on PRINTCIPHER.

**Invariant Subspace Description.** Let us briefly rephrase the attack in terms of an invariant subspace attack. For this we fix a permutation key of the above form. Remember that the inner state at the beginning and the end of each round was

$$\text{Start} = \texttt{00* *10 *** *** 00* *10 *** *** 00* *10 *** *** 00* *10 *** ***}.$$

This means that the relevant subspace $U \subset \mathbb{F}_2^{48}$ is defined by

$$U = \{\texttt{00* *00 *** *** 00* *00 *** *** 00* *00 *** *** 00* *00 *** ***}\}, \tag{1}$$

and that the affine subspace is defined by any fixed vector $d$ of the form

$$d = \texttt{00* *10 *** *** 00* *10 *** *** 00* *10 *** *** 00* *10 *** ***}. \tag{2}$$

Then for any fixed vector $c$ of the form

$$c = \texttt{01* *01 *** *** 01* *01 *** *** 01* *01 *** *** 01* *01 *** ***}, \tag{3}$$

and any xor-key $k \in (U + c + d)$, the round function does indeed map $U + d$ onto itself.

### 2.3 Other Attack Profiles

In the following we describe other sets of weak keys for PRINTCIPHER-48 and similar ones for PRINTCIPHER-96.

**Other Weak Keys for** PRINTcipher**-48.** As it turns out, there are some more invariant subspaces that also can be used for PRINTcipher-48. They are all of the form

`00* XXX *** 1*1 00* *10 *** *** 00* XXX *** 1*1 00* *10 *** ***,`

where an 'X' marks a bit position where the attacker has to make an arbitrary assignment. Note that each position can be filled independently of the others. Thus, we have $2^6$ possible plaintexts that we can work with, each of which targets another class of weak keys.

For each such assignment, the cipher behaves as follows:

| | | |
|---|---|---|
| Start | (1) | `00* XXX *** 1*1 00* *10 *** *** 00* XXX *** 1*1 00* *10 *** ***` |
| Key xoring | (2) | `0X* X01 *** X*1 01* *0X *** *** 0X* 001 *** X*X 01* *0X *** ***` |
| Lin. layer | (3) | `00* XXX *** X*X 0*0 1*1 *** *** *00 XXX *** 10* 00* 11* *** ***` |
| RC | (4) | `00* XXX *** X*X 0*0 1*1 *** *** *00 XXX *** 10* 00* 11* *** ***` |
| Perm. layer | (5) | `00* XXX *** 1*0 00* *11 *** *** 00* XXX *** 1*0 00* *11 *** ***` |
| S-box layer | (6) | `00* XXX *** 1*1 00* *10 *** *** 00* XXX *** 1*1 00* *10 *** ***` |

The behaviour is best understood by traversing the cipher in the inverse direction, i.e. by starting from the end and then finding the key bits that ensure that all fixed bits in line (1) match their counterparts in line (6).

Let us start with the output of the s-box, i.e. line (6), and let the bit positions marked by 'X' be arbitrarily and independently fixed to either 0 or 1. Then going backwards through the s-box uniquely determines the bits in line (5). We then use a permutation key of the form

Perm. Key = `0* ** ** (00 or 11) 10 01 ** ** 11 ** ** 10 0* 11 ** **`

to obtain line (4), noting that $2^{-13}$ of all permutation keys meet this property. We then apply round counter and linear layer to obtain line (2). Now note that line (2) contains 22 bits that are fixed and that have to match the corresponding bits in line (1). Thus, 22 key bits of the xoring key are determined, meaning that $2^{-22}$ of all xoring keys are suitable for the attack.

Summing up, for each of the $2^6$ possible assignments to the bits marked by 'X' in line (1) or (6), a fraction of exactly $2^{-35}$ keys are weak, meaning that in total, we have found another fraction of $2^{-29}$ weak keys that can be attacked by the above technique.

**Analysis of** PRINTcipher**-96.** As it turns out, the same attack can also be applied to PRINTcipher-96. Again, there are two types of weak keys. The first type is based on 32 active bits and is met by a fraction of $2^{-59}$ of all keys. The second type is based on 44 active bits and has an additional 12 freely chosable input bits. Each of the resulting $2^{12}$ inputs targets a fraction of $2^{-71}$ keys, meaning that this group, too, contains a fraction of $2^{-59}$ weak keys in total. The active bits for these weak keys are given in Table 1.

### 2.4 Protecting Against the Attack

The above attack against PRINTcipher is a special case of the general attack described in the beginning of the section, since the subspace is described by

**Table 1.** Subsets of active bits for PRINTCIPHER-96, grouped according to s-boxes

| | |
|---|---|
| Subset 1 | **Active input bits for linear layer:**<br>(0 1) (4 5) (12 13) (16 17) (24 25) (28 29) (36 37) (40 41)<br>(48 49) (52 53) (60 61) (64 65) (72 73) (76 77) (84 85) (88 89) |
| | **Active output bits for linear layer:**<br>(0 2) (3 5) (12 13) (15 16) (25 26) (28 29) (36 38) (39 41)<br>(48 49) (51 52) (61 62) (64 65) (72 74) (75 77) (84 85) (87 88) |
| Subset 2 | **Active input bits for linear layer:**<br>(0 1) (3 4 5) (9 11) (12 13) (16 17) (24 25) (27 28 29)<br>(33 35) (36 37) (40 41) (48 49) (51 52 53) (57 59) (60 61)<br>(64 65) (72 73) (75 76 77) (81 83) (84 85) (88 89) |
| | **Active output bits for linear layer:**<br>(0 2) (3 4 5) (9 10) (12 13) (15 16) (25 26) (27 28 29)<br>(33 35) (36 38) (39 41) (48 49) (51 52 53) (58 59) (61 62)<br>(64 65) (72 74) (75 76 77) (81 82) (84 85) (87 88) |

simply fixing some of its bits. In theory, describing the subspace by a set of linear equations is possible, opening for a wide range of attacks. The full potential of this generalized attack is yet to be determined.

As for the special case used against PRINTCIPHER, it is relatively easy to protect the design against the attack. Note that the list of attack profiles by fixing bits given here is complete, and that all attack profiles fix two of the bits 39-41 (PRINTCIPHER-48) resp. 87-89 (PRINTCIPHER-96). Thus, it would suffice to spread the round counter over the last three s-boxes, e.g. by assigning two counter bits to each s-box. This would destroy the only attack profiles available, at no extra hardware cost.

We also analysed the block cipher NOEKEON, which was proposed by Daemen et al. in 2000 [9]. NOEKEON is a 16-round block cipher with a constant round key, making it a particularly tempting target for the attack. However, as it turns out, the linear mixing layer of NOEKEON is much more resistant against the above type of attack. Here, the stronger round function (necessary for a cipher with only 16 rounds) works to the advantage of the cipher. As it turns out, even if there was no round counter involved in NOEKEON, the simple attack described above – i.e. where the subspace is defined by fixing certain bits – could not be applied. Whether or not the generalized attack has a better chance of succeeding remains yet to be determined.

## 3 Truncated Differential Attacks

As pointed out by Murphy in [18] the attack complexity for linear attacks is often wrongly stated in the literature. One of the reasons is that it is often easy to compute the average squared bias $\epsilon^2$ when averaging over all keys. However, it is often stated that the average attack complexity is $\frac{\gamma}{\epsilon^2}$ for some small $\gamma$, which,

in general, is wrong. In particular, the average complexity is formally infinite as soon as there exists a single key with no bias, while $\frac{\gamma}{\epsilon^2}$ is finite as soon as there exists a single key with a bias.

Now, to some extent the same is true for (truncated) differential attacks. A truncated differential characteristic on an $n$-bit block cipher can be, in general, described by a set of input and output differences. For $0 \leq i \leq r$ let $U_i \subset F_2^n$ and

$$U_i \xrightarrow{E_i} U_{i+1}$$

be a set of differential characteristics with probability $p_i$.

Assuming independent round keys the average probability, taken over all keys, of the truncated $r$-round differential characteristic

$$U_0 \xrightarrow{E_0} U_1 \xrightarrow{E_1} \cdots \xrightarrow{E_{r-1}} U_r$$

is $p = \prod_i p_i$. One normally assumes (cf. the hypothesis of stochastic equivalence in [16]) that for (almost) all keys it holds that $p_k \approx p$. Here $p_k$ denotes the probability of the truncated differential characteristic for a fixed key $k$.

However, this may be highly incorrect. Indeed PRINTcipher is an example of the extreme opposite. We will show below that for PRINTcipher, the attack discussed in Section 2 implies the existence of a truncated differential characteristic such that

$$p_k \in \{2^{-16}, 0\},$$

for any number of rounds $r \geq 2$. Since a fraction of $2^{-29}$ of all keys is weak, the average probability over all keys is

$$p_{\text{av}} = 2^{-16} \cdot 2^{-29} = 2^{-45},$$

again noting that this holds for any number $r \geq 2$ of rounds. After introducing the invariant subspace attack, the existence of such truncated differential characteristics might not be so surprising, as one basically pays the price for following the characteristic only once. That is to say that pairs that follow the characteristic for two rounds automatically follow the characteristic for any number of rounds.

However, this disproves the naive assumption where multiplying the probabilities for the individual rounds yields an average attack complexity that tends to zero for an increasing number of rounds. Thus, not only is the assumption that all keys behave more or less similar wrong. Also, the assumption that the round keys are independent leads to a very wrong conclusion. Concluding this part, studying the average complexity does not reveal the actual behavior of PRINTcipher. On the contrary, PRINTcipher behaves completely opposite to what is usually assumed.

### 3.1 Rephrasing the Attack in Terms of Truncated Differentials

In this section, we will prove the above claims. To make the description easier, consider a PRINTcipher-48 version where we fix the permutation key to

```
00 11 00 00 10 01 00 00 11 00 00 00 00 11 00 00
```

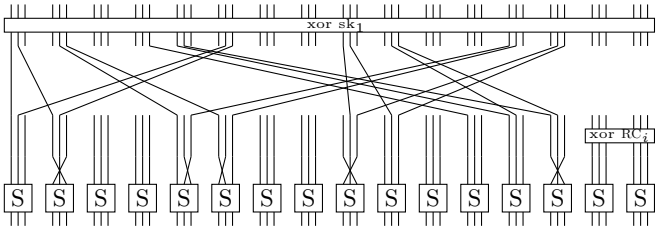


**Fig. 3.** One round of PRINTcipher with fixed permutation key. Only the bits that matter for the differential characteristic are shown in the linear layer.

One round of PRINTcipher with this key is given in Figure 3. Other weak permutation keys behave similarly.
Now, consider an $r$-round truncated differential characteristic[1] of the form

$$\alpha \xrightarrow{E_k} U' \xrightarrow{E_k} U \ldots U \xrightarrow{E_k} U, \tag{4}$$

where $\alpha$ is given by

$$\alpha = \texttt{000 100 011 101 000 100 001 100 000 000 001 110 001 000 101 110},$$

and $U'$ contains all vectors of the form

$$U' = \{\texttt{001 100 **1 1** 001 100 **1 1** 001 100 **1 1** 001 100 **1 1****\}.$$

Finally, as in Section 2, $U$ is defined by

$$U = \{\texttt{00* *00 *** *** 00* *00 *** *** 00* *00 *** *** 00* *00 *** ***}\}.$$

Note that $\alpha \in U' \subset U$. Given these definitions, we can prove the following theorem:

**Theorem 1.** *For a fixed (xor)-key $k$ , denote the probability of the truncated differential characteristic given by Equation (4) by $p_k$. It holds that*

$$p_k = \begin{cases} 2^{-16} & \text{if } k \text{ is weak} \\ 0 & \text{if } k \text{ is not weak} \end{cases}$$

*Here a (xor)-key $k$ is weak if and only if it is of the form*

$$k = \texttt{01* *11 *** *** 01* *11 *** *** 01* *11 *** *** 01* *11 *** ***}.$$

[1] We emphasize that we deal with truncated differential characteristics and not with truncated differentials. In particular, for the characteristic we are using, for the corresponding differential one can expect a probability of $2^{-16}$ even for a random round function.

**Table 2.** This table shows how the differences of the truncated differential characteristic from Section 3 propagate through the round functions. The underlined values are the only positions where the difference propagation has a probability not equal to 1. That is, only those positions pose restrictions on the pairs satisfying the characteristic.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Input 1st Rnd. | 000 | 100 | 011 | 101 | 000 | 100 | 001 | 100 | 000 | 000 | 001 | 110 | 001 | 000 | 101 | 110 |
| After P-layer | 001 | 001 | 001 | 100 | 010 | 010 | 001 | 100 | 100 | 100 | 001 | 100 | 001 | 001 | 001 | 100 |
| After perm.key | 001 | 100 | 001 | 100 | 001 | 100 | 001 | 100 | 001 | 100 | 001 | 100 | 001 | 100 | 001 | 100 |
| S-box layer | 001 | 100 | **1 | 1** | 001 | 100 | **1 | 1** | 001 | 100 | **1 | 1** | 001 | 100 | **1 | 1** |
| Input 2st Rnd. | 001 | 100 | **1 | 1** | 001 | 100 | **1 | 1** | 001 | 100 | **1 | 1** | 001 | 100 | **1 | 1** |
| After P-layer | 001 | 001 | 1** | 1** | 010 | 010 | **1 | **1 | 100 | 100 | *1* | *1* | 001 | 001 | 1** | 1** |
| After perm.key | 001 | 100 | 1** | 1** | 001 | 100 | **1 | **1 | 001 | 100 | *1* | *1* | 001 | 100 | 1** | 1** |
| S-box layer | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** |
| Input 3rd Rnd. | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** |
| After P-layer | 00* | 00* | *** | *** | 0*0 | 0*0 | *** | *** | *00 | *00 | *** | *** | 00* | 00* | *** | *** |
| After perm.key | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** |
| S-box layer | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** | 00* | *00 | *** | *** |

The idea behind this theorem is similar to results from [10]. Before we prove this result, we have to introduce some notation similar to that in [10]. Given a round function $E : \mathbb{F}_2^n \to \mathbb{F}_2^n$ excluding the initial key addition and two subsets $A, B \subseteq F_2^n$ we denote by

$$F_{(A,B)} = \{x \mid E(x) + E(x + \alpha) = \beta,\ \alpha \in A, \beta \in B\}$$

the set of input pairs satisfying the truncated differential $A \overset{E_k}{\to} B$.

As observed in [10] in many cases $F(A, B)$ is a coset of a subspace. This is also the case here. More precisely, we have the following lemma.

**Lemma 1.** *Let $E$ be the round function of* PRINTcipher *excluding the inital key addition, $\alpha$, $U'$ and $U$ as defined above. Then*

$$F_{(\{\alpha\},U')} = F_{(U',U)} = U + c,$$

*where $c$ is defined as in Equation (3).*

*Proof.* This can be seen from Table 2, where it is shown how the differences propagate through the round functions. □

Now to prove Theorem 1, from the invariant subspace attack on PRINTcipher, as discussed in Section 2.2, we know that $E(U + c) = U + d$. It follows that for the good pairs for

$$\alpha \overset{E_k}{\to} U' \overset{E_k}{\to} U$$

the inputs to the second round are given by

$$(U + d + k)\ \cap\ (U + c) = \begin{cases} U + c & \text{if } k \in U + d + c \\ \emptyset & \text{if } k \neq U + d + c \end{cases}$$

This already proves that for a non-weak key the probability of the truncated differential characteristic given in the above theorem is 0. For a weak key it holds that $E_k(U + d) = U + d$ and therefore, for any $x$ in $F_{a,b} + k = U + d$ the pair $(x, x + \alpha)$ fullfils the whole $r$ round truncated differential characteristic. □

# 4 Statistical Saturation Attacks and Multidimensional Linear Attacks

The attack on PRINTcipher discussed in Section 2.2 is clearly strongly related to statistical saturation attacks as described in [7]. In this section, after briefly recalling some of the principles of statistical saturation attacks, we elaborate on the details of this relation. Maybe the most interesting finding here is that for PRINTcipher there exist strongly biased linear approximations for any number of rounds, if the key is weak in the sense of the invariant subspace attack. This result follows using a link between statistical saturation attacks and multidimensional linear attacks (see [17]). Understanding these strongly biased linear approximations by studying the linear hulls directly is an interesting problem that we leave open for further investigation.

## 4.1 Necessary Background Information

**Notations.** The canonical inner product on $\mathbb{F}_2^n$ is denoted by $\langle \cdot, \cdot \rangle$, i.e.

$$\langle (a_0, \ldots, a_{n-1}), (b_0, \ldots, b_{n-1}) \rangle := \sum_{i=0}^{n-1} a_i b_i.$$

We note that all linear forms, i.e. all linear functions $l : \mathbb{F}_2^n \to \mathbb{F}_2$, can be described as $\ell(x) = \langle a, x \rangle$ for a suitable $a \in \mathbb{F}_2^n$. Given a (vectorial Boolean) function $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$ the *Fourier coefficient* of $F$ at the pair $(a, b) \in \mathbb{F}_2^n \times \mathbb{F}_2^m$ is defined by

$$\widehat{F}(a, b) = \sum_{x \in \mathbb{F}_2^n} (-1)^{\langle b, F(x) \rangle + \langle a, x \rangle}.$$

The *bias* $\epsilon_F(a, b)$ of the linear approximation $\langle a, x \rangle$ of $\langle b, F(x) \rangle$ is defined as

$$\epsilon_F(a, b) := \frac{|\{x \mid \langle b, F(x) \rangle + \langle a, x \rangle = 0\}|}{2^n} - \frac{1}{2}.$$

The fundamental relation between the Fourier transformation of $F$ and the bias of a linear approximation is given by

$$\epsilon_F(a, b) = \frac{\widehat{F}(a, b)}{2^{n+1}} \tag{5}$$

Given $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$, the value used to determine the complexity of both multidimensional linear attacks and statistical saturation attacks is the capacity of $F$ given by

$$\mathrm{Cap}(F) = \sum_{z \in \mathbb{F}_2^m} \frac{(2^{-n} \cdot |\{x \in F_2^n \mid F(x) = z\}| - 2^{-m})^2}{2^{-m}}.$$

**Statistical Saturation Attacks.** Let us first briefly recall some concepts from statistical saturation attacks. We refer to [7] for details. Given an encryption function

$$e : \mathbb{F}_2^n \to \mathbb{F}_2^n,$$

statistical saturation attacks study the distribution of $e$ when some of its input bits are fixed. Up to a fixed bijective linear transformation before and after the cipher, we can restrict ourselves without loss of generality to the case where one fixes the first $r$ bits in the inputs and considers only the first $t$ bits of the output[2]. Thus we write

$$\begin{aligned} e : \mathbb{F}_2^r \times \mathbb{F}_2^s &\to \mathbb{F}_2^t \times \mathbb{F}_2^u \\ e(y,x) &= \left(e^{(1)}(y,x), e^{(2)}(y,x)\right), \end{aligned}$$

where $r+s = t+u = n$ and $e^{(1)}(y,x) \in \mathbb{F}_2^t$, $e^{(2)}(y,x) \in \mathbb{F}_2^u$. For convenience we denote by $h_y$ the restriction of $e$ by fixing the first $r$ bits to $y$ and considering only the first $t$ bits of the output, that is

$$\begin{aligned} h_y : \mathbb{F}_2^s &\to \mathbb{F}_2^t \\ h_y(x) &= e^{(1)}(y,x). \end{aligned}$$

In a statistical saturation attack one considers the capacity of $h_y$, and the attack complexity is usually a constant times $1/\operatorname{Cap}(h_y)$. Computing this capacity is difficult in general. However, when averaging over all possible fixings $y$ the following has been proven in [17]:

**Theorem 2.** *The average capacity in statistical saturation attacks where the average is taken over all possible fixations $y$ is given by*

$$\overline{\operatorname{Cap}(h_y)} = 2^{-r} \sum_{y \in \mathbb{F}_2^r} \operatorname{Cap}(h_y) = 2^{-2n} \sum_{\substack{a \in \mathbb{F}_2^r \times \{0\} \\ b \in \mathbb{F}_2^t \times \{0\}, b \neq 0}} \left(\widehat{e}(a,b)\right)^2 \tag{6}$$

### 4.2 On the Choice of the Values of the Fixed Bits

We now focus on the case where $r = t$, that is the number of fixed bits is the same as the number of bits considered at the output.

Assume a cipher is vulnerable to an invariant subspace attack. As for statistical saturation attacks, up to a fixed bijective linear transformation before and after the cipher, we can assume that, for a weak key, the affine subspace of the form $\{d\} \times \mathbb{F}_2^s$ is mapped to an affine subspace of the form $\{d\} \times \mathbb{F}_2^s$. It then follows immediately that (for a weak key) the function of the restriction $h_y$ for $y = d$ is a constant, more precisely

$$h_d(x) = e^{(1)}(d,x) = d.$$

[2] This differs slightly from the notation in [17].

For the special choice of the values of the fixed bits the capacity is maximal. Hence for a weak key this special fixing of the bits leads to an optimal statistical saturation attack. Note that Theorem 2 does not reveal the existence of such extreme cases, as it only considers the average capacity of the restrictions.

While in an invariant subspace attack, given the subspace, the choice of the coset is crucial, for statistical saturation attacks the fixed bits are usually assigned with random values. As the invariant subspace attack on PRINTCIPHER does not imply that PRINTCIPHER is in general vulnerable to a statistical saturation attack, it does not come as a surprise that the experiments in [15] did not reveal any weakness of PRINTCIPHER with respect to those attacks.

### 4.3 On the Existence of Highly Biased Approximations

Theorem 2 was used to compute the average capacity using the Fourier coefficients. However, for us, the reciprocal is of interest as it implies the following corollary.

**Corollary 1.** *Assume an $n$-bit block cipher $E_k$ is vulnerable to an invariant subspace attack, that is there exist a subspace $U$, a constant $d$ and keys $k$ such that*

$$E_k(U + d) = U + d.$$

*Then, for those keys, there exist linear approximations with a bias $\epsilon$ such that*

$$\epsilon \geq 2^{\dim(U)-n-1} - 2^{2(\dim(U)-n)-1}.$$

*Proof.* With the notation as in Section 4.2, $h_d$ is a constant function. Thus $\mathrm{Cap}(h_d) = 2^r - 1$ and furthermore

$$\sum_{y \in \mathbb{F}_2^r} \mathrm{Cap}(h_y) \geq \mathrm{Cap}(h_d) = 2^r - 1.$$

Considering Equation (6) it follows that

$$\sum_{\substack{a \in \mathbb{F}_2^r \times \{0\} \\ b \in \mathbb{F}_2^t \times \{0\}, b \neq 0}} \left(\widehat{e}(a,b)\right)^2 \geq 2^{2n}(1 - 2^{-r})$$

Lower bounding the maximal value by the average (and recalling that $r = t$), we compute

$$\max_{a,b \neq 0} \left(\widehat{e}(a,b)\right)^2 \geq 2^{-2r} \sum_{\substack{a \in \mathbb{F}_2^r \times \{0\} \\ b \in \mathbb{F}_2^t \times \{0\}, b \neq 0}} \left(\widehat{e}(a,b)\right)^2 \geq 2^{2n-2r}(1 - 2^{-r})$$

Thus there exists at least one Fourier coefficient such that

$$|\widehat{e}(a,b)| \geq 2^{n-r}\sqrt{1 - 2^{-r}} \geq 2^{n-r} - 2^{n-2r}$$

Applying identity (5) and remembering that $r = n - \dim U$, the theorem follows. □

Clearly, this Theorem is only interesting for the case where $\dim(U) > n/2$ as the existence of the stated approximations otherwise is trivial. For the case of PRINTcipher-48 we summarize the findings below

**Corollary 2.** *Given a weak key for any round $r \leq 48$ there exists at least one linear approximation for* PRINTcipher-*48 with bias at least $2^{-17} - 2^{-33}$.*

## 5 Conclusions

We have presented a new attack against iterative block ciphers named *invariant subspace attack* and demonstrated its validity by breaking PRINTcipher for a significant fraction of its keys. The presented *invariant subspace attack* shows that $2^{52}$ keys (out of $2^{80}$) for PRINTcipher-48 and $2^{102}$ keys (out of $2^{160}$) for PRINTcipher-96 are weak. In addition, we have shown the relationship of the *invariant subspace attack* to other classes of attacks such as truncated differential attack, multi-dimentional attack linear attack and statistical saturation attack. In doing this, we could provide an example for a truncated differential attack whose success probability is round-independent, disproving the common implicit assumptions that the total success probability is the product of the individual round probabilities and that the overall success probability against a cipher can be averaged over all keys. The probability of this truncated differential characteristic is $2^{-16}$ for weak keys and zero for non-weak keys given that the number of rounds is greater than or equal to two. Moreover, for PRINTcipher there are strongly biased linear approximations for any number of rounds, if a weak key is chosen. For example, there is at least one linear approximation for PRINTcipher-48 with bias at least $2^{-17}$.

*Open Questions and Future Work.* The attack presented against PRINTcipher is a special case of the general *invariant subspace attack*. It should be evaluated whether the generalised attack provides even better results against PRINTcipher and other potentially vulnerable ciphers. Hence, the possibility of extending the presented distinguishing attack on weak keys classes into a key recovery attack is an open problem that needs to be further analysed. Understanding the strongly biased linear approximations by studying the linear hulls directly is another interesting open problem. We believe that it will increase our general understanding of linear hulls and how (very simple) key scheduling algorithms influence the distribution of biases.

## References

1. Abdelraheem, M.A., Leander, G., Zenner, E.: Differential cryptanalysis of round-reduced PRINTcipher: Computing roots of permutations. In: Joux, A. (ed.) FSE 2011. LNCS, vol. 6733, pp. 1–17. Springer, Heidelberg (2011)

2. Ben-Aroya, I., Biham, E.: Differential cryptanalysis of Lucifer. Journal of Cryptology 9(1), 21–34 (1996)
3. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: An Ultra-Lightweight Block Cipher. In: Paillier, P., Verbauwhede, I. (eds.) CHES 2007. LNCS, vol. 4727, pp. 450–466. Springer, Heidelberg (2007)
4. Bogdanov, A., Rechberger, C.: A 3-Subset Meet-in-the-Middle Attack: Cryptanalysis of the Lightweight Block Cipher KTANTAN. In: Biryukov, A., Gong, G., Stinson, D.R. (eds.) SAC 2010. LNCS, vol. 6544, pp. 229–240. Springer, Heidelberg (2011)
5. De Cannière, C., Dunkelman, O., Knežević, 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, Heidelberg (2009)
6. Cho, J.Y.: Linear Cryptanalysis of Reduced-Round PRESENT. In: Pieprzyk, J. (ed.) CT-RSA 2010. LNCS, vol. 5985, pp. 302–317. Springer, Heidelberg (2010)
7. Collard, B., Standaert, F.-X.: A Statistical Saturation Attack against the Block Cipher PRESENT. In: Fischlin, M. (ed.) CT-RSA 2009. LNCS, vol. 5473, pp. 195–210. Springer, Heidelberg (2009)
8. Collard, B., Standaert, F.-X.: Multi-trail Statistical Saturation Attacks. In: Zhou, J., Yung, M. (eds.) ACNS 2010. LNCS, vol. 6123, pp. 123–138. Springer, Heidelberg (2010)
9. Daemen, J., Peeters, M., van Assche, G., Rijmen, V.: Nessie proposal: NOEKEON (2000), `http://gro.noekeon.org/Noekeon-spec.pdf`
10. Daemen, J., Rijmen, V.: Plateau characteristics. Information Security, IET 1(1), 11–17 (2007)
11. Dinur, I., Shamir, A.: Breaking grain-128 with dynamic cube attacks. In: Joux, A. (ed.) FSE 2011. LNCS, vol. 6733, pp. 167–187. Springer, Heidelberg (2011)
12. Harpes, C., Massey, J.L.: Partitioning Cryptanalysis. In: Biham, E. (ed.) FSE 1997. LNCS, vol. 1267, pp. 13–27. Springer, Heidelberg (1997)
13. Hong, D., Sung, J., Hong, S.H., Lim, J.-I., Lee, S.-J., Koo, B.-S., Lee, C.-H., Chang, D., Lee, J., Jeong, K., Kim, H., Kim, J.-S., Chee, S.: HIGHT: A New Block Cipher Suitable for Low-Resource Device. In: Goubin, L., Matsui, M. (eds.) CHES 2006. LNCS, vol. 4249, pp. 46–59. Springer, Heidelberg (2006)
14. Knellwolf, S., Meier, W., Naya-Plasencia, M.: Conditional Differential Cryptanalysis of NLFSR-Based Cryptosystems. In: Abe, M. (ed.) ASIACRYPT 2010. LNCS, vol. 6477, pp. 130–145. Springer, Heidelberg (2010)
15. Knudsen, L.R., Leander, G., Poschmann, A., Robshaw, M.J.B.: PRINTcipher: A Block Cipher for IC-Printing. In: Mangard, S., Standaert, F.-X. (eds.) CHES 2010. LNCS, vol. 6225, pp. 16–32. Springer, Heidelberg (2010)
16. Lai, X., Massey, J.L., Murphy, S.: Markov Ciphers and Differential Cryptanalysis. In: Davies, D.W. (ed.) EUROCRYPT 1991. LNCS, vol. 547, pp. 17–38. Springer, Heidelberg (1991)
17. Leander, G.: On Linear Hulls, Statistical Saturation Attacks, PRESENT and a Cryptanalysis of PUFFIN. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 303–322. Springer, Heidelberg (2011)
18. Murphy, S.: The Effectiveness of the Linear Hull Effect. Technical report, RHUL-MA-2009-19 (2009)