Improving the Upper Bound on the Maximum Average Linear Hull Probability for Rijndael

Liam Keliher[1], Henk Meijer[1], and Stafford Tavares[2]

[1] Department of Computing and Information Science,
Queen's University at Kingston, Ontario, Canada, K7L 3N6
{keliher,henk}@cs.queensu.ca
[2] Department of Electrical and Computer Engineering,
Queen's University at Kingston, Ontario, Canada, K7L 3N6
tavares@ee.queensu.ca

Abstract. In [15], Keliher et al. present a new method for upper bounding the maximum average linear hull probability (MALHP) for SPNs, a value which is required to make claims about provable security against linear cryptanalysis. Application of this method to Rijndael (AES) yields an upper bound of $UB = 2^{-75}$ when 7 or more rounds are approximated, corresponding to a lower bound on the data complexity of $\frac{32}{UB} = 2^{80}$ (for a 96.7% success rate). In the current paper, we improve this upper bound for Rijndael by taking into consideration the distribution of linear probability values for the (unique) Rijndael 8×8 s-box. Our new upper bound on the MALHP when 9 rounds are approximated is 2^{-92}, corresponding to a lower bound on the data complexity of 2^{97} (again for a 96.7% success rate). [This is after completing 43% of the computation; however, we believe that values have stabilized—see Section 7.]

Keywords: linear cryptanalysis, maximum average linear hull probability, provable security, Rijndael, AES

1 Introduction

The *substitution-permutation network* (SPN) [9,1,12] is a fundamental block cipher architecture based on Shannon's principles of *confusion* and *diffusion* [22]. These principles are implemented through substitution and linear transformation (LT), respectively. Recently, SPNs have been the focus of increased attention. This is due in part to the selection of the SPN Rijndael [6] as the U.S. Government Advanced Encryption Standard (AES).

Linear cryptanalysis (LC) [18] and differential cryptanalysis (DC) [4] are generally considered to be the two most powerful cryptanalytic attacks on block ciphers. In this paper we focus on the linear cryptanalysis of SPNs. As a first attempt to quantify the resistance of a block cipher to LC, the *expected linear characteristic probability* (ELCP) of the *best linear characteristic* often is evaluated. However, Nyberg [21] showed that the use of linear characteristics can underestimate the success of LC. To guarantee *provable security*, a block cipher

S. Vaudenay and A. Youssef (Eds.): SAC 2001, LNCS 2259, pp. 112–128, 2001.

designer needs to consider *linear hulls* instead of linear characteristics, and the *maximum average linear hull probability* (MALHP) instead of the ELCP of the best linear characteristic.

Since the MALHP is difficult, if not infeasible, to compute exactly, researchers have adopted the approach of upper bounding it [2,13,15]. In [15], Keliher et al. present a new general method for upper bounding the MALHP for SPNs. They apply their method to Rijndael, obtaining an upper bound on the MALHP of $UB = 2^{-75}$ when 7 or more rounds are approximated, corresponding to a lower bound on the data complexity of $\frac{32}{UB} = 2^{80}$ (for a 96.7% success rate—see Table 1).[1]

The current paper is based on the following observation: *the general method of Keliher et al. in [15] can potentially be improved by incorporating specific information about the distribution of linear probability (LP) values for the SPN s-boxes.* Due to the fact that Rijndael has only one (repeated) s-box, and because of the structure of this s-box, this observation applies readily to Rijndael, and allows us to improve the upper bound on the MALHP to $UB = 2^{-92}$ when 9 rounds are approximated, for a lower bound on the data complexity of 2^{97} (again for a 96.7% success rate). (This value is based on completion of 43% of the computation, although we believe that the values have stabilized—see Section 7.

Conventions
The Hamming weight of a binary vector $\mathbf{x}$ is written $wt(\mathbf{x})$. If $\mathbf{Z}$ is a random variable, $E[\mathbf{Z}]$ denotes the expected value of $\mathbf{Z}$. And we use $\#\mathcal{A}$ to indicate the number of elements in the set $\mathcal{A}$.

2 Substitution-Permutation Networks

A block cipher is a bijective mapping from N bits to N bits (N is the *block size*) parameterized by a bitstring called a *key*, denoted $\mathbf{k}$. Common block sizes are 64 and 128 bits (we consider Rijndael with a block size of 128 bits). The input to a block cipher is called a *plaintext*, and the output is called a *ciphertext*.

An SPN encrypts a plaintext through a series of R simpler encryption steps called *rounds*. (Rijndael with a key size of 128 bits consists of 10 rounds.) The input to round r ($1 \leq r \leq R$) is first bitwise XOR'd with an N-bit *subkey*, denoted $\mathbf{k}^r$, which is typically derived from the key, $\mathbf{k}$, via a separate *key-scheduling algorithm*. The *substitution stage* then partitions the resulting vector into M sub-blocks of size n ($N = Mn$), which become the inputs to a row of bijective $n \times n$ *substitution boxes* (*s-boxes*)—bijective mappings from $\{0,1\}^n$ to $\{0,1\}^n$. Finally, the *permutation stage* applies an invertible linear transformation (LT) to the output of the s-boxes (classically, a bitwise permutation). Often the permutation stage is omitted from the last round. A final subkey, $\mathbf{k}^{R+1}$, is XOR'd with

[1] In [15], the value 2^{80} was incorrectly given as 2^{78} due to an error in the table corresponding to Table 1. See Remark 2 for clarification.

the output of round R to form the ciphertext. Figure 1 depicts an example SPN with $N = 16$, $M = n = 4$, and $R = 3$.

We assume the most general situation for the key, namely, that $\mathbf{k}$ is an *independent key* [3], a concatenation of $(R + 1)$ subkeys chosen independently from the uniform distribution on $\{0,1\}^N$—symbolically, $\mathbf{k} = \langle \mathbf{k}^1, \mathbf{k}^2, \ldots, \mathbf{k}^{R+1} \rangle$. We use $\mathcal{K}$ to denote the set of all independent keys.

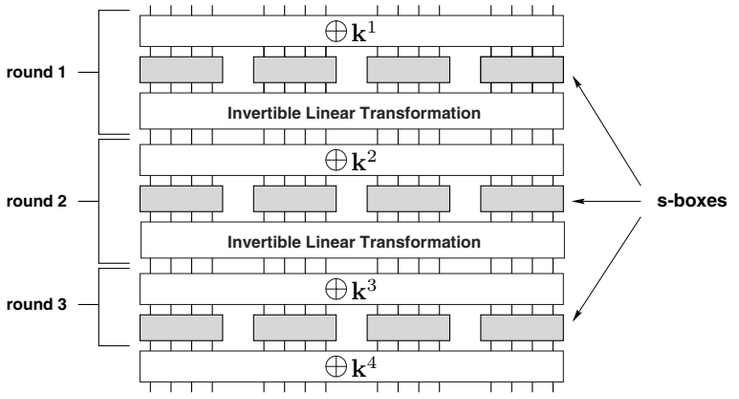

Fig. 1. SPN with $N = 16$, $M = n = 4$, $R = 3$

3 Linear Probability

In this section, and in Section 4, we make use of some of the treatment and notation from Vaudenay [23].

Definition 1. *Suppose* $B : \{0,1\}^d \to \{0,1\}^d$ *is a bijective mapping. Let* $\mathbf{a}, \mathbf{b} \in \{0,1\}^d$ *be fixed, and let* $\mathbf{X} \in \{0,1\}^d$ *be a uniformly distributed random variable. The linear probability* $LP(\mathbf{a}, \mathbf{b})$ *is defined as*

$$LP(\mathbf{a}, \mathbf{b}) \stackrel{\text{def}}{=} \left(2 \cdot \text{Prob}_{\mathbf{X}} \left\{\mathbf{a} \bullet \mathbf{X} = \mathbf{b} \bullet B(\mathbf{X})\right\} - 1\right)^2. \tag{1}$$

If B *is parameterized by a key,* $\mathbf{k}$*, we write* $LP(\mathbf{a}, \mathbf{b}; \mathbf{k})$*, and the* expected *LP (ELP) is defined as*

$$ELP(\mathbf{a}, \mathbf{b}) \stackrel{\text{def}}{=} E\left[LP(\mathbf{a}, \mathbf{b}; \mathbf{K})\right],$$

where $\mathbf{K}$ *is a random variable uniformly distributed over the space of keys.*

Note that LP values lie in the interval $[0, 1]$. A nonzero LP value indicates a correlation between the input and output of B, with a higher value indicating a stronger correlation (in fact, $LP(\mathbf{a}, \mathbf{b})$ is the square of entry $[\mathbf{a}, \mathbf{b}]$ in the *correlation matrix* for B [5]).

The values $\mathbf{a}/\mathbf{b}$ in Definition 1 are referred to as input/output *masks*. For our purposes, the bijective mapping B may be an s-box, a single encryption round, or a sequence of consecutive encryption rounds.

The following lemma derives immediately from Parseval's Theorem [20].

Lemma 1. *Let $B : \{0,1\}^d \to \{0,1\}^d$ be a bijective mapping parameterized by a key, $\mathbf{k}$, and let $\mathbf{a}, \mathbf{b} \in \{0,1\}^d$. Then*

$$\sum_{\mathbf{x} \in \{0,1\}^d} LP(\mathbf{a}, \mathbf{x}; \mathbf{k}) = \sum_{\mathbf{x} \in \{0,1\}^d} LP(\mathbf{x}, \mathbf{b}; \mathbf{k}) = 1$$
$$\sum_{\mathbf{x} \in \{0,1\}^d} ELP(\mathbf{a}, \mathbf{x}) = \sum_{\mathbf{x} \in \{0,1\}^d} ELP(\mathbf{x}, \mathbf{b}) = 1.$$

3.1 LP Values for the Rijndael S-box

Consider the (unique) Rijndael 8×8 s-box (see the Rijndael reference code [7]) as the bijective mapping B in Definition 1. A short computation yields the following interesting fact.

Lemma 2. *Let the bijective mapping under consideration be the 8×8 Rijndael s-box. If $\mathbf{a} \in \{0,1\}^8 \setminus \mathbf{0}$ is fixed, and $\mathbf{b}$ varies over $\{0,1\}^8$, then the distribution of values $LP(\mathbf{a}, \mathbf{b})$ is constant, and is given in the following table (ρ_i is the LP value, and ϕ_i is the number of times it occurs, for $1 \leq i \leq 9$). The same distribution is obtained if $\mathbf{b} \in \{0,1\}^8 \setminus \mathbf{0}$ is fixed, and $\mathbf{a}$ varies over $\{0,1\}^8$.*

i	1	2	3	4	5	6	7	8	9
ρ_i	$\left(\frac{8}{64}\right)^2$	$\left(\frac{7}{64}\right)^2$	$\left(\frac{6}{64}\right)^2$	$\left(\frac{5}{64}\right)^2$	$\left(\frac{4}{64}\right)^2$	$\left(\frac{3}{64}\right)^2$	$\left(\frac{2}{64}\right)^2$	$\left(\frac{1}{64}\right)^2$	0
ϕ_i	5	16	36	24	34	40	36	48	17

4 Linear Cryptanalysis of Markov Ciphers

It will be useful to consider linear cryptanalysis (LC) in the general context of Markov ciphers [17].

4.1 Markov Ciphers

Let $\mathcal{E} : \{0,1\}^N \to \{0,1\}^N$ be an R-round cipher, for which round r is given by the function $\mathbf{y} = \epsilon_r(\mathbf{x}; \mathbf{k}^r)$ ($\mathbf{x} \in \{0,1\}^N$ is the round input, and $\mathbf{k}^r \in \{0,1\}^N$ is the round-r subkey). Then $\mathcal{E}$ is a Markov cipher with respect to the XOR group operation ($\oplus$) on $\{0,1\}^N$ if, for $1 \leq r \leq R$, and any $\mathbf{x}$, $\Delta\mathbf{x}$, $\Delta\mathbf{y} \in \{0,1\}^N$,

$$\begin{aligned}\text{Prob}_{\mathbf{K}} \{\epsilon_r(\mathbf{x}; \mathbf{K}) \oplus \epsilon_r(\mathbf{x} \oplus \Delta\mathbf{x}; \mathbf{K}) = \Delta\mathbf{y}\} = \\ \text{Prob}_{\mathbf{K},\mathbf{X}} \{\epsilon_r(\mathbf{X}; \mathbf{K}) \oplus \epsilon_r(\mathbf{X} \oplus \Delta\mathbf{x}; \mathbf{K}) = \Delta\mathbf{y}\} \qquad (2)\end{aligned}$$

(where $\mathbf{X}$ and $\mathbf{K}$ are uniformly distributed and independent). That is, the probability over the key that a fixed input difference produces a fixed output difference is independent of the round input.

It is easy to show that the SPN model we are using is a Markov cipher, as are certain Feistel ciphers [10], such as DES [8].

Remark 1. The material in the remainder of Section 4 applies to any Markov cipher. Although we are dealing with LC, which ostensibly does not involve the $\oplus$ operation, the relevance of the Markov property given in (2) is via an interesting connection between linear probability and *differential probability* (see, for example, equations (3) and (4) in [23]).

4.2 Linear Cryptanalysis

Linear cryptanalysis (LC) is a known-plaintext attack (ciphertext-only in some cases) introduced by Matsui [18]. The more powerful version is known as Algorithm 2 (Algorithm 1 extracts only a single subkey bit). Algorithm 2 can be used to extract (pieces of) the round-1 subkey, $\mathbf{k}^1$. Once $\mathbf{k}^1$ is known, round 1 can be stripped off, and LC can be reapplied to obtain $\mathbf{k}^2$, and so on.

We do not give the details of LC here, as it is treated in many papers [18,3,14,15]. It suffices to say that the attacker wants to find input/output masks $\mathbf{a}, \mathbf{b} \in \{0,1\}^N$ for the bijective mapping consisting of rounds $2 \dots R$, for which $LP(\mathbf{a}, \mathbf{b}; \mathbf{k})$ is maximal. Based on this value, the attacker can determine the number of known $\langle \text{plaintext}, \text{ciphertext} \rangle$ pairs, $\mathcal{N}_L$ (called the *data complexity*), required for a successful attack. Given an assumption about the behavior of round-1 output [18], Matsui shows that if

$$\mathcal{N}_L = \frac{c}{LP(\mathbf{a}, \mathbf{b}; \mathbf{k})},$$

then Algorithm 2 has the success rates in Table 1, for various values of the constant, c. Note that this is the same as Table 3 in [18], except that the constant values differ by a factor of 4, since Matsui uses *bias* values, not LP values.

Remark 2. The table in [15] corresponding to Table 1 has an error, in that the constants have *not* been multiplied by 4 to reflect the use of LP values.

Notational Issues. Above, we have discussed input and output masks and the associated LP values for rounds $2 \dots R$ of an R-round cipher. It is useful to consider these and other related concepts as applying to any $T \geq 2$ consecutive

Table 1. Success rates for LC Algorithm 2

c	8	16	32	64
Success rate	48.6%	78.5%	96.7%	99.9%

"core" rounds (we say that these are the rounds being *approximated*). For Algorithm 2 as outlined above, $T = R - 1$, and the "first round," or "round 1," is actually round 2 of the cipher.

We use superscripts for individual rounds, so $LP^t(\mathbf{a}, \mathbf{b}; \mathbf{k}^t)$ and $ELP^t(\mathbf{a}, \mathbf{b})$ are LP and ELP values, respectively, for round t. On the other hand, we use t as a *subscript* to refer to values which apply to the first t rounds as a unit, so, for example, $ELP_t(\mathbf{a}, \mathbf{b})$ is an ELP value over rounds $1 \ldots t$.

4.3 Linear Characteristics

For fixed $\mathbf{a}, \mathbf{b} \in \{0,1\}^N$, direct computation of $LP_T(\mathbf{a}, \mathbf{b}; \mathbf{k})$ for T core rounds is generally infeasible, first since it requires encrypting all N-bit vectors through rounds $1 \ldots T$, and second because of the dependence on an unknown key. The latter difficulty is usually handled by working instead with the expected value $ELP_T(\mathbf{a}, \mathbf{b})$. The data complexity of Algorithm 2 for masks $\mathbf{a}$ and $\mathbf{b}$ is now taken to be

$$\mathcal{N}_L = \frac{c}{ELP_T(\mathbf{a}, \mathbf{b})} \,. \tag{3}$$

The implicit assumption is that $LP_T(\mathbf{a},\mathbf{b};\mathbf{k})$ is approximately equal to $ELP_T(\mathbf{a},\mathbf{b})$ for almost all values of $\mathbf{k}$ (this derives from the *Hypothesis of Stochastic Equivalence* in [17]).

The problem of computational complexity is usually treated by approximating $ELP_T(\mathbf{a}, \mathbf{b})$ through the use of *linear characteristics* (or simply *characteristics*). A T-round characteristic is a $(T+1)$-tuple $\Omega = \langle \mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^T, \mathbf{a}^{T+1} \rangle$. We view $\mathbf{a}^t$ and $\mathbf{a}^{t+1}$ as input and output masks, respectively, for round t.

Definition 2. *Let $\Omega = \langle \mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^T, \mathbf{a}^{T+1} \rangle$ be a T-round characteristic. The* linear characteristic probability *(LCP) and* expected *LCP (ELCP) of Ω are defined as*

$$LCP(\Omega; \mathbf{k}) = \prod_{t=1}^{T} LP^t(\mathbf{a}^t, \mathbf{a}^{t+1}; \mathbf{k}^t)$$

$$ELCP(\Omega) = \prod_{t=1}^{T} ELP^t(\mathbf{a}^t, \mathbf{a}^{t+1}).$$

4.4 Choosing the Best Characteristic

In carrying out LC, the attacker typically runs an algorithm to find the T-round characteristic, Ω, for which $ELCP(\Omega)$ is maximal; such a characteristic (not necessarily unique) is called the *best characteristic* [19]. If $\Omega = \langle \mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^T, \mathbf{a}^{T+1} \rangle$, and if the input and output masks used in Algorithm 2 are taken to be $\mathbf{a} = \mathbf{a}^1$ and $\mathbf{b} = \mathbf{a}^{T+1}$, respectively, then $ELP_T(\mathbf{a}, \mathbf{b})$ (used to determine $\mathcal{N}_L$ in (3)) is approximated by

$$ELP_T(\mathbf{a}, \mathbf{b}) \approx ELCP(\Omega) \,. \tag{4}$$

The approximation in (4) has been widely used to evaluate the security of block ciphers against LC [12,14]. Knudsen calls a block cipher *practically secure* if the data complexity determined by this method is prohibitive [16]. However, by introducing the concept of *linear hulls*, Nyberg demonstrated that the above approach can underestimate the success of LC [21].

4.5 Linear Hulls

Definition 3 (Nyberg). *Given N-bit masks $\mathbf{a}, \mathbf{b}$, the corresponding linear hull, denoted* $\mathrm{ALH}(\mathbf{a}, \mathbf{b})$,[2] *is the set of all T-round characteristics (for the T rounds under consideration) having $\mathbf{a}$ as the input mask for round 1 and $\mathbf{b}$ as the output mask for round T, i.e., all characteristics of the form*

$$\Omega = \left\langle \mathbf{a}, \mathbf{a}^2, \mathbf{a}^3, \ldots, \mathbf{a}^T, \mathbf{b} \right\rangle .$$

Theorem 1 (Nyberg). *Let $\mathbf{a}, \mathbf{b} \in \{0,1\}^N$. Then*

$$ELP_T(\mathbf{a}, \mathbf{b}) = \sum_{\Omega \in \mathrm{ALH}(\mathbf{a}, \mathbf{b})} ELCP(\Omega) .$$

It follows immediately from Theorem 1 that (4) does not hold in general, since $ELP_T(\mathbf{a}, \mathbf{b})$ is seen to be equal to a sum of terms $ELCP(\Omega)$ over a (large) set of characteristics, and therefore, in general, the ELCP of any characteristic will be strictly *less than* the corresponding ELP value. This is referred to as the *linear hull effect.* An important consequence is that an attacker may overestimate the number of $\langle \text{plaintext}, \text{ciphertext} \rangle$ pairs required for a given success rate.

Remark 3. It can be shown that the linear hull effect is significant for Rijndael, since, for example, the ELCP of any characteristic over $T = 8$ rounds is upper bounded by 2^{-300} [6],[3] but the largest ELP value has 2^{-128} as a trivial lower bound.[4]

The next lemma follows easily from Theorem 1 and Definition 2 (recall the conventions for superscripts and subscripts).

Lemma 3. *Let $T \geq 2$, and let $\mathbf{a}, \mathbf{b} \in \{0,1\}^N$. Then*

$$ELP_T(\mathbf{a}, \mathbf{b}) = \sum_{\mathbf{x} \in \{0,1\}^N} ELP_{T-1}(\mathbf{a}, \mathbf{x}) \cdot ELP^T(\mathbf{x}, \mathbf{b}) .$$

[2] Nyberg [21] originally used the term *approximate linear hull,* hence the abbreviation ALH, which we retain for consistency with [15].

[3] Any 8-round characteristic, Ω, has a minimum of 50 active s-boxes, and the maximum LP value for the Rijndael s-box is 2^{-6}, so $ELCP(\Omega) \leq \left(2^{-6}\right)^{50} = 2^{-300}$.

[4] This follows by observing that Lemma 1 is contradicted if the maximum ELP value is less than 2^{-d}.

4.6 Maximum Average Linear Hull Probability

An SPN is considered to be *provably secure* against LC if the maximum ELP,

$$\max_{\mathbf{a},\mathbf{b}\in\{0,1\}^N\setminus\mathbf{0}} ELP_T(\mathbf{a},\mathbf{b}), \tag{5}$$

is sufficiently small that the resulting data complexity is prohibitive for any conceivable attacker.[5] The value in (5) is also called the *maximum average linear hull probability* (MALHP). We retain this terminology for consistency with [15].

Since evaluation of the MALHP appears to be infeasible in general, researchers have adopted the approach of upper bounding this value [2,13,15]. If such an upper bound is sufficiently small, provable security can be claimed.

5 SPN-Specific Considerations

In the current section, we adapt certain results from Section 4 to the SPN model. Note that where matrix multiplication is involved, we view all vectors as column vectors. Also, if $\mathcal{M}$ is a matrix, $\mathcal{M}'$ denotes the transpose of $\mathcal{M}$.

Lemma 4. *Consider T core SPN rounds. Let $1 \le t \le T$, and $\mathbf{a},\mathbf{b},\mathbf{k}^t \in \{0,1\}^N$. Then $LP^t(\mathbf{a},\mathbf{b};\mathbf{k}^t)$ is independent of $\mathbf{k}^t$, and therefore*

$$LP^t(\mathbf{a},\mathbf{b};\mathbf{k}^t) = ELP^t(\mathbf{a},\mathbf{b}).$$

Proof. Follows by observing the interchangeable roles of the round input, $\mathbf{x}$, and $\mathbf{k}^t$, and from a simple change of variables $\hat{\mathbf{x}} = \mathbf{x} \oplus \mathbf{k}^t$ when evaluating (1).

Corollary 1. *Let Ω be a T-round characteristic for an SPN. Then $LCP(\Omega) = ELCP(\Omega)$.*

Definition 4. *Let $\mathbf{L}$ denote the N-bit LT of the SPN represented as a binary $N \times N$ matrix, i.e., if $\mathbf{x},\mathbf{y} \in \{0,1\}^N$ are the input and output, respectively, for the LT, then $\mathbf{y} = \mathbf{L}\mathbf{x}$.*

Lemma 5 ([5]). *If $\mathbf{b} \in \{0,1\}^N$ and $\mathbf{a} = \mathbf{L}'\mathbf{b}$, then $\mathbf{a} \bullet \mathbf{x} = \mathbf{b} \bullet \mathbf{y}$ for all N-bit inputs to the LT, $\mathbf{x}$, and corresponding outputs, $\mathbf{y}$ (i.e., if $\mathbf{b}$ is an output mask for the LT, then $\mathbf{a} = \mathbf{L}'\mathbf{b}$ is the (unique) corresponding input mask).*

It follows from Lemma 5 that if $\mathbf{a}^t$ and $\mathbf{a}^{t+1}$ are input and output masks for round t, respectively, then the resulting input and output masks for the *substitution stage* of round t are $\mathbf{a}^t$ and $\mathbf{b}^t = \mathbf{L}'\mathbf{a}^{t+1}$. Further, $\mathbf{a}^t$ and $\mathbf{b}^t$ determine input and output masks for each s-box in round t. Let the masks for S_i^t be

[5] For Algorithm 2 as described above, this must hold for $T = R - 1$. Since variations of LC can be used to attack the first and last SPN rounds simultaneously, it may also be important that the data complexity remain prohibitive for $T = R - 2$.

denoted $\mathbf{a}_i^t$ and $\mathbf{b}_i^t$, for $1 \leq i \leq M$ (we number s-boxes from left to right). Then from Matsui's Piling-up Lemma [18] and Lemma 4,

$$ELP^t(\mathbf{a}^t, \mathbf{a}^{t+1}) = \prod_{i=1}^{M} LP^{S_i^t}(\mathbf{a}_i^t, \mathbf{b}_i^t). \tag{6}$$

From the above, any characteristic $\Omega \in \mathrm{ALH}(\mathbf{a}, \mathbf{b})$ determines an input and an output mask for each s-box in rounds $1 \ldots T$. If this yields at least one s-box for which the input mask is zero and the output mask is nonzero, or vice versa, the linear probability associated with that s-box will trivially be 0, and therefore $ELCP(\Omega) = 0$ by (6) and Definition 2. We exclude such characteristics from consideration via the following definition.

Definition 5. *For $\mathbf{a}, \mathbf{b} \in \{0,1\}^N$, let $\mathrm{ALH}(\mathbf{a}, \mathbf{b})^*$ consist of the elements $\Omega \in \mathrm{ALH}(\mathbf{a}, \mathbf{b})$ such that for each s-box in rounds $1 \ldots T$, the input and output masks determined by Ω for that s-box are either both zero or both nonzero.*

Remark 4. In [23], the characteristics in $\mathrm{ALH}(\mathbf{a}, \mathbf{b})^*$ are called *consistent.*

Definition 6 ([3]). *Any T-round characteristic, Ω, determines an input and an output mask for each s-box in rounds $1 \ldots T$. Those s-boxes having nonzero input and output masks are called* active.

Definition 7. *Let $\mathbf{v}$ be an input or an output mask for the substitution stage of round t. Then the active s-boxes in round t can be determined from $\mathbf{v}$ (without knowing the corresponding output/input mask). We define $\gamma_{\mathbf{v}}$ to be the M-bit vector which encodes the pattern of active s-boxes: $\gamma_{\mathbf{v}} = \gamma_1 \gamma_2 \ldots \gamma_M$, where $\gamma_i = 1$ if the i^{th} s-box is active, and $\gamma_i = 0$ otherwise, for $1 \leq i \leq M$.*

Definition 8 ([15]). *Let $\gamma, \hat{\gamma} \in \{0,1\}^M$. Then*

$$W[\gamma, \hat{\gamma}] \stackrel{\text{def}}{=} \#\left\{\mathbf{y} \in \{0,1\}^N : \gamma_{\mathbf{x}} = \gamma, \gamma_{\mathbf{y}} = \hat{\gamma}, \quad \text{where } \mathbf{x} = \mathbf{L}'\mathbf{y}\right\} .$$

Remark 5. Informally, the value $W[\gamma, \hat{\gamma}]$ represents the number of ways the LT can "connect" a pattern of active s-boxes in one round (γ) to a pattern of active s-boxes in the next round ($\hat{\gamma}$).

We now proceed to our improved method for upper bounding the MALHP for Rijndael.

6 Improved Upper Bound on MALHP for Rijndael

6.1 Technical Lemmas

Lemma 6 ([15]). *Let $m \geq 2$, and suppose $\{c_i\}_{i=1}^m$, $\{d_i\}_{i=1}^m$ are sequences of nonnegative values. Let $\{\dot{c}_i\}_{i=1}^m$, $\left\{\dot{d}_i\right\}_{i=1}^m$ be the sequences obtained by sorting $\{c_i\}$ and $\{d_i\}$, respectively, in nonincreasing order. Then $\sum_{i=1}^m c_i d_i \leq \sum_{i=1}^m \dot{c}_i \dot{d}_i$.*

Lemma 7 ([15]). *Suppose* $\{\dot{c}_i\}_{i=1}^m$, $\{\ddot{c}_i\}_{i=1}^m$, *and* $\left\{\dot{d}_i\right\}_{i=1}^m$ *are sequences of nonnegative values, with* $\left\{\dot{d}_i\right\}$ *sorted in nonincreasing order. Suppose there exists* $\tilde{m}$, $1 \le \tilde{m} \le m$, *such that*

(a) $\ddot{c}_i \ge \dot{c}_i$, *for* $1 \le i \le \tilde{m}$
(b) $\ddot{c}_i \le \dot{c}_i$, *for* $(\tilde{m}+1) \le i \le m$
(c) $\sum_{i=1}^m \dot{c}_i \le \sum_{i=1}^m \ddot{c}_i$

Then $\sum_{i=1}^m \dot{c}_i\dot{d}_i \le \sum_{i=1}^m \ddot{c}_i\dot{d}_i$.

6.2 Distribution of LP Values for Multiple Active S-boxes

Definition 9. *Let* $\mathbf{a} \in \{0,1\}^{128} \setminus \mathbf{0}$ *be a fixed input mask for the substitution stage of Rijndael, and let* $\mathbf{b}$ *be an output mask which varies over* $\{0,1\}^{128}$, *with the restriction that* $\gamma_{\mathbf{a}} = \gamma_{\mathbf{b}}$. *If* A *is the number of s-boxes made active* ($A = wt(\gamma_{\mathbf{a}})$), *define* $\mathcal{D}_A$ *to be the set of distinct LP values produced as* $\mathbf{b}$ *varies, and let* $D_A = \#\mathcal{D}_A$. *Define* $\langle \rho_1^A, \rho_2^A, \ldots, \rho_{D_A}^A \rangle$ *to be the sequence obtained by sorting* $\mathcal{D}_A$ *in decreasing order, and let* ϕ_j^A *be the number of occurrences of the value* ρ_j^A, *for* $1 \le j \le D_A$.

Note that if $A = 1$, then $D_A = 9$, and ρ_j^1 and ϕ_j^1 are as given in Lemma 2.

Lemma 8. *For* $A \ge 2$,

$$\mathcal{D}_A = \left\{\rho_s^1 \cdot \rho_t^{A-1} : 1 \le s \le D_1,\ 1 \le t \le D_{A-1}\right\},$$

and for each j, $1 \le j \le D_A = \#\mathcal{D}_A$,

$$\phi_j^A \ = \ \sum \left\{\phi_s^1 \cdot \phi_t^{A-1} \ : \ \rho_s^1 \cdot \rho_t^{A-1} = \rho_j^A,\ 1 \le s \le D_1,\ 1 \le t \le D_{A-1}\right\}.$$

Proof. Follows easily from Lemma 4 and (6).

Definition 10. *For* $A \ge 1$ *and* $1 \le J \le D_A$, *we define the partial sums*

$$\Phi_J^A = \sum_{j=1}^{J} \phi_j^A$$

$$\Lambda_J^A = \sum_{j=1}^{J} \rho_j^A \cdot \phi_j^A.$$

Also, we define $\mathcal{S}_A$ *to be the sequence*

$$\underbrace{\rho_1^A, \ldots, \rho_1^A}_{\phi_1^A \ terms},\ \underbrace{\rho_2^A, \ldots, \rho_2^A}_{\phi_2^A \ terms},\ \ldots, \underbrace{\rho_{D_A}^A, \ldots, \rho_{D_A}^A}_{\phi_{D_A}^A \ terms}.$$

Remark 6. For $1 \le A \le M$, $\Lambda_{D_A}^A = 1$ by Lemma 1.

6.3 Derivation of Improved Upper Bound

Convention: In this subsection, whenever we deal with values of the form $ELP_t(\mathbf{a}, \mathbf{b})$ or $ELP^t(\mathbf{a}, \mathbf{b})$ $(1 \le t \le T)$, we omit the LT from round t. This is simply a technical matter that simplifies the proofs which follow.

Let $T \ge 2$. As in [15], our approach is to compute an upper bound for each nonzero pattern of active s-boxes in round 1 and round T—that is, we compute $UB_T[\gamma, \hat{\gamma}]$, for $\gamma, \hat{\gamma} \in \{0,1\}^M \setminus \mathbf{0}$, such that the following holds:

UB Property for T. For all $\mathbf{a}, \mathbf{b} \in \{0,1\}^N \setminus \mathbf{0}$, $ELP_T(\mathbf{a}, \mathbf{b}) \le UB_T[\gamma_{\mathbf{a}}, \gamma_{\mathbf{b}}]$.

If the *UB Property for T* holds, then the MALHP is upper bounded by

$$\max_{\gamma, \hat{\gamma} \in \{0,1\}^M \setminus \mathbf{0}} UB_T[\gamma, \hat{\gamma}].$$

The case $T = 2$ is handled in Theorem 2, and the case $T \ge 3$ in Theorem 3.

Theorem 2. *Let the values $UB_2[\gamma, \hat{\gamma}]$ be computed using the algorithm in Figure 2. Then the* UB Property for 2 *holds.*

Proof. In this proof, "Line X" refers to the X^{th} line in Figure 2. Let $\gamma, \hat{\gamma} \in \{0,1\}^M \setminus \mathbf{0}$ be fixed, and let $\mathbf{a}, \mathbf{b} \in \{0,1\}^N \setminus \mathbf{0}$ such that $\gamma_{\mathbf{a}} = \gamma$ and $\gamma_{\mathbf{b}} = \hat{\gamma}$. We want to show that $ELP_2(\mathbf{a}, \mathbf{b}) \le UB_2[\gamma, \hat{\gamma}]$. There are $W = W[\gamma, \hat{\gamma}]$ ways that the LT can "connect" the f active s-boxes in round 1 to the ℓ active s-boxes in round 2. Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_W$ be the corresponding output masks for the substitution stage of round 1 (and therefore the input masks for the round-1 LT), and let $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_W$ be the respective output masks for the round-1 LT (and therefore the input masks for the substitution stage of round 2). So $\gamma_{\mathbf{x}_i} = \gamma$ and $\gamma_{\mathbf{y}_i} = \hat{\gamma}$, for $1 \le i \le W$. Let $c_i = ELP^1(\mathbf{a}, \mathbf{x}_i)$ and $d_i = ELP^2(\mathbf{y}_i, \mathbf{b})$, for $1 \le i \le W$. It follows from Lemma 3 that $ELP_2(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{W} c_i d_i$.

Without loss of generality, $f \le \ell$, so $A_{\min} = f$ and $A_{\max} = \ell$. Let $\{\dot{c}_i\}$ $(\{\dot{d}_i\})$ be the sequence obtained by sorting $\{c_i\}$ $(\{d_i\})$ in nonincreasing order. Then $\sum_{i=1}^{W} c_i d_i \le \sum_{i=1}^{W} \dot{c}_i \dot{d}_i$ by Lemma 6. Let $\{\ddot{c}_i\}$ $(\{\ddot{d}_i\})$ consist of the first W terms of $\mathcal{S}_f$ $(\mathcal{S}_\ell)$. Since the terms $\dot{c}_i$ $(\dot{d}_i)$ are elements of $\mathcal{S}_f$ $(\mathcal{S}_\ell)$, it follows that $\dot{c}_i \le \ddot{c}_i$ $(\dot{d}_i \le \ddot{d}_i)$, for $1 \le i \le W$, so

$$ELP_2(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{W} c_i d_i \le \sum_{i=1}^{W} \dot{c}_i \dot{d}_i \le \sum_{i=1}^{W} \ddot{c}_i \ddot{d}_i \,.$$

It is not hard to see that the value $UB_2[\gamma, \hat{\gamma}]$ computed in Figure 2 is exactly $\sum_{i=1}^{W} \ddot{c}_i \ddot{d}_i$. For computational efficiency, we do not sum "element-by-element" (i.e., for each i), but instead take advantage of the fact that $\{\ddot{c}_i\}$ has the form

$$\underbrace{\rho_1^f, \ldots, \rho_1^f}_{\phi_1^f \text{ terms}}, \underbrace{\rho_2^f, \ldots, \rho_2^f}_{\phi_2^f \text{ terms}}, \underbrace{\rho_3^f, \ldots, \rho_3^f}_{\phi_3^f \text{ terms}}, \ \ldots,$$

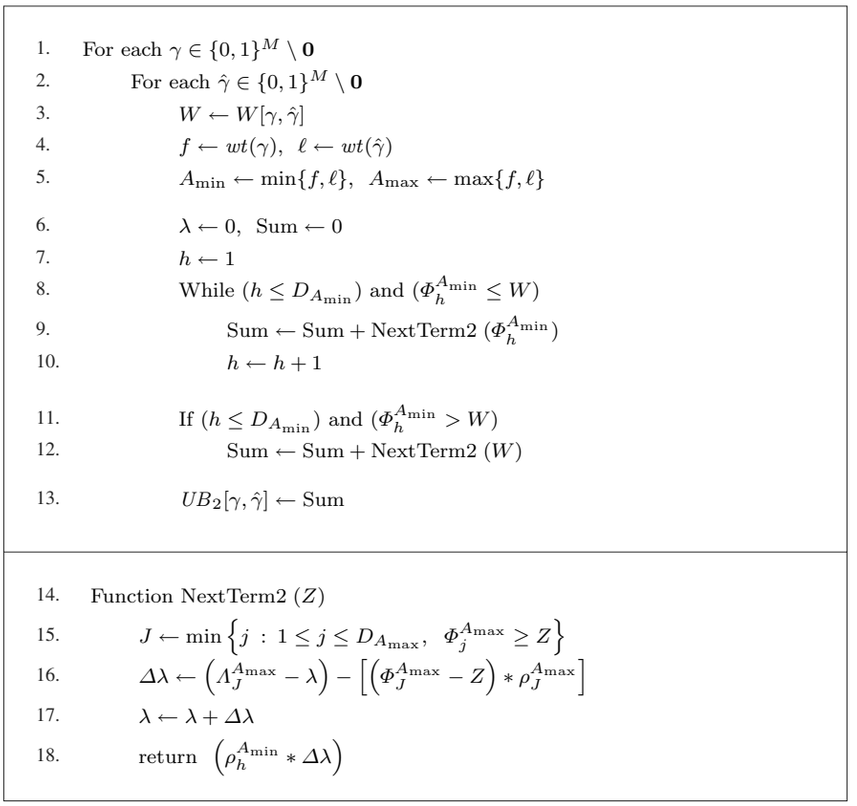

Fig. 2. Algorithm to compute $UB_2[\,]$

and similarly for $\{\ddot{d}_i\}$ (replace f with ℓ). Viewing these sequences as "groups" of consecutive identical elements, the algorithm in Figure 2 proceeds "group-by-group." The variable h is the index of the current group in $\{\ddot{c}_i\}$. The function NextTerm2() identifies the corresponding elements in $\{\ddot{d}_i\}$, and computes the equivalent of the element-by-element product, which is added to the growing sum in Line 9. The situation in which $\{\ddot{c}_i\}_{i=1}^{W}$ is a truncated version of $\mathcal{S}_f$ is handled by the conditional statement in Lines 11–12.

Theorem 3. *Let $T \ge 3$. Assume that the values $UB_{T-1}[\gamma, \hat{\gamma}]$ have been computed for all $\gamma, \hat{\gamma} \in \{0,1\}^M \setminus \mathbf{0}$ such that the* UB Property for $(T-1)$ *holds. Let the values $UB_T[\gamma, \hat{\gamma}]$ be computed using the algorithm in Figure 3. Then the* UB Property for T *holds.*

Proof. Throughout this proof, "Line X" refers to the X^{th} line in Figure 3. Let $\mathbf{a}, \mathbf{b} \in \{0,1\}^N \setminus \mathbf{0}$. It suffices to show that if $\gamma = \gamma_{\mathbf{a}}$ in Line 1 and $\hat{\gamma} = \gamma_{\mathbf{b}}$ in

```
1.  For each γ ∈ {0,1}^M \ 0
2.      For each γ̂ ∈ {0,1}^M \ 0
3.          ℓ ← wt(γ̂)
4.          Γ ← {ξ ∈ {0,1}^M \ 0 : W[ξ, γ̂] ≠ 0}
5.          Order the H = #Γ elements of Γ as γ_1, γ_2, ..., γ_H such that
6.              UB_{T-1}[γ, γ_1] ≥ UB_{T-1}[γ, γ_2] ≥ ··· ≥ UB_{T-1}[γ, γ_H]
7.          U_h ← UB_{T-1}[γ, γ_h],  for 1 ≤ h ≤ H
8.          W_h ← W[γ_h, γ̂],  for 1 ≤ h ≤ H
9.          Ψ ← 0,   λ ← 0,   W_total ← 0,   Sum ← 0
10.         h ← 1
11.         While (h ≤ H) and (U_h > 0) and (Ψ + (U_h * W_h) ≤ 1) and (λ < 1)
12.             W_total ← W_total + W_h
13.             Sum ← Sum + NextTermT (W_total)
14.             h ← h + 1
15.         If (h ≤ H) and (U_h > 0) and (Ψ + (U_h * W_h) > 1) and (λ < 1)
16.             W_total ← W_total + (1 − Ψ)/U_h
17.             Sum ← Sum + NextTermT (W_total)
18.         UB_T[γ, γ̂] ← Sum

19. Function NextTermT (Z)
20.     J ← min { j : 1 ≤ j ≤ D_ℓ,  Φ_j^ℓ ≥ Z }
21.     Δλ ← (Λ_J^ℓ − λ) − [(Φ_J^ℓ − Z) * ρ_J^ℓ]
22.     Ψ ← Ψ + (U_h * W_h)
23.     λ ← λ + Δλ
24.     return ( ρ_h^{A_min} * Δλ )
```

Fig. 3. Algorithm to compute $UB_T[\,]$ for $T \geq 3$

Line 2, then the value $UB_T[\gamma, \hat{\gamma}]$ computed in Figure 3 satisfies $ELP_T(\mathbf{a}, \mathbf{b}) \leq UB_T[\gamma, \hat{\gamma}]$. Enumerate the elements of $\{0,1\}^N \setminus \mathbf{0}$ as $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{2^N-1}$. We view these as input masks for round T, and hence as *output* masks for the LT of round $(T-1)$. For each $\mathbf{y}_i$, let $\mathbf{x}_i$ be the corresponding input mask for the LT. It follows from Lemma 3 that $ELP_T(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{2^N-1} ELP_{T-1}(\mathbf{a}, \mathbf{x}_i) \cdot ELP^T(\mathbf{y}_i, \mathbf{b})$. If $\gamma_{\mathbf{y}_i} \neq \gamma_{\mathbf{b}}$ $(= \hat{\gamma})$, then $ELP^T(\mathbf{y}_i, \mathbf{b}) = 0$ (this follows from (6)), so we remove these $\mathbf{y}_i$ from consideration, leaving $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \ldots, \bar{\mathbf{y}}_L$ (for some L), and corresponding input masks, $\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \ldots, \bar{\mathbf{x}}_L$, respectively.

Let $c_i = ELP_{T-1}(\mathbf{a}, \bar{\mathbf{x}}_i)$ and $d_i = ELP^T(\bar{\mathbf{y}}_i, \mathbf{b})$, for $1 \leq i \leq L$. Then $ELP_T(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^{L} c_i d_i$. Note that $\sum c_i \leq 1$, $\sum d_i \leq 1$ by Lemma 1. Let

$\{\dot{c}_i\}$ ($\{\dot{d}_i\}$) be the sequence obtained by sorting $\{c_i\}$ ($\{d_i\}$) in nonincreasing order. Then $\sum_{i=1}^{L} c_i d_i \le \sum_{i=1}^{L} \dot{c}_i \dot{d}_i$ by Lemma 6. If $\{\ddot{d}_i\}$ consists of the first L terms of $\mathcal{S}_\ell$ ($\ell = wt(\hat{\gamma})$ as in Line 3), then $\dot{d}_i \le \ddot{d}_i$, for $1 \le i \le L$ (since the $\dot{d}_i$ are elements of $\mathcal{S}_\ell$), so $\sum_{i=1}^{L} \dot{c}_i \dot{d}_i \le \sum_{i=1}^{L} \dot{c}_i \ddot{d}_i$.

Let $u_i = UB_{T-1}[\mathbf{a}, \bar{\mathbf{x}}_i]$, for $1 \le i \le L$, and let $\{\dot{u}_i\}$ be obtained by sorting $\{u_i\}$ in nonincreasing order. Clearly $\dot{c}_i \le \dot{u}_i$, for $1 \le i \le L$. Using notation from Lines 4–8, $\{\dot{u}_i\}$ has the form

$$\underbrace{U_1, \ldots, U_1}_{W_1 \text{ terms}}, \underbrace{U_2, \ldots, U_2}_{W_2 \text{ terms}}, \underbrace{U_3, \ldots, U_3}_{W_3 \text{ terms}}, \ \ldots . \tag{7}$$

If $\sum_{i=1}^{L} \dot{u}_i \le 1$, let $\{\ddot{c}_i\}$ be identical to the sequence $\{\dot{u}_i\}$. If $\sum_{i=1}^{L} \dot{u}_i > 1$, let L_u ($1 \le L_u \le L$) be minimum such that $\sum_{i=1}^{L_u} \dot{u}_i > 1$, and let $\{\ddot{c}_i\}$ consist of the first L terms of

$$\dot{u}_1, \dot{u}_2, \ldots, \dot{u}_{L_u - 1}, \left(1 - \sum_{i=1}^{L_u - 1} \dot{u}_i\right), 0, 0, 0, \ldots . \tag{8}$$

It follows that $\sum_{i=1}^{L} \dot{c}_i \ddot{d}_i \le \sum_{i=1}^{L} \ddot{c}_i \ddot{d}_i$ by Lemma 7 (with $\{\ddot{d}_i\}$ playing the role of $\{\dot{d}_i\}$ in the statement of the lemma). Combining inequalities gives

$$ELP_T(\mathbf{a}, \mathbf{b}) \le \sum_{i=1}^{L} \ddot{c}_i \ddot{d}_i \, . \tag{9}$$

The value $\sum_{i=1}^{L} \ddot{c}_i \ddot{d}_i$ in (9) is exactly the upper bound computed in Figure 3. We argue similarly to the $T = 2$ case. Since $\{\ddot{c}_i\}$ and $\{\ddot{d}_i\}$ are derived from sequences which consist of groups of consecutive identical elements (the sequence in (7) and $\mathcal{S}_\ell$, respectively), the algorithm operates group-by-group, not element-by-element. Beginning at Line 10, the variable h is the index of the current group in $\{\ddot{c}_i\}$ (having element value U_h and size W_h). Function NextTermT() identifies the corresponding elements in $\{\ddot{d}_i\}$, and computes the equivalent of the element-by-element product.

If the terms in $\{\ddot{c}_i\}$ (resp. $\{\ddot{d}_i\}$) shrink to 0 because the corresponding terms in (7) (resp. $\mathcal{S}_\ell$) become 0, the check ($U_h > 0$) (resp. ($\lambda < 1$)) in Line 11 or Line 15 will fail, and the algorithm will exit. The check ($\Psi + (U_h * W_h) > 1$) in Line 15 detects the case that in the derivation of $\{\ddot{c}_i\}$ from $\{\dot{u}_i\}$ above, $\sum_{i=1}^{L} \dot{u}_i > 1$, and therefore $\{\ddot{c}_i\}$ is based on the truncated sequence in (8).

7 Computational Results

We estimate that running the above algorithm to completion will take up to 200,000 hours on a single Sun Ultra 5. We are currently running on about 50 CPUs, and have completed 43% of the computation for $2 \le T \le 10$.

It is worth noting that in progressing from 11% to 43% of the computation, there was no change in the upper bound for $2 \leq T \leq 10$. Combined with our experience in running the algorithm of [15], for which the numbers also stabilized quickly, we expect that the final results will be the same as those presented below.

In Figure 4, we plot our improved upper bound against that of [15] for $2 \leq T \leq 10$. Note that the new bound is noticeably superior to that of [15] for $T \geq 4$. When $T = 9$ rounds are being approximated, the upper bound value is $UB = 2^{-92}$. For a success rate of 96.7%, this corresponds to a data complexity of $\frac{32}{UB} = 2^{97}$ (Table 1). The corresponding upper bound value from [15] is 2^{-75}, for a data complexity of 2^{80}. This represents a significant improvement in the calculation of the provable security of Rijndael against linear cryptanalysis.

We also plot *very* preliminary results for $11 \leq T \leq 15$, in order to gain a sense of the behavior of the upper bound (for these values of T, we have completed only 1.5% of the necessary computation, hence the label "Extrapolation"). Unlike the upper bound in [15], the new upper bound does not appear to flatten out, but continues a downward progression as T increases.

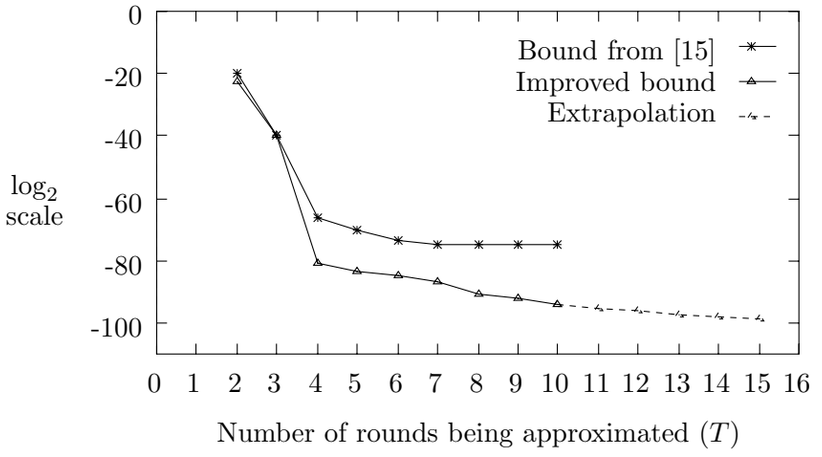

Fig. 4. Improved upper bound on MALHP for Rijndael

7.1 Presentation of Final Results

Upon completion of computation, we will post our final results in the IACR Cryptology ePrint Archive (`eprint.iacr.org`) under the title *Completion of Computation of Improved Upper Bound on the Maximum Average Linear Hull Probability for Rijndael.*

8 Conclusion

We have presented an improved version of the algorithm given in [15] (which computes an upper bound on the maximum average linear hull probability (MALHP) for SPNs) in the case of Rijndael. The improvement is achieved by taking into account the distribution of linear probability values for the (unique) Rijndael s-box. When 9 rounds of Rijndael are approximated, the new upper bound is 2^{-92}, which corresponds to a lower bound on the data complexity of 2^{97}, for a 96.7% success rate. (This is based on completion of 43% of the computation. However, we expect that the values obtained so far for $2 \leq T \leq 10$ core rounds will remain unchanged—see Section 7.) This is a significant improvement over the corresponding upper bound from [15], namely 2^{-75}, for a data complexity of 2^{80} (also for a 96.7% success rate). The new result strengthens our confidence in the provable security of Rijndael against linear cryptanalysis.

Acknowledgments

We are grateful to the reviewers for comments which improved the content and presentation of this paper. We are also grateful to the following for help in obtaining access to significant computational resources: the High Performance Computing Virtual Laboratory (Canada), the San Diego Supercomputer Center, Tom Bradshaw, Randy Ellis, Peter Hellekalek, Alex MacPherson, and Gerhard Wesp.

References

1. C.M. Adams, *A formal and practical design procedure for substitution-permutation network cryptosystems,* Ph.D. Thesis, Queen's University, Kingston, Canada, 1990.
2. K. Aoki and K. Ohta, *Strict evaluation of the maximum average of differential probability and the maximum average of linear probability,* IEICE Trans. Fundamentals, Vol. E80-A, No. 1, January 1997.
3. E. Biham, *On Matsui's linear cryptanalysis,* Advances in Cryptology—EUROCRYPT'94, LNCS 950, Springer-Verlag, pp. 341–355, 1995.
4. E. Biham and A. Shamir, *Differential cryptanalysis of DES-like cryptosystems,* Journal of Cryptology, Vol. 4, No. 1, pp. 3–72, 1991.
5. J. Daemen, R. Govaerts, and J. Vandewalle, *Correlation matrices,* Fast Software Encryption : Second International Workshop, LNCS 1008, Springer-Verlag, pp. 275–285, 1995.
6. J. Daemen and V. Rijmen, *AES proposal: Rijndael,* `http://csrc.nist.gov/encryption/aes/rijndael/Rijndael.pdf`.
7. J. Daemen and V. Rijmen, *AES (Rijndael) reference code in ANSI C,* `http://csrc.nist.gov/encryption/aes/rijndael/`.
8. *Data Encryption Standard (DES),* National Bureau of Standards FIPS Publication 46, 1977.
9. H. Feistel, *Cryptography and computer privacy,* Scientific American, Vol. 228, No. 5, pp. 15–23, May 1973.

10. H. Feistel, W.A. Notz, and J.L. Smith, *Some cryptographic techniques for machine to machine data communications,* Proceedings of the IEEE, Vol. 63, No. 11, pp. 1545–1554, November 1975.
11. C. Harpes, G. Kramer, and J. Massey, *A generalization of linear cryptanalysis and the applicability of Matsui's piling-up lemma,* Advances in Cryptology—EUROCRYPT'95, LNCS 921, Springer-Verlag, pp. 24–38, 1995.
12. H.M. Heys and S.E. Tavares, *Substitution-permutation networks resistant to differential and linear cryptanalysis,* Journal of Cryptology, Vol. 9, No. 1, pp. 1–19, 1996.
13. S. Hong, S. Lee, J. Lim, J. Sung, and D. Cheon, *Provable security against differential and linear cryptanalysis for the SPN structure,* Fast Software Encryption (FSE 2000), LNCS 1978, Springer-Verlag, pp. 273–283, 2001.
14. L. Keliher, H. Meijer, and S. Tavares, *Modeling linear characteristics of substitution-permutation networks,* Sixth Annual International Workshop on Selected Areas in Cryptography (SAC'99), LNCS 1758, Springer-Verlag, pp. 78–91, 2000.
15. L. Keliher, H. Meijer, and S. Tavares, *New method for upper bounding the maximum average linear hull probability for SPNs,* Advances in Cryptology—EUROCRYPT 2001, LNCS 2045, Springer-Verlag, pp. 420–436, 2001.
16. L.R. Knudsen, *Practically secure Feistel ciphers,* Fast Software Encryption, LNCS 809, Springer-Verlag, pp. 211–221, 1994.
17. X. Lai, J. Massey, and S. Murphy, *Markov ciphers and differential cryptanalysis,* Advances in Cryptology—EUROCRYPT'91, LNCS 547, Springer-Verlag, pp. 17–38, 1991.
18. M. Matsui, *Linear cryptanalysis method for DES cipher,* Advances in Cryptology—EUROCRYPT'93, LNCS 765, Springer-Verlag, pp. 386–397, 1994.
19. M. Matsui, *On correlation between the order of s-boxes and the strength of DES,* Advances in Cryptology—EUROCRYPT'94, LNCS 950, Springer-Verlag, pp. 366–375, 1995.
20. W. Meier and O. Staffelbach, *Nonlinearity criteria for cryptographic functions,* Advances in Cryptology—EUROCRYPT'89, LNCS 434, Springer-Verlag, pp. 549–562, 1990.
21. K. Nyberg, *Linear approximation of block ciphers,* Advances in Cryptology—EUROCRYPT'94, LNCS 950, LNCS 950, Springer-Verlag, pp. 439–444, 1995.
22. C.E. Shannon, *Communication theory of secrecy systems,* Bell System Technical Journal, Vol. 28, no. 4, pp. 656–715, 1949.
23. S. Vaudenay, *On the security of CS-Cipher,* Fast Software Encryption (FSE'99), LNCS 1636, Springer-Verlag, pp. 260–274, 1999.