Higher Order Derivatives and Differential Cryptanalysis

Article · January 1994

DOI: 10.1007/978-1-4615-2694-0_23

1 author:

Xuejia Lai
Shanghai Jiao Tong University
112 PUBLICATIONS **2,829** CITATIONS

Reprint of pp.227-233, "Communications and Cryptography", (Ed. R.Blahut et al), Kluwer Academic Publishers, 1994.

Higher Order Derivatives and Differential Cryptanalysis

Xuejia Lai [1]
R^3 Security Engineering AG
CH-8607 Aathal, Switzerland
lai@r3.ch

Abstract

High-order derivatives of multi-variable functions are studied in this paper as a natural generalization of the basic concept used in differential cryptanalysis. Possible applications of such derivatives in cryptology are discussed.

I. Introduction

In 1990, just after Jim and I published at Eurocrypt'90 our new block cipher PES [1], the previous version of the IDEA cipher, differential cryptanalysis was proposed by Biham and Shamir [2] as a chosen-plaintext attack to find the secret-key of an iterated block cipher. In [3], Biham and Shamir showed that the full 16-round DES can be broken by differential cryptanalysis using only 2^{47} encryptions, which is the first reported attack that finds the secret-key of DES with fewer encryptions than exhaustive key-search. Besides DES, differential cryptanalysis has been applied successfully to many other ciphers [4, 5], for example, the block ciphers FEAL, Khafre, REDOC, LOKI and Lucifer. Like every cipher designer who heard a new attack that can be applied to his cipher, we started to apply the differential cryptanalysis to our own proposal. As I talked to Jim about the first result of the analysis, he noticed immediately that the use of the properties of a Markov chain is essential to differential cryptanalysis. This observation led to the concept of "Markov Cipher" which we presented in [6], where it was shown that the differential cryptanalysis of many (if not all) practical block ciphers can be formulated in terms of the theory of Markov chains. For example, the implication of the only Lemma in Biham and Shamir's paper is that DES is a Markov cipher. In particular, a fairly tight lower bound on the complexity of a differential cryptanalysis attack on a cipher can be derived in terms of parameters of the Markov chain. Although differential cryptanalysis requires far too much chosen plaintexts to be a practical attack on a cipher, the complexity of a differential cryptanalysis attack is the best measure of its cryptographic strength that is known today.

In this paper, we consider a possible generalization of the original (first-order) differential cryptanalysis in terms of higher-order derivative, which is defined in Section 2, where some general properties of derivatives are studied. In Section 3 we consider some

[1] This work was carried out at the Signal and Information Processing Laboratory of the Swiss Federal Institute of Technology, Zürich, Switzerland

special properties of derivatives of binary functions. Section 4 discusses the applications of such higher order derivatives in cryptology. In particular, we consider the relationship among derivatives, differential cryptanalysis and linear structure of cryptographic functions.

II. Higher order derivatives

Definition Let $(S,+)$ and $(T,+)$ be Abelian groups. For a function $f: S \to T$, the *derivatives of the f at point $a \in S$* is defined as

$$\Delta_a f(x) = f(x+a) - f(x).$$

Note that the derivative of f is itself a function from S to T, we can define the *i-th ($i>1$) derivative of f at $(a_1, a_2, \ldots, a_i)$* as

$$\Delta^{(i)}_{a_1,\ldots,a_i} f(x) = \Delta_{a_i}(\Delta^{(i-1)}_{a_1,\ldots,a_{i-1}} f(x))$$

where $\Delta^{(i-1)}_{a_1,\ldots,a_{i-1}} f(x)$ being the $(i-1)$-th derivative of f at $(a_1, a_2, ..., a_{i-1})$. The 0-th derivative of $f(x)$ is defined to be $f(x)$ itself.

For example, for $i=2$, we have

$$\begin{aligned}
\Delta^{(2)}_{a_1,a_2} f(x) &= \Delta_{a_2}(\Delta_{a_1} f(x)) \\
&= \Delta_{a_2}(f(x+a_1) - f(x)) \\
&= (f(x+a_1+a_2) - f(x+a_2)) - (f(x+a_1) - f(x)) \\
&= f(x+a_1+a_2) - f(x+a_1) - f(x+a_2) + f(x).
\end{aligned}$$

It then follows that

$$f(x+a_1+a_2) = \Delta^{(2)}_{a_1,a_2} f(x) + \Delta_{a_1} f(x) + \Delta_{a_2} f(x) + f(x).$$

In general, one can show the following result.

Proposition 1

$$f(x+a_1+a_2+\cdots+a_n) = \sum_{i=0}^{n} \sum_{1 \le j_1 < \cdots < j_i \le n} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x). \tag{1}$$

Proof. For $n=1$ it follows from the definition. Suppose (1) holds for $n-1$. Then

$$\Delta^{(n)}_{a_1,\ldots,a_n} f(x) \tag{2}$$

$$= \Delta^{(n-1)}_{a_1,\ldots,a_{n-1}} f(x \oplus a_n) - \Delta^{(n-1)}_{a_1,\ldots,a_{n-1}} f(x) \tag{3}$$

$$= \left(f(x+a_n+a_1\cdots+a_{n-1}) - \sum_{i=0}^{n-2} \sum_{1 \le j_1 < \cdots < j_i \le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x+a_n) \right) \tag{4}$$

$$-\Delta^{(n-1)}_{a_1,\ldots,a_{n-1}} f(x) \tag{5}$$

$$= f(x+a_1+\cdots+a_n) - \left(\sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x+a_n) \right. \tag{6}$$

$$\left. - \sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x) + \sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x) + \Delta^{(n-1)}_{a_1,\ldots,a_{n-1}} f(x) \right) \tag{7}$$

$$= f(x+a_1+\cdots+a_n) - \left(\sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} (f(x+a_n) - f(x)) \right. \tag{8}$$

$$\left. + \sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x) + \Delta^{(n-1)}_{a_1,\ldots,a_{n-1}} f(x) \right) \tag{9}$$

$$= f(x+a_1+\cdots+a_n) - \left(\sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} (\Delta_{a_n} f(x)) \right. \tag{10}$$

$$\left. + \sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x) + \Delta^{(n-1)}_{a_1,\ldots,a_{n-1}} f(x) \right) \tag{11}$$

$$= f(x+a_1+\cdots+a_n) - \left(\sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i+1)}_{a_{j_1},\ldots,a_{j_i},a_n} f(x) \right. \tag{12}$$

$$\left. + \sum_{i=0}^{n-2} \sum_{1\le j_1<\cdots<j_i\le n-1} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i}} f(x) + \Delta^{(n-1)}_{a_1,\ldots,a_{n-1}} f(x) \right) \tag{13}$$

$$= f(x+a_1+\cdots+a_n) - \left(\sum_{i=0}^{n-1} \sum_{1\le j_1<\cdots<j_i\le n} \Delta^{(i)}_{a_{j_1},\ldots,a_{j_i},a_n} f(x) \right) \tag{14}$$

□

Some basic properties of the derivative are as follows.

$$\Delta_a(f+g) = \Delta_a f + \Delta_a g \tag{15}$$

$$\Delta_a(f(x)g(x)) = f(x+a)\Delta_a g(x) + (\Delta_a f(x))g(x) \tag{16}$$

Equation (15) is rather obvious and Equation (16) can be obtained as follows:

$$\begin{aligned} \Delta_a(f(x)g(x)) &= f(x+a)g(x+a) - f(x)g(x) \\ &= f(x+a)(g(x+a)-g(x)) + (f(x+a)-f(x))g(x) \\ &= f(x+a)\Delta_a g(x) + (\Delta_a f(x))g(x). \end{aligned}$$

□

Proposition 2 *Let* $\deg(f)$ *denote the nonlinear degree of a multi-variable polynomial function* $f(x)$*. Then*

$$\deg(\Delta_a f(x)) \le \deg(f(x)) - 1. \tag{17}$$

Proof. It follows from equation (15) and (16), from the facts that $\deg(f+g) \leq \max(\deg(f), \deg(g))$ and $\deg(fg) \leq \deg(f) + \deg(g)$ and that the derivative of a linear function is a constant. □

III. Derivatives of binary functions

In what follows, we will consider only the binary functions and the group operation is the *bitwise XOR*, denoted by $\oplus$.

Proposition 3 *Let $L[a_1, a_2, ..., a_i]$ be the list of all 2^i possible linear combinations of $a_1, a_2, ..., a_i$. Then,*

$$\Delta^{(i)}_{a_1,...,a_i} f(x) = \sum_{c \in L[a_1,a_2,...,a_i]} f(x \oplus c) \tag{18}$$

Proof. We prove it by induction. For $i = 1$ it is obvious. Suppose (18) hold for $i-1$, then by definition,

$$\begin{aligned}
\Delta^{(i)}_{a_1,...,a_i} f(x) &= \Delta_{a_i}(\Delta^{(i-1)}_{a_1,...,a_{i-1}} f(x)) && (19)\\
&= \left(\Delta^{(i-1)}_{a_1,...,a_{i-1}} f(x \oplus a_i)\right) \oplus \left(\Delta^{(i-1)}_{a_1,...,a_{i-1}} f(x)\right) && (20)\\
&= \left(\sum_{c \in L[a_1,a_2,...,a_{i-1}]} f(x \oplus c \oplus a_i)\right) \oplus \left(\sum_{c \in L[a_1,a_2,...,a_{i-1}]} f(x \oplus c)\right) && (21)\\
&= \sum_{c \in L[a_1,a_2,...,a_i]} f(x \oplus c). && (22)
\end{aligned}$$

□

Corollary 4 *Derivatives of binary function is independent of the order in which the derivation is taken, i.e., for any permutation $p(j)$ of index j,*

$$\Delta^{(i)}_{a_1,...,a_i} f(x) = \Delta^{(i)}_{a_{p(1)},...,a_{p(i)}} f(x) \tag{23}$$

The above results lead to the lower bound on the probability of derivative function taking on a certain value. This probability is essential in cryptanalysis using such derivatives.

Proposition 5 *For function $f : F_2^n \to F_2^n$ and linearly independent $a_1, a_2, \ldots, a_i$ in F_2^n and for any $b \in F_2^n$, $P\left(\Delta^{(i)}_{a_1,...,a_i} f(x) = b\right)$ is either 0 or at least 2^{i-n} if x is uniformly random.*

Proof. From equation (18), if at input x_0 the derivative takes on value b, then at all (2^i) inputs $x_0 \oplus c$, $c \in L[a_1, a_2, ..., a_{i-1}]$, the derivative will take on value b. □

Proposition 6 *If a_i is linearly dependent of $a_1, a_2, \ldots, a_{i-1}$, then $\Delta^{(i)}_{a_1,...,a_i} f(x) = 0$.*

Proof. If a_i is linearly dependent of $a_1, ..., a_{i-1}$, then a_i is contained in the list $L[a_1, a_2, ..., a_{i-1}]$. Thus, in (21) we have

$$\sum_{c \in L[a_1, a_2, ..., a_{i-1}]} f(x \oplus c \oplus a_i) = \sum_{c \in L[a_1, a_2, ..., a_{i-1}]} f(x \oplus c),$$

which implies that the derivative is zero. □

The above result shows that derivatives should be computed at the points that are linearly independent. Otherwise the higher order derivatives will be trivially zero, such cases are of no interest for our purpose.

Proposition 7 *For any function $f : F_2^n \to F_2^m$, the n-th derivative of f is a constant. If $f : F_2^n \to F_2^n$ is invertible, then $(n-1)$-th derivative of f is a constant.*

Proof. Each component function of f can have a nonlinear degree at most n. When f is invertible, the nonlinear degree of its component function is at most $n-1$. The proof then follows from Proposition 2. □

Example. For

$$f(x_1, x_2, x_3, x_4) = x_1x_2x_3 \oplus x_1x_2x_4 \oplus x_2x_3x_4,$$

we compute the 2-nd derivative at (0001,1010).

$$\Delta_{0001} f(x_1, x_2, x_3, x_4) = f(x_1, x_2, x_3, x_4 \oplus 1) \oplus f(x_1, x_2, x_3, x_4) = x_1x_2 \oplus x_2x_3,$$

$$\Delta_{1010}(x_1x_2 \oplus x_2x_3) = x_2 \oplus x_2 = 0.$$

Thus, $\Delta^{(2)}_{(0001,1010)} f(x_1, x_2, x_3, x_4) = 0$. Note that (0001) and (1010) are linearly independent and function f has a nonlinear degree 3.

IV. Cryptographic significance of derivatives

Differential cryptanalysis and derivatives The basic concept of differential cryptanalysis is the probability of differentials. An *differential* is a couple (a, b), where a is the difference of a pair of distinct inputs x and x^* and where b is a possible difference for the resulting outputs $y = f(x)$ and $y^* = f(x^*)$. The *probability of an differential* (a, b) is the conditional probability that b is the difference Δy of the outputs given that the input pair (x, x^*) has difference $\Delta x = a$ when the x is uniformly random. We denote this differential probability by $P(\Delta y = b | \Delta x = a)$. If the "difference" is defined by the group operation"+", i.e., if $\Delta x = x - x^*$, then

$$P(\Delta y = b | \Delta x = a) = P(f(x + a) - f(x) = b) = P(\Delta_a f = b). \tag{24}$$

Proposition 8 *The probability of a differential (a, b) is the probability that the first derivative of function $f(x)$ at point a takes on value b when x is uniformly random.*

The success of differential cryptanalysis is based on the fact that many practical block ciphers are obtained from iterating a cryptographically weak round function. If the difference of a pair of inputs to the last round can be anticipated with a high probability, then the secret key used in the last round can usually be derived from the pair of outputs and from the difference at the input. By using high-order derivatives, the basic idea of differential cryptanalysis can be generalized to the case when more than two inputs are used simultaneously for deriving the secret key: *if a (nontrivial) i-th derivative of (r-1) round function takes on a value with high probability, then it is possible to derive the key for the last round from the known 2^i outputs and from the anticipated derivative value.* Although some independent preliminary experiments [7, 8] indicated that cryptanalysis using high order derivative may not be more powerful than the first order differential cryptanalysis, we expect, however, that the derivative will provide new measurement for the strength of cryptographic functions.

Linear structure and derivatives A function f is said to have a *linear structure* if there is a nonzero a, such that $f(x+a)-f(x)$ remains invariant for all x. The study of linear structure has lead to attacks on cipher functions [9, 10] and to the nonlinearity criteria for cryptographic functions [11, 12]. From the definitions, it is easy to see that function $f(x)$ has a linear structure if and only if there is a nonzero a such that the derivative of $f(x)$ at a is a constant, or equivalently, if and only if function f has a differential of probability one. The relationship among the concepts of differential, linear structure and derivatives then suggests the following:

A new design principle for cryptographic functions. For each small i, the nontrivial i-th derivatives of function should take on each possible value roughly uniform. In particular a binary function from F_2^n to F_2^n, the nontrivial i-th derivatives should take on each possible value with probability about 2^{i-n}.

References

[1] X. Lai and J. L. Massey, "A Proposal for a New Block Encryption Standard", Advances in Cryptology – EUROCRYPT'90, Proceedings, LNCS 473, pp. 389-404, Springer-Verlag, Berlin, 1991.

[2] E. Biham and A. Shamir, "Differential Cryptanalysis of DES–like Cryptosystems", Advances in Cryptology – CRYPTO'90, Proceedings, LNCS 537, pp. 2-21, Springer-Verlag, Berlin 1991.

[3] E. Biham and A. Shamir, "Differential Cryptanalysis of the full 16-round DES", Abstracts of CRYPTO'92.

[4] E. Biham and A. Shamir, "Differential Cryptanalysis of FEAL and N-Hash", Advances in Cryptology – EUROCRYPT'91, Proceedings, LNCS 547, pp. 1-16, Springer-Verlag, Berlin 1991.

[5] E. Biham and A. Shamir, "Differential Cryptanalysis of Snefru, Khafre, REDOC-II, LOKI and Lucifer", Advances in Cryptology – CRYPTO'91, Proceedings, LNCS 576, pp. 156-171, Springer-Verlag, Berlin 1992.

[6] X. Lai, J. L. Massey and S. Murphy, "Markov Ciphers and Differential Cryptanalysis", Advances in Cryptology – EUROCRYPT'91, Proceedings, LNCS 547, pp. 17-38, Springer-Verlag, Berlin, 1991.

[7] C. Harpes, "Notes on High Order Differential Cryptanalysis of DES," Internal report, Signal and Information Processing Laboratory, Swiss Federal Institute of Technology, August 12, 1993.

[8] E. Biham, "Higher Order Differential Cryptanalysis," (Preliminary draft) August 13, 1993.

[9] D. Chaum, J.H. Evertse, Cryptanalysis of DES with a reduced number of rounds, Advances in Cryptology - CRYPTO'85, Proceedings, pp. 192–211, Springer-Verlag, 1986.

[10] J.H. Evertse, Linear structures in block ciphers, Advances in Cryptology - EUROCRYPT'87, Proceedings, pp. 249–266, Springer-Verlag, 1988.

[11] W. Meier, O. Staffelbach, Nonlinearity criteria for cryptographic functions, Advances in Cryptology - EUROCRYPT'89, Proceedings, pp. 549–562, Springer-Verlag, 1990.

[12] K. Nyberg, On the construction of highly nonlinear permutations, Advances in Cryptology - EUROCRYPT'92, Proceedings, pp. 92–98, Springer-Verlag, 1993.