Toward Provable Security Against Differential and Linear Cryptanalysis for Camellia and Related Ciphers

Liam Keliher

Department of Mathematics and Computer Science, Mount Allison University

Sackville, New Brunswick, Canada (Email: lkeliher@mta.ca)

(Received Mar. 14, 2006; revised and accepted July 31, 2006)

Abstract

We present a new algorithm that evaluates provable security against differential and linear cryptanalysis for Feistel ciphers with invertible substitution-diffusion (SD)-based round functions. This algorithm computes an upper bound on the maximum expected differential or linear probability (MEDP or MELP) based on the number of rounds. We then apply our algorithm to Camellia (minus FL/FL^{-1}). Previously, the best upper bounds for Camellia were 2^{-12} (both MEDP and MELP) for 3+ rounds. Our algorithm improves these bounds to 1.065×2^{-28} (MEDP) and 1.161×2^{-27} (MELP) for 6+ rounds. This is a first step toward establishing the provable security of Camellia and related ciphers against differential and linear cryptanalysis.

Keywords: Camellia, differential cryptanalysis, Feistel cipher, linear cryptanalysis, provable security

1 Introduction

Differential cryptanalysis and linear cryptanalysis, generally considered the most powerful attacks on block ciphers, were first proposed in the context of Feistel ciphers, specifically DES [7], in the early 1990s [4, 16]. Many papers followed that investigated the theoretical underpinnings of these attacks, and explored nuances and extensions. In particular, Lai et al. [15] and Nyberg [20] defined the concepts necessary for establishing *provable security* [9, 21] against differential and linear cryptanalysis, namely *differentials* and *linear hulls,* respectively.[1] However, the application of these concepts to most block ciphers is nontrivial, so resistance to differential and linear cryptanalysis is often claimed when only the less stringent criterion of *practical security* [14] has been established.

Since the work of Lai et al. and Nyberg, there have been relatively few publications dealing with provable security for Feistel ciphers [1, 18, 19, 21]. On the other hand, during the past few years many papers have appeared dealing with provable security for block ciphers based on the substitution-permutation network (SPN) structure [5, 9, 11, 12, 13, 22, 23, 24]. This flurry of results has produced a growing "toolbox" of techniques for such analysis of SPNs. (The recent bias toward SPNs is no doubt due in large part to the adoption of the SPN *Rijndael* as the Advanced Encryption Standard (AES) [6].) However, there are almost no comparable techniques for Feistel ciphers. It is our hope that the work of this paper will constitute the first in a series of such techniques.

We present a new algorithm for deriving an upper bound on the *maximum expected differential probability* (MEDP) or *maximum expected linear probability* (MELP) for Feistel ciphers with invertible *substitution-diffusion* (SD)-based round functions. This upper bound is computed as a function of the number of rounds under consideration. (Our approach has elements in common with the algorithm KMT2/KMT2-DC of Keliher et al. [12] (see also [10]), which upper bounds the MEDP and MELP for SPNs.) We then apply our algorithm to Camellia (minus FL/FL^{-1}) [2]. Prior to this paper, the best upper bounds on both the MEDP and MELP for Camellia were 2^{-12} for 3 or more rounds; this follows from a result of Aoki and Ohta [1]. Our algorithm improves these bounds to 1.065×2^{-28} (MEDP) and 1.161×2^{-27} (MELP) for 6 or more rounds.

[1] The definition of provable security used in this paper is well established. However, note that there is at least one other widely used definition of this term (see [26]).

2 Background Concepts

2.1 Block Ciphers

A *block cipher* is a bijective mapping $E_{\mathbf{k}} : \{0,1\}^N \rightarrow \{0,1\}^N$, where N is the *block size* and $\mathbf{k}$ is a *key*. The input to a block cipher is called a *plaintext*, and the output is called a *ciphertext*. Most block ciphers consist of a sequence of R weaker encryption steps called *rounds*, where round r ($1 \leq r \leq R$), which also maps $\{0,1\}^N \rightarrow \{0,1\}^N$, is parameterized by a *subkey* $\mathbf{k}^r$. The vector of subkeys $\langle \mathbf{k}^1, \mathbf{k}^2, \ldots, \mathbf{k}^R \rangle$ is derived from $\mathbf{k}$ via a separate *key-scheduling algorithm*.

2.2 Feistel Ciphers

A standard Feistel cipher[2] is a block cipher that modifies *half* of the current block in each round (so N must be even) [8]. Denote the left and right halves of the N-bit input to round r by $\mathbf{x}_{\mathrm{L}}^r$ and $\mathbf{x}_{\mathrm{R}}^r$, respectively. The left half, $\mathbf{x}_{\mathrm{L}}^r$, becomes the input to a *round function*, $F^r : \{0,1\}^{N/2} \rightarrow \{0,1\}^{N/2}$, which also takes $\mathbf{k}^r$ as a parameter. The output from F^r is XORed with $\mathbf{x}_{\mathrm{R}}^r$ to form $\mathbf{x}_{\mathrm{L}}^{r+1}$ (the left half of the input to the next round), while $\mathbf{x}_{\mathrm{L}}^r$ is preserved unchanged as $\mathbf{x}_{\mathrm{R}}^{r+1}$. This swapping of half blocks occurs in every round except the last. Figure 1 depicts the basic Feistel cipher structure.

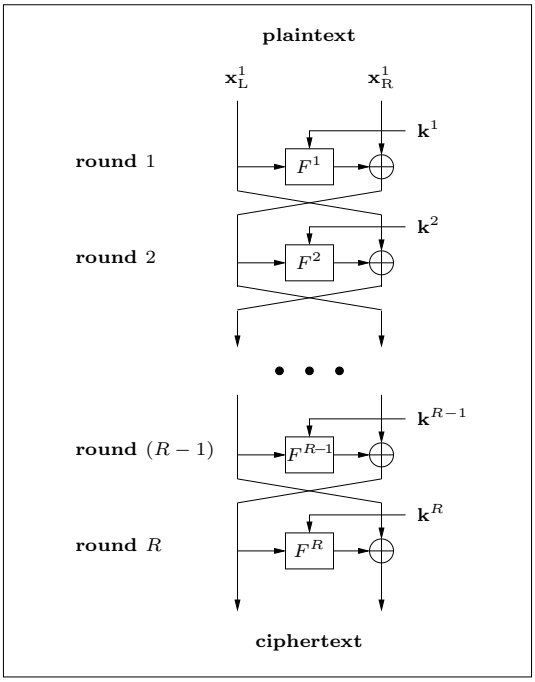

Figure 1: Basic Feistel cipher structure

In terms of implementation, one of the main advantages of Feistel ciphers is that encryption and decryption are essentially identical operations—a ciphertext is decrypted by processing it through the encryption algorithm with the order of the round functions and corresponding subkeys reversed. If the same round function is used in every round (a common approach), only the order of the subkeys must be reversed. A related observation is that the round functions are not required to be invertible. Overall, there is no need to generate/store inverse cipher components.

Obviously the choice of round function(s) is critical, and many designs have been proposed and studied. We will focus on round functions consisting of the three stages of a single SPN round:

1) *key-mixing stage* – round function input is XORed with $N/2$-bit subkey, $\mathbf{k}^r$

2) *substitution stage* – current block is partitioned into M sub-blocks of size n ($n = N/2M$), and each sub-block becomes the input to an $n \times n$ *substitution box* (*s-box*), i.e., a mapping $\{0,1\}^n \rightarrow \{0,1\}^n$

3) *diffusion stage* – output from the substitution stage is processed through a linear transformation $\mathcal{L} : \{0,1\}^{N/2} \rightarrow \{0,1\}^{N/2}$ (historically, a bitwise permutation)

Remark 1. We adopt two simplifying assumptions that hold for Camellia-like ciphers. First, all round functions are identical; denote this unique round function by F. Second, the s-boxes and linear transformation are invertible (this is a requirement for SPNs, but can be relaxed for general Feistel ciphers).

2.3 Camellia

Camellia is a Feistel cipher developed by NTT and Mitsubishi Corporation that was introduced in 2000 [2]. It has been adopted by numerous international standards bodies, including ISO/IEC and IETF, and is generally viewed as being competitive with the AES in terms of security and performance.

Camellia has a block size of $N = 128$, and accepts keys of sizes 128, 192, and 256 bits. The number of rounds is 18 for 128-bit keys, and 24 for 192- and 256-bit keys. Camellia adheres to the Feistel cipher structure described above—including the use of an invertible SD-based round function that is identical in all rounds—except for two features:

- a 128-bit *whitening key* is XORed to the plaintext before the first round, and another is XORed to the output of the last round to form the ciphertext
- key-dependent linear functions FL and FL^{-1} are applied to the left and right halves, respectively, of the current 128-bit block after rounds 6 and 12 for 18-round Camellia, and after rounds 6, 12, and 18 for 24-round Camellia

The substitution stage of the round function uses four different 8×8 s-boxes, denoted $s_1, \ldots, s_4$, each of which appears twice ($M = 8$). All four s-boxes are variations of the

[2]There are many variations on the basic Fiestel structure—for example, see [25].

inversion mapping in $GF(2^8)$. If the input to the linear transformation stage is viewed as a vector $(x_1, x_2, \ldots, x_8)$ over $GF(2^8)$, then the linear transformation can be represented by an 8×8 matrix, denoted P, all of whose entries are 0 or 1:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \end{pmatrix}$$

2.4 Assumption of Independent Subkeys

In analyzing the resistance of block ciphers to certain attacks, including differential and linear cryptanalysis, it is standard to assume that each subkey is chosen independently and uniformly from the set of all possible subkeys, and to work with average values over this distribution. We adopt this approach. Issues concerning the "non-averaged" analysis of block ciphers remain largely unexplored, as do issues related to subkey distributions generated by specific key-scheduling algorithms.

3 Differential (and Linear) Cryptanalysis

Differential cryptanalysis, due to Biham and Shamir [4], is a chosen-plaintext attack that exploits the existence of relatively large *expected differential probability* (EDP) values over T core cipher rounds ($T \leq R$). Linear cryptanalysis, introduced by Matsui [16], is a known-plaintext attack that exploits relatively large *expected linear probability* (ELP) values over T core rounds. Because of the well-known duality between differential and linear cryptanalysis [3, 17, 27], concepts related to one attack often have counterparts for the other attack (see [13])—this holds for the results of the current paper. In light of this, we limit our exposition to differential cryptanalysis.

We first define concepts relevant to differential cryptanalysis in the context of a general block cipher,[3] and then focus on Feistel ciphers with invertible SD-based round functions.

3.1 Differential Probability

Definition 1. Let $B : \{0,1\}^d \to \{0,1\}^d$, and let $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^d$ be fixed. If $\mathbf{X} \in \{0,1\}^d$ is a uniformly distributed random variable, then the *differential probability* $DP(\Delta\mathbf{x}, \Delta\mathbf{y})$ is defined as

$$DP(\Delta\mathbf{x}, \Delta\mathbf{y}) = \mathrm{Prob}_{\mathbf{X}} \left\{ B(\mathbf{X}) \oplus B(\mathbf{X} \oplus \Delta\mathbf{x}) = \Delta\mathbf{y} \right\}.$$

If B is parameterized by a key, $\mathbf{k}$, we write $DP(\Delta\mathbf{x}, \Delta\mathbf{y}; \mathbf{k})$, and the *expected differential probability* $EDP(\Delta\mathbf{x}, \Delta\mathbf{y})$ is defined as

$$EDP(\Delta\mathbf{x}, \Delta\mathbf{y}) = E_{\mathbf{K}} \left[DP(\Delta\mathbf{x}, \Delta\mathbf{y}; \mathbf{K}) \right],$$

where $\mathbf{K}$ is a random variable uniformly distributed over the space of keys, and $E[\,]$ denotes expectation.

The values $\Delta\mathbf{x}$ / $\Delta\mathbf{y}$ in Definition 1 are called input/output *differences*. We can view DP and EDP values as entries in a $2^d \times 2^d$ table in the obvious way. The following lemma is trivial.

Lemma 1. Let $B : \{0,1\}^d \to \{0,1\}^d$, and let $\Delta\mathbf{x} \in \{0,1\}^d$. Then

$$\sum_{\Delta\mathbf{y} \in \{0,1\}^d} DP(\Delta\mathbf{x}, \Delta\mathbf{y}) \;=\; 1\,.$$

Remark 2. In what follows, terms such as "first round" and "last round" are relative to the T rounds under consideration. Single-variable superscripts refer to individual rounds, e.g., $DP^t(\Delta\mathbf{x}, \Delta\mathbf{y}; \mathbf{k}^t)$ and $EDP^t(\Delta\mathbf{x}, \Delta\mathbf{y})$ are DP and EDP values, respectively, for round t ($1 \leq t \leq T$). Superscripts of the form $[i \ldots j]$ (with $i < j$) refer to a sequence of consecutive rounds viewed as a single unit, e.g., $EDP^{[1\ldots3]}(\Delta\mathbf{x}, \Delta\mathbf{y})$ is an EDP value over rounds $1 \ldots 3$. For Feistel ciphers, we assume that swapping occurs in all T rounds.

3.2 Provable Security (MEDP)

Given $T \geq 2$ core rounds, the critical value for differential cryptanalysis is the *maximum expected differential probability* (MEDP):

$$MEDP^{[1\ldots T]} \;=\; \max_{\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N \setminus \mathbf{0}} EDP^{[1\ldots T]}(\Delta\mathbf{x}, \Delta\mathbf{y})\,. \tag{1}$$

The data complexity of the attack (the number of chosen plaintext-ciphertext pairs required) is proportional to the inverse of the MEDP. Therefore *provable security* can be claimed if the MEDP is sufficiently small that the corresponding data complexity is prohibitive [21].

3.3 Differential Characteristics

For most block ciphers, it is difficult to compute the MEDP exactly. A traditional method of approximation involves the use of *characteristics*.

Definition 2. A *differential characteristic* for rounds $1 \ldots T$ is a $(T+1)$-tuple

$$\Omega = \left\langle \Delta\mathbf{x}^1, \Delta\mathbf{x}^2, \ldots, \Delta\mathbf{x}^T, \Delta\mathbf{x}^{T+1} \right\rangle,$$

[3] Technically, for all the following results to hold the block cipher must be a *Markov cipher*—see [15].

where $\Delta\mathbf{x}^t$ and $\Delta\mathbf{x}^{t+1}$ are N-bit input and output differences, respectively, for round t ($1 \le t \le T$). The corresponding *expected differential characteristic probability* (EDCP) is defined as

$$EDCP^{[1\ldots T]}(\Omega) = \prod_{t=1}^{T} EDP^t(\Delta\mathbf{x}^t, \Delta\mathbf{x}^{t+1}). \quad (2)$$

Remark 3. For many ciphers, it is feasible to compute $EDP^t(\Delta\mathbf{x}^t, \Delta\mathbf{x}^{t+1})$ for each round t, and therefore to compute $EDCP^{[1\ldots T]}(\Omega)$.

3.3.1 Using the Best Characteristic (Practical Security)

A *best* characteristic (not necessarily unique) is one that maximizes $EDCP^{[1\ldots T]}(\Omega)$. Denote a best characteristic by $\hat{\Omega} = \langle \Delta\hat{\mathbf{x}}^1, \Delta\hat{\mathbf{x}}^2, \ldots, \Delta\hat{\mathbf{x}}^T, \Delta\hat{\mathbf{x}}^{T+1} \rangle$. The data complexity of differential cryptanalysis is often estimated by assuming that

$$\begin{aligned} MEDP^{[1\ldots T]} &\approx EDP^{[1\ldots T]}(\Delta\hat{\mathbf{x}}^1, \Delta\hat{\mathbf{x}}^{T+1}) \\ &\approx EDCP^{[1\ldots T]}(\hat{\Omega}). \end{aligned} \quad (3)$$

If the resulting data complexity is prohibitive, the cipher is *practically secure* [14].

3.4 Differentials

Definition 3 (Lai et al. [15]). If $T \ge 2$ and $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N$, then the corresponding *differential*, denoted $DIFF(\Delta\mathbf{x}, \Delta\mathbf{y})$, is the set of all characteristics for rounds $1 \ldots T$ having $\Delta\mathbf{x}$ as the first difference and $\Delta\mathbf{y}$ as the last difference, i.e., all characteristics of the form

$$\Omega = \langle \Delta\mathbf{x}, \Delta\mathbf{x}^2, \Delta\mathbf{x}^3, \ldots, \Delta\mathbf{x}^T, \Delta\mathbf{y} \rangle.$$

Theorem 1 ([15]). Let $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N$. Then

$$EDP^{[1\ldots T]}(\Delta\mathbf{x}, \Delta\mathbf{y}) = \sum_{\Omega \in DIFF(\Delta\mathbf{x}, \Delta\mathbf{y})} EDCP^{[1\ldots T]}(\Omega).$$

It follows from Theorem 1 that the approximation in Equation (3) does not hold in general, since $EDP^{[1\ldots T]}(\Delta\mathbf{x}, \Delta\mathbf{y})$ is seen to be the sum of (a large number of) terms $EDCP^{[1\ldots T]}(\Omega)$, and therefore, in general, the EDCP of any characteristic will be strictly *less than* the corresponding EDP value. Further, the MEDP may not be equal to (i.e., may be strictly greater than) the EDP associated with any best characteristic. This situation may result in an overestimation of the data complexity—beneficial for an attacker, but problematic for a cipher designer.

3.5 Expected Differential Probability for One SD-based Round

Consider input/output differences for one SD-based round of a Feistel cipher, as in Figure 2. We are interested in $EDP^t(\Delta\mathbf{x}, \Delta\mathbf{y})$, where $\Delta\mathbf{x} = (\Delta\mathbf{x}_\mathrm{L} \parallel \Delta\mathbf{x}_\mathrm{R})$ and $\Delta\mathbf{y} = (\Delta\mathbf{y}_\mathrm{L} \parallel \Delta\mathbf{y}_\mathrm{R})$ (the symbol $\parallel$ denotes concatenation). Clearly

$$EDP^t(\Delta\mathbf{x}, \Delta\mathbf{y}) = EDP^F(\Delta\mathbf{x}_\mathrm{L}, \Delta\mathbf{x}_\mathrm{R} \oplus \Delta\mathbf{y}_\mathrm{L}), \quad (4)$$

where $EDP^F(\cdot,\cdot)$ is an EDP value over the round function, F.

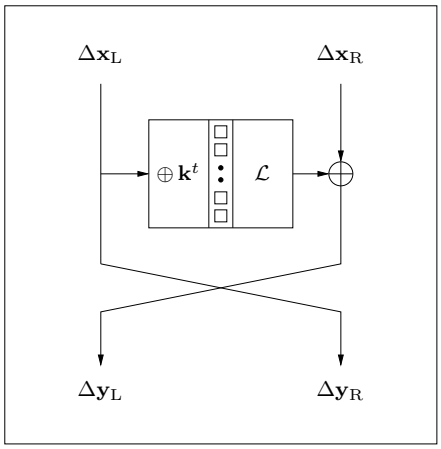

Figure 2: Input/output differences for one SD-based Feistel round

Let $\Delta\mathbf{a}$ and $\Delta\mathbf{b}$ be the input and output differences, respectively, for the *substitution stage* of F, i.e., $\Delta\mathbf{a} = \Delta\mathbf{x}_\mathrm{L}$ and $\Delta\mathbf{b} = \mathcal{L}^{-1}(\Delta\mathbf{x}_\mathrm{R} \oplus \Delta\mathbf{y}_\mathrm{L})$. Enumerate the s-boxes as $S_1, \ldots, S_M$. Subdividing $\Delta\mathbf{a}$ and $\Delta\mathbf{b}$ gives input/output differences for each s-box, denoted $\Delta\mathbf{a}_i$ / $\Delta\mathbf{b}_i$ ($1 \le i \le M$). It is easy to show that

$$EDP^t(\Delta\mathbf{x}, \Delta\mathbf{y}) = \prod_{i=1}^{M} DP^{S_i}(\Delta\mathbf{a}_i, \Delta\mathbf{b}_i). \quad (5)$$

A characteristic is *consistent* if, for any round with input/output differences $\Delta\mathbf{x} = (\Delta\mathbf{x}_\mathrm{L} \parallel \Delta\mathbf{x}_\mathrm{R})$ / $\Delta\mathbf{y} = (\Delta\mathbf{y}_\mathrm{L} \parallel \Delta\mathbf{y}_\mathrm{R})$, the following two conditions hold:

1) $\Delta\mathbf{y}_\mathrm{R} = \Delta\mathbf{x}_\mathrm{L}$
2) the input and output differences for any s-box are either both zero or both nonzero

Given any consistent characteristic, an s-box with nonzero input and output differences is called *active*. We limit consideration to consistent characteristics.

3.6 Distribution of DP Values for Multiple Active s-boxes

Definition 4. Let $\mathbf{v}$ be a binary vector of length $\{0,1\}^{N/2}$ (resp. $\{0,1\}^N$). Then $\gamma_\mathbf{v}$ is a binary vector of length $L = M$ (resp. $L = 2M$) formed as follows: partition $\mathbf{v}$ into L sub-vectors of length $n = N/2M$ (the size of the s-box input/output), denoted $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_L$; define $\gamma_i = 0$ if $\mathbf{v}_i = \mathbf{0}$, and $\gamma_i = 1$ otherwise, for $1 \le i \le L$; and let $\gamma_\mathbf{v} = \gamma_1\gamma_2\cdots\gamma_L$.

Remark 4. We think of $\gamma_{\mathbf{v}}$ as encoding the "pattern" of nonzero sub-vectors in $\mathbf{v}$. If $\mathbf{v}$ is an input or output difference for the substitution stage of F, taken from a consistent characteristic, then the active s-boxes are given by $\gamma_{\mathbf{v}} = \gamma_1\gamma_2\cdots\gamma_M$, i.e., S_i is active if and only if $\gamma_i = 1$; the number of active s-boxes is $wt(\gamma_{\mathbf{v}})$, where $wt(\,)$ denotes Hamming weight.

For simplicity, we assume that all nontrivial rows and columns of the DP table for any s-box have the same distribution of values. This property holds for Camellia and the AES—in fact, they share the same distribution [23]. (Dealing with the more general situation is straightforward, but detailed—see Section 6.5 in [10].) Let D_1 be the number of distinct values, let $\rho_1^1, \rho_2^1, \ldots, \rho_{D_1}^1$ be these distinct values in decreasing order, and let ϕ_i^1 be the frequency with which ρ_i^1 occurs, for $1 \le i \le D_1$.

Definition 5. Let $\Delta\mathbf{x} \in \{0,1\}^{N/2} \setminus \mathbf{0}$ be a fixed input difference for the substitution stage of F, let $\Delta\mathbf{y}$ be an output difference for the same substitution stage, and suppose $\Delta\mathbf{y}$ varies over $\{0,1\}^{N/2}$, with the restriction that $\gamma_{\Delta\mathbf{x}} = \gamma_{\Delta\mathbf{y}}$. If A is the number of active s-boxes ($A = wt(\gamma_{\Delta\mathbf{x}}) = wt(\gamma_{\Delta\mathbf{y}})$), define $\mathcal{D}_A$ to be the set of distinct one-round DP values produced as $\Delta\mathbf{y}$ 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_i^A be the number of occurrences of the value ρ_i^A, for $1 \le i \le D_A$.

The proof of the following lemma is straightforward.

Lemma 2. 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 i, $1 \le i \le D_A = \#\mathcal{D}_A$,

$$\phi_i^A \;=\; \sum_{\substack{\rho_s^1 \cdot \rho_t^{A-1} = \rho_i^A \\ 1 \le s \le D_1,\ 1 \le t \le D_{A-1}}} \phi_s^1 \cdot \phi_t^{A-1}\ .$$

Definition 6. For $A \ge 1$ and $1 \le J \le D_A$, define the partial sums

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

Remark 5. It follows from Lemma 1 that $\Lambda_{D_A}^A = 1$, for $A \ge 1$.

4 New Algorithm

Consider a Feistel cipher with an SD-based round function. For any $T \ge 2$ and $\gamma, \hat{\gamma} \in \{0,1\}^{2M} \setminus \mathbf{0}$, our algorithm computes a value $UB^{[1\ldots T]}(\gamma, \hat{\gamma})$ such that for all $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N \setminus \mathbf{0}$ satisfying $\gamma_{\Delta\mathbf{x}} = \gamma$, $\gamma_{\Delta\mathbf{y}} = \hat{\gamma}$, the following holds:

$$EDP^{[1\ldots T]}(\Delta\mathbf{x}, \Delta\mathbf{y}) \;\le\; UB^{[1\ldots T]}(\gamma, \hat{\gamma})\,. \qquad (6)$$

This immediately yields an upper bound on the MEDP:

$$MEDP^{[1\ldots T]} \;\le \max_{\gamma,\hat{\gamma} \in \{0,1\}^{2M} \setminus \mathbf{0}} UB^{[1\ldots T]}(\gamma, \hat{\gamma})\,. \qquad (7)$$

For many ciphers, if we can compute the individual values $UB^{[1\ldots T]}(\gamma, \hat{\gamma})$, then it is feasible to evaluate the right-hand side of Equation (7) (e.g., $2M = 16$ for Camellia). This reduction of computational complexity relative to direct computation of the MEDP (as in Equation (1)) is the main reason for adopting a "pattern-oriented" approach. We divide our algorithm into a base case ($T = 2$) and an inductive step ($T \ge 3$); these are given in Sections 4.2 and 4.3, respectively. We start with some technical definitions.

4.1 Technical Definitions

The following table of values extracts useful pattern-based information from the linear transformation, $\mathcal{L}$.

Definition 7. Let $\gamma, \hat{\gamma} \in \{0,1\}^M$. Then

$$W_d[\gamma, \hat{\gamma}] = \#\left\{ \Delta\mathbf{x} \in \{0,1\}^{N/2} : \gamma_{\Delta\mathbf{x}} = \gamma,\ \gamma_{\mathcal{L}(\Delta\mathbf{x})} = \hat{\gamma} \right\}.$$

Remark 6. For efficiency, $W_d[\,]$ should be pre-computed for use in the algorithm below.

Definition 8. Let $\gamma, \hat{\gamma}, \hat{\hat{\gamma}} \in \{0,1\}^M$. We say that γ and $\hat{\gamma}$ *produce* $\hat{\hat{\gamma}}$, denoted $(\gamma, \hat{\gamma}) \Rightarrow \hat{\hat{\gamma}}$, if there exist $\mathbf{v}, \hat{\mathbf{v}} \in \{0,1\}^{N/2}$ with $\gamma_{\mathbf{v}} = \gamma$, $\gamma_{\hat{\mathbf{v}}} = \hat{\gamma}$, and $\gamma_{\mathbf{v}\oplus\hat{\mathbf{v}}} = \hat{\hat{\gamma}}$.

Definition 9. Let $\gamma, \hat{\gamma}, \hat{\hat{\gamma}} \in \{0,1\}^M$ such that $(\gamma, \hat{\gamma}) \Rightarrow \hat{\hat{\gamma}}$. Fix any $\hat{\mathbf{v}} \in \{0,1\}^{N/2}$ satisfying $\gamma_{\hat{\mathbf{v}}} = \hat{\gamma}$. Define $\#[(\gamma, \hat{\gamma}) \Rightarrow \hat{\hat{\gamma}}]$ to be the number of $\mathbf{v} \in \{0,1\}^{N/2}$ for which $\gamma_{\mathbf{v}} = \gamma$ and $\gamma_{\mathbf{v}\oplus\hat{\mathbf{v}}} = \hat{\hat{\gamma}}$.

Remark 7. The value $\#[(\gamma, \hat{\gamma}) \Rightarrow \hat{\hat{\gamma}}]$ in Definition 9 is independent of the specific choice of $\hat{\mathbf{v}} \in \{0,1\}^{N/2}$. Note that the ordering of γ and $\hat{\gamma}$ is not important in Definition 8, i.e., $(\gamma, \hat{\gamma}) \Rightarrow \hat{\hat{\gamma}}$ if and only if $(\hat{\gamma}, \gamma) \Rightarrow \hat{\hat{\gamma}}$, but it *is* important in Definition 9, i.e., in general $\#[(\gamma, \hat{\gamma}) \Rightarrow \hat{\hat{\gamma}}] \ne \#[(\hat{\gamma}, \gamma) \Rightarrow \hat{\hat{\gamma}}]$.

Definition 10. Let $\gamma, \hat{\gamma}, \hat{\hat{\gamma}} \in \{0,1\}^M$. We say that γ *F-matches* $\hat{\gamma}$ and $\hat{\hat{\gamma}}$, denoted $\gamma \dashv (\hat{\gamma}, \hat{\hat{\gamma}})$, if there exists $\delta \in \{0,1\}^M$ such that $(\hat{\gamma}, \hat{\hat{\gamma}}) \Rightarrow \delta$ and $W_d[\gamma, \delta] > 0$.

Remark 8. The symbol $\dashv$ is chosen to be reminiscent of the XOR junction in each round. The pattern γ (resp. δ) corresponds to the input difference (resp. output difference) of the round function (or, equivalently, of the linear transformation, $\mathcal{L}$).

Definition 11. Let $DP_{\max}$ be the maximum nontrivial DP value over the round function s-boxes, i.e.,

$$DP_{\max} \;=\; \max_{1 \le i \le M} \ \max_{\Delta\mathbf{a}, \Delta\mathbf{b} \in \{0,1\}^n \setminus \mathbf{0}} DP^{S_i}(\Delta\mathbf{a}, \Delta\mathbf{b})\,.$$

4.2 Base Case ($T = 2$)

Theorem 2. Let $\gamma, \hat{\gamma} \in \{0,1\}^{2M} \setminus \mathbf{0}$, and partition these patterns into left and right halves as $\gamma = (\gamma_{\mathrm{L}} \parallel \gamma_{\mathrm{R}})$ and $\hat{\gamma} = (\hat{\gamma}_{\mathrm{L}} \parallel \hat{\gamma}_{\mathrm{R}})$. Let $A = wt(\gamma_{\mathrm{L}}) + wt(\hat{\gamma}_{\mathrm{R}})$. If

$$UB^{[1\ldots2]}(\gamma,\hat{\gamma}) = \begin{cases} (DP_{\max})^A & \text{if } \gamma_{\mathrm{L}} \dashv (\gamma_{\mathrm{R}}, \hat{\gamma}_{\mathrm{R}}) \text{ and} \\ & \quad \hat{\gamma}_{\mathrm{R}} \dashv (\gamma_{\mathrm{L}}, \hat{\gamma}_{\mathrm{L}}) \\ 0 & \text{otherwise} \end{cases}$$

then Equation (6) holds for $T = 2$.

Proof. Let $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N \setminus \mathbf{0}$ be any vectors satisfying $\gamma_{\Delta\mathbf{x}} = \gamma$, $\gamma_{\Delta\mathbf{y}} = \hat{\gamma}$, and suppose $\Omega = \langle \Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{y} \rangle \in DIFF(\Delta\mathbf{x}, \Delta\mathbf{y})$. If $\Delta\mathbf{x}$, $\Delta\mathbf{y}$, and $\Delta\mathbf{z}$ are partitioned into halves as $(\Delta\mathbf{x}_{\mathrm{L}} \parallel \Delta\mathbf{x}_{\mathrm{R}})$, $(\Delta\mathbf{y}_{\mathrm{L}} \parallel \Delta\mathbf{y}_{\mathrm{R}})$, and $(\Delta\mathbf{z}_{\mathrm{L}} \parallel \Delta\mathbf{z}_{\mathrm{R}})$, respectively, then $\Delta\mathbf{z}_{\mathrm{L}} = \Delta\mathbf{y}_{\mathrm{R}}$ and $\Delta\mathbf{z}_{\mathrm{R}} = \Delta\mathbf{x}_{\mathrm{L}}$, *so Ω is the only characteristic in $DIFF(\Delta\mathbf{x}, \Delta\mathbf{y})$.* It follows from Theorem 1, Equations (2) and (4) that $EDP^{[1\ldots2]}(\Delta\mathbf{x}, \Delta\mathbf{y})$ is equal to

$$EDP^F(\Delta\mathbf{x}_{\mathrm{L}},\, \Delta\mathbf{x}_{\mathrm{R}} \oplus \Delta\mathbf{y}_{\mathrm{R}}) \cdot EDP^F(\Delta\mathbf{y}_{\mathrm{R}},\, \Delta\mathbf{x}_{\mathrm{L}} \oplus \Delta\mathbf{y}_{\mathrm{L}})\,. \tag{8}$$

And it follows from Equation (5) and Definition 11 that Equation (8) is upper bounded by $(DP_{\max})^A$, where $A = wt(\gamma_{\mathrm{L}}) + wt(\hat{\gamma}_{\mathrm{R}})$ is the total number of active s-boxes in rounds 1 and 2.

The above argument demonstrates that $UB^{[1\ldots2]}(\gamma,\hat{\gamma}) = (DP_{\max})^A$ will always satisfy Equation (6). Note, however, that for certain $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N \setminus \mathbf{0}$, $DIFF(\Delta\mathbf{x}, \Delta\mathbf{y})$ may be *empty*, i.e., the characteristic $\langle \Delta\mathbf{x}, (\Delta\mathbf{y}_{\mathrm{R}} \parallel \Delta\mathbf{x}_{\mathrm{L}}), \Delta\mathbf{y} \rangle$ may not be consistent. Further, given $\gamma, \hat{\gamma} \in \{0,1\}^{2M} \setminus \mathbf{0}$, $DIFF(\Delta\mathbf{x}, \Delta\mathbf{y})$ may be empty for *all* $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N \setminus \mathbf{0}$ satisfying $\gamma_{\Delta\mathbf{x}} = \gamma$, $\gamma_{\Delta\mathbf{y}} = \hat{\gamma}$, and therefore we can use the trivial value $UB^{[1\ldots2]}(\gamma,\hat{\gamma}) = 0$. If either $\gamma_{\mathrm{L}} \dashv (\gamma_{\mathrm{R}}, \hat{\gamma}_{\mathrm{R}})$ or $\hat{\gamma}_{\mathrm{R}} \dashv (\gamma_{\mathrm{L}}, \hat{\gamma}_{\mathrm{L}})$ fails to hold, then we are in this degenerate case. □

4.3 Inductive Step ($T \geq 3$)

Theorem 3. Let $T \geq 3$, and assume that the values $UB^{[1\ldots(T-1)]}(\gamma,\hat{\gamma})$ have been computed for all $\gamma, \hat{\gamma} \in \{0,1\}^{2M} \setminus \mathbf{0}$. If the values $UB^{[1\ldots T]}(\gamma,\hat{\gamma})$ are computed using the algorithm in Figure 3, then Equation (6) holds.

Proof. In what follows, "Line X" refers to the X^{th} line in Figure 3. Let $T \geq 3$, and let $\gamma, \hat{\gamma} = (\hat{\gamma}_L, \hat{\gamma}_R) \in \{0,1\}^{2M} \setminus \mathbf{0}$. Fix any $\Delta\mathbf{x}, \Delta\mathbf{y} \in \{0,1\}^N$ satisfying $\gamma_{\Delta\mathbf{x}} = \gamma$, $\gamma_{\Delta\mathbf{y}} = \hat{\gamma}$. It follows from Theorem 1 and Equation (2) that

$$\begin{aligned} & EDP^{[1\ldots T]}(\Delta\mathbf{x}, \Delta\mathbf{y}) \\ = & \sum_{\Delta\mathbf{z} \in \{0,1\}^N} EDP^{[1\ldots(T-1)]}(\Delta\mathbf{x}, \Delta\mathbf{z}) \cdot EDP^T(\Delta\mathbf{z}, \Delta\mathbf{y}). \end{aligned} \tag{9}$$

If we partition $\Delta\mathbf{y}$ and $\Delta\mathbf{z}$ into halves as $(\Delta\mathbf{y}_L \parallel \Delta\mathbf{y}_R)$ and $(\Delta\mathbf{z}_L \parallel \Delta\mathbf{z}_R)$, respectively, then $\Delta\mathbf{z}_L = \Delta\mathbf{y}_R$, so we can limit consideration in Equation (9) to $\Delta\mathbf{z} = (\Delta\mathbf{y}_R \parallel \Delta\mathbf{z}_R)$, i.e., it suffices to vary over $\Delta\mathbf{z}_R \in \{0,1\}^{N/2}$. It follows that $\gamma_{\Delta\mathbf{z}} = (\hat{\gamma}_R \parallel \varepsilon)$, for some $\varepsilon \in \{0,1\}^M$. The heart of the algorithm is the replacement of each term $EDP^{[1\ldots(T-1)]}(\Delta\mathbf{x}, \Delta\mathbf{z})$ in Equation (9) with a previously computed upper bound value $UB^{[1\ldots(T-1)]}(\gamma, (\hat{\gamma}_R \parallel \varepsilon))$, along with the observation that each term $EDP^T(\Delta\mathbf{z}, \Delta\mathbf{y})$ is an element of the sequence

$$\underbrace{\rho_1^A, \ldots, \rho_1^A}_{\phi_1^A \text{ terms}},\ \underbrace{\rho_2^A, \ldots, \rho_2^A}_{\phi_2^A \text{ terms}},\ \ldots,\ \underbrace{\rho_{D_A}^A, \ldots, \rho_{D_A}^A}_{\phi_{D_A}^A \text{ terms}}, \tag{10}$$

where $A = wt(\hat{\gamma}_{\mathrm{R}})$ (i.e., A is the number of active s-boxes in round T). The selected values $UB^{[1\ldots(T-1)]}(\gamma, (\hat{\gamma}_R \parallel \varepsilon))$ in nonincreasing order comprise the sequence

$$\underbrace{v_1, \ldots, v_1}_{C[\varepsilon_1] \text{ terms}},\ \underbrace{v_2, \ldots, v_2}_{C[\varepsilon_2] \text{ terms}},\ \ldots,\ \underbrace{v_J, \ldots, v_J}_{C[\varepsilon_J] \text{ terms}} \tag{11}$$

(see Line 15). In Lines 16–30 we compute the sum of the term-by-term product of Equations (10) and (11), and this sum becomes the bound $UB^{[1\ldots T]}(\gamma,\hat{\gamma})$ in Line 24. More specifically, Lines 17–20 handle each group of identical terms

$$\underbrace{v_j, \ldots, v_j}_{C[\varepsilon_j] \text{ terms}},$$

calling MatchGroup to calculate the sum of the matching terms in Equation (10)—this sum, stored in $\Delta\lambda$, is multiplied by v_j in Line 30, and the product is added to the growing value of Sum in Line 19. As a refinement, since

$$\sum_{\Delta\mathbf{z} \in \{0,1\}^N} EDP^{[1\ldots(T-1)]}(\Delta\mathbf{x}, \Delta\mathbf{z}) = 1$$

(this follow from Lemma 1), it is easy to show that the sequence in Equation (11) can be truncated so that the terms sum exactly to 1, without compromising the upper bound that is being computed (see Lemma 6 in [11]). Lines 21–23 handle this truncation.

It remains to explain the selection of the values $v_1, v_2, \ldots, v_J$. First note that Lines 3–4 handle the trivial situation in which we "fall through" round T because the input difference (and therefore the output difference) for F is $\mathbf{0}$, so the EDP for round T is 1. Now consider Lines 5–14. Recall that we are varying over all $\Delta\mathbf{z}_R \in \{0,1\}^{N/2}$ as we upper bound Equation (9). Equivalently, we can think of varying over all output differences $\Delta\mathbf{w} \in \{0,1\}^{N/2}$ for F, and setting $\Delta\mathbf{z}_R = \Delta\mathbf{w} \oplus \Delta\mathbf{y}_L$. Further, we can limit consideration to those $\Delta\mathbf{w}$ for which $W_d[\hat{\gamma}_R, \gamma_{\Delta\mathbf{w}}] > 0$, since if $W_d[\hat{\gamma}_R, \gamma_{\Delta\mathbf{w}}] = 0$, then the corresponding $\Delta\mathbf{z}_R$ does not belong to a consistent characteristic (for output difference $\Delta\mathbf{y}$). For each $\varepsilon \in \{0,1\}^M$, $C[\varepsilon]$ is a counter that will hold the number of $\Delta\mathbf{w}$ such that $UB^{[1\ldots(T-1)]}(\gamma, (\hat{\gamma}_R \parallel \varepsilon))$ is chosen by the algorithm as the upper bound for $EDP^{[1\ldots(T-1)]}(\Delta\mathbf{x}, (\Delta\mathbf{y}_R \parallel \Delta\mathbf{z}_R))$.

For each $\delta \in \{0,1\}^M$ satisfying $W \stackrel{\text{def}}{=} W_d[\hat{\gamma}_R, \delta] > 0$, there are W output differences $\Delta\mathbf{w}$ for F with pattern δ when the pattern of active s-boxes is $\hat{\gamma}_R$. The possible

1. For $\gamma \in \{0,1\}^{2M} \setminus \mathbf{0}$
2. For $\hat{\gamma} = (\hat{\gamma}_{\mathrm{L}} \parallel \hat{\gamma}_{\mathrm{R}}) \in \{0,1\}^{2M} \setminus \mathbf{0}$
3. If $(\hat{\gamma}_{\mathrm{R}} = 0)$
4. $UB^{[1\ldots T]}(\gamma, \hat{\gamma}) = UB^{[1\ldots(T-1)]}(\gamma, (\hat{\gamma}_{\mathrm{R}} \parallel \hat{\gamma}_{\mathrm{L}}))$
5. Else
6. Initialize counters $C[\varepsilon]$ to 0 for all $\varepsilon \in \{0,1\}^M$
7. For $\delta \in \{0,1\}^M$ such that $W_d[\hat{\gamma}_{\mathrm{R}}, \delta] > 0$
8. Let $\mu_1, \mu_2, \ldots, \mu_I$ be the elements μ of $\{0,1\}^M$ satisfying $(\delta, \hat{\gamma}_{\mathrm{L}}) \Rightarrow \mu$, ordered such that if $u_i = UB^{[1\ldots(T-1)]}(\gamma, (\hat{\gamma}_{\mathrm{R}} \parallel \mu_i))$, then $u_1, \ldots, u_I$ is a nonincreasing sequence
9. $W \leftarrow W_d[\hat{\gamma}_{\mathrm{R}}, \delta], \quad i \leftarrow 1$
10. While $(W > 0)$ and $(i \leq I)$
11. $c \leftarrow \min\{W,\ \#[(\delta, \hat{\gamma}_{\mathrm{L}}) \Rightarrow \mu_i]\}$
12. $C[\mu_i] \leftarrow C[\mu_i] + c$
13. $W \leftarrow W - c$
14. $i \leftarrow i + 1$
15. Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_J$ be the elements ε of $\{0,1\}^M$ satisfying $C[\varepsilon] > 0$, ordered such that if $v_j = UB^{[1\ldots(T-1)]}(\gamma, (\hat{\gamma}_{\mathrm{R}} \parallel \varepsilon_j))$, then $v_1, \ldots, v_J$ is a nonincreasing sequence
16. $\Psi \leftarrow 0, \quad \lambda \leftarrow 0, \quad C_{\mathrm{total}} \leftarrow 0, \quad \mathrm{Sum} \leftarrow 0, \quad j \leftarrow 1$
17. While $(j \leq J)$ and $(v_j > 0)$ and $(\Psi + (v_j * C[\varepsilon_j]) \leq 1)$ and $(\lambda < 1)$
18. $C_{\mathrm{total}} \leftarrow C_{\mathrm{total}} + C[\varepsilon_j]$
19. $\mathrm{Sum} \leftarrow \mathrm{Sum} + \mathrm{MatchGroup}(C_{\mathrm{total}})$
20. $j \leftarrow j + 1$
21. If $(j \leq J)$ and $(v_j > 0)$ and $(\Psi + (v_j * C[\varepsilon_j]) > 1)$ and $(\lambda < 1)$
22. $C_{\mathrm{total}} \leftarrow C_{\mathrm{total}} + (1 - \Psi)/v_j$
23. $\mathrm{Sum} \leftarrow \mathrm{Sum} + \mathrm{MatchGroup}(C_{\mathrm{total}})$
24. $UB^{[1\ldots T]}(\gamma, \hat{\gamma}) \leftarrow \mathrm{Sum}$

25. Function MatchGroup (Z)
26. $H \leftarrow \min\{h : 1 \leq h \leq D_A,\ \Phi_h^A \geq Z\}, \quad$ where $A = wt(\hat{\gamma}_{\mathrm{R}})$
27. $\Delta\lambda \leftarrow (\Lambda_H^A - \lambda) - [(\Phi_H^A - Z) * \rho_H^A]$
28. $\Psi \leftarrow \Psi + (v_j * C[\varepsilon_j])$
29. $\lambda \leftarrow \lambda + \Delta\lambda$
30. return $(v_j * \Delta\lambda)$

Figure 3: Pseudocode for inductive step ($T \geq 3$)

patterns for $\Delta\mathbf{z}_R = \Delta\mathbf{w} \oplus \Delta\mathbf{y}_L$ are those $\mu \in \{0,1\}^M$ for which $(\delta, \hat{\gamma}_L) \Rightarrow \mu$ (Definition 8). We think of distributing W among the counters $C[\mu]$ for all such μ, ensuring that we ultimately produce an upper bound by biasing this distribution toward those μ with larger associated values $UB^{[1\ldots(T-1)]}(\gamma, (\hat{\gamma}_R \parallel \mu))$ via the sorting step in Line 8. On the other hand, we guard against "over-biasing" by observing that the maximum possible number of $\Delta\mathbf{w}$ for which $\gamma_{\Delta\mathbf{w}\oplus\Delta\mathbf{y}_L} = \mu$ is $\#[(\delta, \hat{\gamma}_L) \Rightarrow \mu]$ (Definition 9); we use this maximum to cap the contribution to $C[\mu]$ in Line 11. □

Table 1: New upper bounds on MEDP and MELP for Camellia

T	MEDP	MELP
2	2^{-6}	2^{-6}
3	2^{-12}	2^{-12}
4	1.732×2^{-23}	1.313×2^{-22}
5	1.124×2^{-26}	1.243×2^{-25}
6	1.065×2^{-28}	1.161×2^{-27}
7	1.033×2^{-28}	1.378×2^{-28}
8	1.032×2^{-28}	1.952×2^{-29}

5 Application of New Algorithm to Camellia

Prior to this paper, the best provable security bounds for Camellia were based on a result of Aoki and Ohta [1]: if F is invertible (which it is for Camellia), then $MEDP^{[1\ldots T]} \leq (DP_{\max})^2$ and $MELP^{[1\ldots T]} \leq (LP_{\max})^2$ for $T \geq 3$ ($LP_{\max}$ is the natural counterpart of $DP_{\max}$). Since $DP_{\max} = LP_{\max} = 2^{-6}$ for Camellia, we obtain upper bounds of 2^{-12}.

We applied our algorithm to Camellia (minus $\mathrm{FL}/\mathrm{FL}^{-1}$; also, since we are dealing with expected values, the key whitening can be ignored). We ran both the MEDP and MELP versions of the algorithm for $2 \leq T \leq 24$. We present the bounds for $2 \leq T \leq 8$ in Table 1. The MEDP bounds have levelled off at $T = 8$, but the MELP values continue a slow progression downward, reaching 1.947×2^{-31} at $T = 20$, for example. Notice that our algorithm replicates the result of Aoki and Ohta for $T = 3$.

These new bounds represent a significant improvement over the previous value of 2^{-12}. However, since the data complexity of differential/linear cryptanalysis is proportional to the inverse of the MEDP/MELP (with relatively small constants), these bounds are not yet sufficient to establish the provable security of Camellia against differential and linear cryptanalysis. Nonetheless, we believe that our algorithm represents an important step forward in the analysis of Feistel ciphers with SD-based round functions, and we anticipate that further refinements of our approach, as well as the introduction of other new techniques, will soon enable the assertion of provable security for Camellia and related ciphers.

Acknowledgements

The author is grateful to Jiayuan Sui for careful proofreading of an earlier version of this paper, and to the anonymous referees for comments that clarified the final version. This work was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] K. Aoki and K. Ohta, "Strict evaluation of the maximum average of differential probability and the maximum average of linear probability," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences,* vol. E80-A, no. 1, pp. 1–8, 1997.

[2] K. Aoki, T. Ichikawa, M. Kanda, M. Matsui, S. Moriai, J. Nakajima, and T. Tokita, "Camellia: A 128-bit block cipher suitable for multiple platforms—design and analysis," in *Seventh Annual International Workshop on Selected Areas in Cryptography (SAC 2000),* LNCS 2012, pp. 39–56, Springer-Verlag, 2001.

[3] E. Biham, "On Matsui's linear cryptanalysis," in *Advances in Cryptology (EUROCRYPT'94),* LNCS 950, pp. 341–355, Springer-Verlag, 1995.

[4] E. Biham and A. Shamir, "Differential cryptanalysis of DES-like cryptosystems," in *Advances in Cryptology (CRYPTO'90),* LNCS 537, pp. 2–21, Springer-Verlag, 1991.

[5] K. Chun, S. Kim, S. Lee, S.H. Sung, S. Yoon, "Differential and linear cryptanalysis for 2-round SPNs," *Information Processing Letters,* vol. 87, pp. 277–282, 2003.

[6] J. Daemen and V. Rijmen, *The Design of Rijndael: AES—The Advanced Encryption Standard.* Berlin: Springer-Verlag, 2002.

[7] *Data Encryption Standard,* Federal Information Processing Standards Publication 46, U.S. Department of Commerce, National Bureau of Standards, National Technical Information Service, Springfield, Virginia, 1977.

[8] 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, 1975.

[9] S. Hong, S. Lee, J. Lim, J. Sung, and D. Cheon, "Provable security against differential and linear cryptanalysis for the SPN structure," in *Fast Software Encryption (FSE 2000),* LNCS 1978, pp. 273–283, Springer-Verlag, 2001.

[10] L. Keliher, *Linear cryptanalysis of substitution-permutation networks.* Ph.D. Thesis, Queen's University, Kingston, Canada, 2003.

[11] L. Keliher, H. Meijer, and S. Tavares, "New method for upper bounding the maximum average linear hull probability for SPNs," in *Advances in Cryptology (EUROCRYPT 2001),* LNCS 2045, pp. 420–436, Springer-Verlag, 2001.

[12] L. Keliher, H. Meijer, and S. Tavares, "Improving the upper bound on the maximum average linear hull

probability for Rijndael," in *Eighth Annual International Workshop on Selected Areas in Cryptography (SAC 2001),* LNCS 2259, pp. 112–128, Springer-Verlag, 2001.

[13] L. Keliher and J. Sui, "Exact maximum expected differential and linear probability for 2-round Advanced Encryption Standard (AES)," Technical Report, IACR ePrint Archive (`http://eprint.iacr.org`, Paper # 2005/321), 2005.

[14] L. Knudsen, "Practically secure Feistel ciphers," in *Fast Software Encryption,* LNCS 809, pp. 211–221, Springer-Verlag, 1994.

[15] X. Lai, J. Massey, and S. Murphy, "Markov ciphers and differential cryptanalysis," in *Advances in Cryptology (EUROCRYPT'91),* LNCS 547, pp. 17–38, Springer-Verlag, 1991.

[16] M. Matsui, "Linear cryptanalysis method for DES cipher," in *Advances in Cryptology (EUROCRYPT'93),* LNCS 765, pp. 386–397, Springer-Verlag, 1994.

[17] M. Matsui, "On correlation between the order of s-boxes and the strength of DES," in *Advances in Cryptology (EUROCRYPT'94),* LNCS 950, pp. 366–375, Springer-Verlag, 1995.

[18] M. Matsui, "New structure of block ciphers with provable security against differential and linear cryptanalysis," in *Fast Software Encryption (FSE'96),* LNCS 1039, pp.205–218, Springer-Verlag, 1996.

[19] M. Matsui, "New block encryption algorithm MISTY," in *Fast Software Encryption (FSE'97),* LNCS 1267, pp. 54–68, Springer-Verlag, 1997.

[20] K. Nyberg, "Linear approximation of block ciphers," in *Advances in Cryptology (EUROCRYPT'94),* LNCS 950, pp. 439–444, Springer-Verlag, 1995.

[21] K. Nyberg and L. Knudsen, "Provable security against a differential attack," *Journal of Cryptology,* vol. 8, no. 1, pp. 27–37, 1995.

[22] S. Park, S.H. Sung, S. Chee, E-J. Yoon, and J. Lim, "On the security of Rijndael-like structures against differential and linear cryptanalysis," in *Advances in Cryptology (ASIACRYPT 2002),* LNCS 2501, pp. 176–191, Springer-Verlag, 2002.

[23] S. Park, S.H. Sung, S. Lee, J. Lim, "Improving the upper bound on the maximum differential and the maximum linear hull probability for SPN structures and AES," in *Fast Software Encryption (FSE 2003),* LNCS 2887, pp. 247–260, Springer-Verlag, 2003.

[24] F. Sano, K. Ohkuma, H. Shimizu, and S. Kawamura, "On the security of nested SPN cipher against the differential and linear cryptanalysis," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences,* vol. E86-A, no. 1, pp. 37–46, 2003.

[25] B. Schneier and J. Kelsey, "Unbalanced Feistel networks and block-cipher design," in *Fast Software Encryption (FSE'96),* LNCS 1039, pp. 121–144, Springer-Verlag, 1996.

[26] D.R. Stinson, *Cryptography: Theory and Practice, Third Edition.* Boca Raton: Chapman & Hall/CRC, 2006.

[27] S. Vaudenay, "On the security of CS-Cipher," in *Fast Software Encryption (FSE'99),* LNCS 1636, pp. 260–274, Springer-Verlag, 1999.

Liam Keliher is an Assistant Professor in the Department of Mathematics and Computer Science at Mount Allison University. He received a B.Sc. in Mathematics from St. Francis Xavier University in 1993, an M.Sc. in Mathematics from McGill University in 1996, an M.Sc. in Computer Science from Queen's University in 1997, and a Ph.D. in Computer Science from Queen's University in 2003. His current interests include symmetric-key cryptography and theoretical computer science.