New Applications of Differential Bounds of the SDS Structure[⋆]

Jiali Choy and Khoongming Khoo

DSO National Laboratories
20 Science Park Drive, Singapore 118230
Email: cjiali,kkhoongm@dso.org.sg

Abstract. In this paper, we present some new applications of the bounds for the differential probability of a SDS (Substitution-Diffusion-Substitution) structure by Park et al. at FSE 2003. Park et al. have applied their result on the AES cipher which uses the SDS structure based on MDS matrices. We shall apply their result to practical ciphers that use SDS structures based on $\{0,1\}$-matrices of size $n \times n$. These structures are useful because they can be efficiently implemented in hardware. We prove a bound on $\{0,1\}$-matrices to show that they cannot be MDS and are almost-MDS only when $n = 2, 3$, or 4. Thus we have to apply Park's result whenever $\{0,1\}$-matrices where $n \geq 5$ are used because previous results only hold for MDS and almost-MDS diffusion matrices. Based on our bound, we also show that the $\{0,1\}$-matrix used in E2 is almost-optimal among $\{0,1\}$-matrices. Using Park's result, we prove differential bounds for E2 and an MCrypton-like cipher, from which we can deduce their security against boomerang attack and some of its variants. At ICCSA 2006, Khoo and Heng constructed block cipher-based universal hash functions, from which they derived Message Authentication Codes (MACs) which are faster than CBC-MAC. Park's result provides us with the means to obtain a more accurate bound for their universal hash function. With this bound, we can restrict the number of MAC's performed before a change of MAC key is needed.

Keywords SPN, branch number, differential, $\{0,1\}$-matrices, universal hash functions.

1 Introduction

Differential cryptanalysis is one of the most well-known attacks on block ciphers. It exploits differential characteristics, which consist of a sequence of difference patterns in consecutive rounds, with high probability. However, even if the maximal characteristic probability is low, one cannot conclude that the cipher is secure against differential attack as it may not be necessary to fix the values of the input and output differences for intermediate rounds to perform the attack. Instead, one must turn to the concept of a *differential*, which is the set of all differential characteristics with the same initial and terminal difference. To be provably secure against differential cryptanalysis, the differential probability of all differentials must be low enough.

[⋆] This is a corrected version of a paper presented at the ISC 2008 conference. It was claimed in the conference paper that we proved the security of MCrypton against boomerang attack. In this paper, we corrected the claim to say that we prove the security of a variant of MCrypton, which we call MCrypton-x, against boomerang attack. Moreover, some typos were also corrected.

Another motivation for studying differential probability is to analyze a cipher's security against boomerang attacks [21] and its variants such as amplified boomerang attacks [13] and rectangle attacks [3]. In usual differential cryptanalysis, it is not easy to determine the differential probability of a cipher if it has too many rounds. In boomerang-based attacks, a cipher is split into two shorter sub-ciphers from which it is easier to find a differential with high probability for each half. These differentials are then combined to form boomerang attacks based on adaptive chosen plaintext-ciphertexts, or amplified boomerang and rectangle attacks based on chosen plaintexts. If we can prove that a cipher has low differential probability over reduced rounds, then we can prove security against these attacks.

S. Hong et al. [11] analyzed the provable security of the SPN (Substitution-Permutation Network) structure depicted in Figure 1 in Appendix C. This structure is widely used in many block cipher designs as it is highly parallelizable and its security is more easily analyzed. Each round consists of key addition, substitution, and permutation of bits. The diffusion layer is paramount to the whole design as it provides the avalanche effect to ensure good randomization. An SPN cipher with a low branch number associated with its diffusion layer is regarded as weak against differential and linear cryptanalysis. In particular, their paper dealt with the provable security against differential and linear cryptanalysis of an SPN structure with a maximal distance separable (MDS) diffusion layer and an almost-MDS diffusion layer.

In [18], Park et al. extended Hong's results in two directions:

(i) Improvement 1: They took into account the differential and linear probability distribution of the S-boxes involved as compared to Hong et al. who considered the maximal differential and linear probability of the S-box. This enabled them to derive differential and linear probability bounds which are better (lower) than previously known bounds in [8] and [18].

(ii) Improvement 2: They derived the differential and linear probability of the SDS (Substitution-Diffusion-Substitution) structure for diffusion layers with any specified differential or linear branch number respectively, as opposed to Hong et al.'s results which are only applicable for MDS and almost-MDS diffusion layers.

They then went on to prove differential and linear bounds for the AES cipher which are better than the known bounds by Rijmen and Daemen in [8]. This demonstrates the advantage of their result for Improvement 1. However, the second advantage of their analysis is that we can derive the differential and linear probability of SDS structures where the diffusion layer is not MDS or almost-MDS (Improvement 2 above) . We shall demonstrate the practicability of their results by applying these results to SDS structures based on diffusion layers which are $\{0, 1\}$-matrices. They are widely used in ciphers like Camellia and E2 [1, 12] because they require less gates when implemented in hardware and will be well-suited to constrained environments such as RFID tags.

First, we prove an upper bound for the branch numbers of $\{0, 1\}$-matrices of size $n \times n$. This will provide us with an idea of $\{0, 1\}$-matrices which are optimal or almost-optimal. We show that $\{0, 1\}$-matrices are never MDS and they are almost-MDS only when $n = 2, 3$, or 4. Thus we need to apply Park's result whenever $n \geq 5$. The E2 cipher uses $\{0, 1\}$-matrices. With our bounds, we can show that the diffusion mapping in E2 is almost-optimal among $\{0, 1\}$-matrices.

Second, we apply Park's result [18] to derive a general formula for the differential probability of the SDS structure which uses an affine transform of the inverse S-box over $GF(2^m)$ and a diffusion mapping with any arbitrary branch number. This will allow us to prove the differential probability of E2 [12] and an MCrypton-like cipher, which we call MCrypton-x, formed by replacing the linear diffusion layer in MCrypton with a $\{0,1\}$ matrix [15]. Furthermore, we are also able to prove the resistance of these ciphers against boomerang attack [21] and some of its variants [3, 13].

Third, we improve on an upper bound for a universal hash construction by Khoo and Heng [14]. In their paper, the authors showed that we can implement a block cipher-based universal hash function which uses reduced rounds and is parallelizable. This results in a universal hash function-based message authentication code (MAC) which is faster than CBC-MAC. However, we show that the upper bound in [14] is approximate and is, in fact, higher than the actual bound. By applying on Park's result [18], we can give a more accurate bound for the MAC. Based on this bound, we can restrict the number of authentications performed before a change of MAC-key is needed.

2 Definitions

A measure of the efficiency of block ciphers against differential cryptanalysis is to have low maximal differential.

Definition 1 *The* maximal differential *of a function $f : GF(2)^w \rightarrow GF(2)^w$ is defined as*

$$\Delta_f =^{max}_{\Delta x \neq 0, \Delta y} \mid \{x \in GF(2)^w | f(x) \oplus f(x \oplus \Delta x) = \Delta y\} \mid$$

In the subsequent sections, we consider an SPN structure with an mn-bit round function. Let each S-box S_i be an $m \times m$ bijective function

$$S_i : GF(2)^m \rightarrow GF(2)^m \qquad (1 \leq i \leq n).$$

Also we assume that the round keys, which are XORed with the input data at each round, are independent and uniformly random.

Definition 2 *For any given $\Delta x, \Delta y \in GF(2)^m$, the* differential probability *of each S_i is defined as*

$$DP^{S_i}(\Delta x \rightarrow \Delta y) = \frac{\#\{x \in GF(2)^m | S_i(x) \oplus S_i(x \oplus \Delta x) = \Delta y\}}{2^m},$$

where we consider Δx to be the input difference and Δy the output difference. The maximal differential probability *of S_i is defined as*

$$DP((S_i)_{max}) =^{max}_{\Delta x \neq 0, \Delta y} DP^{S_i}(\Delta x \rightarrow \Delta y).$$

The maximal values *of $DP((S_i)_{max})$ for $1 \leq i \leq n$ is defined as*

$$p =^{max}_{1 \leq i \leq n} (DP(S_i)_{max}).$$

S_i is strong against differential cryptanalysis if $DP((S_i)_{max})$ is low enough, while a substitution layer is strong if p is low enough. However, it is important to note that a strong S-box and a strong substitution layer do not ensure a secure SPN structure against differential attacks. To evaluate provable security, one has to take the diffusion layer into account as well. The latter is an invertible linear mapping, the purpose of which is to provide an avalanche effect, both in the context of differential and linear approximations. In the differential context, this means that small input changes should cause large output changes, and conversely, to produce a small output change, a large input change should be necessary.

A *differentially active* S-box is an S-box given a non-zero input difference. Differentially inactive S-boxes with zero input difference always have zero output difference with probability 1. Due to the independent round key assumption, the key addition layer in Figure 1 has no effect on the number of active S-boxes.

Definition 3 *Let $x = (x_1, \cdots, x_n)^t \in GF(2^m)^n$. The* Hamming weight *of x is defined as*

$$wt(x) = \#\{i | x_i \neq 0\},$$

which is the number of non-zero m-bit characters in x.

Now define a SDS (Substitution-Diffusion-Substitution) function as shown in Figure 2 in Appendix C. Let the diffusion layer of the SDS function be denoted by D, its input difference by $\Delta x = x \oplus x^*$, its output difference by $\Delta y = y \oplus y^* = D(x) \oplus D(x^*)$. If D is linear, we have $\Delta y = D(\Delta x)$. The minimum number of differentially active S-boxes of the SDS function is given by the branch number of the diffusion layer.

Definition 4 *The* branch number *of a diffusion layer D is defined as:*

$$Br(D) = \min_{v \neq 0} \{wt(v) + wt(D(v))\} \tag{1}$$

If we want to find the number of active S-boxes in two consecutive rounds of the SPN structure, we may disregard the two key addition layers since they have no influence on the number. Consequently, we only need to consider the SDS function. Therefore, $Br(D)$ gives a measure of the worst case diffusion: it is a lower bound for the number of active S-boxes in two consecutive rounds of a differential characteristic approximation. Since a cryptanalyst will always exploit the worst case, this is a good measure of the diffusion property.

Definition 5 *A diffusion layer is* maximal distance separable (MDS) *if $Br(D)$ is $n+1$; it is called* almost-MDS *if $Br(D)$ is equal to n.*

Proposition 1 *([11, Theorem 1,Theorem 3]) Assume that the round keys, which are XORed to the input data at each round, are independent and uniformly random. The probability of each differential of the SDS structure (and consequently, the SPN structure)*

(i) is bounded by p^n if D is MDS, i.e. if $Br(D) = n+1$;
(ii) is bounded by p^{n-1} if D is almost-MDS, i.e. if $Br(D) = n$.

However, the above lemma has been improved by Park et al. to apply to SDS structures where the diffusion layer need not be MDS or almost-MDS.

Proposition 2 *([18, Theorem 1]) Assume that the round keys, which are XORed to the input data at each round, are independent and uniformly random. If $Br(D) = k$, the probability of each differential of the SDS structure (and consequently, the SPN structure) is bounded by:*

$$\max\left(\max_{1\le i\le n} \max_{1\le u\le 2^m-1} \sum_{j=1}^{2^m-1} DP^{S_i}(u \to j)^k, \max_{1\le i\le n} \max_{1\le u\le 2^m-1} \sum_{j=1}^{2^m-1} DP^{S_i}(j \to u)^k \right)$$

As a corollary, Park et al. obtained the following result which can be viewed as a direct generalization of Hong et al.'s result.

Proposition 3 *([18, Corollary 1]) Assume that the round keys, which are XORed to the input data at each round, are independent and uniformly random. The probability of each differential of the SDS structure (and consequently, the SPN structure) is bounded by p^{k-1} if $Br(D) = k$.*

3 Branch Number of $\{0,1\}$-Matrices

We shall look at the differential probability of the SDS structure where the diffusion layer is a matrix with entries 0 or 1, which we call $\{0,1\}$-matrices. The reason we study $\{0,1\}$-matrices is because they are faster to compute than MDS transforms which are used in many block ciphers like Rijndael, Square and Shark [8, 10, 19]. Another reason is that in hardware implementation, they will take up less space and thus allow for more compact implementation.

In this section, we consider a closely related problem: The study of the branch number of such matrices. The proofs of results in this section can be found in Appendix A.1, A.2, and A.3.

Theorem 1 *Let $A : GF(2^m)^n \to GF(2^m)^n$ be an $n \times n$ $\{0,1\}$-matrix over $GF(2^m)$. Then the branch number of A is at most $\frac{2n+4}{3}$.*

By studying the upper bound of Theorem 1, it is easy to deduce the following Corollary.

Corollary 1 *Let $A : GF(2^m)^n \to GF(2^m)^n$ be an $n \times n$ $\{0,1\}$-matrix over $GF(2^m)$. Then A is not a MDS matrix and it can be an almost-MDS matrix only when $n = 2, 3,$ or 4.*

In Table 1 in Appendix B, we list the upper bounds for the branch number of $\{0,1\}$-matrices for different n. We see from Corollary 1 that when we want to deduce the true differential probability of SDS structures, where the diffusion layer is represented by a $\{0,1\}$-matrix, we can only apply the known results (on almost-MDS matrices) from [11] for $n = 2, 3,$ or 4. For $n \geq 5$, we have to apply Theorem 1 from [18].

3.1 Some $\{0,1\}$-Matrices with Optimal Branch Numbers

Based on Theorem 1, we give the following definition.

Definition 6 *A $\{0,1\}$-matrix A of size $n \times n$ is called* optimal *(w.r.t. Theorem 1) if its branch number is $\lfloor \frac{2n+4}{3} \rfloor$. It is called* almost-optimal *(w.r.t. Theorem 1) if its branch number is $\lfloor \frac{2n+4}{3} \rfloor - 1$.*

We shall look at some $\{0,1\}$-matrices with optimal or almost optimal branch numbers. The first matrix we shall study has 0 on the diagonal and 1 elsewhere.

Proposition 4 *(i) Consider the following $n \times n$ matrix A, $n \geq 2$, which maps from $GF(2^m)^n \to GF(2^m)^n$.*

$$A = \begin{pmatrix} 0\,1\,1 \dots 1 \\ 1\,0\,1 \dots 1 \\ 1\,1\,0 \dots 1 \\ \vdots\,\vdots\,\vdots\,\ddots\,\vdots \\ 1\,1\,1 \dots 0 \end{pmatrix}. \tag{2}$$

The branch number of A is $\min(n, 4)$.

(ii) *More generally, a $n \times n$ matrix $A : GF(2^m)^n \rightarrow GF(2^m)^n$ where each row and each column has 1 occurrence of 0 and $n-1$ occurrences of 1, has branch number $\min(n, 4)$.*

(iii) *When $n = 4$, the matrices in part (ii) are the only 4×4 $\{0,1\}$-matrices over $GF(2^m)$ with optimal branch number 4.*

By referring to Table 1, we see that the matrices in Proposition 4 part (ii) are optimal among $\{0,1\}$-matrices when $n = 2, 3, 4, 5$ and almost-optimal when $n = 6$.

The following $\{0,1\}$-matrix acting on 8 bytes is used in the E2 cipher:

$$A = \begin{pmatrix} 0\,1\,1\,1\,1\,1\,1\,0 \\ 1\,0\,1\,1\,0\,1\,1\,1 \\ 1\,1\,0\,1\,1\,0\,1\,1 \\ 1\,1\,1\,0\,1\,1\,0\,1 \\ 1\,1\,0\,1\,1\,1\,0\,0 \\ 1\,1\,1\,0\,0\,1\,1\,0 \\ 0\,1\,1\,1\,0\,0\,1\,1 \\ 1\,0\,1\,1\,1\,0\,0\,1 \end{pmatrix}. \tag{3}$$

It is known that this diffusion layer has branch number 5 (see [12]). By referring to Table 1, it is an almost-optimal $\{0,1\}$-matrix.

4 Differential Bounds and Security Against Boomerang Attacks for Ciphers based on $\{0,1\}$-Matrices

We shall now study SDS structures where the S-boxes are affine transforms of the inversion function over $GF(2^m)$ defined by: $S(x) = x^{-1}$ when $x \neq 0$ and $S(0) = 0$.

In order to apply Proposition 2, we need to find the difference distribution $DP^S(u \rightarrow j)$ where $u \in GF(2^m)$ is fixed and j varies over all of $GF(2^m)$, i.e. the difference distribution for each row of the difference table. Likewise, we need to find the difference distribution of the columns. It is well-known that the difference distribution table of $S^{-1}(x)$ is the transpose of the original difference distribution table. Since $S^{-1}(x) = S(x)$ for the inversion function, the difference distribution of the columns will be the same as that for the rows.

From the proof of [17, Proposition 6], we see that $S(x) + S(x+u) = j$, where u is fixed and j varies over $GF(2^m)$, can be reduced to a quadratic equation and it has 0 or 2 solutions for x in general. Only when m is even, we have 4 solutions for one value of j. Thus the inversion mapping has maximal differential 4 when m is even and 2 when m is odd. It is also well-known that the sum of each row of the difference distribution table should be 2^m. From this, we can easily get the difference distribution for each row of the inversion mapping shown in Table 2 in Appendix B.

Theorem 2 *Consider a SDS structure where the S-boxes are affine transforms of the inversion function over $GF(2^m)$ defined by: $S(x) = x^{-1}$ when $x \neq 0$ and $S(0) = 0$, and the diffusion mapping has branch number $Br(D) = k$. Then the differential probability of the SDS structure is bounded by:*

(i) $2^{(1-m)(k-1)} - 2^{(1-m)k+1} + 2^{(2-m)k}$ *when m is even;*

(ii) $2^{(1-m)(k-1)}$ *when* m *is odd.*

The proof of Theorem 2 can be found in Appendix A.4.

Remark 1. Note that when n is even, the upper bound for the differential probability from Proposition 3 is $(4/2^m)^{k-1} = 2^{(2-m)(k-1)}$. By simplifying the inequality:

$$2^{(2-m)k} - 2^{(1-m)k+1} < 2^{m-2}(2^{(2-m)k} - 2^{(1-m)k+1}),$$

we get the inequality:

$$2^{(1-m)(k-1)} - 2^{(1-m)k+1} + 2^{(2-m)k} < 2^{(2-m)(k-1)}.$$

This shows that the bound in Theorem 2 is better (lower) than that obtained from Proposition 3 when m is even. On the other hand, the upper bound of Theorem 2 and Proposition 3 is the same when m is odd.

4.1 Application on the E2 Cipher

The E2 cipher is a 12-round Feistel block cipher with a block size of 128 bits and a key size of 128, 192, or 256 bits [12]. There is also an initial transform (IT) that XORs a subkey, multiplies by a subkey and performs a byte permutation BP; and a final transformation (FT) that performs an inverse byte permutation BP^{-1}, divides by a subkey and XORs a subkey.

The nonlinear F-function in each Feistel round maps 64-bit to 64-bit with the help of two 64-bit subkeys $K^{(1)}$ and $K^{(2)}$. It is defined by:

$$F(x, K^{(1)}, K^{(2)}) = L(S(D(S(x \oplus K^{(1)})) \oplus K^{(2)})),$$

where $S : GF(2^8)^8 \to GF(2^8)^8$ is defined as:

$$S(x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7) = (s(x_0), s(x_1), s(x_2), s(x_3), s(x_4), s(x_5), s(x_6), s(x_7)).$$

Each $s : GF(2^8) \to GF(2^8)$ is affinely equivalent to the inversion function on $GF(2^8)$ and D is a 8-byte linear transform defined by the matrix in equation (3), which is known to have branch number 5 [1, 12]. The final linear transform L is a 1-byte cyclic rotation over 8 bytes.

Before we go on to the analysis, we shall need the following standard result on the differential probability of Feistel ciphers.

Proposition 5 *[22] Consider a* 3*-round Feistel cipher with* $2w$*-bit block size and nonlinear function* $F : GF(2)^w \to GF(2)^w$*. If the maximal differential probability of* $F(x)$ *is* p*, then the maximal differential probability of the* 3*-round cipher is* p^2*.*

In [12], the authors derived the differential characteristic probability of E2 and concluded that it has practical security against differential cryptanalysis. Here we can give upper bounds for its differential probability in Theorem 3. The proof of the result can be found in Appendix A.5.

Theorem 3 *The differential probability of* 3 *rounds of the E2 cipher is at most* $2^{-55.39}$.

Thus to defend E2 against a stronger form of differential attack using true differentials, we can recommend a change of key after every 2^{55} encryptions.

Security of E2 against Boomerang Attack There is also a stronger form of differential attack called boomerang attack [21]. It splits $R-1$ rounds ($R-2$ for Feistel ciphers) of an R-round block cipher into 2 shorter ciphers such that the differential probability of each part is known to be large, say with probability p for the differential $\alpha \to \beta$ for the first part and probability q for the differential $\gamma \to \delta$ for the second part. The distinguisher is the following boomerang process:

(i) Ask for the encryption of a pair of plaintexts (P_1, P_2) such that $P_1 \oplus P_2 = \alpha$ and denote the corresponding ciphertexts by (C_1, C_2).
(ii) Calculate $C_3 = C_1 \oplus \delta$ and $C_4 = C_2 \oplus \delta$, and ask for the decryption of the pair (C_3, C_4). Denote the corresponding plaintexts by (P_3, P_4).
(iii) Check whether $P_3 \oplus P_4 = \alpha$.

For a random permutation, the probability that the last condition is satisfied is $2^{-blocksize}$. The probability that a quartet of plaintexts and ciphertexts satisfies the boomerang conditions is $(pq)^2$. Therefore, we have a distinguisher which distinguishes between the cipher being attacked and a random cipher if $(pq)^2 > 2^{-blocksize}$.

For the E2 cipher, 8 rounds of the cipher already has maximal differential characteristic probability $((2^{-6})^5)^2 \times ((2^{-6})^5)^2 \times (2^{-6})^5 = 2^{-150}$ which is less than 2^{-128}. Here we use the easily proven fact that there are at least 2 active F-functions for every 3 rounds and at least 1 active F-function every 2 rounds. Thus it is unlikely that an adversary can find a good differential over 8 rounds and any good differential is likely to involve 7 or less rounds. Thus when the adversary splits $12-2=10$ rounds into two sub-ciphers, they will each contain at least 3 rounds. By Theorem 3, we see that the differential probabilities p, q of the 2 sub-ciphers are at most $2^{-55.39}$. Thus $(pq)^2 \leq 2^{-221.57} < 2^{-128}$ and E2 is secure against boomerang attack.

Remark 2. We have used the assumption that if the differential characteristic probability of R' rounds of a cipher is less than $2^{-blocksize}$, then it is not likely that a good differential over R' rounds can be found. This is in line with the common approach of practical provable security against differential cryptanalysis employed in the proofs of security of many ciphers like Camellia [1], AES [8], SQUARE [10], E2 [12], MCrypton [15] and SHARK [19]. Thus if our assumption is not true, then the approach is wrong because although we can prove that the differential characteristic probability is less than $2^{-blocksize}$, we can still find a differential with high probability to launch differential cryptanalysis. We shall also make the same assumption in the analysis of MCrypton against boomerang attack in Section 4.2.

Security of E2 against Variants of Boomerang Attack Since the boomerang attack requires adaptively chosen plaintexts and ciphertexts, many of the techniques that were developed for using distinguishers in key recovery attacks cannot be applied. As an alternative, the amplified boomerang attack [13] encrypts many plaintext pairs with the same input difference α and looks for right quartets which satisfy the requirements of the boomerang process. Out of x plaintext pairs, the number of right quartets is expected to be $x^2 \cdot 2^{(-blocksize+1)} p^2 q^2$ [4]. For a random permutation, the expected number of right quartets is $x^2 \cdot 2^{-2 \cdot blocksize}$. Therefore, if $(pq)^2 > 2^{-blocksize+1}$, then we would count more quartets than random noise. For protection against the amplified boomerang attack, we would want to show that $(pq)^2 < 2^{-blocksize+1}$. Following the above argument, we know that $(pq)^2 \leq 2^{-221.57} < 2^{-128+1} = 2^{-127}$. Thus, E2 is also secure against the amplified boomerang attack.

Another variant of the boomerang attack is: suppose the initial and terminal differences, α and δ, are fixed while the intermediate differences, β and γ, are allowed to vary over index sets $\Lambda, \Omega \subseteq GF(2)^{blocksize}$ respectively, i.e. the attacker tries to find several differential paths with the same initial and terminal differences of high probability. Then he needs $\sum_{\beta \in \Lambda} \Pr^2(\alpha \to \beta) \sum_{\gamma \in \Omega} \Pr^2(\gamma \to \delta) > 2^{-blocksize}$ to get a good distinguisher for the attack to succeed. For E2, we know that for any $\alpha, \beta, \gamma, \delta$, $\Pr^2(\alpha \to \beta) \cdot \Pr^2(\gamma \to \delta) \leq 2^{-221.570}$. Thus the attacker must obtain at least $\lceil 2^{221.570}/2^{128} \rceil + 1 = \lceil 2^{93.570} \rceil + 1 > 10^{28}$ high probability differential paths for the attack to work. It is highly improbable that the attacker will be able to find such a large number of useful differential paths and hence, this attack is unlikely to succeed.

4.2 Application on an MCrypton-like Cipher

The MCrypton cipher is a 12-round block cipher with a block size of 64 bits and a key size of 64, 96, or 128 bits [15]. Its structure is similar to that of the AES cipher [8] but it uses lightweight components suited for RFID applications. The plaintext is first written as a 4 by 4 array of nibbles and XORed to a 64-bit subkey. It then goes through 12 rounds of transformation where every round is composed of four operations:

$$\rho_K = \sigma_K \circ \tau \circ \pi \circ \lambda,$$

The substitution operation λ transforms each nibble by an affine transform of the inversion map on $GF(2^4)$. The linear map π multiplies each column of the array by a matrix. The linear map τ transposes the 4 by 4 array. Finally, a 64-bit subkey is XORed with the array. In this section, we examine the security of an MCrypton-like cipher, MCrypton-x, which modifies the original MCrypton by replacing the matrix in π by one of the 4×4 $\{0,1\}$-matrices from Proposition 4 part (ii). Based on our results on the differential probability of MCrypton-x, we show it is secure against boomerang attack and its variants. The proof of the following Theorem 4 is given in Appendix A.6.

Theorem 4 *The differential probability of* 4 *rounds of the MCrypton-x cipher is at most* $2^{-22.62}$.

Thus, to defend MCrypton-x against a stronger form of differential attack using true differentials, we can recommend a change of key after every 2^{22} encryptions.

Security of MCrypton-x against Boomerang Attack For MCrypton-x, 8 rounds of the cipher should have differential characteristic probability at most $(2^{-2})^{32} = 2^{-64}$ if we follow the approach of AES. However, due to the careful choice of S-boxes and diffusion mappings of MCrypton-x in [15, Section 3], the authors proved that the differential characteristic probability is at most $2^{-96} < 2^{-64}$. Therefore it is unlikely that an adversary can find a good differential over 8 rounds and any good differential is likely to involve 7 or less rounds. Thus when the adversary splits $12 - 1 = 11$ rounds into two sub-ciphers, they will each contain at least 4 rounds. By Theorem 4, we see that the differential probabilities p, q of the 2 sub-ciphers are at most $2^{-22.62}$. Thus $(pq)^2 \leq 2^{-90.49} < 2^{-64}$ and MCrypton-x is secure against boomerang attack.

Security of MCrypton-x against Variants of Boomerang Attack Since $(pq)^2 \leq 2^{-90.49} < 2^{-64+1} = 2^{-63}$, MCrypton-x is also secure against the amplified boomerang attack. For the second variant of the boomerang attack where only the initial and terminal differences are fixed while the intermediate differences are allowed to vary, at least $\lceil 2^{90.49}/2^{64} \rceil + 1 = \lceil 2^{26.49} \rceil + 1 \approx 10^7$ high probability differential paths are required. Again, it is unlikely that the attacker will be able to find that many useful differential paths.

5 On Differential Probability, Universal Hash Functions and Message Authentication Codes

Let $H : [GF(2)^w]^* \rightarrow GF(2)^w$ be a family of functions. The probabilities below, denoted by $\Pr_{h\in H}[\cdot]$, are taken over the choice of $h \in H$.

Definition 7 *H is a Δ-*universal family of hash functions *if for all $x, y \in [GF(2)^w]^*$ with $x \neq y$ and all $a \in GF(2)^w$, $\Pr_{h\in H}[h(x) - h(y) = a] = 2^{-w}$.*

Definition 8 *H is an* ϵ-almost-Δ-universal *(ϵ-$A\Delta U$)* family of hash functions *if* $\Pr_{h\in H}[h(x) - h(y) = a] \leq \epsilon$.

It is well-known that $\epsilon \geq 2^{-w}$ (see [20]). Universal hash functions can be used to construct *Message Authentication Codes* (MAC) via the Wegman-Carter approach [23]. The MAC tag is given by the value $h(msg)$ exclusive-or-ed with the one-time-pad OTP as follows:

$$MAC_{h,OTP}(msg) = h(msg) \oplus OTP$$

where h is a randomly chosen hash function from the family H and OTP is a random one-time-pad. The communicating parties must share the secret key pair (h, OTP) in this scenario. However, it is not practical to generate one-time-pads long enough to handle long messages. In [6], Brassard proposed that we substitute the one-time-pad encryption with a computationally secure encryption scheme, for example, AES.

In [14], the authors constructed the following universal hash functions based on functions with low maximal differential.

Proposition 6 *([14, Theorem 1]) Let $f : GF(2)^w \rightarrow GF(2)^w$ have maximal differential Δ_f. Let $x = (x_1, \ldots, x_r)$ and $msg = (msg_1, \ldots, msg_r)$ where $x_i, msg_i \in GF(2)^w$. The* function sum hash (FSH) family of functions *defined by $FSH = \{h_x : [GF(2)^w]^r \rightarrow GF(2)^w | x \in [GF(2)^w]^r\}$ where $h_x(msg) = \sum_{i=1}^r f(msg_i + x_i)$ is an ϵ-$A\Delta$ universal family of hash functions with $\epsilon \leq \frac{\Delta_f}{2^w}$.*

By applying Proposition 6 to Proposition 3 and Theorem 2, we get the following result.

Proposition 7 *Let $h_x : GF(2)^{mnr} \rightarrow GF(2)^{mn}$ be a FSH based on a SDS structure defined by*

$$h_x(msg) = \sum_{i=1}^{r} (S(D(S(msg_i + x_i)))),$$

where $x = (x_1, \ldots, x_r)$, $msg = (msg_1, \ldots, msg_r)$, $D : GF(2^m)^n \rightarrow GF(2^m)^n$ has branch number k and $S(\cdot)$ is a layer of n m-bit S-boxes. Then $h_x(msg)$ is an ϵ-$A\Delta$ universal family of hash functions with:

(i) $\epsilon \leq p^{k-1}$ *if the maximal differential probability of the S-boxes is p;*
(ii) $\epsilon \leq 2^{(1-m)(k-1)} - 2^{(1-m)k+1} + 2^{(2-m)k}$ *if the inversion S-box on* $GF(2^m)$ *is used where* m *is even;*
(iii) $\epsilon \leq 2^{(1-m)(k-1)}$ *if the inversion S-box on* $GF(2^m)$ *is used where* m *is odd.*

Proposition 7 can be viewed as an improvement over [14, Theorem 6] where the authors only approximated the upper bound of the universal hash function by the differential characteristic probability p^k, which is too low. By using the above two-round structure, we may be able to use it to compute a message authentication code (MAC) based on:

$$MAC_{K,x}(msg) = Enc_K(\sum_{i=1}^{r} S(D(S(msg_i + x_i)))),$$

where $Enc(\cdot)$ is a R-round block cipher where each round XORs a subkey, applies the S-box layer S and then, the diffusion layer D. In this way, we get a MAC which is $R/2$ times faster than encryption and is easily parallelizable. Based on the upper bound on ϵ, we can restrict the number of MAC computations to significantly less than $\sqrt{\epsilon^{-1}}$ when a change of MAC key is needed, so as to protect against forgery attacks. In this case, the MAC key consists of the encryption key K and a sequence of secret values $x_i \in GF(2)^{mn}$ which are to be shared between the sender and receiver. It may not be feasible to generate and share long strings of secret values $x_1, x_2, x_3, \dots$ In [14], it is suggested that a single secret value $x_1 \in GF(2)^{mn}$ be chosen to seed a LFSR of length mn bits to produce $x_2, x_3, \dots$

5.1 Applications to Ciphers Used in Practice

By applying Proposition 6 where f is taken to be the 3-round E2 cipher excluding the key addition layer in each F function, we can construct an E2-based MAC as follows:

$$MAC_{K,x}(msg) = \text{E2}_K(\sum_{i=1}^{r} \text{3-Round-E2}(msg_i + x_i)),$$

where $x = (x_1, \dots, x_r)$ and $msg = (msg_1, \dots, msg_r)$. From Theorem 3, we see that the above MAC is based on a universal hash function with collision probability at most $\epsilon \leq 2^{-55.39}$. Thus we require a change of MAC key before $\sqrt{\epsilon^{-1}} \approx 2^{27.7}$ MAC computations. Moreover, the above MAC is $12/3 = 4$ times faster than CBC-MAC based on 12-round E2.

Similarly, by applying Proposition 6 where f is taken to be the 4-round MCrypton-x cipher excluding the key addition layer in each round, we can also construct a MCrypton-x-based MAC as follows:

$$MAC_{K,x}(msg) = \text{MCrypton-x}_K(\sum_{i=1}^{r} \text{4-Round-MCrypton-x}(msg_i + x_i)),$$

From Theorem 4, we see that the above MAC is based on a universal hash function with collision probability at most $\epsilon \leq 2^{-22.62}$. Thus we require a change of MAC key before $\sqrt{\epsilon^{-1}} \approx 2^{11.3}$ MAC computations. Moreover, the above MAC is $12/4 = 3$ times faster than CBC-MAC based on 12-round MCrypton-x.

Remark 3. Note that our universal-hash based MAC is parallelizable while CBC-MAC is not. So if we can have N copies of the E2 cipher, for example, then our MAC will be $4N$ times faster than E2-based CBC-MAC.

Remark 4. Note that we could have included the subkeys in the reduced-round ciphers used to define the universal hash function. This will not affect the probability bound of the universal hash function and heuristically, do give a more secure MAC as the secret key K is not just used in the final encryption, but also in the compression of every message block msg_i.

6 Conclusion

In this paper, we proved an upper bound for the branch numbers of $\{0,1\}$-matrices. Furthermore, we showed that they are never MDS and are almost-MDS only when $n = 2, 3$, or 4. The cipher E2 was found to employ a $\{0,1\}$-matrix which is almost-optimal. We also used Park's result on the differential probability of the SDS structure in [18] to obtain a general formula for the differential probability of such a structure which uses an affine transform of the inverse S-box in its S-box layer. With this formula, we were able to prove the differential probability of E2 and MCrypton-x, as well as their resistance against boomerang attack and its variants. Our results provide new direction in the analysis of block ciphers based on $\{0,1\}$-matrices or non-MDS matrices and affine transforms of inverse S-boxes against differential-based attacks. Finally, we improved on the upper bound for a FSH based on a SDS structure. Two MACs based on E2 and MCrypton-x, faster than CBC-MAC and with provable collision probability, were also proposed.

References

1. 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", *Selected Areas in Cryptography 2000*, LNCS 2012, pp. 39-56, Springer-Verlag, 2000.
2. P.S.L.M. Barreto and V. Rijmen. "The WHIRLPOOL Hashing Function". Primitive submitted to NESSIE, September 2000, revised on May 2003, http://paginas.terra.com.br/informatica/paulobarreto/WhirlpoolPage.html.
3. E. Biham, O. Dunkelman, and N. Keller, "The Rectangle Attack - Rectangling the Serpent", *Eurocrypt 2001*, LNCS 2045, pp. 340-357, Springer-Verlag, 2001.
4. E. Biham, O. Dunkelman, and N. Keller, "Related-Key Boomerang and Rectangle Attack", *Eurocrypt 2005*, LNCS 3494, pp. 507-525, 2005.
5. E. Biham and A. Shamir, "Differential Cryptanalysis of DES-like Cryptosystems", *Journal of Cryptology*, vol. 4, 1991.
6. G. Brassard. "On computationally secure authentication tags requiring short secret shared keys", *Crypto 1983*, pp. 79-86, Springer-Verlag, 1983.
7. J. Daemen and V. Rijmen, "The Wide Trail Strategy", 8^{th} *IMA International Conference 2001*, LNCS 2260, pp. 222-238, Springer, 2001.
8. J. Daemen and V. Rijmen, *The Design of Rijndael: AES, The Advanced Encryption Standard*, Springer, 2002.
9. J. Daemen, R. Govaerts, and J. Vandewalle, "Correlation Matrices", *Fast Software Encryption 1994*, LNCS 1008, pp. 275-285, Springer-Verlag, 1995.
10. J. Daemen, L. Knudsen, and V. Rijmen, "The Block Cipher Square", *Fast Software Encryption 1997*, LNCS 1267, pp. 149-165, Springer-Verlag, 1997.

11. S. Hong, S. Lee, J. Lim, J. Sung, D. Cheong, and I. Cho, “Provable Security against Differential and Linear Cryptanalysis for the SPN Structure”, *Fast Software Encryption 2000*, LNCS 1978, pp. 273-283, Springer-Verlag, 2001.
12. M. Kanda, S. Moriai, K. Aoki, H. Ueda, Y. Takashima, K. Ohta, and T. Matsumoto, “E2 - A New 128-bit Block Cipher”, *IEICE Transactions Fundamentals - Special Section on Cryptography and Information Security*, vol. E83-A no. 1, pp. 48-59, 2000.
13. J. Kelsey, T. Kohno, and B. Schneier, “Amplified Boomerang Attacks Against Reduced-Round MARS and Serpent”, *proceedings of Fast Software Encryption 7*, LNCS 1978, pp. 75-93, Springer-Verlag, 2000.
14. K. Khoo and S.H. Heng, “New Constructions of Universal Hash Functions based on Function Sum”, *ICCSA 2006*, LNCS 3982, pp. 416-425, Springer-Verlag, 2006.
15. C.H. Lim and T. Korkishko, “mCrypton - A Lightweight Block Cipher for Security of Low-Cost RFID Tags and Sensors”, *WISA 2005*, LNCS 3786, pp. 243-258, Springer-Verlag, 2006.
16. M. Matsui, “Linear Cryptanalysis Method for DES Cipher”, *Eurocrypt 1993*, LNCS 765, pp. 386-397, Springer-Verlag, 1994.
17. K. Nyberg, “Differentially Uniform Mappings for Cryptography”, *Eurocrypt 1993*, LNCS 765, pp. 55-64, Springer-Verlag, 1994.
18. S. Park, S.H. Sang, S. Lee, and J. Lim, “Improving the Upper Bound on the Maximum Differential and the Maximum Linear Hull Probability for SPN Structures and AES”, *Fast Software Encryption 2003*, LNCS 2887, pp. 247-260, Springer-Verlag, 2003.
19. V. Rijmen, J. Daemen, B. Preneel, A. Bosselars, and E.D. Win, “The Cipher Shark”, *Fast Software Encryption 1996*, LNCS 1039, pp. 99-111, Springer-Verlag, 1996.
20. D. R. Stinson, “On the connections between universal hashing, combinatorial designs and error-correcting codes”, *Congressus Numerantium 114*, pp. 7-27, 1996.
21. D. Wagner, “The Boomerang Attack”, *Fast Software Encryption 1999*, LNCS 1636, pp. 156-170, Springer-Verlag, 1999.
22. J. Wallen, “Design Principles of the KASUMI Block Cipher”. http://citeseer.ist.psu.edu/wallen00design.html.
23. M. N. Wegman and J. L. Carter, “New hash functions and their use in authentication and set equality”, *Journal of Computer and System Sciences*, vol. 22, no. 3, pp. 265–279, 1981.

A Proofs

A.1 Proof of Theorem 1

Proof. Suppose A has branch number k, where $k > (2n+4)/3$. Let v be a vector of weight 1. Then

$$wt(Av) \geq k - 1 > (2n+1)/3.$$

On the other hand, if v is a vector of weight 2, then

$$wt(Av) \geq k - 2 > (2n-2)/3.$$

Thus the XOR-sum of any two columns of A has weight greater than $(2n-2)/3$.

These two facts contradict each other: if v_1 and v_2 have weight greater $(2n+1)/3$, then each of v_1 and v_2 has less than $(n-1)/3$ zeroes. For a bit of $v_1 \oplus v_2$ to be 1, exactly one of the corresponding bits of v_1 or v_2 must be 0. Thus $v_1 \oplus v_2$ has less than $2(n-1)/3 = (2n-2)/3$ ones, which is a contradiction. □

A.2 Proof of Corollary 1

Proof. For $n \geq 2$, it is easy to see that the upper bound from Theorem 1: $\frac{2n+4}{3}$ is always less than $n+1$, the MDS bound. It is also easy to show that the upper bound $\frac{2n+4}{3}$ is at least as large as n, the almost-MDS bound, when $n \leq 4$. □

A.3 Proof of Proposition 4

Proof. To prove part (i). It is easy to see that A has branch number 2 when $n = 2$.

For the case $n = 3$: If the input has weight 1, then the output which corresponds to a column of the matrix will have weight 2. If the input has weight 2, then the output which corresponds to a linear combination of two columns will have weight at least 2. If the input has weight 3, then there will definitely be ≥ 4 non-zero entries in the total (input and output) as the output must be non-zero. Thus the branch number of A is 3.

For the case $n \geq 4$: If the input has weight 1, then the output which corresponds to a column of the matrix will have weight $n - 1 \geq 3$. If the input has weight 2, then the output which corresponds to a linear combination of two columns will have weight at least 2. If the input has weight 3, then the output which corresponds to a linear combination of three columns will have weight at least 1. If the input has weight ≥ 4, then there will be ≥ 5 non-zero entries in total since the output is non-zero. Thus the branch number of A is 4.

It is easy to see that permuting the rows and columns of a matrix will preserve its branch number, thus part (ii) follows naturally from part (i).

Part (iii) is verified by a computer search over all 4×4 $\{0, 1\}$-matrices. □

A.4 Proof of Theorem 2

Proof. Because the S-boxes used are affine transforms of the inversion function, they have the same difference distribution. Moreover, every row and every column of the inversion function have the same distribution as explained above; thus we just need to consider one row in the difference table. This allows us to simplify the upper bound in Proposition 2 to $\sum_{j=1}^{2^m-1} DP^{S_i}(u \to j)^k$, $u \neq 0$.

(i) When m is even:

$$\sum_{j=1}^{2^m-1} DP^{S_i}(u \to j)^k \leq (2/2^m)^k(2^{m-1} - 2) + (4/2^m)^k$$
$$= 2^{(1-m)(k-1)} - 2^{(1-m)k+1} + 2^{(2-m)k}.$$

(ii) When m is odd:

$$\sum_{j=1}^{2^m-1} DP^{S_i}(u \to j)^k \leq (2/2^m)^k \times 2^{m-1}$$
$$= 2^{(1-m)(k-1)}.$$

□

A.5 Proof of Theorem 3

Proof. Because the inversion map is used and the diffusion map has branch number 5, we can apply Theorem 2 with $m = 8$ and $k = 5$ to get an upper bound for the differential probability of the SDS structure in the F-function:

$$DP_{SDS} \leq 2^{(1-8)(5-1)} - 2^{(1-8)5+1} + 2^{(2-8)5} \approx 2^{-27.696}.$$

The final linear transform L does not influence the maximal differential probability of the F function since it simply rotates the bytes. Therefore, the maximal differential probability of F is $2^{-27.696}$. By applying Proposition 5, the maximal differential probability of 3 rounds of E2 is at most $(2^{-27.696})^2 = 2^{-55.392}$. □

A.6 Proof of Theorem 4

Proof. Here we base much of the proof that follows on the wide trail strategy (see [7, Theorem 3], [8, Theorem 9.5.1]). The design of MCrypton-x follows the design principle in AES [8] where τ is a diffusion optimal mapping. After an admissible re-arrangement of the operations in MCrypton-x, we can view the cipher as alternating between $\pi \circ \lambda$ and $\tau \circ \pi \circ \tau \circ \lambda$. The linear map $\tau \circ \pi \circ \tau$ has the same branch number as π but it acts on bundles of size 4-nibble (16-bit). Since the branch number of π is 4 and τ is a diffusion optimal transposition of bundles, $\tau \circ \pi \circ \tau$ also has branch number 4.

Each bundle over a $\lambda \circ \pi \circ \lambda$ transformation is a 16-bit SDS structure consisting of the linear map π sandwiched between two layers of four 4-bit S-boxes. Since π has branch number 4 by Proposition 4 part (ii) and each S-box is an affine transform of the inversion function on $GF(2^4)$, we can apply Theorem 2 with $m = 4$, $k = 4$. The maximal differential probability is:

$$DP_{SDS} \leq 2^{(1-4)(4-1)} - 2^{(1-4)4+1} + 2^{(2-4)4} \approx 2^{-7.54}.$$

Next, we can view 4 rounds of MCrypton-x as $\tau \circ \pi \circ \tau$ sandwiched between two layers of four bundles where each bundle has differential probability at most $2^{-7.54}$. By Proposition 3, the differential probability of 4 rounds is upper bounded by $(2^{-7.54})^{4-1} = 2^{-22.62}$. □

B Tables

Table 1. Upper Bound for the Branch Number in Theorem 1

Size of n for $n \times n$ $\{0,1\}$-Matrix	2	3	4	5	6	7	8	9	10
Upper Bound of Branch Number	2	3	4	4	5	6	6	7	8

Table 2. Difference Distribution for Each Row of the Difference Distribution Table for the Inversion Mapping on $GF(2^m)$

Difference	Frequency (m even)	Frequency (m odd)
0	$2^{m-1}+1$	2^{m-1}
2	$2^{m-1}-2$	2^{m-1}
4	1	0

C Diagrams

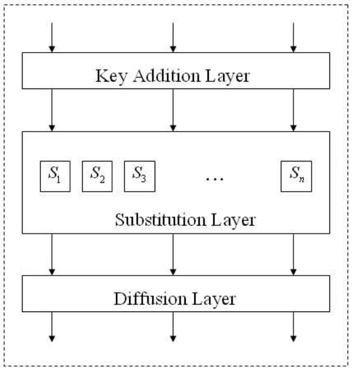

Fig. 1. One round of a SPN structure

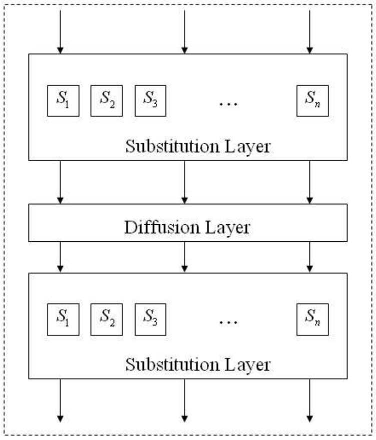

Fig. 2. SDS function