Provable Security of Substitution-Permutation Networks

Yevgeniy Dodis[1], Jonathan Katz[2], John Steinberger[3], Aishwarya Thiruvengadam[4,⋆], and Zhe Zhang[3]

[1] Dept. of Computer Science, New York University, USA
dodis@cs.nyu.edu
[2] Dept. of Computer Science, University of Maryland, USA
jkatz@cs.umd.edu
[3] Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing
{jpsteinb,leomwzz}@gmail.com
[4] Dept. of Computer Science, University of California, Santa Barbara, USA
aish@cs.ucsb.edu

Abstract. Many modern block ciphers are constructed based on the paradigm of substitution-permutation networks (SPNs). But, somewhat surprisingly—especially in comparison with Feistel networks, which have been analyzed by dozens of papers going back to the seminal work of Luby and Rackoff—there are essentially no provable-security results about SPNs. In this work, we initiate a comprehensive study of the security of SPNs as strong pseudorandom permutations when the underlying "S-box" is modeled as a public random permutation. We show that 3 rounds of S-boxes are necessary and sufficient for secure *linear* SPNs, but that even 1-round SPNs can be secure when non-linearity is allowed. Additionally, our results imply security in settings where an SPN structure is used for domain extension of a block cipher, even when the attacker has direct access to the small-domain block cipher.

1 Introduction

Modern block ciphers are generally constructed using two main paradigms [22]: Feistel networks [15] or substitution-permutation networks (SPNs) [15, 38]. These two approaches share the same goal: namely, to extend a "pseudorandom object" on a small domain to a (keyed) pseudorandom permutation on a larger domain by repeating a few, relatively simple operations several times across multiple rounds. Simplifying somewhat, Feistel networks begin with a keyed pseudorandom function on n-bit inputs and extend this to give a keyed pseudorandom permutation on $2n$-bit inputs; SPNs start with one or more public "random permutations" on n-bit inputs and extend them to give a keyed pseudorandom permutation on wn-bit inputs for some w. Examples of block ciphers based on Feistel networks include DES, FEAL, MISTY and KASUMI; block ciphers based on SPNs include AES, Serpent, and PRESENT.

⋆ Work done while at the University of Maryland.

Although proving security unconditionally for any concrete block cipher is beyond current techniques, we can still hope to prove that particular approaches to constructing block ciphers are sound in an appropriate theoretical framework. Starting with the seminal paper by Luby and Rackoff [25] in the 1980s, there are by now dozens of papers that use this approach to prove security of Feistel networks (of sufficiently many rounds), i.e., showing that if the underlying building block is an n-bit pseudorandom function then the resulting construction is a $2n$-bit pseudorandom permutation. In contrast, it is somewhat surprising that there are almost no results about provable security of SPNs. (We discuss relevant prior work below.) Here, we address this gap and explore conditions under which SPNs can be proven secure.

1.1 Our Model and Results

An SPN on wn-bit inputs is computed via repeated invocation of two basic steps: a *substitution step* in which a public (unkeyed) "cryptographic" permutation $S : \{0,1\}^n \to \{0,1\}^n$, called an *$S$-box*, is computed in a blockwise fashion over the wn-bit intermediate state, and a *permutation step* in which a keyed but "non-cryptographic" permutation on $\{0,1\}^{wn}$ is applied. An r-round SPN uses r rounds of S-boxes (plus an additional permutation step at the beginning). We consider both linear and non-linear SPNs; in a linear SPN, the keyed permutation in each round is linear (or, more generally, affine) in both the key and the intermediate state.

The only "cryptographic hardness" comes from the S-box, and we capture this by modeling S as a public, random permutation available to all parties as an oracle. This is a key difference between our work and most prior work analyzing security of SPNs (see Section 1.3), which treated the S-box as a key-dependent, random permutation inaccessible to the adversary. In practical constructions of block ciphers, however, the S-box is typically unkeyed.

We analyze SPNs in the standard sense of security against adaptive chosen-plaintext and chosen-ciphertext attacks; that is, we analyze the SPN as a strong pseudorandom permutation [25]. We first characterize the security of *linear* SPNs in Section 3. Extending and generalizing an attack by Halevi and Rogaway [18] (credited there to [21]), we show that no 2-round linear SPN with $w \geq 2$ can be secure in Section 3.1. (Even a 1-round linear SPN can be secure when $w = 1$; see remark below for the significance of this result.) Complementing this, and giving a tight characterization, in Section 3.2, we show that 3-round linear SPNs *are* secure, for any w, if the keyed permutations satisfy some mild technical requirements. Moreover, a simple variant of our 3-round linear SPN construction can be used to reduce the key-length to n bits.

In an effort to reduce the number of rounds, we then turn our attention to non-linear SPNs in Section 4. Here we show that even a 1-round SPN can be secure if appropriate keyed permutations are used. In Section 4.1, we identify a combinatorial property on the permutations that suffices for security in this case, called *blockwise universality*. Informally, a keyed permutation π_k is blockwise universal if, for any distinct inputs x, x' and any constant c, the probability

(taken over uniform k) of each of the following events is low: (1) a block of $\pi(k, x)$ is equal to a block of $\pi(k, x')$, (2) two different blocks of $\pi(k, x)$ are equal, (3) a block of $\pi(k, x)$ is equal to c. Then, in Section 4.2, we study the efficiency of constructing permutations satisfying this property. Specifically, we show a construction of a satisfactory permutation with n-bit keys (but having high degree), and another construction with longer keys but having degree 3.

Remarks. Although an SPN with $w = 1$ (i.e., no domain extension) may seem uninteresting, it captures the well-known Even-Mansour construction [14] of a keyed pseudorandom permutation from a public random permutation. Our positive results imply security of the Even-Mansour construction (with similar concrete security bounds) as a special case, even in the setting where the pre- and post-whitening keys are identical [13, 23].

1.2 Implications

We view our results as providing support for the SPN approach to constructing block ciphers, but we caution that they say little about any concrete SPN-based block cipher. This is especially true since our positive results (inherently) achieve security only for $q < 2^n$ queries, yet practical constructions tend to use S-boxes with a very small domain (e.g., $n \leq 32$).

We notice that conceptually similar limitations — such as getting exact security which is too low for current settings of parameters, or assuming some building block to satisfy some clean security property not satisfied in practice — applies to most of the provable security work in the area of symmetric-key cryptography. For example, at the time of the original Luby-Rackoff paper [25] the main Feistel candidate DES had $n = 32$. Instead of discounting such "practice-oriented" provable security results as unimportant, it was an understanding of the community that such results should be interpreted in several constructive ways: (1) future work might be able to improve the exact security, possibly by increasing number of rounds and/or analysis; (2) a provable security result rules out certain generic, structural attacks on the construction. Hence, we believe such structural results on SPNs are useful and serve as valid sanity checks, despite the fact that they do not technically apply to existing constructions.

An SPN structure can also be used for domain extension of a block cipher, i.e., for building a block cipher F' with wn-bit block length from a block cipher F having n-bit block length. Our results imply security in that setting as well. (The above-mentioned limitation of n being too low in practice no longer applies in that setting.) Although several prior papers have considered this problem (see below), unlike most prior work, we analyze it in the case where the attacker is given access to the internal function F, which would correspond in practice to publicly fixing the key to F (thus potentially improving the efficiency of F').

1.3 Related Work

There are only a few prior papers looking at provable security of SPNs. The vast majority of such work analyzes the case of secret, key-dependent S-boxes (rather than public S-boxes as we consider here), and so we survey that work first.

SPNs with secret S-boxes. Naor and Reingold [30] prove security for what can be viewed as a non-linear, 1-round SPN. Their ideas were further developed, in the context of domain extension for block ciphers (see further discussion below), by Chakraborty and Sarkar [6] and Halevi [16].

Iwata and Kurosawa [20] analyze SPNs in which the linear permutation step is based on the specific permutations used in the block cipher Serpent. They show an attack against 2-round SPNs of this form, and prove security for 3-round SPNs against non-adaptive adversaries. In addition to the fact that we consider public S-boxes, our linear SPN model considers generic linear permutations and we prove security against adaptive attackers.

Miles and Viola [29] study SPNs from a complexity-theoretic viewpoint. Two of their results are relevant here. First, they analyze the security of linear SPNs using S-boxes that are not necessarily injective (so the resulting keyed functions are not, in general, invertible). They show that r-round SPNs of this type (for $r \geq 2$) are secure against chosen-plaintext attacks. (In contrast, our results show that 2-round, linear SPNs are not secure against a combination of chosen-plaintext and chosen-ciphertext attacks when $w \geq 2$.) They also analyze SPNs based on a concrete set of S-boxes, but in this case they only show security against linear/differential attacks (a form of chosen-plaintext attack), rather than all possible attacks, and only when the number of rounds is $r = \Theta(\log n)$.

SPNs with public S-boxes. A difference between our work and all the work discussed above is that we treat the S-boxes as public. We are aware of only one prior work analyzing the provable security of SPNs in this setting. Dodis et al. [12] recently studied the *indifferentiability* [27] of confusion-diffusion networks, which can be viewed as *unkeyed* SPNs. One could translate their results to the keyed setting, but that would require using multiple, key-dependent S-boxes (rather than a fixed, public S-box) and so would not imply our results. We remark further that they show positive results only for 5 rounds and above.

As observed earlier, the Even-Mansour construction [14] of a (keyed) pseudorandom permutation from a public random permutation can be viewed as a 1-round, linear SPN in the degenerate case where $w = 1$ (i.e., no domain extension) and all round permutations are instantiated using simple key mixing. Security of the 1-round Even-Mansour construction against adaptive chosen-plaintext/ciphertext attacks, using independent keys for the initial and final key mixing, was shown in the original paper [14]. Kilian and Rogaway [23] and Dunkelman, Keller, and Shamir [13] showed that security holds even if the keys used are the same. Our positive results imply security of the 1-round Even-Mansour construction (with similar concrete security bounds) as a special case.

Cryptanalysis of SPNs. Researchers have also explored cryptanalytic attacks on generic SPNs [1–3, 11]. These works generally consider a model of SPNs in

which round permutations are secret, random (invertible) linear transformations, and S-boxes may be secret as well; this makes the attacks stronger but positive results weaker. In many cases the complexities of the attacks are exponential in n (though still faster than a brute-force search for the key), and hence do not rule out asymptotic security results. On the positive side, Biryukov et al. [1] show that 2-round SPNs (of the stronger form just mentioned) are secure against some specific types of attacks, but other attacks on such schemes have recently been identified [11].

Domain extension of block ciphers. It is worth noting that our results also address the problem of *domain extension* for block ciphers. That is, we may view a block cipher F having an n-bit block length as an n-bit S-box (either viewing F as an ideal cipher, fixing its key, and thus viewing it as a public S-box, or viewing F as a pseudorandom permutation, keeping the key secret, and so viewing it as a secret S-box), and then use it in a SPN construction to obtain a block cipher on wn-bit inputs for $w > 1$. As mentioned earlier, non-linear, 1-round SPNs with secret S-boxes have been used for domain extension of block ciphers before [6, 16]. Other approaches for domain extension, not relying on (pure) SPNs, have also been considered [4, 9, 17, 18, 28]. Most prior work differ from ours as, to the best of our knowledge, they do not consider a stronger setting where the attacker has access to the underlying n-bit permutation. Exceptions that we are aware of are works that achieve domain extension by using Feistel networks [9, 37]. However, these only achieve a "native" domain extension of two, and are not as efficient for larger domains (see Coron et al. [9] for some discussion). Our main motivation, however, is not to beat prior approaches to domain extension, but rather to study the security of SPNs as used in practice to construct block ciphers.

2 Preliminaries

For $i \in \mathbb{N}$, we let $[i]$ denote the set $\{1, \ldots, i\}$. We write $\mathsf{Perm}(m)$ for the set of permutations of $\{0,1\}^m$. We view n as a cryptographic security parameter and let $\mathbb{F} \stackrel{\text{def}}{=} \mathrm{GF}(2^n)$, which is identified with $\{0,1\}^n$. If $x \in \mathbb{F}^w = \{0,1\}^{wn}$, then we denote the jth entry of x (for $j \in [w]$) by $x[j]$.

2.1 Substitution-Permutation Networks

A *substitution-permutation network (SPN)* defines a keyed permutation via repeated invocation of two transformations: blockwise computation of a public, cryptographic permutation called an "S-box," and application of a keyed, non-cryptographic permutation. Formally, an r-round SPN taking inputs of length wn where $w \in \mathbb{N}$ is the *width* of the network, is defined by $r+1$ keyed permutations $\{\pi_i : K_i \times \{0,1\}^{wn} \to \{0,1\}^{wn}\}_{i=0}^{r}$, a distribution $\mathcal{K}$ over[1]

[1] In practice, the *round keys* are derived from a single, master key using a prescribed *key schedule*, but for our purposes we leave this process implicit in the distribution $\mathcal{K}$.

$K_0 \times \cdots \times K_r$, and a permutation $S : \{0,1\}^n \to \{0,1\}^n$. Given round keys $(k_0, \ldots, k_r) \in K_0 \times \cdots \times K_r$ and an input $x \in \{0,1\}^{wn}$, the output of the SPN is computed as follows (cf. Fig. 1):

- Let[2] $x_1 := \pi_0(k_0, x)$.
- For $i = 1$ to r do:
  1. $y_i := \overline{S}(x_i)$, where $\overline{S}(x[1] \,\|\, \cdots \,\|\, x[w]) \stackrel{\text{def}}{=} S(x[1]) \,\|\, \cdots \,\|\, S(x[w])$.
  2. $x_{i+1} := \pi_i(k_i, y_i)$.
- The output is x_{r+1}.

If S is efficiently invertible and each π_i is efficiently invertible (given the appropriate key), then the above process is reversible given the round keys $k_0, \ldots, k_r$.

In our definition of an SPN, we apply a fixed permutation S to all w blocks of the intermediate state in each round. More generally, one could consider using w different functions $S_1, \ldots, S_w$ in each round, or even different S-boxes in different rounds. Our positive results hold even when a single permutation S is used, and our negative result holds even if multiple permutations are used.

Linear SPNs. We are interested in understanding the security of both linear and non-linear SPNs. We now define what we mean by these terms.

Definition 1. *A keyed permutation $\pi : \mathbb{F}^w \times \mathbb{F}^w \to \mathbb{F}^w$ is* linear[3] *if*

$$\pi(k, x) = (T_k \cdot k) + (T_x \cdot x) + \Delta,$$

where $T_k, T_x \in \mathbb{F}^{w \times w}$ are linear transformations, T_x is invertible, and $\Delta \in \mathbb{F}^w$. An SPN is linear *if all its round permutations $\{\pi_i\}_{i=0}^r$ are linear.*

If $\pi(k, x) = (T_k \cdot k) + (T \cdot x) + \Delta$ is linear, then we may write

$$\pi(k, x) = T \cdot (T^{-1}T_k \cdot k) + (T^{-1} \cdot \Delta) + x);$$

thus, by setting $k' = (T^{-1}T_k \cdot k) + T^{-1}\Delta$ we can express π as

$$\pi(k', x) = T \cdot (k' + x). \tag{1}$$

In other words, if an SPN is linear, then we may assume (by redefining the distribution $\mathcal{K}$ on keys appropriately) that each of its permutations π_i takes the form (1). This matches what is often done in practice (e.g., in AES, Serpent, PRESENT, etc.), where round permutations are computed by first performing a key-mixing step followed by an invertible linear transformation [22]. Since the linear transformation and key mixing commute, and the adversary can compute T and T^{-1} on its own (as T is public), we may further assume without loss of generality that the first and last permutations involve only a key-mixing step. In other words, we may set $\pi_i(k_i, x) = k_i + x$ for $i \in \{0, r\}$.

[2] Initial application of a keyed permutation is necessary, since otherwise the attacker could compute $\overline{S}$ in round 1 on its own, making that round ineffective.

[3] We could also call it *affine*, but we show shortly that $\Delta = 0$ without loss of generality.

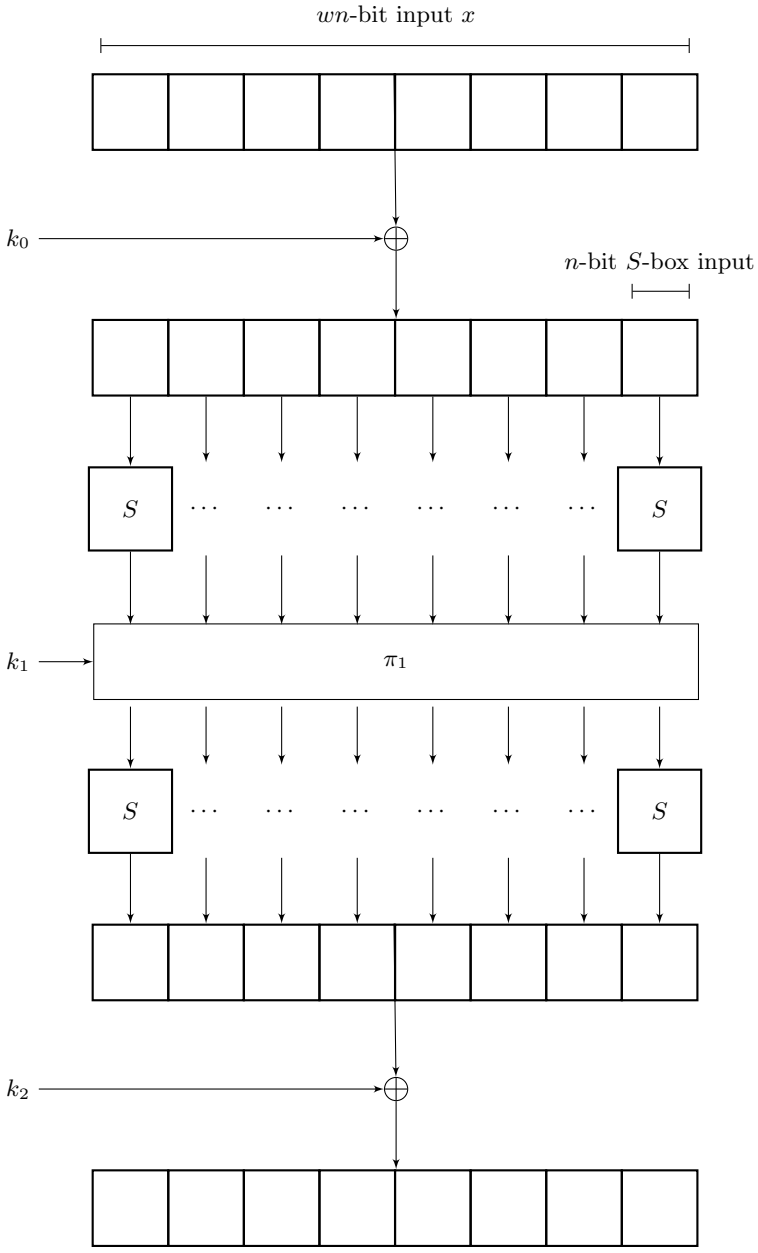

Fig. 1. A 2-round, linear SPN.

Security of an SPN. Our goal is to prove that certain SPNs are strong pseudorandom permutations. We cannot hope to prove such a result unconditionally; instead, we look at SPN *constructions* that are defined by permutations $\{\pi_i : K_i \times \{0,1\}^{wn} \to \{0,1\}^{wn}\}_{i=0}^{r}$ and a distribution $\mathcal{K}$, and that take *oracle access* to a public, random permutation $S : \{0,1\}^n \to \{0,1\}^n$; we write this as $\mathcal{C}^S$. We then analyze security of the construction against unbounded-time attackers making a bounded number of queries to the construction and to S. Formally, we consider the ability of an adversary D to distinguish two worlds: the "real world," in which it is given oracle access to S and $\mathcal{C}^S_{k_0,\ldots,k_r}$ (for unknown keys $k_0,\ldots,k_r$ sampled according to $\mathcal{K}$), and an "ideal world" in which it has access to S and an independent, random permutation $P : \{0,1\}^{wn} \to \{0,1\}^{wn}$. By default, we always allow D to make forward and inverse queries to both its oracles (though we do not write this explicitly). We have:

Definition 2. *The* strong-PRP advantage *of a distinguisher D against SPN construction $\mathcal{C}$ is:*

$$\mathsf{Adv}_{\mathcal{C}}(D) \stackrel{\text{def}}{=} \left|\Pr\left[(k_0,\ldots,k_r) \leftarrow \mathcal{K} : D^{\mathcal{C}^S_{k_0,\ldots,k_r},S} = 1\right] - \Pr\left[D^{P,S} = 1\right]\right|,$$

where P and S are independent, uniform permutations on $\{0,1\}^{wn}$ and $\{0,1\}^n$, respectively. The strong-PRP security *of $\mathcal{C}$ is*

$$\mathsf{Adv}_{\mathcal{C}}(q_C, q_S) \stackrel{\text{def}}{=} \max_D\{\mathsf{Adv}_{\mathcal{C}}(D)\},$$

where the maximum is taken over all distinguishers that make at most q_C queries to their left oracle and q_S queries to their right oracle.

2.2 The H-coefficient Technique

We use the H-coefficient technique [31, 32, 34] to prove various indistinguishability results. We provide a quick overview of its main ingredients here. Our presentation is essentially that of Chen and Steinberger [8]; for further details, refer there or to the tutorial by Patarin [36].

Fix a distinguisher D that makes at most q queries to its oracles. As in the security definition presented above, D's aim is to distinguish between two worlds: a "real world" and an "ideal world". Assume without loss of generality that D is deterministic. The execution of D defines a *transcript* that includes the sequence of queries and answers received from its oracles; D's output is a deterministic function of its transcript. Thus if X, Y denote the probability distributions on transcripts induced by the real and ideal worlds, respectively, then D's distinguishing advantage is upper bounded by the statistical distance

$$\Delta(X,Y) := \frac{1}{2}\sum_\tau |\Pr[X=\tau] - \Pr[Y=\tau]|,$$

where the sum is taken over all possible transcripts τ.

Let $\mathcal{T}$ denote the set of all transcripts that can be generated by D in either world. We look for a partition of $\mathcal{T}$ into two sets $\mathcal{T}_1$ and $\mathcal{T}_2$ of "good" and "bad" transcripts, respectively, along with a constant $\epsilon_1 \in [0, 1)$ such that

$$\tau \in \mathcal{T}_1 \implies \Pr[X = \tau] / \Pr[Y = \tau] \geq 1 - \epsilon_1. \tag{2}$$

It is then possible to show (see [8] for details) that

$$\Delta(X, Y) \leq \epsilon_1 + \Pr[Y \in \mathcal{T}_2] \tag{3}$$

where the right hand side of the inequality thus becomes an upper bound on the distinguisher's advantage. One should think of ϵ_1 and $\Pr[Y \in \mathcal{T}_2]$ as small, so "good" transcripts have nearly the same probability of appearing in the real world and the ideal world, whereas "bad" transcripts have low probability of occurring in the ideal world.

3 Linear SPNs

We begin by exploring the security of linear SPNs. We first show that 2-round linear SPNs cannot be secure against adaptive chosen-plaintext/ciphertext attacks when $w \geq 2$. Complementing this result, and giving a tight characterization, we then prove that 3-round linear SPNs can be secure when the round permutations and keys are chosen appropriately.

3.1 Insecurity of 2-Round, Linear SPNs

We present an attack showing that 2-round, linear SPNs (cf. Figure 1) cannot be secure for $w \geq 2$. The attack is based on one shown by Halevi and Rogaway [18] in a different context (and is a simple application of the boomerang technique [39]); our contribution here is to observe that the attack is applicable to any 2-round, linear SPN. The attack relies on the fact that the field $\mathbb{F} = \mathrm{GF}(2^n)$ is of characteristic 2.[4] In Appendix A, we present an attack that works for fields of arbitrary characteristic.

Recall from the previous section that any 2-round, linear SPN can be expressed in the following form. On input $x_0 \in \mathbb{F}^w$ and keys $k_0, k_1, k_2 \in \mathbb{F}^w$ do:

1. Compute $x_1 := x_0 \oplus k_0$ followed by $y_1 := \overline{S}(x_1)$.
2. Compute $x_2 := \pi_1(k_1, y_1) = T \cdot (y_1 \oplus k_1)$ for some invertible linear transformation T.
3. Compute $y_2 := \overline{S}(x_2)$ followed by $x_3 := y_2 \oplus k_2$, and return x_3.

[4] In a field of characteristic 2, any two field elements x and y can be swapped by adding the same nonzero constant c to both, i.e., $\{x, y\} = \{x \oplus c, y \oplus c\}$ for $c = x \oplus y$. This property, which has no analogue in fields of higher characteristic, is the central "trick" used at the heart of the Halevi-Rogaway attack.

We show an attacker D, given access to an oracle $\mathcal{O} : \{0,1\}^{wn} \to \{0,1\}^{wn}$ (with $w \geq 2$), that distinguishes whether $\mathcal{O}$ is an instance of the above construction (using uniform keys k_0, k_1, k_2) or a random permutation on $\{0,1\}^{wn}$. The attacker D proceeds as follows:

1. Choose inputs x_0, x_0' that are equal on all blocks except the first. with $x_0[1] \neq x_0'[1]$.
2. Query $\mathcal{O}(x_0)$ and $\mathcal{O}(x_0')$ to obtain x_3 and x_3' respectively.
3. Set $\hat{x}_3 := x_3'[1] \,\|\, x_3[2] \,\|\, \cdots \,\|\, x_3[w]$ and $\hat{x}_3' := x_3[1] \,\|\, x_3'[2] \,\|\, \cdots \,\|\, x_3'[w]$. Then query $\mathcal{O}^{-1}(\hat{x}_3)$ and $\mathcal{O}^{-1}(\hat{x}_3')$ to obtain $\hat{x}_0$ and $\hat{x}_0'$ respectively.
4. If $\hat{x}_0[2] \,\|\, \cdots \,\|\, \hat{x}_0[w] = \hat{x}_0'[2] \,\|\, \cdots \,\|\, \hat{x}_0'[w]$, then output 1; otherwise, output 0.

It is not hard to see that if $\mathcal{O}$ is a random permutation on $\{0,1\}^{wn}$ then D outputs 1 with negligible probability. On the other hand, we claim that when $\mathcal{O}$ is an instance of the above construction then D always outputs 1. To see this, let y_1, x_2 be the intermediate values during evaluation of $\mathcal{O}(x_0)$; let y_1', x_2' be the intermediate values during evaluation of $\mathcal{O}(x_0')$; let $\hat{y}_1, \hat{x}_2$ be the intermediate values during evaluation of $\mathcal{O}^{-1}(\hat{x}_3)$; and let $\hat{y}_1', \hat{x}_2'$ be the intermediate values during evaluation of $\mathcal{O}^{-1}(\hat{x}_3')$. Observe first that $\hat{x}_2 \oplus \hat{x}_2' = x_2 \oplus x_2'$. From this it follows that

$$\begin{aligned}
\hat{y}_1 \oplus \hat{y}_1' &= \left(T^{-1}\hat{x}_2 \oplus k_1\right) \oplus \left(T^{-1}\hat{x}_2' \oplus k_1\right) \\
&= T^{-1} \cdot (\hat{x}_2 \oplus \hat{x}_2') \\
&= T^{-1} \cdot (x_2 \oplus x_2') \\
&= T^{-1} \cdot (Ty_1 \oplus Ty_1') \\
&= T^{-1} \cdot (T \cdot (y_1 \oplus y_1')) \\
&= y_1 \oplus y_1'.
\end{aligned}$$

Since y_1 and y_1' are equal on all but their first blocks (by construction of x_0, x_0'), we conclude that $\hat{y}_1$ and $\hat{y}_1'$ also agree everywhere but in their first blocks. But this implies that $\hat{x}_0$ and $\hat{x}_0'$ are equal everywhere except in their first blocks, and so D outputs 1.

We remark that D makes only four queries to the construction—two in the forward direction, and two in the inverse direction—and no queries to S. We also note that the attack does not require linearity of T; it suffices for the permutation T to be *additive*, i.e., for T to satisfy

$$\forall x, y \in \{0,1\}^{wn} : \; T(x+y) = T(x) + T(y).$$

(Note that additivity of T implies additivity of T^{-1}.) Since $\mathbb{F}$ has characteristic 2, the "blockwise squaring" transformation $T : \mathbb{F}^w \to \mathbb{F}^w$ where $T(x)[j] = x[j]^2$ for all $j \in [w]$ is an example of a transformation that is additive but nonlinear.

3.2 Security of 3-Round, Linear SPNs

We now explore conditions under which 3-round, linear SPNs are secure. Recall from Section 2.1 that a 3-round SPN has four round permutations $\{\pi_i\}_{i=0}^{3}$, and

without loss of generality we may assume

$$\pi_i(k_i, x) = \begin{cases} x \oplus k_i & i \in \{0, 3\} \\ T_i \cdot (x \oplus k_i) & i \in \{1, 2\} \end{cases}, \tag{4}$$

where $T_1, T_2 \in \mathbb{F}^{w \times w}$ are invertible linear transformations. We prove that a 3-round, linear SPN is secure so long as (i) T_1 and T_2^{-1} contain no zero entries (Miles and Viola [29] show that matrices with maximal branch number [10] satisfy this property), and (ii) round keys k_0 and k_3 are (individually) uniform.

Theorem 1. *Assume $w > 1$. Let $\mathcal{C}$ be a 3-round, linear SPN with round permutations as in (4) and with distribution $\mathcal{K}$ over keys k_0, k_1, k_2, k_3. If k_0 and k_3 are uniformly distributed and the matrices T_1, T_2^{-1} contain no zero entries, then*

$$\mathsf{Adv}_{\mathcal{C}}(q_C, q_S) \leq \frac{5w^2 q_C^2 + 4w q_C q_S}{2^n - q_S - 2w} + \frac{q_C^2}{2^{wn}}.$$

Proof. Fix a deterministic distinguisher D. Without loss of generality, we assume D makes exactly q_C (non-redundant) forward/inverse queries to its left oracle that is either $\mathcal{C}_k^S$ or the ideal permutation P, and exactly q_S (non-redundant) forward/inverse queries to its right oracle that is the S-box. We call a query from D to its left oracle a *construction query* (even though in the ideal world the oracle is P), and a query from D to its right oracle an *S-box query*.

The interaction between D and its oracles can be recorded in the form of two sets of pairs $Q_C \subseteq \{0,1\}^{wn} \times \{0,1\}^{wn}$ and $Q_S \subseteq \{0,1\}^n \times \{0,1\}^n$, where Q_C contains every pair (x, y) for which D made a construction query x that was answered by y or an inverse query y that was answered by x, and Q_S is defined similarly with respect to S-box queries. Note that D's interaction with its oracles can be unambiguously reconstructed from these sets since D is deterministic.

Following [8], we augment the transcript (Q_C, Q_S) with a *key value* $k = (k_0, k_1, k_2, k_3)$. In the real world, k is the actual key used by the construction. In the ideal world, k is a dummy key sampled independently from all other values according to the prescribed key distribution $\mathcal{K}$. Thus, a transcript τ has the final form $\tau = (Q_C, Q_S, k_0, k_1, k_2, k_3)$.

Let $\mathcal{T}$ be set of all transcripts that can be generated with nonzero probability in the ideal world. (This includes all transcripts that can be generated with nonzero probability in the real world.) As in Section 2.2, let X, Y be the distributions over transcripts in the real and ideal worlds, respectively.

We define a set $\mathcal{T}_2 \subseteq \mathcal{T}$ of *bad transcripts* as follows: a transcript $\tau = (Q_C, Q_S, k_0, k_1, k_2, k_3)$ is bad if and only if one of these events occurs:

1. There exist a pair $(x, y) \in Q_C$, a pair $(a, b) \in Q_S$, and an index $j \in [w]$ such that $(x \oplus k_0)[j] = a$ or $(y \oplus k_3)[j] = b$.
2. There exist a pair $(x, y) \in Q_C$ and distinct indices $j, j' \in [w]$ such that $(x \oplus k_0)[j] = (x \oplus k_0)[j']$ or $(y \oplus k_3)[j] = (y \oplus k_3)[j']$.
3. There exist distinct pairs $(x, y), (x', y') \in Q_C$ and distinct indices $j, j' \in [w]$ such that $(x \oplus k_0)[j] = (x' \oplus k_0)[j']$ or $(y \oplus k_3)[j] = (y' \oplus k_3)[j']$.

As in Section 2.2, $\mathcal{T}_1 := \mathcal{T} \backslash \mathcal{T}_2$ denotes the set of *good* transcripts.

Since, in the ideal world, the values k_0, k_3 are independent of Q_C, Q_S and (individually) uniform in $\{0,1\}^{wn}$, a simple union bound shows that

$$\Pr[Y \in \mathcal{T}_2] < 2wq_Cq_S/2^n + w(w-1)q_C/2^n + w(w-1)q_C(q_C-1)/2^n$$

where the three terms account for the three events above, in that order. Thus

$$\Pr[Y \in \mathcal{T}_2] \leq 2wq_Cq_S/2^n + w(w-1)q_C^2/2^n. \tag{5}$$

This gives us "one half" of (3).

In order to finish applying the H-coefficient technique, it remains to lower bound the ratio

$$\frac{\Pr[X = \tau]}{\Pr[Y = \tau]}$$

for $\tau \in \mathcal{T}_1$. Let $\Omega_X = \mathsf{Perm}(n) \times K$ be the probability space underlying the real world, whose measure is the product of the uniform measure on $\mathsf{Perm}(n)$ and the measure induced by the distribution $\mathcal{K}$ on keys. (Thus, each element of Ω_X is a pair (S,k) with $S \in \mathsf{Perm}(n)$ and $k = (k_0,k_1,k_2,k_3) \in K$.) Also let

$$\Omega_Y = \mathsf{Perm}(wn) \times \mathsf{Perm}(n) \times K$$

be the probability space underlying the ideal world, whose measure is the product of the uniform measure on $\mathsf{Perm}(wn)$ with the measure on Ω_X.

Let $\nu = (Q_C^\nu, Q_S^\nu, k^\nu)$ be a transcript. An element $\omega = (S^*, k^*) \in \Omega_X$ is *compatible with* ν if $k^* = k^\nu$, if $S^*(a) = b$ for all $(a,b) \in Q_S^\nu$, and if $\mathcal{C}_{k^*}^{S^*}(x) = y$ for all $(x,y) \in Q_C^\nu$. An element $\omega = (P^*, S^*, k^*) \in \Omega_Y$ is *compatible with* ν if $k^* = k^\nu$, if $S^*(a) = b$ for all $(a,b) \in Q_S^\nu$, and if $P^*(x) = y$ for all $(x,y) \in Q_C^\nu$. We write

$$\omega \downarrow \nu$$

to indicate that an element $\omega \in \Omega_X \cup \Omega_Y$ is compatible with ν.

For the rest of the proof we fix a transcript $\tau = (Q_C, Q_S, k) \in \mathcal{T}_1$ so, in particular, Q_C, Q_S, and k will be fixed for the rest of the proof. Since $\tau \in \mathcal{T}$, it is easy to see (cf. [8]) that

$$\Pr[X = \tau] = \Pr_{w \leftarrow \Omega_X}[\omega \downarrow \tau] \tag{6}$$

$$\Pr[Y = \tau] = \Pr_{w \leftarrow \Omega_Y}[\omega \downarrow \tau], \tag{7}$$

where the notation indicates that ω is sampled from the relevant probability space according to that space's probability measure. (In other words, the probability of obtaining τ in each world is just the probability that the random coins of that world are "compatible" with τ in the sense outlined above.) We bound $\Pr[X = \tau]/\Pr[Y = \tau]$ by reasoning about the latter probabilities.

As additional notation/terminology:

- Say that $S^* \in \mathsf{Perm}(n)$ is *compatible* with a transcript $\nu = (Q_C^\nu, Q_S^\nu, k^\nu)$, and write $S^* \downarrow \nu$, if $(S^*, k) \in \Omega_X$ is compatible with ν, where k is the key value of the fixed transcript τ.

– Likewise, say that $(P^*, S^*) \in \mathsf{Perm}(wn) \times \mathsf{Perm}(n)$ is *compatible* with a transcript $\nu = (Q_C^\nu, Q_S^\nu, k^\nu)$, and write $(P^*, S^*) \downarrow \nu$, if $(P^*, S^*, k^\nu) \downarrow \nu$.

Incorporating these notations, the product structure of Ω_X, Ω_Y implies

$$\begin{aligned}\Pr_{\omega \leftarrow \Omega_X}[\omega \downarrow \tau] &= \Pr[\mathcal{K} = k] \cdot \Pr_{S^*}[S^* \downarrow \tau] \\ \Pr_{\omega \leftarrow \Omega_Y}[\omega \downarrow \tau] &= \Pr[\mathcal{K} = k] \cdot \Pr_{P^*,S^*}[(P^*, S^*) \downarrow \tau],\end{aligned}$$

where S^* and (P^*, S^*) are sampled uniformly from $\mathsf{Perm}(n)$ and $\mathsf{Perm}(wn) \times \mathsf{Perm}(n)$, respectively. Thus,

$$\frac{\Pr[X = \tau]}{\Pr[Y = \tau]} = \frac{\Pr_{S^*}[S^* \downarrow \tau]}{\Pr_{P^*,S^*}[(P^*, S^*) \downarrow \tau]}\,. \tag{8}$$

It is immediate that

$$\Pr_{P^*,S^*}[(P^*, S^*) \downarrow \tau] = \frac{(2^{wn} - q_C)!}{2^{wn}!} \cdot \frac{(2^n - q_S)!}{2^n!} \tag{9}$$

since Q_C, Q_S have size exactly q_C, q_S, respectively.

To compute $\Pr_{S^*}[S^* \downarrow \tau]$ we start by writing

$$\begin{aligned}\Pr_{S^*}[S^* \downarrow \tau] &= \Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k)] \\ &= \Pr_{S^*}[S^* \downarrow (\emptyset, Q_S, k)] \cdot \Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid S^* \downarrow (\emptyset, Q_S, k)] \\ &= \frac{(2^n - q_S)!}{2^n!} \cdot \Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid S^* \downarrow (\emptyset, Q_S, k)]\,. \end{aligned} \tag{10}$$

Define

$$\begin{aligned}\mathrm{Dom}(\tau) &\stackrel{\text{def}}{=} \{a \in \{0,1\}^n : (a, b) \in Q_S \text{ for some } b \in \{0,1\}^n\} \\ \mathrm{Range}(\tau) &\stackrel{\text{def}}{=} \{b \in \{0,1\}^n : (a, b) \in Q_S \text{ for some } a \in \{0,1\}^n\} \\ \mathrm{ExtDom}(\tau) &\stackrel{\text{def}}{=} \{(x \oplus k_0)[j] : (x, y) \in Q_C, j \in [w]\} \\ \mathrm{ExtRange}(\tau) &\stackrel{\text{def}}{=} \{(y \oplus k_3)[j] : (x, y) \in Q_C, j \in [w]\}.\end{aligned}$$

Note that $\mathrm{ExtDom}(\tau)$ (resp., $\mathrm{ExtRange}(\tau)$) contains all the first-round S-box inputs (resp., third-round S-box outputs) corresponding to the construction queries in Q_C. We let $\mathsf{Good}(S^*)$ be a predicate of S^* that holds if and only if all the following conditions are met (as usual, $\overline{S}^*$ denotes blockwise evaluation of S^* on a wn-bit string):

1. $S^* \downarrow (\emptyset, Q_S, k)$.
2. $T_1(\overline{S}^*(x \oplus k_0) \oplus k_1)[j] \notin \mathrm{Dom}(\tau) \cup \mathrm{ExtDom}(\tau)$ for all $(x, y) \in Q_C$ and all $j \in [w]$.
3. $(T_2^{-1}\overline{S}^{*-1}(y \oplus k_3) \oplus k_2)[j] \notin \mathrm{Range}(\tau) \cup \mathrm{ExtRange}(\tau)$ for all $(x, y) \in Q_C$ and all $j \in [w]$.

4. $T_1(\overline{S}^*(x \oplus k_0) \oplus k_1)[j] \neq T_1(\overline{S}^*(x' \oplus k_0) \oplus k_1)[j']$ for all distinct tuples $(x, y, j), (x', y', j') \in Q_C \times [w]$.
5. $(T_2^{-1}\overline{S}^{*-1}(y \oplus k_3) \oplus k_2)[j] \neq (T_2^{-1}\overline{S}^{*-1}(y' \oplus k_3) \oplus k_2)[j']$ for all distinct tuples (x, y, j), $(x', y', j') \in Q_C \times [w]$.

The second condition requires that no second-round S-box inputs are in $\mathrm{Dom}(\tau) \cup \mathrm{ExtDom}(\tau)$, and the fourth condition requires that all second-round S-box inputs are distinct; the third and fifth conditions parallel these, but for second-round S-box *outputs*.

We have

$$\begin{aligned}
&\Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid S^* \downarrow (\emptyset, Q_S, k)] \\
&\geq \Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \wedge \mathsf{Good}(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)] \\
&= \Pr_{S^*}[\mathsf{Good}(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)] \cdot \Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid \mathsf{Good}(S^*)], \quad (11)
\end{aligned}$$

using the fact that $\mathsf{Good}(S^*) \implies S^* \downarrow (\emptyset, Q_S, k)$ for the final equality. Thus, all that remains is to lower bound the two terms in the product of (11).

Let $\mathsf{Bad}_i(S^*)$ be the predicate that is true if and only if condition i in the definition of $\mathsf{Good}(S^*)$ is violated. Note that

$$\Pr_{S^*}\left[\overline{\mathsf{Good}(S^*)} \mid S^* \downarrow (\emptyset, Q_S, k)\right] \leq \sum_{i=2}^{5} \Pr_{S^*}\left[\mathsf{Bad}_i(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)\right],$$

since the first condition in the definition of $\mathsf{Good}(S^*)$ cannot be violated here. We now upper bound $\Pr_{S^*}[\mathsf{Bad}_i(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)]$ for $i = 2, 3, 4, 5$.

Lemma 1. $\Pr_{S^*}[\mathsf{Bad}_2(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)] \leq wq_C(q_S + wq_C)/(2^n - q_S - w)$.

Proof. Fix some $(x, y) \in Q_C$ and an index $j \in [w]$. Since τ is a good transcript, $(x \oplus k_0)[i] \notin \mathrm{Dom}(\tau)$ for all $i \in [w]$, and $(x \oplus k_0)[i] \neq (x \oplus k_0)[i']$ for $i' \neq i$. So after conditioning on $S^* \downarrow (\emptyset, Q_S, k)$ and the values of $S^*((x \oplus k_0)[i])$ for $i \neq 1$, the value $S^*((x \oplus k_0)[1])$ is uniform in a set of size $2^n - q_S - w + 1$. Because every entry in the first column of T_1 is nonzero, we have

$$\begin{aligned}
&\Pr_{S^*}[T_1(\overline{S}^*(x \oplus k_0) \oplus k_1)[j] \in \mathrm{Dom}(\tau) \cup \mathrm{ExtDom}(\tau) \mid S^* \downarrow (\emptyset, Q_S, k)] \\
&\qquad \leq \frac{|\mathrm{Dom}(\tau)| + |\mathrm{ExtDom}(\tau)|}{2^n - q_S - w + 1} \\
&\qquad \leq \frac{q_S + wq_C}{2^n - q_S - w}.
\end{aligned}$$

The statement follows by a union bound over all $(x, y) \in Q_C$, $j \in [w]$. □

Lemma 2. $\Pr_{S^*}[\mathsf{Bad}_3(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)] \leq wq_C(q_S + wq_C)/(2^n - q_S - w)$.

Proof. (Symmetric to Lemma 1.) □

Lemma 3. $\Pr_{S^*}[\mathsf{Bad}_4(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)] \le w^2 q_C^2/(2^n - q_S - 2w)$.

Proof. Fix distinct $(x, y, j), (x', y', j') \in Q_C \times [w]$. Then either $(x, y) \neq (x', y')$ (which implies $x \neq x'$) or $x = x'$ but $j \neq j'$.

Assume first that $x \neq x'$. Then $x[i_0] \neq x'[i_0]$ for some $i_0 \in [w]$. By the definition of a good transcript, $(x \oplus k_0)[i_0] \neq (x \oplus k_0)[i]$ for all $i \neq i_0$ and $(x \oplus k_0)[i_0] \neq (x' \oplus k_0)[i]$ for all i. So after conditioning on $S^* \downarrow (\emptyset, Q_S, k)$ and the values of $S^*((x \oplus k_0)[i])$ for $i \neq i_0$ and $S^*((x' \oplus k_0)[i])$ for $i \in [w]$, the value of $S^*((x \oplus k_0)[i_0])$ is uniform in a set of size at least $2^n - q_S - 2w + 1$. Because every entry in the i_0th column of T_1 is nonzero, we have

$$\Pr_{S^*}[T_1(\overline{S}^*(x \oplus k_0) \oplus k_1)[j] = T_1(\overline{S}^*(x' \oplus k_0) \oplus k_1)[j'] \mid S^* \downarrow (\emptyset, Q_S, k)] \le \frac{1}{2^n - q_S - 2w}\,.$$

Assume next that $x = x'$ and so $j \neq j'$. Since T_1 is invertible, the jth and j'th rows of T_1 are linearly independent and, in particular, there exists an index $i_0 \in [w]$ such that the (j, i_0)th and (j', i_0)th entries of T_1 are not equal. After conditioning on $S^* \downarrow (\emptyset, Q_S, k)$ and the values of $S^*((x \oplus k_0)[i])$ for $i \neq i_0$, the value of $S^*((x \oplus k_0)[i_0])$ is uniform in a set of size $2^n - q_S - w + 1$. It follows that

$$\Pr_{S^*}[T_1(\overline{S}^*(x \oplus k_0) \oplus k_1)[j] = T_1(\overline{S}^*(x \oplus k_0) \oplus k_1)[j'] \mid S^* \downarrow (\emptyset, Q_S, k)] \le \frac{1}{2^n - q_S - w}.$$

The statement now follows by taking a union bound over all possible pairs of distinct elements $(x, y, j), (x', y', j') \in Q_C \times [w]$. □

Lemma 4. $\Pr_{S^*}[\mathsf{Bad}_5(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)] \le w^2 q_C^2/(2^n - q_S - 2w)$.

Proof. (Symmetric to Lemma 3.) □

Combining the bounds of Lemmas 1–4, we find

$$\begin{aligned}\Pr_{S^*}[\mathsf{Good}(S^*) \mid S^* \downarrow (\emptyset, Q_S, k)] &\ge 1 - \frac{2w^2 q_C^2}{2^n - q_S - 2w} - \frac{2w q_C (q_S + w q_C)}{2^n - q_S - w} \\ &\ge 1 - \frac{4w^2 q_C^2 + 2w q_C q_S}{2^n - q_S - 2w}\,. \qquad (12)\end{aligned}$$

Our next step is to lower bound the term $\Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid \mathsf{Good}(S^*)]$ from (11). For any $S_\perp \in \mathsf{Perm}(n)$ such that $\mathsf{Good}(S_\perp)$ is true, define

$$Q^{\mathrm{Ext}}(S_\perp) = \{(a, S_\perp(a))\}_{a \in \mathrm{ExtDom}(\tau)} \cup \{(S_\perp^{-1}(b), b)\}_{b \in \mathrm{ExtRange}(\tau)}.$$

Let $\mathsf{Perm}_\perp(n) \subset \mathsf{Perm}(n)$ be maximal such that the predicate $\mathsf{Good}(S_\perp)$ holds for every $S_\perp \in \mathsf{Perm}_\perp(n)$ and such that $Q^{\mathrm{Ext}}(S_\perp) \neq Q^{\mathrm{Ext}}(S'_\perp)$ for distinct

$S_\perp$, $S'_\perp \in \mathsf{Perm}_\perp(n)$. (I.e., $\mathsf{Perm}_\perp(n)$ is a system of representatives, with one representative per distinct value of $Q^{\mathrm{Ext}}(S_\perp)$.) Then the event

$$\mathsf{Good}(S^*) \wedge (S^* \downarrow (Q_C, Q_S, k))$$

is the disjoint union of the events

$$\left\{\mathsf{Good}(S^*) \wedge \left(S^* \downarrow (Q_C, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k)\right)\right\}_{S_\perp \in \mathsf{Perm}_\perp(n)},$$

and so

$$\begin{aligned}
&\Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid \mathsf{Good}(S^*)] \\
&= \sum_{S_\perp \in \mathsf{Perm}_\perp(n)} \Pr_{S^*}[S^* \downarrow (Q_C, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k) \mid \mathsf{Good}(S^*)] \\
&= \sum_{S_\perp \in \mathsf{Perm}_\perp(n)} \Big(\Pr_{S^*}[S^* \downarrow (\emptyset, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k) \mid \mathsf{Good}(S^*)] \ \times \\
&\quad \Pr_{S^*}[S^* \downarrow (Q_C, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k) \mid S^* \downarrow (\emptyset, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k)]\Big). \qquad (13)
\end{aligned}$$

Fixing an arbitrary $S_\perp \in \mathsf{Perm}_\perp(n)$, parts 4 and 5 of the definition of $\mathsf{Good}(S_\perp)$ imply that the sets

$$\begin{aligned}
A &= \left\{T_1(\overline{S}_\perp(x \oplus k_0) \oplus k_1)[j] : (x, y) \in Q_C, j \in [w]\right\} \\
B &= \left\{(T_2^{-1}\overline{S}_\perp^{-1}(y \oplus k_3) \oplus k_2)[j] : (x, y) \in Q_C, j \in [w]\right\}
\end{aligned}$$

each consist of wq_C *distinct* elements, whereas parts 2 and 3 of the same definition imply that

$$\begin{aligned}
A \cap (\mathrm{Dom}(\tau) \cup \mathrm{ExtDom}(\tau)) &= \emptyset, \\
B \cap (\mathrm{Range}(\tau) \cup \mathrm{ExtRange}(\tau)) &= \emptyset.
\end{aligned}$$

It follows that $S^* \downarrow (Q_C, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k)$ iff $S^* \downarrow (\emptyset, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k)$ and $S^*(a) = b$ for all wq_C "matching" pairs (a, b) in $A \times B$ (that is, we match the element $a \in A$ associated with $(x, y) \in Q_C$ and $j \in [w]$ with the element $b \in B$ associated to the same (x, y) and j). Thus,

$$\begin{aligned}
\Pr_{S^*}[S^* \downarrow (Q_C, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k) \ \mid \ S^* \downarrow (\emptyset, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k)] \\
&= \frac{(2^n - q_S - |Q^{\mathrm{Ext}}(S_\perp)| - wq_C)!}{(2^n - q_S - |Q^{\mathrm{Ext}}(S_\perp)|)!} \\
&\geq \frac{(2^n - q_S - wq_C)!}{(2^n - q_S)!}.
\end{aligned}$$

Then since

$$\sum_{S_\perp \in \mathsf{Perm}_\perp(n)} \Pr_{S^*}[S^* \downarrow (\emptyset, Q_S \cup Q^{\mathrm{Ext}}(S_\perp), k) \mid \mathsf{Good}(S^*)] \ = \ 1,$$

(13) implies that

$$\Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid \mathsf{Good}(S^*)] \geq \frac{(2^n - q_S - wq_C)!}{(2^n - q_S)!}\,. \tag{14}$$

Combining (11), (12), and (14), we thus obtain

$$\Pr_{S^*}[S^* \downarrow (Q_C, Q_S, k) \mid S^* \downarrow (\emptyset, Q_S, k)] \\ \geq \left(1 - \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w}\right) \frac{(2^n - q_S - wq_C)!}{(2^n - q_S)!}\,.$$

By (10) we therefore have

$$\Pr_{S^*}[S^* \downarrow \tau] \geq \frac{(2^n - q_S)!}{2^n!} \left(1 - \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w}\right) \frac{(2^n - q_S - wq_C)!}{(2^n - q_S)!},$$

and so (using (8) and (9))

$$\begin{aligned}
\frac{\Pr[X = \tau]}{\Pr[Y = \tau]} &= \frac{\Pr_{S^*}[S^* \downarrow \tau]}{\Pr_{P^*,S^*}[(P^*, S^*) \downarrow \tau]} \\
&\geq \left(1 - \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w}\right) \frac{(2^n - q_S - wq_C)!}{(2^n - q_S)!} \Big/ \frac{(2^{wn} - q_C)!}{2^{wn}!} \\
&\geq \left(1 - \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w}\right) \left(\frac{1}{2^n - q_S}\right)^{wq_C} \Big/ \left(\frac{1}{2^{wn} - q_C}\right)^{q_C} \\
&\geq \left(1 - \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w}\right) \left(\frac{1}{2^n}\right)^{wq_C} \Big/ \left(\frac{1}{2^{wn} - q_C}\right)^{q_C} \\
&= \left(1 - \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w}\right) \left(1 - \frac{q_C}{2^{wn}}\right)^{q_C} \\
&= 1 - \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w} - \frac{q_C^2}{2^{wn}}.
\end{aligned}$$

Combining with (5), which gave an upper bound on the probability of obtaining a bad transcript in the ideal world, yields a final upper bound of

$$\frac{2wq_Cq_S}{2^n} + \frac{w(w-1)q_C^2}{2^n} + \frac{4w^2q_C^2 + 2wq_Cq_S}{2^n - q_S - 2w} + \frac{q_C^2}{2^{wn}} \\ \leq \frac{5w^2q_C^2 + 4wq_Cq_S}{2^n - q_S - 2w} + \frac{q_C^2}{2^{wn}}$$

on the distinguisher's advantage, as per (3). This completes the proof. □

A minimal secure (linear) SPN. We proved that a 3-round, linear SPN is secure if the keys k_0 and k_3 are individually uniform and T_1, T_2^{-1} contain no 0-entries. No assumptions were made about independence of k_0, k_3, nor were any assumptions made about the distributions of k_1, k_2. So the theorem implies

security for the following "minimal" 3-round, linear SPN: Let $k_0 = k_3 = k$, where k is uniform, set $k_1 = k_2 = 0^{wn}$, and let $T_1 = {T_2}^{-1} = T$ be invertible with no 0-entries. Define keyed permutations

$$\pi_i(k, x) = \begin{cases} x \oplus k & i \in \{0, 3\} \\ Tx & i = 1 \\ T^{-1}x & i = 2. \end{cases} \tag{15}$$

We have:

Corollary 1. *Assume $w > 1$. Let $\mathcal{C}$ be a 3-round, linear SPN with round permutations as in (15) and $\mathcal{K}$ choosing uniform $k_0 = k_3$ and $k_1 = k_2 = 0^{wn}$. Then*

$$\mathsf{Adv}_{\mathcal{C}}(q_C, q_S) \leq \frac{5w^2 q_C^2 + 4w q_C q_S}{2^n - q_S - 2w} + \frac{q_C^2}{2^{wn}}.$$

Reducing key-length. An inspection of the proof of Theorem 1 reveals that it is sufficient for the wn-bit key k ($= k_0 = k_3$) in Corollary 1 to satisfy the following conditions: informally, for any n-bit constant c and distinct indices i, i', (a) $k[i]$ equals c with negligible probability, and (b) the sum of $k[i]$ and $k[i']$ equals c with negligible probability. This can be achieved by choosing a uniform n-bit key k' and letting $k[i] = a_i \cdot k'$ where a_i are distinct non-zero elements of $\mathbb{F}$. Thus, one can make do with a "master key" of only n bits, while preserving the same security as in Corollary 1. We state this in the following corollary.

Corollary 2. *Assume $w > 1$. Let $\mathcal{C}$ be a 3-round, linear SPN with round permutations as in (15). Let a_i for $i = 1, \ldots, w$ be distinct non-zero elements of $\mathbb{F}$. Then, the distribution over the keys $\mathcal{K}$ is defined by choosing a uniform n-bit key k' and setting $k_0[i] = k_3[i] = a_i \cdot k'$ and $k_1 = k_2 = 0^{wn}$. Then*

$$\mathsf{Adv}_{\mathcal{C}}(q_C, q_S) \leq \frac{5w^2 q_C^2 + 4w q_C q_S}{2^n - q_S - 2w} + \frac{q_C^2}{2^{wn}}.$$

4 Non-Linear SPNs

In the previous section we considered linear SPNs and showed that 3 rounds are necessary for security to hold. In this section, we show that by allowing non-linear permutations, the number of rounds (i.e., the number of applications of the S-boxes) needed for constructing a secure SPN can be reduced to 1.

This section is organized as follows. We first define what it means for a keyed permutation over $\{0,1\}^{wn}$ to be *blockwise universal*.[5] We then show how to use blockwise-universal permutations to construct a 1-round SPN. Finally, we explore various constructions of blockwise-universal permutations.

[5] A similar notion was defined in [16, 17, 30] and even called blockwise universal in [16, 17]. The definition we give here is related, but different.

Interestingly, for $w = 1$ we show that simple key addition gives a blockwise-universal permutation; as a corollary of our work, we thus obtain a proof of security for the classical Even-Mansour construction [14], even in the case where the pre- and post-whitening keys are the same [13, 23]. For $w \geq 2$ our results from Section 3.1 imply that no linear function can be blockwise universal, but we show that several non-linear constructions are possible.

4.1 Secure 1-Round SPNs via Blockwise-Universal Permutations

We begin by defining the notion of a blockwise-universal (keyed) permutation over $\mathbb{F}^w$. Let the *blocks* of $y \in \mathbb{F}^w$ be $y[1], \ldots, y[w] \in \mathbb{F}$. Informally, a keyed permutation $\pi : K \times \mathbb{F}^w \to \mathbb{F}^w$ is blockwise universal if, for any distinct x, $x' \in \mathbb{F}^w$ and any $c \in \mathbb{F}$, the probability (taken over uniform $k \in K$) of each of the following events is low: (1) a block of $\pi(k, x)$ is equal to a block of $\pi(k, x')$, (2) two different blocks of $\pi(k, x)$ are equal, (3) a block of $\pi(k, x)$ is equal to c. Formally:

Definition 3. *A keyed permutation* $\{\pi : K \times \mathbb{F}^w \to \mathbb{F}^w\}$ *is* (ϵ, ϵ')-blockwise universal *if the following hold:*

1. *For all distinct* $(x, i), (x', i') \in \mathbb{F}^w \times [w]$*, we have*

$$\mathrm{Pr}_{k \leftarrow K}\left[\pi(k, x)[i] = \pi(k, x')[i']\right] \leq \epsilon.$$

2. *For all* $(x, i, c) \in \mathbb{F}^w \times [w] \times \mathbb{F}$*, we have* $\mathrm{Pr}_{k \leftarrow K}\left[\pi(k, x)[i] = c\right] \leq \epsilon'$.

If $\epsilon = \epsilon'$*, we simply call the keyed permutation* ϵ*-blockwise universal.*

Examples of (ϵ, ϵ')-blockwise universal permutations with small ϵ, ϵ' can be found in Section 4.2.

We now show that if π is blockwise universal, then a 1-round SPN construction $\mathcal{C}$ using round permutations $\pi_0 = \pi$ and $\pi_1 = \pi^{-1}$ is secure. In fact, $\mathcal{C}$ is secure even when π_0, π_1 share the same (uniform) key. I.e., we may define $\mathcal{C}$ as

$$\mathcal{C}_k^S(x) \stackrel{\mathrm{def}}{=} \pi^{-1}(k, \overline{S}(\pi(k, x))).$$

Theorem 2. *Let* $\pi : K \times \mathbb{F}^w \to \mathbb{F}^w$ *be* (ϵ, ϵ')*-blockwise universal. Then*

$$\mathsf{Adv}_{\mathcal{C}}(q_C, q_S) \leq w^2 q_C^2 \epsilon + 2w\, q_C\, q_S\, \epsilon',$$

where $\mathcal{C}$ *is the 1-round SPN construction in which* $\pi_0 = \pi$*,* $\pi_1 = \pi^{-1}$*, and in which* $\mathcal{K}$ *samples a uniform* $k \in K$ *and sets* $k_0 = k_1 = k$*.*

Proof. The proof is conceptually similar to the proof of Theorem 1, though the technical details are simpler here. Fix a deterministic distinguisher D. Without loss of generality, assume D makes exactly q_C (non-redundant) forward/inverse queries to its left oracle that is either $\mathcal{C}_k$ or P, and exactly q_S (non-redundant) forward/inverse queries to its right oracle that is the S-box. We call a query

from D to its left oracle a *construction query* (even though in the ideal world the oracle is P), and a query from D to its right oracle an *S-box query*.

The interaction between D and its oracles is recorded in two sets of pairs $Q_C \subseteq \{0,1\}^{wn} \times \{0,1\}^{wn}$ and $Q_S \subseteq \{0,1\}^n \times \{0,1\}^n$, where Q_C contains every pair (x, y) for which D made a construction query x that was answered by y or an inverse query y that was answered by x, and Q_S is defined similarly with respect to S-box queries. As in the proof of Theorem 1, we also append the key value k to the transcript in the real world and append a (uniform) dummy key value $k \in K$ to the transcript in the ideal world; a transcript τ thus has the form of a triple $\tau = (Q_C, Q_S, k)$.

Let $\mathcal{T}$ be the set of all possible transcripts that can be generated by D in the ideal world. We say that a transcript $\tau = (Q_C, Q_S, k)$ is *bad* if either:

1. There exist a pair $(x, y) \in Q_C$, a pair $(a, b) \in Q_S$, and an index $j \in [w]$ such that $\pi(k, x)[j] = a$ or $\pi(k, y)[j] = b$.
2. There exist distinct tuples $(x, y, j), (x', y', j') \in Q_C \times [w]$ such that $\pi(k, x)[j] = \pi(k, x')[j']$ or $\pi(k, y)[j] = \pi(k, y')[j']$.

We let $\mathcal{T}_2 \subseteq \mathcal{T}$ denote the set of bad transcripts. Transcripts in $\mathcal{T}_1 = \mathcal{T} \backslash \mathcal{T}_2$ are called *good* transcripts.

Let X, Y be the distributions over transcripts in the real and ideal worlds, respectively. Because the key value k is independent of Q_C, Q_S in the ideal world, the definition of (ϵ, ϵ')-blockwise universality and two applications of a union bound give

$$\Pr[Y \in \mathcal{T}_2] \leq 2wq_C q_S \epsilon' + w^2 q_C^2 \epsilon.$$

We next lower bound the ratio

$$\frac{\Pr[X = \tau]}{\Pr[Y = \tau]}$$

for transcripts $\tau \in \mathcal{T}_1$. Let Ω_X, Ω_Y be the probability spaces underlying the real and ideal worlds respectively. That is,

$$\Omega_X = \{(S, k) : S \in \mathsf{Perm}(n), k \in K\}$$
$$\Omega_Y = \{(P, S, k) : P \in \mathsf{Perm}(wn), S \in \mathsf{Perm}(n), k \in K\},$$

each with uniform measure.

For the rest of the proof, fix a transcript $\tau = (Q_C, Q_S, k) \in \mathcal{T}_1$. We say an element $\omega = (S^*, k^*) \in \Omega_X$ is *compatible with* τ if $k^* = k$, if $S^*(a) = b$ for all $(a, b) \in Q_S$, and if $\mathcal{C}_k^{S^*}(x) = y$ for all $(x, y) \in Q_C$. Analogously, an element $\omega = (P^*, S^*, k^*) \in \Omega_Y$ is *compatible with* τ if $k^* = k$, if $S^*(a) = b$ for all $(a, b) \in Q_S$, and if $P^*(x) = y$ for all $(x, y) \in Q_C$. We write $\omega \downarrow \tau$ to denote compatibility between $\omega \in \Omega_X \cup \Omega_Y$ and τ.

Since

$$\Pr[X = \tau] = |\{\omega \in \Omega_X : \omega \downarrow \tau\}| / |\Omega_X|$$
$$\Pr[Y = \tau] = |\{\omega \in \Omega_Y : \omega \downarrow \tau\}| / |\Omega_Y|$$

(compare with (6), (7)), we have

$$\begin{aligned}\frac{\Pr[X=\tau]}{\Pr[Y=\tau]} &= \frac{|\{\omega \in \Omega_X : \omega \downarrow \tau\}|}{|\Omega_X|} \cdot \frac{|\Omega_Y|}{|\{\omega \in \Omega_Y : \omega \downarrow \tau\}|} \\ &= \frac{(2^n - q_S - wq_C)!}{2^n!\,|K|} \cdot \frac{2^{wn}!\,2^n!\,|K|}{(2^{wn} - q_C)!\,(2^n - q_S)!} \\ &= \frac{(2^n - q_S - wq_C)!\,2^{wn}!}{(2^n - q_S)!\,(2^{wn} - q_C)!} \geq 1,\end{aligned}$$

where the fact that $\tau \in \mathcal{T}_1$ is used in the second equality. Hence we may apply the H-coefficient technique with $\epsilon_1 = 0$ (cf. Section 2.2), and D's distinguishing advantage is upper bounded by

$$\epsilon_1 + \Pr[Y \in \mathcal{T}_2] \leq 2wq_Cq_S\epsilon' + w^2q_C^2\epsilon$$

as claimed. □

4.2 Constructing Blockwise-Universal Permutations

We now show several constructions of blockwise-universal permutations.

Key mixing ($w = 1$). Let $w = 1$ and $k \in \{0,1\}^n$. Define $\pi(k, x) = x \oplus k$. It is trivial to see that this construction is $(0, 2^{-n})$-blockwise universal: if $x \neq x'$, then $x \oplus k \neq x' \oplus k$ (meaning $\epsilon = 0$); also, for any c we have $x \oplus k = c$ with probability $\epsilon' = 2^{-n}$ when k is uniform.

Instantiating the construction analyzed in Theorem 2 with this permutation yields the Even-Mansour construction [14] with the same key used for both pre- and post-whitening. Our results thus imply that the Even-Mansour construction in this case has a concrete security bound of $2q_Cq_S/2^n$, matching the security bound given in prior works [13, 23].

A construction of degree 3. A consequence of our attack in Section 3.1 is that blockwise-universal permutations with $w > 1$ must be non-linear. We now show a construction of degree 3. This construction is inspired by (though distinct from and considerably simpler than) a construction of a *non-keyed* permutation given by Dodis et al. [12] that achieves (in their terminology) good "entry-wise random collision resistance".

Let $w \geq 2$ and let $T \in \mathbb{F}^{w \times w}$ be an invertible matrix of the form

$$T = \left[\begin{array}{c|c} 1 & \\ 1 & \\ \vdots & * \\ 1 & \end{array}\right],$$

i.e., the first column of T is all 1s. We define a polynomial $p : \mathbb{F}^w \to \mathbb{F}$ by

$$p(x) = \bigoplus_{j=2}^{w} x[j]^3$$

and then define a transformation $\eta : \mathbb{F}^w \to \mathbb{F}^w$ by

$$\eta(x)[i] = \begin{cases} x[i] \oplus p(x) & \text{if } i = 1, \\ x[i] & \text{otherwise.} \end{cases}$$

It is easy to see that η is a permutation of $\mathbb{F}^w$. (In fact, since $\mathbb{F}$ has characteristic 2, the permutation η is an *involution*, i.e., it is its own inverse.) Finally, define the keyed permutation $\pi : \mathbb{F}^w \times \mathbb{F}^w \to \mathbb{F}^w$ as

$$\pi(k, x) = T \cdot \eta(x \oplus k). \tag{16}$$

Note that π is the composition of three invertible transformations, where the first transformation consists of the map $x \to x \oplus k$; hence, π is a keyed permutation.

Theorem 3. *The keyed permutation $\pi : \mathbb{F}^w \times \mathbb{F}^w \to \mathbb{F}^w$ defined in* (16) *is $(2/2^n, 1/2^n)$-blockwise universal.*

Proof. We start by noting that

$$\forall i : \ \pi(k, x)[i] = (Tx)[i] \oplus (Tk)[i] \oplus p(x \oplus k), \tag{17}$$

by the structure of T. The term $(Tk)[i]$ is a linear combination of the variables $k[1], \ldots, k[w]$ in which $k[1]$ has coefficient 1, whereas

$$p(x \oplus k) = (x[2] \oplus k[2])^3 \oplus \cdots \oplus (x[w] \oplus k[w])^3$$

can be viewed as the sum of $w - 1$ degree-3 polynomials in the variables $k[2]$, $\ldots$, $k[w]$ whose coefficients depend on x. Since $(Tk)[i]$ is the only term in (17) that depends on $k[1]$, it is easy to see that the second condition in the definition of blockwise universality is satisfied with $\epsilon' = 1/2^n$.

To verify the first condition of Definition 3, let (x, i), $(x', i') \in \mathbb{F}^w \times [w]$ be distinct. Then

$$\begin{aligned} &\pi(k, x)[i] \oplus \pi(k, x')[i'] \\ &= (Tx)[i] \oplus (Tx')[i] \oplus (Tk)[i] \oplus (Tk)[i'] \oplus p(x \oplus k) \oplus p(x' \oplus k). \end{aligned}$$

If the last $w-1$ entries of x, x' are equal, then $p(x \oplus k) = p(x' \oplus k)$ and moreover $(Tx)[i] \oplus (Tx')[i] = x[1] \oplus x'[1]$; thus,

$$\pi(k, x)[i] \oplus \pi(k, x')[i'] = x[1] \oplus x'[1] \oplus (Tk)[i] \oplus (Tk)[i'].$$

If $i = i'$ then $x[1] \oplus x'[1] \neq 0$ (otherwise (x, i) would be equal to (x', i')), so $\pi(k, x)[i] \neq \pi(k, x')[i']$. If $i \neq i'$ then $(Tk)[i] \oplus (Tk)[i']$ is uniform (by linear independence of the rows of T), so $\pi(k, x)[i] = \pi(k, x')[i']$ with probability $1/2^n$. In any case,

$$\Pr_k[\pi(k, x)[i] = \pi(k, x')[i']] \leq 1/2^n.$$

On the other hand, if the last $w - 1$ entries of x, x' are not all equal then there exists a $j \in \{2, \ldots, w\}$ such that $x[j] \neq x'[j]$. Then

$$(x[j] \oplus k[j])^3 \oplus (x'[j] \oplus k[j])^3$$

is a (nonzero) polynomial of degree 2 in $k[j]$. By extension, $p(x,k) \oplus p(x',k)$ and hence $\pi(k,x)[i] \oplus \pi(k,x')[i']$ are nonzero polynomials of degree 2 in $k[j]$ as well. Fixing arbitrary values of $k[j']$ for $j' \neq j$, the probability that this polynomial evaluates to zero over uniform choice of $k[j]$ is thus at most $2/2^n$; hence

$$\Pr_k[\pi(k,x)[i] = \pi(k,x')[i']] \leq 2/2^n$$

in this case. This concludes the proof. □

A construction with short keys. Let $T \in \mathbb{F}^{w \times w}$ be an invertible matrix, with entries $T_{i,j}$, such that (i) $T_{i,j} \neq 0$ for all $i,j \in [w]$, (ii) $\oplus_{j=1}^{w} T_{i,j} \neq 0$ for all $i \in [w]$, and (iii) $\oplus_{j=1}^{w} T_{i,j} \neq \oplus_{j=1}^{w} T_{i',j}$ if $i \neq i'$. (A random matrix has all the required properties with high probability.) Then we define the keyed permutation $\pi : \mathbb{F} \setminus \{0\} \times \mathbb{F}^w \to \mathbb{F}^w$ by $\pi(k,x) \coloneqq T z(k,x)$, where

$$z(k,x)[i] \coloneqq x[i] \cdot k^{i+1} \oplus k$$

for $i = 1, \ldots, w$. It is easy to verify that π is invertible when k is nonzero.

This construction can be viewed as individually applying a key-dependent affine transformation to each block $x[i]$, and then applying a linear transformation to the result (for a fixed k, the result is affine in x). Here we use a much shorter key k as compared to the previous construction, but at the expense of the non-linear transformation having degree $w+1$.

Theorem 4. *The keyed permutation $\pi : \mathbb{F} \setminus \{0\} \times \mathbb{F}^w \to \mathbb{F}^w$ defined above is $((w+1)/(2^n-1))$-blockwise universal.*

Proof. We first verify property 1 of Definition 3. Let $(x,i),(x',i') \in \mathbb{F}^w \times [w]$ be distinct. By definition of π, we have $\pi(k,x)[i] = \pi(k,x')[i']$ if and only if

$$\bigoplus_{j=1}^{w} T_{i,j}(x[j] \cdot k^{j+1} \oplus k) = \bigoplus_{j=1}^{w} T_{i',j}(x'[j] \cdot k^{j+1} \oplus k).$$

Rewriting the equation, this is the same as

$$\left(\bigoplus_{j=1}^{w} (T_{i,j} \oplus T_{i',j})\right) k \oplus \bigoplus_{j=1}^{w} (T_{i,j}\, x[j] \oplus T_{i',j}\, x'[j])\, k^{j+1} = 0. \tag{18}$$

If $i \neq i'$ then, since $\sum_{j=1}^{w} T_{i,j} \neq \sum_{j=1}^{w} T_{i',j}$, (18) is a polynomial equation in k of degree at least 1 and at most $w+1$. If $i = i'$ then $x \neq x'$ and so there exists $i_0 \in [w]$ such that $x[i_0] \neq x'[i_0]$. Again, this means (18) is a polynomial equation in k of degree at least $i_0 + 1$ and at most $w+1$. In either case, the probability over uniform choice of k that the equation holds is at most $(w+1)/(2^n-1)$.

To see property 2, observe that $\pi(k,x)[i] = c$ for some $c \in \mathbb{F}$ if and only if $\bigoplus_{j=1}^{w} T_{i,j}(x[j] \cdot k^{j+1} \oplus k) = c$. This is a polynomial equation in k of degree at least 1 (since $\bigoplus_{j=1}^{w} a_{i,j} \neq 0$) and at most $w+1$. So the probability this equation holds for uniform k is at most $(w+1)/(2^n-1)$. This concludes the proof. □

A second construction with short keys. We give another instantiation with n-bit keys that achieves the same parameters as above, but is (arguably) more direct and intuitive. Let $k \in \mathbb{F} \setminus \{0\}$. Let $C_1, \ldots, C_w \in \mathbb{F}$ be any w distinct nonzero elements. Then, we define

$$\pi(k, x)[i] \coloneqq \bigoplus_{j=1}^{w} (k \cdot C_i)^{j+1} x[j] \oplus k \cdot C_i, \tag{19}$$

for $i = 1, \ldots, w$. It is easy to verify that π is invertible when k is nonzero.

Theorem 5. *The keyed permutation $\pi : \mathbb{F} \setminus \{0\} \times \mathbb{F}^w \to \mathbb{F}^w$ as in (19) is $((w+1)/(2^n-1))$-blockwise universal.*

Proof. The proof is quite similar to that of Theorem 4. We first show that π satisfies property 1 of Definition 3. By definition of π, we have $\pi(k, x)[i] = \pi(k, x')[i']$ if and only if

$$\bigoplus_{j=1}^{w} (k \cdot C_i)^{j+1} x[j] \oplus k \cdot C_i = \bigoplus_{j=1}^{w} (k \cdot C_{i'})^{j+1} x'[j] \oplus k \cdot C_{i'}.$$

This is equivalent to

$$(C_i - C_{i'})k \oplus \bigoplus_{j=1}^{w} ({C_i}^{j+1} x[j] \oplus {C_{i'}}^{j+1} x'[j]) k^{j+1} = 0. \tag{20}$$

Fix distinct $(x, i) \neq (x', i') \in \mathbb{F}^w \times [w]$. If $i \neq i'$ then $C_i \neq C_{i'}$ and so (20) is a polynomial equation in k of degree at least 1 and at most $w+1$. If $i = i'$ then $x \neq x'$, and so there exists $i_0 \in [w]$ such that $x[i_0] \neq x'[i_0]$. Again, this implies that (20) is a polynomial equation in k of degree at least $t+1$ and at most $w+1$. In either case, the probability over uniform choice of k that the equation holds is at most $(w+1)/(2^n-1)$.

To see property 2, observe that $\pi(k, x)[i] = c$ for some $c \in \mathbb{F}$ if and only if $\bigoplus_{j=1}^{w} (k \cdot C_i)^{j+1} x[j] \oplus k \cdot C_i = c$. This is a polynomial equation in k of degree at least 1 (since C_i is nonzero by assumption) and at most $w+1$. The probability with which this equation holds for uniform k is at most $(w+1)/(2^n-1)$. This concludes the proof. □

Relation to our 3-round, linear SPN construction. We conclude with the following informal observation relating the result of Theorem 2 to our 3-round, linear SPN construction from Section 3.2: The initial round of the "minimal" linear SPN discussed there (i.e., key mixing, followed by evaluation of $\overline{S}$, and then finally a linear transformation) can be shown to be a blockwise-universal permutation if the S-box is viewed as part of the key of the permutation. To see this, consider any $(x, i) \neq (x', i')$. If $x = x'$ but $i \neq i'$, then key mixing with a uniform wn-bit key ensures that the output is uniform, so two different blocks collide with probability $1/2^n$. On the other hand, if $x \neq x'$ then $x[j] \neq x'[j]$ for

some j. After the key mixing, the S-boxes will be applied to two unequal values $x[j] \oplus k[j]$ and $x'[j] \oplus k[j]$, which means the resulting values $y[j]$ and $y'[j]$ will be uniform and (essentially) independent. Subsequent application of the linear transformation T ensures that any blocks i and i' of the outputs will also be uniform and uncorrelated, hence unlikely to collide.

Thus, the 3-round, linear SPN construction can be viewed almost as a "special case" of our 1-round, non-linear SPN, in which the blockwise-universal permutations are implemented via a linear SPN round. We stress that this is only intuition, and formally we cannot derive Theorem 1 as a corollary of Theorem 2 because (i) the same S-boxes are shared in the permutations and the middle layer, and (ii) the S-boxes are *public*, which is not taken into account in the definition of blockwise universality.

5 Conclusion and Open Problems

We study the security of SPNs as strong pseudorandom permutations when the S-box is modeled as a public random permutation. This model captures the design approach of most block ciphers following the SPN paradigm. Within this model, we give an exact characterization of the properties required to achieve security in both the linear and non-linear settings.

A number of interesting open questions remain. For instance, while generic information-theoretic attacks show that constructions with at most 1 round cannot surpass birthday security (presuming a key of $O(wn)$ bits) we are not aware of matching birthday attacks at 3 rounds. Hence the question of determining the exact security of linear 3-round SPNs (and in particular, whether the security goes beyond birthday or not) remains open. Moreover, proving beyond-birthday security at *any* number of rounds remains open as well. This question has been analyzed extensively for Feistel networks [26, 33–35] and iterated Even-Mansour constructions [5, 7, 8, 19, 24].

Another, more technical question concerns SPNs with a limited number of nonlinear rounds. Specifically, if we limit consideration to SPNs using exactly one nonlinear keyed permutation, then we do not know if 2 rounds suffice for security. (Note that a 1-round network with this structure is easily attacked. On the other hand, 3 rounds suffice by the results of this paper.)

Acknowledgments

Work of the second and fourth authors was performed under financial assistance award 70NANB15H328 from the U.S. Department of Commerce, National Institute of Standards and Technology.

References

1. Alex Biryukov, Charles Bouillaguet, and Dmitry Khovratovich. Cryptographic schemes based on the ASASA structure: Black-box, white-box, and public-key

(extended abstract). In *Advances in Cryptology—Asiacrypt 2014, Part I*, volume 8873 of *LNCS*, pages 63–84. Springer, 2014.
2. Alex Biryukov and Dmitry Khovratovich. Decomposition attack on SASASASAS. Available at http://eprint.iacr.org/2015/646.
3. Alex Biryukov and Adi Shamir. Structural cryptanalysis of SASAS. *Journal of Cryptology*, 23(4):505–518, 2010.
4. Daniel Bleichenbacher and Anand Desai. A construction of a super-pseudorandom cipher, February 1999. Unpublished manuscript.
5. Andrey Bogdanov, Lars R. Knudsen, Gregor Leander, François-Xavier Standaert, John P. Steinberger, and Elmar Tischhauser. Key-alternating ciphers in a provable setting: Encryption using a small number of public permutations. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology—Eurocrypt 2012*, volume 7237 of *LNCS*, pages 45–62. Springer, 2012.
6. Debrup Chakraborty and Palash Sarkar. A new mode of encryption providing a tweakable strong pseudo-random permutation. In *Fast Software Encryption—FSE 2006*, volume 4047 of *LNCS*, pages 293–309. Springer, 2006.
7. Shan Chen, Rodolphe Lampe, Jooyoung Lee, Yannick Seurin, and John P. Steinberger. Minimizing the two-round Even-Mansour cipher. In *Advances in Cryptology—Crypto 2014, Part I*, volume 8616 of *LNCS*, pages 39–56. Springer, 2014.
8. Shan Chen and John P. Steinberger. Tight security bounds for key-alternating ciphers. In *Advances in Cryptology—Eurocrypt 2014*, volume 8441 of *LNCS*, pages 327–350. Springer, 2014.
9. Jean-Sébastien Coron, Yevgeniy Dodis, Avradip Mandal, and Yannick Seurin. A domain extender for the ideal cipher. In *7th Theory of Cryptography Conference—TCC 2010*, volume 5978 of *LNCS*, pages 273–289. Springer, 2010.
10. Joan Daemen. *Cipher and Hash Function Design Strategies Based on Linear and Differential Cryptanalysis*. PhD thesis, Katholieke Universiteit Leuven, 1995.
11. Itai Dinur, Orr Dunkelman, Thorsten Kranz, and Gregor Leander. Decomposing the ASASA block cipher construction. Available at http://eprint.iacr.org/2015/507.
12. Yevgeniy Dodis, Martijn Stam, John P. Steinberger, and Tianren Liu. Indifferentiability of confusion-diffusion networks. In *Advances in Cryptology—Eurocrypt 2016, Part II*, volume 9666 of *LNCS*, pages 679–704. Springer, 2016.
13. Orr Dunkelman, Nathan Keller, and Adi Shamir. Minimalism in cryptography: The Even-Mansour scheme revisited. In *Advances in Cryptology—Eurocrypt 2012*, volume 7237 of *LNCS*, pages 336–354. Springer, 2012.
14. Shimon Even and Yishay Mansour. A construction of a cipher from a single pseudorandom permutation. In *Advances in Cryptology—Asiacrypt'91*, volume 739 of *LNCS*, pages 210–224. Springer, 1991.
15. Horst Feistel. Cryptography and computer privacy. *Scientific American*, 228(5):15–23, 1973.
16. Shai Halevi. Invertible universal hashing and the TET encryption mode. In *Advances in Cryptology—Crypto 2007*, volume 4622 of *LNCS*, pages 412–429. Springer, 2007.
17. Shai Halevi and Phillip Rogaway. A tweakable enciphering mode. In *Advances in Cryptology—Crypto 2003*, volume 2729 of *LNCS*, pages 482–499. Springer, 2003.
18. Shai Halevi and Phillip Rogaway. A parallelizable enciphering mode. In *Cryptographers' Track—RSA 2004*, volume 2964 of *LNCS*, pages 292–304. Springer, 2004.

19. Viet Tung Hoang and Stefano Tessaro. Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In *Advances in Cryptology—Crypto 2016, Part I*, volume 9814 of *LNCS*, pages 3–32. Springer, 2016.
20. Tetsu Iwata and Kaoru Kurosawa. On the pseudorandomness of the AES finalists—RC6 and Serpent. In *Fast Software Encryption—FSE 2000*, volume 1978 of *LNCS*, pages 231–243. Springer, 2000.
21. Antoine Joux. Cryptanalysis of the EMD mode of operation. In *Advances in Cryptology—Eurocrypt 2003*, volume 2656 of *LNCS*, pages 1–16. Springer, 2003.
22. J. Katz and Y. Lindell. *Introduction to Modern Cryptography, 2nd edition*. Chapman & Hall/CRC Press, 2015.
23. Joe Kilian and Phillip Rogaway. How to protect DES against exhaustive key search (an analysis of DESX). *Journal of Cryptology*, 14(1):17–35, 2001.
24. Rodolphe Lampe, Jacques Patarin, and Yannick Seurin. An asymptotically tight security analysis of the iterated Even-Mansour cipher. In Xiaoyun Wang and Kazue Sako, editors, *Advances in Cryptology—Asiacrypt 2012*, volume 7658 of *LNCS*, pages 278–295. Springer, 2012.
25. Michael Luby and Charles Rackoff. How to construct pseudorandom permutations from pseudorandom functions. *SIAM Journal on Computing*, 17(2):373–386, 1988.
26. Ueli M. Maurer and Krzysztof Pietrzak. The security of many-round Luby-Rackoff pseudo-random permutations. In Eli Biham, editor, *Advances in Cryptology—Eurocrypt 2003*, volume 2656 of *LNCS*, pages 544–561. Springer, 2003.
27. Ueli M. Maurer, Renato Renner, and Clemens Holenstein. Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology. In *1st Theory of Cryptography Conference—TCC 2004*, volume 2951 of *LNCS*, pages 21–39. Springer, 2004.
28. David A. McGrew and Scott R. Fluhrer. The security of the extended codebook (XCB) mode of operation. In *14th Annual Intl. Workshop on Selected Areas in Cryptography (SAC)*, volume 4876 of *LNCS*, pages 311–327. Springer, 2007.
29. Eric Miles and Emanuele Viola. Substitution-permutation networks, pseudorandom functions, and natural proofs. *J. ACM*, 62(6):46, 2015.
30. Moni Naor and Omer Reingold. On the construction of pseudorandom permutations: Luby-Rackoff revisited. *Journal of Cryptology*, 12(1):29–66, 1999.
31. Jacques Patarin. Pseudorandom permutations based on the D.E.S. scheme. In *1st European Symposium on Research in Computer Security (ESORICS)*, LNCS, pages 185–187, 1990.
32. Jacques Patarin. *Etude de Générateurs de Permutations Basés sur les Schémas du DES*. PhD thesis, INRIA, 1991.
33. Jacques Patarin. About Feistel schemes with six (or more) rounds. In Serge Vaudenay, editor, *Fast Software Encryption—FSE '98*, volume 1372 of *LNCS*, pages 103–121. Springer, March 1998.
34. Jacques Patarin. Luby-Rackoff: 7 rounds are enough for $2^{n(1-\epsilon)}$ security. In *Advances in Cryptology—Crypto 2003*, volume 2729 of *LNCS*, pages 513–529. Springer, 2003.
35. Jacques Patarin. Security of random Feistel schemes with 5 or more rounds. In Matthew Franklin, editor, *Advances in Cryptology—Crypto 2004*, volume 3152 of *LNCS*, pages 106–122. Springer, 2004.
36. Jacques Patarin. The "coefficients-H" technique (invited talk). In *SAC 2008: 15th Annual Intl. Workshop on Selected Areas in Cryptography (SAC)*, LNCS, pages 328–345. Springer, 2009.

37. Zulfikar Ramzan and Leonid Reyzin. On the round security of symmetric-key cryptographic primitives. In Mihir Bellare, editor, *Advances in Cryptology—Crypto 2000*, volume 1880 of *LNCS*, pages 376–393. Springer, 2000.
38. Claude E. Shannon. Communication theory of secrecy systems. *Bell Systems Technical Journal*, 28(4):656–715, 1949.
39. David Wagner. The boomerang attack. In Lars R. Knudsen, editor, *Fast Software Encryption—FSE '99*, volume 1636 of *LNCS*, pages 156–170. Springer, March 1999.

A A General Attack on 2-Round Linear SPNs

We present an attack showing that 2-round, linear SPNs are insecure for $w \geq 2$, regardless of the characteristic of $\mathbb{F}$. (In particular, this means that we do not define $\mathbb{F}$ to be $\mathrm{GF}(2^n)$ in this section as was the convention in the main body of the paper.) As will be seen, the ideas for this attack are fairly different than that of the attack presented in Section 3.1.

Let $\mathbb{F}$ be a finite field such that $\log(|\mathbb{F}|) = O(n)$. We use "+" to denote field addition. The definitions in Section 2.1 can be modified in a straightforward manner to be defined over a field $\mathbb{F}$ as defined above (instead of by setting $\mathbb{F} = \mathrm{GF}(2^n)$ as in that section). Then, any 2-round, linear SPN can be expressed in the following form. On input $x \in \mathbb{F}^w$ and for subkeys $k_0, k_1, k_2 \in \mathbb{F}^w$, do:

1. Compute $x_1 := x + k_0$ followed by $y_1 := \overline{S}(x_1)$.
2. Compute $x_2 := \pi_1(k_1, y_1) = T \cdot (y_1 + k_1)$ for some invertible linear transformation T.
3. Compute $y_2 := \overline{S}(x_2)$ followed by $y := y_2 + k_2$, and return y.

In fact, we will describe our attack for the following slightly more general scheme that replaces the inner key addition and linear map T with an arbitrary key-dependent affine map $U : \mathbb{F}^w \to \mathbb{F}^w$. The more general scheme is as follows:

1. Compute $x_1 := x + k_0$ followed by $y_1 := \overline{S}(x_1)$.
2. Compute $x_2 := U(y_1)$ for some invertible $\mathbb{F}$-affine map $U : \mathbb{F}^w \to \mathbb{F}^w$ (possibly depending on k_0, k_1, k_2).
3. Compute $y_2 := \overline{S}(x_2)$ followed by $y := y_2 + k_2$, and return y.

Obviously, the original scheme is recovered as a special case of the second scheme by setting $U(y_1) = T \cdot (y_1 + k_1)$. We will write the above scheme as $2\mathrm{rnd}^S_{k_0,k_1,k_2}$. Thus

$$2\mathrm{rnd}^S_{k_0,k_1,k_2}(x) = \overline{S}(U(\overline{S}(x + k_0))) + k_2$$

for all $x \in \mathbb{F}^w$. The inverse map $(2\mathrm{rnd}^S_{k_0,k_1,k_2})^{-1}$ is obtained by replacing U with U^{-1}, $\overline{S}$ with $\overline{S}^{-1}$, and by reversing the order of the subkey additions.

To further simplify the description of our attack we will assume $w = 2$. (The general attack for $w \geq 2$ can be easily recovered from this description.) The pseudocode of our distinguisher D, that distinguishes $\mathcal{O} = 2\mathrm{rnd}^S_{k_0,k_1,k_2}$ with $w = 2$ from a uniform random permutation $\mathcal{O}$ of $\mathbb{F}^2$ (indeed, our distinguisher makes

no use of S-box queries), is given in Algorithm 1. However, some preliminary description of high-level ideas is necessary to understand how the attack works.

We start by establishing some terminology. Removing ourselves a moment from the distinguisher D described in Algorithm 1, consider a generic distinguisher D with access to a permutation oracle $\mathcal{O} : \mathbb{F}^2 \to \mathbb{F}^2$ and its inverse $\mathcal{O}^{-1}$. We say that D makes a *forward* query when it calls $\mathcal{O}(\cdot)$ and that it makes a *backward* query when it calls $\mathcal{O}^{-1}(\cdot)$. For the purpose of the discussion we assume that D never makes a redundant query to $\mathcal{O}$. The latter includes calling $\mathcal{O}^{-1}(y)$ where y is some value previously obtained by a forward query $\mathcal{O}(x)$, or calling $\mathcal{O}(x)$ where x is some value previously returned by a backward query $\mathcal{O}^{-1}(y)$.

The queries made by D, and answers obtained, can be kept in a *query history* of the form $(\sigma^i, x^i, y^i)_{i=1}^q$ where q is the number of queries made by D and where $\sigma^i \in \{+, -\}$ records the direction of the i-th query. In more detail, if $\sigma^i = +$ then D's i-th query was $\mathcal{O}(x^i)$ and it was answered by y^i, while if $\sigma^i = -$ then D's i-th query was $\mathcal{O}^{-1}(y^i)$ and it was answered by x^i.

To keep the following discussion formal we introduce the notion of a *token*. This is an element of the set $\mathbb{F} \times \{\mathrm{TL}, \mathrm{BL}, \mathrm{TR}, \mathrm{BR}\}$ where TL, BL, TR and BR are formal symbols that stand for "top left", "bottom left" "top right" and "bottom right" respectively. We will be using tokens of the type $(x^i[1], \mathrm{TL})$, $(x^i[2], \mathrm{BL})$, $(y^i[1], \mathrm{TR})$ and $(y^i[2], \mathrm{BR})$ for elements (σ^i, x^i, y^i) of the query history. We refer to the first coordinate of a token as its *value* and to the second coordinate of a token as its *position*. For the sake of expediency we will often describe a token by its value when the position can be inferred from the notation. (Such as, specifically, by writing "$x^i[1]$" instead of "$(x^i[1], \mathrm{TL})$".)

The tokens induced by a query history $(\sigma^i, x^i, y^i)_{i=1}^q$ have a natural topological structure that we will encode in terms of certain binary trees. The building blocks for these binary trees will be *bipods*, that we define next.

A bipod is a three-node graph with a middle node (the "child") connected to two outer nodes (the "parents"). We write a bipod with child v and parents u, w as "u—v—w". The nodes of a bipod will be labeled with tokens. Specifically, each forward query $(+, x^i, y^i)$ in the query history gives rise to two bipods, $x^i[1]$—$y^i[1]$—$x^i[2]$ and $x^i[1]$—$y^i[2]$—$x^i[2]$ (or

$$(x^i[1], \mathrm{TL})\text{—}(y^i[1], \mathrm{TR})\text{—}(x^i[2], \mathrm{BL})$$

and

$$(x^i[1], \mathrm{TL})\text{—}(y^i[2], \mathrm{BR})\text{—}(x^i[2], \mathrm{BL})$$

if we write out the tokens fully). Likewise, each backward query $(-, x^i, y^i)$ gives rise to the two opposite bipods, namely $y^i[1]$—$x^i[1]$—$y^i[2]$ and $y^i[1]$—$x^i[2]$—$y^i[2]$.

One can observe that the parent/child terminology is chosen to make parents come chronologically before children. That is, when the adversary makes (say) a forward query $\mathcal{O}(x^i)$, and obtains answers y^i, the "parents" $x^i[1]$ and $x^i[2]$ were known to the adversary before the "children" $y^i[1]$ and $y^i[2]$. (Referring to the two bipods created by the resulting query $(+, x^i, y^i)$.)

We say that a bipod α is *earlier* than a bipod β if the query associated to α is earlier in the query history than the query associated to β.

A *bipod tree* is a rooted binary tree, recursively defined as follows: (i) every bipod is a bipod tree rooted at the child, (ii) if T is a bipod tree, ℓ is a leaf of T, and B is a bipod that is earlier than the bipod containing ℓ in T such that the child of B is labeled by the same token as ℓ, attaching B to T by identifying ℓ with the child (root) of B yields another bipod tree.

For example, if $(+, x^i, y^i)$ and $(-, x^j, y^j)$ are two query history elements such that $i < j$ and such that $y^i[1] = y^j[1]$, then we may identify the child of the bipod $x^i[1]$—$y^i[1]$—$x^i[2]$ (coming from $(+, x^i, y^i)$) with the first parent of, say, $y^j[1]$—$x^j[1]$—$y^j[2]$ (coming from $(-, x^j, y^j)$) such as to form a binary tree with 5 nodes rooted at $x^j[1]$.

If a bipod tree has been maximally extended[6] (i.e., nothing can be further attached to the leaves) then one may think of the tree as being some kind of "genealogical certificate" of the root (i.e., it shows a query that gave rise to the token at the root, it shows queries that gave rise to its parents within that query if such exist, and so on). Intuitively, two distinct maximally extended trees should have distinct values at the root. (At least, this is intuitively the case if the oracle $\mathcal{O}(\cdot)$ is a random permutation of $\mathbb{F}^2$, and the number of queries is polynomial.) Our attack essentially boils down to showing that if $\mathcal{O} = 2\mathrm{rnd}^S_{k_0,k_1,k_2}$, then there is a way to build two different bipod trees of depth 4 (comprising a total of $q = 14$ queries to $\mathcal{O}$) whose roots are labeled by the same token.

Assume that $\mathcal{O} = 2\mathrm{rnd}^S_{k_0,k_1,k_2}$. Given a token $t = (u, \mathrm{Pos})$ where $u \in \mathbb{F}$ and $\mathrm{Pos} \in \{\mathrm{TL}, \mathrm{BL}, \mathrm{TR}, \mathrm{BR}\}$, we define the *inner value* $I(t) \in \mathbb{F}$ of t by:

$$I(t) := \begin{cases} S(u + k_0[1]) & \text{if Pos} = \mathrm{TL} \\ S(u + k_0[2]) & \text{if Pos} = \mathrm{BL} \\ S^{-1}(u + k_2[1]) & \text{if Pos} = \mathrm{TR} \\ S^{-1}(u + k_2[2]) & \text{if Pos} = \mathrm{BR} \end{cases}$$

In other words the inner value is the value on the wire adjacent to the affine transformation U, as reached from the token's value by key addition and by application of S or S^{-1}.

Since $U : \mathbb{F}^2 \to \mathbb{F}^2$ is an $\mathbb{F}$-affine function, there exists six (in fact, unique) values a, b, c, d, e, $f \in \mathbb{F}$ such that

$$U(r, s) = (ar + bs + c, dr + es + f)$$

for all $(r, s) \in \mathbb{F}^2$; likewise, since U is invertible, there exist (unique) values A, B, C, D, E, $F \in \mathbb{F}$ such that

$$U^{-1}(r, s) = (Ar + Bs + C, Dr + Es + F)$$

[6] We do *not* make any implicit claim that such a maximal extension is unique for a given root. Indeed, our attack actually hinges on the non-uniqueness of the maximal bipod tree having a given token as the root, when $\mathcal{O} = 2\mathrm{rnd}^S_{k_0,k_1,k_2}$.

for all $(r, s) \in \mathbb{F}^2$. By definition of inner values, and because $\mathcal{O} = 2\mathrm{rnd}^S_{k_0,k_1,k_2}$, we have

$$I(y^i[1]) = aI(x^i[1]) + bI(x^i[2]) + c$$
$$I(y^i[2]) = dI(x^i[1]) + eI(x^i[2]) + f$$

(where $y^i[1]$ represents the token $(y^i[1], \mathrm{TR})$, etc.) for all forward[7] queries $(+, x^i, y^i)$ in the query history, and

$$I(x^i[1]) = AI(y^i[1]) + BI(y^i[2]) + C$$
$$I(x^i[2]) = DI(y^i[1]) + EI(y^i[2]) + F$$

for all backward queries $(-, x^i, y^i)$ in the query history. Thanks to these relations, the inner values of leaf tokens in a bipod tree can be used to determine the inner values of all remaining tokens in the tree—in fact, the inner values of non-leaf tokens are affine combinations of the inner values at the leaves, since an affine combination of affine combinations is an affine combination. Essentially, our attack succeeds in constructing two distinct bipod trees that share the same *set* of tokens at the leaves, albeit arranged in a different order along the leaves, such that the affine combinations at the two roots coincide, and such that the two roots are both in position TL. Since two tokens in the same position have the same value if and only if their inner value is the same, it suffices for the adversary to construct these two trees and to check equality at the roots.

Algorithm 1 "Bipod Tree" Attack for Linear 2-Round SPNs

Uniformly and independently sample $w_1, w_2, z_1, z_2, z_3, z_4, z_5$ and $z_6 \in \mathbb{F}$;
for $i = 1$ to 2 **do**
    $s_1 \leftarrow \mathcal{O}(w_i, z_1)[2]$;
    $s_2 \leftarrow \mathcal{O}^{-1}(z_2, s_1)[1]$;
    $s_3 \leftarrow \mathcal{O}(s_2, z_3)[1]$;
    $s_4 \leftarrow \mathcal{O}(w_{3-i}, z_4)[1]$;
    $s_5 \leftarrow \mathcal{O}^{-1}(s_4, z_5)[1]$;
    $s_6 \leftarrow \mathcal{O}(s_5, z_6)[2]$;
    $r_i \leftarrow \mathcal{O}^{-1}(s_3, s_6)[1]$;
end for
if $r_1 = r_2$ **then**
    return 1
else
    return 0
end if

One of the two bipod trees used by our attack is shown in Fig. 2. In that figure, w_1, w_2, z_1, z_2, z_3, z_4, z_5 and z_6 are independent random elements of $\mathbb{F}$,

[7] Of course the same equations also hold for backward queries, but they *in particular* hold for forward queries.

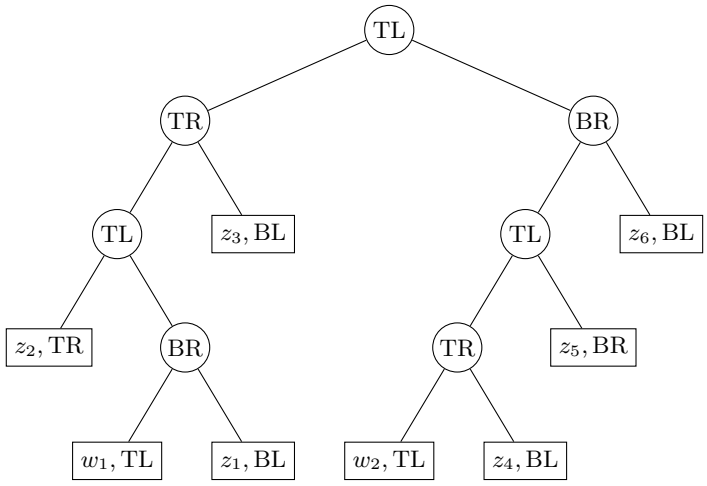

Fig. 2. The first bipod tree used by the Algorithm 1 attack. Non-leaf nodes are labeled by position only.

sampled at the beginning of the attack. (One could in fact choose these elements arbitrarily subject to the condition that $w_1 \neq w_2$, but sampling them at random seems conceptually simplest.) Each leaf is labeled by a "full" token including value and position, while non-leaf nodes only show a position value (necessary to correctly parse the tree).

As explained above, the inner value of the root of the Fig. 2 tree is a fixed affine combination of the inner values of the tokens at the leaves; explicitly, we have

$$\begin{aligned} &AaAI(z_2, \mathrm{TR}) + dBaAI(w_1, \mathrm{TL}) \\ &+ eBaAI(z_1, \mathrm{BL}) + bAI(z_3, \mathrm{BL}) \\ &+ aAdBI(w_2, \mathrm{TL}) + bAdBI(z_4, \mathrm{BL}) \\ &+ BdBI(z_5, \mathrm{BR}) + eBI(z_6, \mathrm{BL}) \end{aligned}$$

as can be checked with not too much effort. The details of this affine combination are unimportant, except that $I(w_1, \mathrm{TL})$ and $I(w_2, \mathrm{TL})$ have the same coefficient, given that $dBaA = aAdB$ by the commutativity of field multiplication. Hence, swapping the positions of w_1 and w_2 results in the same affine combination at the root. In other words, as long as $\mathcal{O} = 2\mathrm{rnd}^S_{k_0,k_1,k_2}$, the inner value of the root is left unchanged by a swap of w_1 and w_2 and, hence, the (non-inner) value of the root is left unchanged as well. Thus our attack consists in evaluating both versions of the tree (with w_1 and w_2 swapped and unswapped), and in comparing whether the two roots have the same value.

As the two roots will have the same value with probability 1 if $\mathcal{O} = 2\mathrm{rnd}^S_{k_0,k_1,k_2}$, all that remains to check is that the two roots are the same with negligible prob-

ability if $\mathcal{O}$ is a random permutation of $\mathbb{F}^2$. This can be checked in a number of ways. For example, the event that $w_1 = w_2$ (which can, indeed, entirely be avoided by slightly modifying the attack) happens with probability $1/|\mathbb{F}|$. Assuming this event doesn't occur, the 14 queries made by the attacker to $\mathcal{O}$ are distinct unless some query returns an answer $u \in \mathbb{F}^2$ such that either $u[1]$ or $u[2]$ is equal to a "previously seen" element of $\mathbb{F}$, i.e., is equal to one of $x^j[1]$, $x^j[2]$, $y^j[1]$, $y^j[2]$ for some earlier element (σ^j, x^j, y^j) of the query history. As long as this bad event has not happened for queries number 1, ..., $i-1$, however, then query i is fresh and the probability that this event occurs at query i, $i \leq 13$, is at most $2 \cdot 4 \cdot 12 \cdot |\mathbb{F}|/(|\mathbb{F}|^2 - 12) \leq 100/|\mathbb{F}|$ (assuming $|\mathbb{F}| \geq 100$), which is negligible. Finally, if the 14-th (i.e., last) query is fresh, the probability of it returning the same first half as the 7-th query is at most $|\mathbb{F}|/(|\mathbb{F}|^2 - 13)$, which is negligible as well. This concludes the analysis of the attack.