On the Additive Differential Probability of Exclusive-Or

Helger Lipmaa[1], Johan Wallén[1], and Philippe Dumas[2]

[1] Laboratory for Theoretical Computer Science
Helsinki University of Technology
P.O.Box 5400, FIN-02015 HUT, Espoo, Finland
{helger,johan}@tcs.hut.fi
[2] Algorithms Project
INRIA Rocquencourt, 78153 Le Chesnay Cedex, France
Philippe.Dumas@inria.fr

Abstract. We study the differential probability $\mathrm{adp}^{\oplus}$ of exclusive-or when differences are expressed using addition modulo 2^N. This function is important when analysing symmetric primitives that mix exclusive-or and addition—especially when addition is used to add in the round keys. (Such primitives include IDEA, Mars, RC6 and Twofish.) We show that $\mathrm{adp}^{\oplus}$ can be viewed as a formal rational series with a linear representation in base 8. This gives a linear-time algorithm for computing $\mathrm{adp}^{\oplus}$, and enables us to compute several interesting properties like the fraction of impossible differentials, and the maximal differential probability for any given output difference. Finally, we compare our results with the dual results of Lipmaa and Moriai on the differential probability of addition modulo 2^N when differences are expressed using exclusive-or.

Keywords: Additive differential probability, differential cryptanalysis, rational series.

1 Introduction

Symmetric cryptographic primitives like block ciphers are typically constructed from a small set of simple building blocks like bitwise exclusive-or and addition modulo 2^N. Surprisingly little is known about how these two operations interact with respect to different cryptanalytic attacks, and some of the fundamental relations between them have been established only recently [LM01, Lip02, Wal03]. Our goal is to share light to this question by studying the interaction of these two operations in one concrete application: differential cryptanalysis [BS91], by studying the differential probability of exclusive-or when differences are expressed using addition modulo 2^N. This problem is dual to the one explored by Lipmaa and Moriai [LM01, Lip02]. We hope that our results will be helpful in evaluating the precise security of ciphers that mix addition and exclusive-or against differential cryptanalysis.

B. Roy and W. Meier (Eds.): FSE 2004, LNCS 3017, pp. 317–331, 2004.

Differential Cryptanalysis. Differential cryptanalysis studies the propagation of differences in functions. Let G, H be Abelian groups and let $f\colon G \to H$ be a function. The input difference $x - x^* \in G$ is said to propagate to the output difference $f(x) - f(x^*) \in H$ through f. A *differential* of f is a pair $(\alpha, \beta) \in G \times H$. This is usually denoted by $\alpha \to \beta$. If the difference between $x, x^* \in G$ is $x - x^* = \alpha$, the differential $\alpha \to \beta$ can be used to predict the corresponding output difference $f(x) - f(x^*)$. It is thus natural to measure the efficiency of a differential by its *differential probability*

$$\mathrm{dp}^f(\alpha \to \beta) = \Pr_{x \in G}[f(x + \alpha) - f(x) = \beta] \enspace .$$

When a cipher uses both bitwise exclusive-or and addition modulo 2^N, both operators are natural choices for expressing differences (depending on how the round keys are added). Depending on this choice, one must either study the differential properties of addition when differences are expressed using exclusive-or, or the dual differential probability of exclusive-or when differences are expressed using addition modulo 2^N. The differential probability of addition was studied in detail by [LM01, Lip02]. However, the dual differential probability of exclusive-or has remained open. This dual case is just as interesting in practise, since most of the popular block ciphers that mix addition and exclusive-or use addition—and not exclusive-or—for adding in the round keys. (Examples include IDEA [LMM91], Mars [BCD$^+$98], RC6 [RRSY98] and Twofish [SKW$^+$99].)

We will exclusively deal with the set $\{0, 1, \dots, 2^N - 1\}$ equipped with two group operations. On one hand, we use the usual addition modulo 2^N, which we denote by $+$. On the other hand, we identify $\{0, 1, \dots, 2^N - 1\}$ and the set $\mathbf{Z}_2^N$ of N-tuples of bits using the natural correspondence that identifies $x_{N-1}2^{N-1} + \dots + x_1 2 + x_0 \in \mathbf{Z}_{2^N}$ with $(x_{N-1}, \dots, x_1, x_0) \in \mathbf{Z}_2^N$. In this way the usual componentwise addition $\oplus$ in $\mathbf{Z}_2^N$ (or bitwise exclusive-or) carries over to a group operation in $\{0, 1, \dots, 2^N - 1\}$. We can thus especially view $\oplus$ as a function $\oplus\colon \mathbf{Z}_{2^N} \times \mathbf{Z}_{2^N} \to \mathbf{Z}_{2^N}$. We call the differential probability of the resulting mapping the *additive differential probability* of exclusive-or and denote it by $\mathrm{adp}^\oplus\colon \mathbf{Z}_{2^N}^3 \to [0, 1]$,

$$\mathrm{adp}^\oplus(\alpha, \beta \to \gamma) = \Pr_{x,y}[((x + \alpha) \oplus (y + \beta)) - (x \oplus y) = \gamma] \enspace . \tag{1}$$

The dual mapping, the *exclusive-or differential probability* of addition, denoted $\mathrm{xdp}^+\colon \mathbf{Z}_{2^N}^3 \to [0, 1]$, is given by

$$\mathrm{xdp}^+(\alpha, \beta \to \gamma) = \Pr_{x,y}[((x \oplus \alpha) + (y \oplus \beta)) \oplus (x + y) = \gamma] \enspace .$$

This dual mapping was studied in detail by Lipmaa and Moriai [LM01, Lip02], who gave a closed formula for xdp^+. Their formula in particular leads to an $\Theta(\log N)$-time algorithm for computing xdp^+ and the differential probability of some related mappings like the pseudo-Hadamard transform [Lip02].

Our contributions. In this paper, we present a detailed analysis of the mapping $\mathrm{adp}^{\oplus}\colon \mathbf{Z}^3_{2^N} \to [0,1]$. This concrete problem has been addressed (and in a rather ad hoc manner) in a few papers, including [Ber92], but it has never been addressed completely—probably because of its "apparent complexity" [Ber92]. We show that $\mathrm{adp}^{\oplus}$ can be expressed as a formal rational series in the sense of formal language theory with a linear representation in base 8. That is, there are eight square matrices A_i, a column vector C and a row vector L, such that if we write the differential $(\alpha, \beta \to \gamma)$ as an octal word $w = w_{N-1} \cdots w_1 w_0$ in a natural way,

$$\mathrm{adp}^{\oplus}(\alpha, \beta \to \gamma) = \mathrm{adp}^{\oplus}(w) = L A_{w_{N-1}} \cdots A_{w_1} A_{w_0} C \enspace .$$

This representation immediately gives a linear-time algorithm for computing $\mathrm{adp}^{\oplus}$. This should be be compared to the naïve $\Theta(2^{2N})$-time algorithm which seems to be the only previously known algorithm for $\mathrm{adp}^{\oplus}$. In addition, we derive some other properties, like the fraction $\frac{3}{7} + \frac{4}{7} \cdot \frac{1}{8^N}$ of differentials with nonzero probability, and determine the maximal differential probability $\max_{\alpha,\beta} \mathrm{adp}^{\oplus}(\alpha, \beta \to \gamma)$ for any given output difference γ. Finally, we show how our approach based on rational series easily can be adapted for studying the dual mapping xdp^{+}.

The paper is organised as follows. We first show that $\mathrm{adp}^{\oplus}$ is a rational series and derive a linear representation for it. This gives an efficient algorithm that computes $\mathrm{adp}^{\oplus}(w)$ in time $O(|w|)$. In Sect. 3, we discuss the distribution of $\mathrm{adp}^{\oplus}$ and differentials with maximal probability. Sect. 4 describes how similar methods can be used to analyse xdp^{+}. The appendix contains some omitted proofs.

2 Rational Series $\mathrm{adp}^{\oplus}$

Throughout this paper, we let N denote the default word length. We will consider $\mathrm{adp}^{\oplus}$ as a function of octal words by writing the differential $(\alpha, \beta \to \gamma)$ as the octal word $w = w_{N-1} \cdots w_0$, where $w_i = \alpha_i 4 + \beta_i 2 + \gamma_i$. This defines $\mathrm{adp}^{\oplus}$ as a function from the octal words of length N to the interval $[0,1]$. As N varies in the set of nonnegative integers, we obtain a function from the set of all octal words to $[0,1]$.

In the terminology of formal language theory, the additive differential probability $\mathrm{adp}^{\oplus}$ is a formal series over the monoid of octal words with coefficients in the field of real numbers. A remarkable subset of these series is the set of *rational series* [BR88]. One possible characterisation of such a rational series S is the following: there exists a square matrix A_k of size $q \times q$ for each letter k in the alphabet, a row matrix L of size $1 \times q$ and a column matrix C of size $q \times 1$ such that for each word $w = w_1 \cdots w_\ell$, the value of the series is $S(w) = L A_{w_1} \cdots A_{w_\ell} C$. The family $L, (A_k)_k$, C is called a *linear representation* of dimension q of the rational series. In our case, the alphabet is the octal alphabet $\{0, 1, \ldots, 7\}$.

Theorem 1 (Linear representation of $\mathrm{adp}^{\oplus}$). *The formal series $\mathrm{adp}^{\oplus}$ has the 8-dimensional linear representation L, $(A_k)_k$, C, where* $L = \begin{pmatrix} 1\,1\,1\,1\,1\,1\,1\,1 \end{pmatrix}$,

$C = \begin{pmatrix}1\,0\,0\,0\,0\,0\,0\,0\end{pmatrix}^{\top}$,

$$A_0 = \frac{1}{4}\begin{pmatrix}4&0&0&1&0&1&1&0\\0&0&0&1&0&1&0&0\\0&0&0&1&0&0&1&0\\0&0&0&1&0&0&0&0\\0&0&0&0&0&1&1&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0\end{pmatrix} ,$$

and A_k, $k \neq 0$, is obtained from A_0 by permuting row i with row $i \oplus k$ and column j with column $j \oplus k$: $(A_k)_{ij} = (A_0)_{i\oplus k, j\oplus k}$. (For completeness, the matrices $A_0, \dots, A_7$ are given in Table 1.) Thus, $\mathrm{adp}^{\oplus}$ *is a rational series.*

For example, the differential $(\alpha, \beta \to \gamma) = (00110, 10100 \to 01110)$ corresponds to the octal word $w = 21750$ and $\mathrm{adp}^{\oplus}(\alpha, \beta \to \gamma) = \mathrm{adp}^{\oplus}(w) = LA_2A_1A_7A_5A_0C = \frac{5}{32}$. The linear representation immediately implies that $\mathrm{adp}^{\oplus}(w)$ can be computed using $O(|w|)$ arithmetic operations. Since the arithmetic operations can be carried out using $2|w|$-bit integer arithmetic, which can be implemented in constant time on a $|w|$-bit RAM model, we have

Corollary 1. *The additive differential probability* $\mathrm{adp}^{\oplus}(w)$ *can be computed in time $O(|w|)$ on a standard unit cost $|w|$-bit* RAM *model of computation.*

This can be compared with the $O(\log|w|)$-time algorithm for computing $\mathrm{xdp}^{+}(w)$ from [LM01].

As a side remark (we will not use this result later), note that the matrices $A_0, \dots A_7$ in the linear representation for $\mathrm{adp}^{\oplus}$ are substochastic. Thus, we could view the linear representation as a inhomogeneous Markov chain by adding a dummy state and dummy state transitions.

The rest of this section is devoted to the technical proof of Theorem 1. To prove this result, we will first give a different formulation of $\mathrm{adp}^{\oplus}$. For $x, y \in \{0, \dots, 2^N - 1\}$, let xy denote their componentwise product in $\mathbf{Z}_2^N$ (equivalently, the bitwise and of two N-bit strings). Let $\mathrm{borrow}(x, y) = x \oplus y \oplus (x - y)$

Table 1. All eight matrices A_i

$$A_0 = \frac{1}{4}\begin{pmatrix}4&0&0&1&0&1&1&0\\0&0&0&1&0&1&0&0\\0&0&0&1&0&0&1&0\\0&0&0&1&0&0&0&0\\0&0&0&0&0&1&1&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0\end{pmatrix} \quad A_1 = \frac{1}{4}\begin{pmatrix}0&0&1&0&1&0&0&0\\0&4&1&0&1&0&0&1\\0&0&1&0&0&0&0&0\\0&0&1&0&0&0&0&1\\0&0&0&0&1&0&0&0\\0&0&0&0&1&0&0&1\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&1\end{pmatrix} \quad A_2 = \frac{1}{4}\begin{pmatrix}0&1&0&0&1&0&0&0\\0&1&0&0&0&0&0&0\\0&1&4&0&1&0&0&1\\0&1&0&0&0&0&0&1\\0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&1&0&0&1\\0&0&0&0&0&0&0&1\end{pmatrix} \quad A_3 = \frac{1}{4}\begin{pmatrix}1&0&0&0&0&0&0&0\\1&0&0&0&0&1&0&0\\1&0&0&0&0&0&1&0\\1&0&0&4&0&1&1&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0\\0&0&0&0&0&1&1&0\end{pmatrix}$$

$$A_4 = \frac{1}{4}\begin{pmatrix}0&1&1&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&1&1&0&4&0&0&1\\0&1&0&0&0&0&0&1\\0&0&1&0&0&0&0&1\\0&0&0&0&0&0&0&1\end{pmatrix} \quad A_5 = \frac{1}{4}\begin{pmatrix}1&0&0&0&0&0&0&0\\1&0&0&1&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0\\1&0&0&0&0&0&1&0\\1&0&0&1&0&4&1&0\\0&0&0&0&0&0&1&0\\0&0&0&1&0&0&1&0\end{pmatrix} \quad A_6 = \frac{1}{4}\begin{pmatrix}1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\1&0&0&1&0&0&0&0\\0&0&0&1&0&0&0&0\\1&0&0&0&0&1&0&0\\0&0&0&0&0&1&0&0\\1&0&0&1&0&1&4&0\\0&0&0&1&0&1&0&0\end{pmatrix} \quad A_7 = \frac{1}{4}\begin{pmatrix}0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&0&0&0\\0&1&1&0&0&0&0&0\\0&0&0&0&1&0&0&0\\0&1&0&0&1&0&0&0\\0&0&1&0&1&0&0&0\\0&1&1&0&1&0&0&4\end{pmatrix}$$

denote the borrows, as an N-tuple of bits, in the subtraction $x - y$. Alternatively, $\operatorname{borrow}(x, y)$ can be recursively defined by $\operatorname{borrow}(x, y)_0 = 0$ and $\operatorname{borrow}(x, y)_{i+1} = 1$ if and only if $x_i - \operatorname{borrow}(x, y)_i < y_i$ as integers. This can be used to *define* $\operatorname{borrow}(x, y)_N = 1$ if and only if $x_{N-1} - \operatorname{borrow}(x, y)_{N-1} < y_{N-1}$ as integers. The borrows can be used to give an alternative formulation of $\operatorname{adp}^{\oplus}$.

Lemma 1. *For all* $\alpha, \beta, \gamma \in \mathbf{Z}_{2^N}$*,*

$$\operatorname{adp}^{\oplus}(w) = \Pr_{x,y}[a \oplus b \oplus c = \alpha \oplus \beta \oplus \gamma] \enspace ,$$

where $a = \operatorname{borrow}(x, \alpha)$*,* $b = \operatorname{borrow}(y, \beta)$ *and* $c = \operatorname{borrow}(x \oplus y, (x - \alpha) \oplus (y - \beta))$*.*

Proof. By replacing x and y with $x - \alpha$ and $y - \beta$ in the definition (1) of $\operatorname{adp}^{\oplus}$, we see that $\operatorname{adp}^{\oplus}(\alpha, \beta \to \gamma) = \Pr_{x,y}[(x \oplus y) - ((x - \alpha) \oplus (y - \beta)) = \gamma]$. Since $(x \oplus y) - ((x - \alpha) \oplus (y - \beta)) = \gamma$ if and only if $\gamma = c \oplus x \oplus y \oplus (x - \alpha) \oplus (y - \beta) = a \enspace \oplus b \oplus c \oplus \alpha \oplus \beta$ if and only if $a \oplus b \oplus c = \alpha \oplus \beta \oplus \gamma$, the result follows. □

We furthermore need the following technical lemma.

Lemma 2. *For all* x*,* y*,* α*,* β*,* γ*,*

$$\begin{aligned} a_{i+1} &= (aa' \oplus \alpha \oplus a'x)_i \enspace , \\ b_{i+1} &= (bb' \oplus \beta \oplus b'y)_i \quad \textit{and} \\ c_{i+1} &= [c \oplus a' \oplus b' \oplus c(a' \oplus b') \oplus (a' \oplus b')(x \oplus y)]_i \enspace , \end{aligned}$$

where $a = \operatorname{borrow}(x, \alpha)$*,* $b = \operatorname{borrow}(y, \beta)$*,* $c = \operatorname{borrow}(x \oplus y, (x - \alpha) \oplus (y - \beta))$*,* $a' = a \enspace \oplus \alpha$ *and* $b' = b \oplus \beta$*.*

Proof. By the recursive definition of $\operatorname{borrow}(x, y)$, $\operatorname{borrow}(x, y)_{i+1} = 1$ if and only if $x_i < y_i + \operatorname{borrow}(x, y)_i$ as integers. The latter event occurs if and only if either $y_i = \operatorname{borrow}(x, y)_i$ and at least two of x_i, y_i and $\operatorname{borrow}(x, y)_i$ are one, or $y_i \neq \operatorname{borrow}(x, y)_i$ and at least two of x_i, y_i and $\operatorname{borrow}(x, y)_i$ are zero. That is, $\operatorname{borrow}(x, y)_{i+1} = 1$ if and only if $y_i \oplus \operatorname{borrow}(x, y)_i \oplus \operatorname{maj}(x_i, y_i, \operatorname{borrow}(x, y)_i) = 1$, where $\operatorname{maj}(u, v, w)$ denotes the majority of the bits u, v, w. Since $\operatorname{maj}(u, v, w) = uv \oplus uw \oplus vw$, we have $\operatorname{borrow}(x, y)_{i+1} = [y \oplus \operatorname{borrow}(x, y) \oplus xy \oplus x \operatorname{borrow}(x, y) \oplus y \operatorname{borrow}(x, y)]_i$.

For a, we thus have $a_{i+1} = (\alpha \oplus a \enspace \oplus x\alpha \oplus xa \enspace \oplus \alpha a)_i = (a' \oplus a \; \alpha \oplus a'x)_i = [a' \oplus a \; (a' \oplus a) \oplus a'x]_i = (aa' \oplus \alpha \oplus a'x)_i$. The formula for b_{i+1} is completely analogous. For c, we have $c_{i+1} = [(x - \alpha) \oplus (y - \beta) \oplus c \oplus (x \oplus y)((x - \alpha) \oplus (y - \beta)) \oplus (x \oplus y)c \oplus ((x - \alpha) \oplus (y - \beta))c]_i = [x \oplus a' \oplus y \oplus b' \oplus c \oplus (x \oplus y)(x \oplus a' \oplus y \oplus b') \oplus (x \oplus y)c \oplus (x \oplus a' \oplus y \oplus b')c]_i = [c \oplus a' \oplus b' \oplus c(a' \oplus b') \oplus (a' \oplus b')(x \oplus y)]_i$. □

Proof (of Theorem 1). Let $(\alpha, \beta \to \gamma)$ be the differential associated with the word w. Denote $N = |w|$ and let x, y be uniformly distributed random variables in $\mathbf{Z}_{2^N}$. Denote $a = \operatorname{borrow}(x, \alpha)$, $b = \operatorname{borrow}(y, \beta)$ and $c = \operatorname{borrow}(x \oplus y, (x - \alpha) \oplus (y - \beta))$. Let ξ be the octal word of borrow triples, $\xi_i = a_i 4 + b_i 2 + c_i$. We define ξ_N in the natural way using $\operatorname{borrow}(u, v)_N = 1$ if and only if $u_{N-1} -$

$\mathrm{borrow}(u,v)_{N-1} < v_{N-1}$ as integers. For compactness, denote $\mathrm{xor}(w) = \alpha \oplus \beta \oplus \gamma$ and $\mathrm{xor}(\xi) = a \oplus b \oplus c$. Let $P(w,k)$ be the 8×1 substochastic matrix

$$P_j(w,k) = \Pr_{x,y}[\mathrm{xor}(\xi) \equiv \mathrm{xor}(w) \pmod{2^k}, \xi_k = j]$$

for $0 \le k \le N$. Let $M(w,k)$ be the 8×8 substochastic transition matrix

$$M_{ij}(w,k) = \Pr_{x,y}[\mathrm{xor}(\xi)_k = \mathrm{xor}(w)_k, \xi_{k+1} = i \mid \\ \mathrm{xor}(\xi) \equiv \mathrm{xor}(w) \pmod{2^k}, \xi_k = j]$$

for $0 \le k < N$. Then $P_i(w,k+1) = \sum_j M_{ij}(w,k) P_j(w,k)$ and thus $P(w,k+1) = M(w,k)P(w,k)$. Note furthermore that $P(w,0) = C$, since $a_0 = b_0 = c_0 = 0$, and that $LP(w,N) = \sum_j \Pr_{x,y}[\mathrm{xor}(\xi) \equiv \mathrm{xor}(w) \pmod{2^N}, \xi_N = j] = \Pr_{x,y}[\mathrm{xor}(\xi) \equiv \mathrm{xor}(w) \pmod{2^N}] = \mathrm{adp}^{\oplus}(w)$, where the last equality is due to Lemma 1. We will show that $M(w,k) = A_{w_k}$ for all k. By induction, it follows that $\mathrm{adp}^{\oplus}(w) = LP(w,N) = LM(w,N-1)\cdots M(w,0)C = LA_{w_{N-1}} \cdots A_{w_0} C$.

It remains to show that $M(w,k) = A_{w_k}$ for all k. Towards this end, let x, y be such that $\mathrm{xor}(\xi) \equiv \mathrm{xor}(w) \pmod{2^k}$ and $\xi_k = j$. We will count the number of ways we can choose (x_k, y_k) such that $\mathrm{xor}(\xi)_k = \mathrm{xor}(w)_k$ and $\xi_{k+1} = i$.

Denote $a' = a \oplus \alpha$, $b' = b \oplus \beta$ and $c' = c \oplus \gamma$. Note that $\mathrm{xor}(\xi)_k = \mathrm{xor}(w)_k$ if and only if $c'_k = (a' \oplus b')_k$. Under the assumption that $\mathrm{xor}(\xi)_k = \mathrm{xor}(w)_k$ we have $(cc' \oplus \gamma)_k = [c(a' \oplus b') \oplus c \oplus a' \oplus b']_k$. By Lemma 2, (x_k, y_k) must thus be a solution to $V \begin{pmatrix} x_k & y_k \end{pmatrix}^\top = U$ in $\mathbf{Z}_2$, where U and V are the matrices

$$U = \begin{pmatrix} (aa' \oplus \alpha)_k \oplus a_{k+1} \\ (bb' \oplus \beta)_k \oplus b_{k+1} \\ (cc' \oplus \gamma)_k \oplus c_{k+1} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} a'_k & 0 \\ 0 & b'_k \\ (a' \oplus b')_k & (a' \oplus b')_k \end{pmatrix}$$

over $\mathbf{Z}_2$. If this equation has a solution, it has exactly $2^{2-\mathrm{rank}(V)}$ solutions. But $\mathrm{rank}(V) = 0$ if and only if $a'_k = b'_k = 0$ (then there are 4 solutions) and $\mathrm{rank}(V) = 2$ otherwise (then there is 1 solution).

The equation has a solution (x_k, y_k) exactly when $\mathrm{rank}(V) = \mathrm{rank}(V\ U)$. From this and from the requirement that $c'_k = (a' \oplus b')_k$, we see that there are solutions exactly in the following cases.

- If $a'_k = b'_k = 0$, then $c'_k = 0$ and $\mathrm{rank}(V) = 0$. There are solutions (4 solutions) if and only if $a_{k+1} = \alpha_k$, $b_{k+1} = \beta_k$ and $c_{k+1} = \gamma_k$.
- If $a'_k = 0$ and $b'_k = 1$ then $c'_k = 1$ and $\mathrm{rank}(V) = 2$. There is a single solution if and only if $a_{k+1} = \alpha_k$.
- If $a'_k = 1$ and $b'_k = 0$, then $c'_k = 1$ and $\mathrm{rank}(V) = 2$. There is a single solution if and only if $b_{k+1} = \beta_k$.
- If $a'_k = 1$ and $b'_k = 1$ then $c'_k = 0$ and $\mathrm{rank}(V) = 2$. There is a single solution if and only if $c_{k+1} = \gamma_k$.

Since $j = \xi_k = a_k 4 + b_k 2 + c_k$ and $i = \xi_{k+1} = a_{k+1}4 + b_{k+1}2 + c_{k+1}$, the derivation so far can be summarised as

$$M_{ij}(w,k) = \begin{cases} 1 \ , & j = (\alpha_k, \beta_k, \gamma_k) \ , \ i = (\alpha_k, \beta_k, \gamma_k) \ , \\ 1/4 \ , & j = (\alpha_k, \beta_k \oplus 1, \gamma_k \oplus 1) \ , \ i = (\alpha_k, *, *) \ , \\ 1/4 \ , & j = (\alpha_k \oplus 1, \beta_k, \gamma_k \oplus 1) \ , \ i = (*, \beta_k, *) \ , \\ 1/4 \ , & j = (\alpha_k \oplus 1, \beta_k \oplus 1, \gamma_k) \ , \ i = (*, *, \gamma_k) \ , \\ 0 \ , & \text{otherwise} \ , \end{cases}$$

where we have identified the integer $r_2 4+r_1 2+r_0$ with the binary tuple (r_2, r_1, r_0) and $*$ represents an arbitrary element of $\{0,1\}$. It follows that $M(w,k) = A_0$ if $w_k = 0$ and $M_{i,j}(w,k) = M_{i\oplus w_k, j\oplus w_k}(0,k)$. That is, $M(w,k) = A_{w_k}$ for all w, k. This completes the proof. □

3 Distribution of adp$^\oplus$ and Maximal Differentials

3.1 Distribution

We will use notation from formal languages to describe octal words (and thus differentials). In particular, we will use concatenation (xy), the corresponding powers ($x^0 = \lambda$ is the empty word and $x^{n+1} = xx^n$), union $(x+y)$ and the Kleene star ($x^* = \sum_{n\geq 0} x^n$). Throughout this section, L, $(A_k)_k$, C is the linear representation of adp$^\oplus$.

We will first consider the effect of tailing and leading zeros.

Corollary 2. *For all octal words w,* $\mathrm{adp}^\oplus(w0^*) = \mathrm{adp}^\oplus(w)$.

This trivial result follows from the observation that $A_0 C = C$.

Corollary 3. *Let w be a word and let $a = \begin{pmatrix} a_0 & \cdots & a_7 \end{pmatrix}^\top = A_{|w|-1} \cdots A_{w_0} C$. Let $\alpha = a_0$ and $\beta = a_3 + a_5 + a_6$. Let w' be a word of the form $w' = 0^* w$. Then* $\mathrm{adp}^\oplus(w') = \alpha + \frac{\beta}{3} + \frac{8}{3} \cdot \beta \cdot 4^{-(|w'|-|w|)}$.

Proof. Using a Jordan form $J = P^{-1} A_0 P$ of A_0, it is easy to see that

$$A_0^k = 4^{-k} \begin{pmatrix} 4^k & 0 & 0 & \frac{4^k-1}{3} & 0 & \frac{4^k-1}{3} & \frac{4^k-1}{3} & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} .$$

If we let $j = |w'| - |w|$, we see that $LA_0^j a = a_0 + \frac{4^j-1}{3\cdot 4^j}(a_3 + a_5 + a_6) + \frac{3}{4^j}(a_3 + a_5 + a_6) = \alpha + \frac{\beta}{3} + \frac{8}{3} \cdot \beta \cdot 4^{-(|w'|-|(|w))}$. □

This means that $\mathrm{adp}^{\oplus}(0^n w)$ decreases with n and $\mathrm{adp}^{\oplus}(0^n w) \to \alpha + \beta/3$ as $n \to \infty$. This can be compared to [LM01], where it was shown that $\mathrm{xdp}^{+}(00w) = \mathrm{xdp}^{+}(0w)$ for all w.

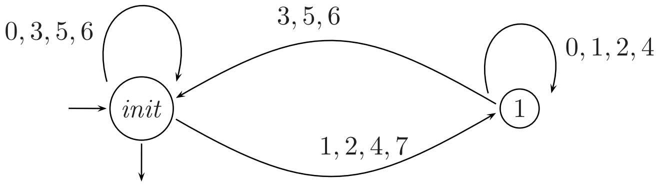
Theorem 2. *The additive differential probability* $\mathrm{adp}^{\oplus}(w)$ *is nonzero if and only if w has the form $w = 0^*$ or $w = w'(3+5+6)0^*$ for any octal word w'.*

Proof. Since $A_1C = A_2C = A_4C = A_7C = 0$, $\mathrm{adp}^{\oplus}(w'(1+2+4+7)0^*) = 0$. Conversely, let w be a word of the form $w = w'(3+5+6)0^*$. Let e_i be the canonical (column) basis vector with a 1 in the ith component and 0 in the others. By direct computation, the kernels are $\ker A_0 = \ker A_3 = \ker A_5 = \ker A_6 = \langle e_1, e_2, e_4, e_7\rangle$ and $\ker A_1 = \ker A_2 = \ker A_4 = \ker A_7 = \langle e_0, e_3, e_5, e_6\rangle$. For all i and $j \neq i$, $e_j \notin \ker A_i$, it can be seen that $A_i e_i = e_i$ and that $A_i e_j$ has the form $A_i e_j = (e_k + e_\ell + e_m + e_n)/4$, where $k \neq \ell$, $m \neq n$, $e_k, e_\ell \in \ker A_0$ and $e_m, e_n \in \ker A_1$. Since $C = e_0$, we see by induction that $A_{w_i} \cdots A_{w_0} C \notin \ker A_{w_{i+1}}$ for all i. Thus, $\mathrm{adp}^{\oplus}(w) \neq 0$.

Since the matrices A_k are nonnegative, an alternative proof is the following. Since

$$LA_{w_{N-1}} \cdots A_{w_0} C = \sum_{i_0,\dots,i_N} L_{i_N} (A_{w_{N-1}})_{i_N, i_{N-1}} \cdots (A_{w_0})_{i_1, i_0} C_{i_0} \ ,$$

we see that $\mathrm{adp}^{\oplus}(w)$ is nonzero if and only if there are indexes $i_0, \dots, i_N$ such that $L_{i_N} (A_{w_{N-1}})_{i_N, i_{N-1}} \cdots (A_{w_0})_{i_1, i_0} C_{i_0} \neq 0$. We construct a nondeterministic finite automaton with state set $\{0, \dots, 7, \mathit{init}\}$ and input alphabet $\{0, \dots, 7\}$ as follows. There is an empty transition from the initial state *init* to state i if and only if $L_i \neq 0$. There is a transition labelled x from state i to state j if and only if $(A_x)_{i,j} \neq 0$. The state i is accepting if and only if $C_i \neq 0$. Clearly, the automaton accepts the word w read from left to right if and only if $\mathrm{adp}^{\oplus}(w) \neq 0$. If we convert the automaton to a minimal deterministic automation, we obtain the following automation.



This automaton clearly accepts the language $0^* + (0+1+\cdots+7)^*(3+5+6)0^*$. □

A complete determination of the distribution of $\mathrm{adp}^{\oplus}$ falls out of the scope of this paper and we will restrict ourselves to some of the most important results. First, we turn to the fraction of *possible* differentials—that is, differentials with $\mathrm{adp}^{\oplus}(w) \neq 0$.

Corollary 4. *For all $N \geq 0$, $\Pr_{|w|=N}[\mathrm{adp}^{\oplus}(w) \neq 0] = \frac{3}{7} + \frac{4}{7} \cdot \frac{1}{8^N}$.*

Proof. According to Theorem 2, $\mathrm{adp}^{\oplus}(w) \neq 0$ if and only if w is the zero word or has form $w = w'\xi 0^k$, where w' is an arbitrary word of length $N-k-1$ and $\xi \in \{3,5,6\}$. For a fixed value of k, we can choose w' and ξ in $3 \cdot 8^{N-k-1}$ ways. Thus, there are $1 + \sum_{k=0}^{N-1} 3 \cdot 8^{N-k-1} = \frac{4}{7} + \frac{3}{7} \cdot 8^N$ words with $\mathrm{adp}^{\oplus}(w) \neq 0$ in total. □

This result can be compared with [LM01, Theorem 2], which states that the corresponding probability for xdp^{+} is $\Pr_{|w|=N}[\mathrm{xdp}^{+}(w) \neq 0] = \frac{4}{7} \cdot \left(\frac{7}{8}\right)^N$. This means, in particular, that $\Pr_{|w|=N}[\mathrm{adp}^{\oplus}(w) \neq 0] \to \frac{3}{7}$ while $\Pr_{|w|=N}[\mathrm{xdp}^{+}(w) \neq 0] \to 0$ as $N \to \infty$. Since the number of possible differentials is larger for $\mathrm{adp}^{\oplus}$ than for xdp^{+}, the average possible differential will obtain a smaller probability.

Next, if $w = (0+3+5+6)0^*$ then clearly $\mathrm{adp}^{\oplus}(w) = 1$. On the other hand, for any $\xi \in \{0,\dots,7\}$, $\mathrm{adp}^{\oplus}(\xi w) \leq 1/2$. It follows that $\Pr_{|w|=N}[\mathrm{adp}^{\oplus}(w) = 1] = 4 \cdot 8^{-N}$, and $\Pr_{|w|=N}[\mathrm{adp}^{\oplus}(w) = k] = 0$ if $1/2 < k < 1$. One can further establish easily that $\mathrm{adp}^{\oplus}(w) = 1/2$ if and only if $w = \Sigma(0+3+5+6)0^*$, where $\Sigma = 0+1+\dots+7$. The following straightforward lemma is useful in such calculations.

Lemma 3. *Let w be an octal word. Denote $\Sigma_0 = \{0,3,5,6\}$, $\Sigma_1 = \{1,2,4,7\}$ and $A_w = A_{w_{|w|-1}} \cdots A_{w_0}$. Then $\mathrm{adp}^{\oplus}(xw) = \mathrm{adp}^{\oplus}(0w)$ for all $x \in \Sigma_0$ and $\mathrm{adp}^{\oplus}(yw) = \mathrm{adp}^{\oplus}(1w)$ for all $y \in \Sigma_1$. Thus, $\mathrm{adp}^{\oplus}(w) = \mathrm{adp}^{\oplus}(xw) + \mathrm{adp}^{\oplus}(yw)$ and $\mathrm{adp}^{\oplus}(xzw) = \mathrm{adp}^{\oplus}(yzw) + (A_w C)_z$ for all $x, z \in \Sigma_b$ and $y \in \Sigma_{1-b}$.*

Proof. By direct calculation, we see that $LA_0 = LA_3 = LA_5 = LA_6 = \begin{pmatrix}1&0&0&1&0&1&1&0\end{pmatrix}$ and $LA_1 = LA_2 = LA_4 = LA_7 = \begin{pmatrix}0&1&1&0&1&0&0&1\end{pmatrix}$. Since $\mathrm{adp}^{\oplus}(vw) = LA_v A_w C$, we have $\mathrm{adp}^{\oplus}(xw) = \mathrm{adp}^{\oplus}(0w)$ and $\mathrm{adp}^{\oplus}(yw) = \mathrm{adp}^{\oplus}(1w)$ for all $x \in \Sigma_0$ and $y \in \Sigma_1$. Finally, if $x, z \in \Sigma_b$ and $y \in \Sigma_{1-b}$, we have $\mathrm{adp}^{\oplus}(w) = LA_w C = (LA_x + LA_y)A_w C = \mathrm{adp}^{\oplus}(xw) + \mathrm{adp}^{\oplus}(yw)$ and $\mathrm{adp}^{\oplus}(xzw) - \mathrm{adp}^{\oplus}(yzw) = (L(A_x - A_y)A_z)A_w C = e_z A_w C = (A_w C)_z$, where e_z is a row vector with a 1 in column z and 0 in the other columns. □

3.2 Maximal Differentials

Although many of the enumerative aspects of $\mathrm{adp}^{\oplus}$ seem infeasible, some optimisation problems are surprisingly simple. For all output differences γ, denote

$$\mathrm{adp}^{\oplus}_{2\max}(\gamma) = \max_{\alpha,\beta} \mathrm{adp}^{\oplus}(\alpha, \beta \to \gamma) \ .$$

For all γ, there is a simple differential with $\mathrm{adp}^{\oplus}(\alpha, \beta \to \gamma) = \mathrm{adp}^{\oplus}_{2\max}(\gamma)$.

Theorem 3. *For all output differences γ, $\mathrm{adp}^{\oplus}(0^N, \gamma \to \gamma) = \mathrm{adp}^{\oplus}_{2\max}(\gamma)$.*

The proof is omitted from the conference version.

4 Rational Series xdp^+

Lipmaa and Moriai [LM01, Lip02] used completely different techniques to analyse the exclusive-or differential probability xdp^+ of addition. We will now demonstrate the power of our approach by showing that it can easily adapt to analyse xdp^+ as well.

4.1 Linear Representation

As for $\mathrm{adp}^\oplus$, we write the differential $(\alpha, \beta \to \gamma)$ as the octal word $w = w_{N-1} \dots w_0$, where $w_i = \alpha_i 4 + \beta_i 2 + \gamma_i$. When N varies, we obtain a rational series xdp^+ with a linear representation of dimension 2.

Theorem 4 (Linear representation of xdp^+). *The formal series xdp^+ has the 2-dimensional linear representation L, $(X_k)_{k=0}^7$, C, where $L = \begin{pmatrix}1 & 1\end{pmatrix}$, $C = \begin{pmatrix}1 & 0\end{pmatrix}^\top$ and X_k is given by*

$$(X_k)_{ij} = \begin{cases} 1 - T(k_2 + k_1 + j) & \text{if } i = 0 \text{ and } k_2 \oplus k_1 \oplus k_0 = j \ , \\ T(k_2 + k_1 + j) & \text{if } i = 1 \text{ and } k_2 \oplus k_1 \oplus k_0 = j \ , \\ 0 & \text{otherwise} \end{cases}$$

for $i, j \in \{0, 1\}$, where $k = k_2 4 + k_1 2 + k_0$ and $T \colon \{0, 1, 2, 3\} \to \mathbf{R}$ is the mapping $T(0) = 0$, $T(1) = T(2) = \frac{1}{2}$ and $T(3) = 1$. (For completeness, all the matrices X_k are given in Table 2.) Thus, xdp^+ is a rational series.

For example, the differential $(\alpha, \beta \to \gamma) = (11100, 00110 \to 10110)$ corresponds to the word $w = 54730$ and $\mathrm{xdp}^+(\alpha, \beta \to \gamma) = \mathrm{xdp}^+(w) = L X_5 X_4 X_7 X_3 X_0 C = \frac{1}{4}$. The proof of this result is given in the appendix.

4.2 Words with a Given Probability

The simplicity of the linear representation of xdp^+ allows us to derive an explicit description of all words with a certain differential probability.

Theorem 5. *For all nonempty words w, $\mathrm{xdp}^+(w) \in \{0\} \cup \{2^{-k} \mid k \in \{0, 1, \dots, |w| - 1\}\}$. The differential probability $\mathrm{xdp}^+(w) = 0$ if and only if w has the form $w = w'(1 + 2 + 4 + 7)$, $w = w'(1 + 2 + 4 + 7)0w''$ or $w = w'(0 + 3 + 5 + 6)7w''$ for arbitrary words w', w'', and $\mathrm{xdp}^+(w) = 2^{-k}$ if and only if $\mathrm{xdp}^+(w) \neq 0$ and $|\{0 \le i < N - 1 \mid w_i \neq 0, 7\}| = k$.*

Table 2. All the eight matrices X_k in Theorem 4

$$X_0 = \tfrac{1}{2}\begin{pmatrix}2 & 0\\ 0 & 0\end{pmatrix} \quad X_1 = \tfrac{1}{2}\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix} \quad X_2 = \tfrac{1}{2}\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix} \quad X_3 = \tfrac{1}{2}\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix}$$

$$X_4 = \tfrac{1}{2}\begin{pmatrix}0 & 1\\ 0 & 1\end{pmatrix} \quad X_5 = \tfrac{1}{2}\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix} \quad X_6 = \tfrac{1}{2}\begin{pmatrix}1 & 0\\ 1 & 0\end{pmatrix} \quad X_7 = \tfrac{1}{2}\begin{pmatrix}0 & 0\\ 0 & 2\end{pmatrix}$$

Proof. Let L, X_k and C be as in Theorem 4 and denote $e_0 = \begin{pmatrix}1\ 0\end{pmatrix}^\top$ and $e_1 = \begin{pmatrix}0\ 1\end{pmatrix}^\top$. Then the kernels of X_i are $\ker X_0 = \ker X_3 = \ker X_5 = \ker X_6 = \langle e_1 \rangle$ and $\ker X_1 = \ker X_2 = \ker X_4 = \ker X_7 = \langle e_0 \rangle$. By direct calculation, $X_0 e_0 = e_0$, $X_3 e_0 = X_5 e_0 = X_6 e_0 = \frac{1}{2}(e_0 + e_1)$, $X_1 e_1 = X_2 e_1 = X_4 e_1 = \frac{1}{2}(e_0 + e_1)$ and $X_7 e_1 = e_1$. Since $C = e_0$, we thus have $\mathrm{xdp}^+(w) = 0$ if and only if w has the form $w = w'(1+2+4+7)$, $w = w'(1+2+4+7)0w''$ or $w = w'(0+3+5+6)7w''$ for arbitrary words w', w''. Similarly, when w is such that $\mathrm{adp}^+(w) \neq 0$, we see that $X_{w_{n-1}} \cdots X_{w_0} C$ has the form $\begin{pmatrix}2^{-\ell}\ 2^{-\ell}\end{pmatrix}^\top$, $\begin{pmatrix}2^{-\ell}\ 0\end{pmatrix}^\top$ or $\begin{pmatrix}0\ 2^{-\ell}\end{pmatrix}^\top$, where $\ell = |\{w_i \mid w_i \notin \{0,7\}, 0 \le i < n\}|$ for all n. Thus, $\mathrm{xdp}^+(w) = 2^{-k}$, where $k = |\{0 \le i < N-1 \mid w_i \neq 0, 7\}|$. □

For example, if w is the word $w = 54730$, we see that $\mathrm{xdp}^+(w) \neq 0$ and $|\{0 \le i < 4\} \mid w_i \neq 0, 7\}| = 2$. Thus, $\mathrm{xdp}^+(w) = 2^{-2}$. This result immediately gives the closed formula from [LM01] and thus the $O(\log N)$-time algorithm.

4.3 Distribution

Based on the explicit description of all words with a certain differential probability, it is easy to determine the distribution of xdp^+. Let $\mathcal{A}(n,k)$, $\mathcal{B}(n,k)$ and $\mathcal{C}(n,k)$ denote the languages that consist of the words of length $n > 0$ with $\mathrm{xdp}^+(w) = 2^{-k}$, and $w_{n-1} = 0$, $w_{n-1} = 7$ and $w_{n-1} \neq 0, 7$, respectively. The languages are clearly given recursively by

$$\begin{aligned}
\mathcal{A}(n,k) &= 0\mathcal{A}(n-1,k) + 0\mathcal{C}(n-1,k-1) \ , \\
\mathcal{B}(n,k) &= 7\mathcal{B}(n-1,k) + 7\mathcal{C}(n-1,k-1) \ , \\
\mathcal{C}(n,k) &= \Sigma_0\mathcal{A}(n-1,k) + \Sigma_1\mathcal{B}(n-1,k) + (\Sigma_0 + \Sigma_1)\mathcal{C}(n-1,k-1) \ ,
\end{aligned}$$

where $\Sigma_0 = 3 + 5 + 6$ and $\Sigma_1 = 1 + 2 + 4$. The base cases are $\mathcal{A}(1,0) = 0$, $\mathcal{B}(1,0) = \emptyset$ and $\mathcal{C}(1,0) = 3 + 5 + 6$. Let $A(z,u) = \sum_{n,k} |\mathcal{A}(n,k)| u^k z^n$, $B(z,u) = \sum_{n,k} |\mathcal{B}(n,k)| u^k z^n$ and $C(z,u) = \sum_{n,k} |\mathcal{C}(n,k)| u^k z^n$ be the corresponding ordinary generating functions. The recursive description of the languages immediately gives the the linear system

$$\begin{cases}
A(z,u) = zA(z,u) + uzC(z,u) + z \ , \\
B(z,u) = zB(z,u) + uzC(z,u) \ , \\
C(z,u) = 3zA(z,u) + 3zB(z,u) + 6uzC(z,u) + 3z \ .
\end{cases}$$

Denote $D(z,u) = A(z,u) + B(z,u) + C(z,u) + 1$. Then the coefficient of $u^k z^n$ in $D(z,u)$, $[u^k z^n]D(z,u)$, gives the number of words of length n with $\mathrm{xdp}^+(w) = 2^{-k}$ (the extra 1 comes from the case $n = 0$). By solving the linear system, we see that

$$D(z,u) = 1 + \frac{4z}{1 - (1 + 6u)z} \ .$$

Since the coefficient of z^n in $D(z,u)$ for $n > 0$ is

$$[z^n]D(z,u) = 4[z^n]z \sum_{m=0}^{\infty} (1 + 6u)^m z^m = 4(1 + 6u)^{n-1} \ ,$$

Table 3. Short comparison of the functions $\mathrm{xdp}^{\oplus}$ and $\mathrm{adp}^{\oplus}$ and of their computational complexity

	xdp^{+}	$\mathrm{adp}^{\oplus}$
Possibility verification		
Algorithm complexity	$\Theta(1)$	$\Theta(\log N)$
Probability of possibility	$\frac{1}{2} \cdot \left(\frac{7}{8}\right)^{N-1}$	$\frac{3}{7} + \frac{4}{7} \cdot 8^{-N}$
Evaluation of a possible differential		
Algorithm complexity	$\Theta(\log N)$	$\Theta(N)$
Maximal differentials $\mathrm{adp}^{\oplus}_{2\max}$ and $\mathrm{xdp}^{+}_{2\max}$		
Finding max. differential (α, β)	$\Theta(\log N)$	$\Theta(1)$
Computing max. differential when (α, β) is known	$\Theta(\log N)$	$\Theta(N)$

we see that

$$[u^k z^n] D(z, u) = 4 \cdot 6^k \binom{n-1}{k}$$

for all $0 \leq k < n$. The coefficient of z^n in $D(z, 1)$ for $n > 0$, $[z^n] D(z, 1) = 4[z^n] \frac{z}{1-7z} = 4 \cdot 7^{n-1}$ gives the number of words of length n with $\mathrm{xdp}^{+}(w) \neq 0$.

Theorem 6 ([LM01, Theorem 2]). *There are $4 \cdot 7^{n-1}$ words of length $n > 0$ with $\mathrm{xdp}^{+}(w) \neq 0$. Of these, $4 \cdot 6^k \binom{n-1}{k}$ have probability 2^{-k} for all $0 \leq k < n$.*

5 Conclusions

We analysed the additive differential probability $\mathrm{adp}^{\oplus}$ of exclusive-or. We expect that our results combined with the work of Lipmaa and Moriai will facilitate advanced differential cryptanalysis of ciphers that mix addition and exclusive-or as well as the design of such ciphers. These results can also be used to guide the choice between addition and exclusive-or as the key-mixing operation.

In general, $\mathrm{adp}^{\oplus}$ is much more difficult to analyse than xdp^{+} (note especially the straightforward analysis of xdp^{+} in Sect. 4). On the other hand, it is easier to find the maximal differentials for $\mathrm{adp}^{\oplus}$, although the maximal differentials for xdp^{+} have higher probability: it can be seen that $\mathrm{adp}^{\oplus}_{2\max}(\gamma) \leq \mathrm{xdp}^{+}_{2\max}(\gamma) = \max_{\alpha,\beta} \mathrm{xdp}^{+}(\alpha, \beta \to \gamma)$ for all γ. (See Fig. 1.) A short comparison of some of the properties of xdp^{+} and $\mathrm{adp}^{\oplus}$ is given in Table 3.

Maybe the main contribution of this paper is the formal series approach. In addition to the new results, we were able to give a simpler proof of the results of Lipmaa and Moriai on xdp^{+}. The results from [Wal03] can also be rephrased using our approach. We expect that our approach of using formal series has also other applications in cryptanalysis.

Acknowledgements

This work was partially supported by the Finnish Defence Forces Research Institute of Technology.

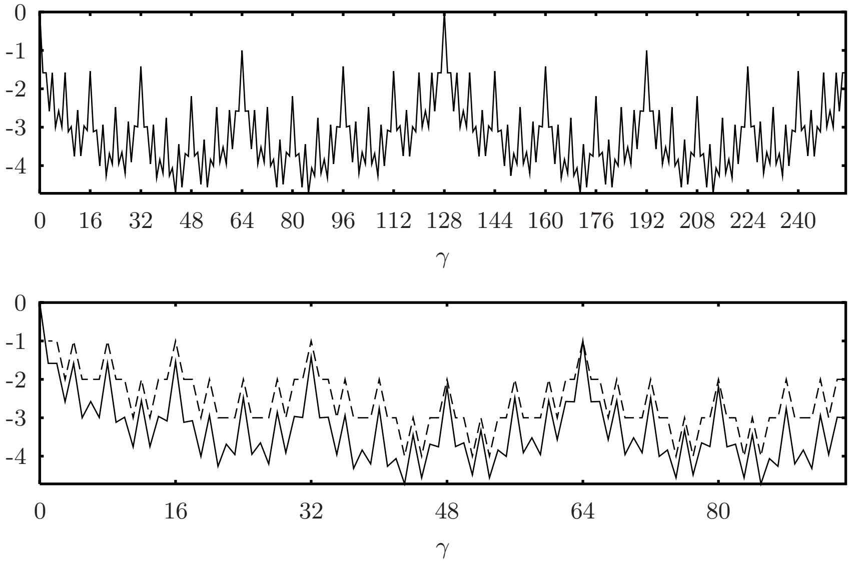


Fig. 1. Top: tabulation of values $\log_2 \mathrm{adp}^{\oplus}_{2\max}(\gamma)$, $0 \le \gamma \le 255$, for $N = 8$. Bottom: partial tabulation of values $\log_2 \mathrm{adp}^{\oplus}_{2\max}(\gamma)$ (solid) and $\log_2 \mathrm{xdp}^{+}_{2\max}(\gamma)$ (dashed), $0 \le \gamma \le 95$, for $N = 8$

References

[BCD$^+$98] Carolynn Burwick, Don Coppersmith, Edward D'Avignon, Rosario Gennaro, Shai Halevi, Charanjit Jutla, Stephen M. Matyas Jr., Luke O'Connor, Mohammad Peyravian, David Safford, and Nevenko Zunic. MARS — A Candidate Cipher for AES. Available from `http://www.research.ibm.com/security/mars.html`, June 1998. 318

[Ber92] Thomas A. Berson. Differential Cryptanalysis Mod 2^{32} with Applications to MD5. In Rainer A. Rueppel, editor, *Advances in Cryptology — EUROCRYPT '92*, volume 658 of *Lecture Notes in Computer Science*, pages 71–80, Balatonfüred, Hungary, 24–28 May 1992. Springer-Verlag. ISBN 3-540-56413-6. 319

[BR88] Jean Berstel and Christophe Reutenauer. *Rational Series and Their Languages*. EATCS Monographs on Theoretical Computer Science. Springer-Verlag, 1988. 319

[BS91] Eli Biham and Adi Shamir. Differential Cryptanalysis of DES-like Cryptosystems. *Journal of Cryptology*, 4(1):3–72, 1991. 317

[Lip02] Helger Lipmaa. On Differential Properties of Pseudo-Hadamard Transform and Related Mappings. In Alfred Menezes and Palash Sarkar, editors, *INDOCRYPT 2002*, volume 2551 of *Lecture Notes in Computer Science*, pages 48–61, Hyderabad, India, 15–18 December 2002. Springer-Verlag. 317, 318, 326

[LM01] Helger Lipmaa and Shiho Moriai. Efficient Algorithms for Computing Differential Properties of Addition. In Mitsuru Matsui, editor, *Fast Soft-*

ware Encryption 2001, volume 2355 of *Lecture Notes in Computer Science*, pages 336–350, Yokohama, Japan, 2–4 April 2001. Springer-Verlag, 2002. 317, 318, 320, 324, 325, 326, 327, 328, 330

[LMM91] Xuejia Lai, James L. Massey, and Sean Murphy. Markov Ciphers and Differential Cryptanalysis. In Donald W. Davies, editor, *Advances in Cryptology — EUROCRYPT '91*, volume 547 of *Lecture Notes in Computer Science*, pages 17–38, Brighton, UK, April 1991. Springer-Verlag. 318

[RRSY98] Ronald L. Rivest, Matt J. B. Robshaw, R. Sidney, and Y. L. Yin. The RC6 Block Cipher. Available from `http://theory.lcs.mit.edu/~rivest/rc6.ps`, June 1998. 318

[SKW+99] Bruce Schneier, John Kelsey, Doug Whiting, David Wagner, Chris Hall, and Niels Ferguson. *The Twofish Encryption Algorithm: A 128-Bit Block Cipher*. John Wiley & Sons, April 1999. ISBN: 0471353817. 318

[Wal03] Johan Wallén. Linear Approximations of Addition Modulo 2^n. In Thomas Johansson, editor, *Fast Software Encryption 2003*, volume 2887 of *Lecture Notes in Computer Science*, pages 261–273, Lund, Sweden, February24–26 2003. Springer-Verlag. 317, 328

A Proof of Theorem 4

In order to prove Theorem 4, we introduce the following notation. Define the carry function carry: $\mathbf{Z}_2^N \times \mathbf{Z}_2^N \to \mathbf{Z}_2^N$ of addition modulo 2^N by $\mathrm{carry}(x, y) = (x + y) \oplus x \oplus y$. It is easy to see that

$$\mathrm{xdp}^+(\alpha, \beta \to \gamma) = \Pr_{x,y}[\mathrm{carry}(x, y) \oplus \mathrm{carry}(x \oplus \alpha, y \oplus \beta) = \alpha \oplus \beta \oplus \gamma] \ .$$

Denote $c = \mathrm{carry}(x, y)$ and $c^* = \mathrm{carry}(x \oplus \alpha, y \oplus \beta)$, where x, y, α and β are understood from context. Note that c_i can be recursively defined as $c_0 = 0$ and $c_{i+1} = 1$ if and only if at least two of x_i, y_i and c_i are 1. To simplify some of the formulae, denote $\mathrm{xor}(x, y, z) = x \oplus y \oplus z$ and $\Delta c = c \oplus c^*$. Then $\mathrm{xdp}^+(\alpha, \beta \to \gamma) = \Pr_{x,y}[\Delta c = \mathrm{xor}(\alpha, \beta, \gamma)]$. Let furthermore xy denote the componentwise product of x and y, $(xy)_i = x_i y_i$.

The linear representation of xdp^+ follows easily from the following result [LM01, Lemma 2].

Lemma 4. *Fix $\alpha, \beta \in \mathbf{Z}_2^n$ and $i \geq 0$. Then*

$$\Pr_{x,y}[\Delta c_{i+1} = 1 \mid \Delta c_i = r] = T(\alpha_i + \beta_i + r) \ ,$$

where T is as in Theorem 4.

This result follows easily from the recursive definition of the carry function and a case-by-case analysis.

Proof (of Theorem 4). Let $(\alpha, \beta \to \gamma)$ be the differential associated to the word w. Let x, y be uniformly distributed independent random variables over

$\mathbf{Z}_2^{|w|}$. For compactness, we denote $\mathrm{xor}(w) = \alpha \oplus \beta \oplus \gamma$. Let $P(w,k)$ be the 2×1 substochastic matrix given by

$$P_j(w,k) = \Pr_{x,y}[\Delta c \equiv \mathrm{xor}(w) \pmod{2^k}, \Delta c_k = j]$$

for $0 \leq k \leq |w|$ and let $M(w,k)$ be the 2×2 substochastic transition matrix

$$M_{ij}(w,k) = \Pr_{x,y}[\Delta c_k = \mathrm{xor}(w)_k, \Delta c_{k+1} = i \mid \Delta c \equiv \mathrm{xor}(w) \pmod{2^k}, \Delta c_k = j]$$

for $0 \leq k < |w|$. Since $P_i(w,k+1) = \sum_j M_{ij}(w,k)P_j(w,k)$, $P(w,k+1) = M(w,k)P(w,k)$. Note furthermore that $P(w,0) = C$ and that $\mathrm{xdp}^+(w) = \sum_j P_j(w,|w|) = LP(w,|w|)$. By Lemma 4, it is clear that

$$M_{ij}(w,k) = \begin{cases} 1 - T(\alpha_k + \beta_k + j) & \text{if } i = 0 \text{ and } \mathrm{xor}(w)_k = j \ , \\ T(\alpha_k + \beta_k + j) & \text{if } i = 1 \text{ and } \mathrm{xor}(w)_k = j \text{ and} \\ 0 & \text{otherwise} \ . \end{cases}$$

That is, $M(w,k) = X_{w_k}$ for all k. It follows by induction that $\mathrm{xdp}^+(w) = LX_{w_{|w|-1}} \cdots X_{w_0}C$. □