Finding Good Linear Approximations of Block Ciphers Application to Cryptanalysis and Side channel attack

Cédric Tavernier

Communication et Systèmes, France

Abstract

We design an algorithm determining the list of linear approximations of a m-variate Boolean function within a given bias. We show how to adapt this algorithm in order to find multiple approximations of 8 rounds of the DES with biases of the same order as the best bias obtained by Matsui. We propose a new very efficient resulting attack based on a soft decision decoding technique of first order Reed-Muller codes.

1 Introduction

Since it was designed by Matsui in 1993 [11] and its success in the cryptanalysis of the DES [12], linear cryptanalysis has become a powerful tool in the analysis of block ciphers. Now conceivers of block ciphers have at least to prove that their cipher is immune to linear cryptanalysis.

One of the crucial steps of linear cryptanalysis in terms of time and memory complexity consists of the quantity of plaintext-ciphertext pairs required so that the attack succeeds with a good probability.

This *data-complexity* can be derived from **linear relations** involving **key bits, plaintext and ciphertext bits**. Suppose that the attacker obtained such a relation which is satisfied with a bias $1/2+\varepsilon$ or $1/2-\varepsilon$, then this data complexity is proportional to $1/\varepsilon^2$ (Matsui).

***Multiple Linear Relations*.**

In 1994, Kaliski and Robshaw showed that the knowledge of several linear relations involving the same key bits and with biases of the same order could reduce significantly the data-complexity of the attack ([10]).

The constraint on the key bits has been erased by Biryukov, De Cannière and Quisquater who showed how to use multiple linear relations to diminish the data-complexity, [1]. They showed that if there are n statistically independent linear relations involving key, plaintext and ciphertext bits satisfied with respective biases ε_j, for $j=1,\ldots,n$, then the data-complexity N becomes

$$N \approx 1/\sum_{j=1}^{n}\varepsilon_j^2$$

These results (and others not mentioned, [13, 2, 3]) point out how crucial it is to be able to compute multiple linear relations between key, plaintext and ciphertext bits which are satisfied with the best possible biases.

Drawbacks of these approaches:

- biases that are obtained can be of a much smaller order of magnitude compared to the best bias obtained by Matsui,
- the method depends heavily on the cipher that is considered.

Our Algorithm : a general purpose algorithm which outputs all linear relations between key, plaintext and ciphertext bits with the best possible biases.

Problem: finding *all* the linear approximations of a m-variable Boolean function which are satisfied with a given bias ε.

Application: we consider the Boolean functions obtained from the inner product between 8-rounds of the DES and a suitable ciphertext mask.

Idea of the algorithm: reconstructing the linear relations variable by variable. At every step, a list of the best linear relations is kept and taken as input for the next step ([4, 5, 8, 9]).

Kabatiansky and Tavernier showed that:

- the maximum size of the list of relations is upper bounded by $1/4\varepsilon^2$,
- the time complexity was upper-bounded by $\mathcal{O}(m^2/\varepsilon^6)$.

We propose a significant improvement of the algorithm by Kabatiansky and Tavernier:

- experimental complexity of $\mathcal{O}(m/\varepsilon^2)$,
- obtained linear relations with biases of the same order as the best bias obtained by Matsui.

Notation:

- P, C, K: the plaintext, ciphertext and key vectors of a block cipher,
- $|K|,\ |P|,\ |C|$ denote their bit-lengths.
- "$<\ ,\ >$": the usual scalar product of binary vectors.

2 How to find many linear approximations?

Problem: Given a bias ε, find the list of all vectors π , κ and γ , and a bit b, such that

$$< P,\pi > \oplus < K,\kappa > \oplus\, b =< C(P,K),\gamma > \qquad (1)$$

is satisfied with probability $\geq 1/2+\varepsilon$, where the probability is taken over the plaintext and key space.

2.1 Multiple linear approximations and polynomial reconstruction

Notation:

- $v \stackrel{def}{=} |P|+|K|$
- $P=(\delta_1,\ldots,\delta_{|P|})$, $K=(\delta_{|P|+1},\ldots,\delta_v)$.

Problem: finding the list $\mathcal{L}$ of all multivariate affine polynomials p of $GF(2)[\delta_1,\ldots,\delta_v]$ and all the vectors γ such that

$$p(\delta_1,\ldots,\delta_v) =< C(\delta_1,\ldots,\delta_v),\gamma >$$

is satisfied with probability greater than $1/2+\varepsilon$.

$\longrightarrow$ *list decoding* problem in the first order Reed-Muller code $RM(1,v)$:

- the linear combination γ is fixed
- noisy codeword = linear combination of the ciphertext bits γ.

Principle of the algorithm (Goldreich, Levin [4, 5]):

- $\mathcal{L}$ = list of solutions (affine polynomials in v variables)
- i-prefix of $p = p$ limited to the first i variables:

$$p(\delta_1,\ldots,\delta_i,0,\ldots,0).$$

- $\mathcal{L}_i$ = list of i-prefix candidates of $\mathcal{L}$.

$\longrightarrow$ *Idea:* construct $\mathcal{L}_i$ from $\mathcal{L}_{i-1}$:

1. Add the i-th variable δ_i : $\mathcal{L}_i=\{s, s+\delta_i \mid s\in\mathcal{L}_{i-1}\}$,
2. *screening process:* eliminate most of the *bad* prefixes candidates.

Problem: find an efficient screening process.

- Johannson and Jönsson: adaptation for fast correlation attacks. ([7]).
- Kabatiansky and Tavernier: screening process shown to be optimal ([8, 9]).

Worst case complexity of these algorithms:

- Memory complexity = maximum size of the lists $\mathcal{L}_i$ (*Johnson bound*): $\mathcal{O}(1/\varepsilon^2)$
- Time complexity at least of $\mathcal{O}(m/\varepsilon^4)$ (see [15]), essentially due to the fact that the list of i-prefixes can reach a size of $\mathcal{O}(1/\varepsilon^2)$ elements.

$\longrightarrow$ *Problem*: complexity $>>$ whole linear cryptanalysis complexity, of order $\mathcal{O}(1/\varepsilon^2)$.

$\longrightarrow$ *Solution*: reduce the size of the lists!

Helleseth-Klove-Levenshtein ([6]) : with high probability (average case), the size of the list is $\mathcal{O}(1)$.

2.2 Design of the algorithm

- $f(\delta_1,\ldots,\delta_v) \stackrel{def}{=} < C(\delta_1,\ldots,\delta_v),\gamma >$
- $\mathcal{L}=\{p \mid d_{\mathcal{H}}(p,f)\leq 2^v(1/2-\varepsilon)$ ($d_{\mathcal{H}}$ = Hamming distance).
- Walsh-Hadamard transform:

$$\widehat{f}(p)=\sum_{x\in GF(2)^v}(-1)^{f(x)+p(x)}=2^v-2d_{\mathcal{H}}(p,f)$$

$\longrightarrow \mathcal{L}=\{p \mid \widehat{f}(p)\geq 2^{v+1}\varepsilon\}$

- **screening process** for i-prefix p^i:

$$\widehat{f}(p) = \sum_{s\in GF(2)^{v-i}}(-1)^{p(0,s)}\sum_{r\in GF(2)^i}(-1)^{f(r,s)+p^i(r)} \qquad (2)$$

$$\geq \sum_{s\in GF(2)^{v-i}}\left|\sum_{r\in GF(2)^i}(-1)^{f(r,s)+p^i(r)}\right| = \sum_s|\widehat{f}_s(p^i)| \qquad (3)$$

$\longrightarrow \mathcal{L}_i \stackrel{def}{=} \{p^i \mid \sum_s|\widehat{f}_s(p)|\geq 2^{v+1}\varepsilon\}$ ($f_s \stackrel{def}{=} f(\cdot,s)$)

- original *probabilistic* algorithm: choose randomly S vectors s and R vectors r.

$\longrightarrow$ *new algorithm*: two steps

1. **Decoding step:** a full decoding on ℓ variables, with a Walsh-Hadamard transform ($\mathcal{O}(\ell 2^\ell)$ complexity).
   - $\mathcal{L}_\ell=\emptyset$;
   - **for** $k=1$ to S:
     - choose $s\in GF(2)^{v-\ell}$ randomly,
     - compute $|\widehat{f}_s|$.
   - **for** each $p\in RM(1,\ell)$:
     - **if** $\sum_s|\widehat{f}_s(p)|\geq 2^{\ell+1}S\times\varepsilon(1-1/c)$
     - **then** $\mathcal{L}_\ell=\mathcal{L}_\ell\cup\{p\}$.
2. **Reconstruction step:** takes $\mathcal{L}_\ell$ as input
   - **for** $i=\ell+1$ to v:
     - $\mathcal{L}_i=\mathcal{L}_{i-1}\cup(\mathcal{L}_{i-1}+\delta_i)$.
     - **for** each $p^i\in\mathcal{L}_i$
       - **for** $k=1$ to S:
         - choose $s\in GF(2)^{v-i}$ randomly,
         - compute $\left|\sum_r(-1)^{f(r,s)+p^i(r)}\right|$ for R random vectors $r\in GF(2)^i$.
       - **if** $\sum_s\left|\sum_r(-1)^{f(r,s)+p^i(r)}\right| < 2SR\times\varepsilon(1-1/c)\}$
       - **then** $\mathcal{L}_i=\mathcal{L}_i\setminus\{p^i\}$
   - **return** $\mathcal{L}=\mathcal{L}_v$.

2.3 Analysis of the algorithm

Complexity:

- *Decoding step*: $\mathcal{O}(S\ell 2^\ell)$.
- *Reconstruction step*: $<\mathcal{O}((v-\ell)STL)$, $L=\max_i|\mathcal{L}_i|$.

Question: since $L=\mathcal{O}(1/4\varepsilon^2)$,why do we add a Fourier transform?

Answer: Helleseth, Kløve and Levenshtein ([6]):

- **if** bias of $f=\varepsilon$ and $2^v>\mathcal{O}(1/\varepsilon^2)$,
- **then** size of list of approximations of f within ε is $\mathcal{O}(1)$ (probability of error $<2^{-2^v\varepsilon^2}$).

Conclusion:

- choose ε large enough,
- fix $\ell\approx 2\log_2(1/\varepsilon)$,
- $S=\mathcal{O}(1)$ (experimentations: $S\approx 20$),
- $R=\mathcal{O}(1/\varepsilon^2)$.

$\longrightarrow$**Complexity** in favourable case:

- Time complexity: $\mathcal{O}\left(\frac{v}{\varepsilon^2}\right)$
- Memory complexity: $\mathcal{O}\left(\frac{1}{\varepsilon^2}\right)$

Acceptable complexity compared to the complexity of linear cryptanalysis. Namely this work of finding good linear approximations needs do be done only once for each considered cipher.

3 Applications: finding approximations on 8 rounds of DES

- $(P,K)=(P_H,P_L,K)$ of length 128 bits in DES (Matsui's notation).
- consider the best combination of ciphertext bits $<C(P,K),\gamma>$, using Matsui's results.
- chosen $\varepsilon = 1.8\times 10^{-4}$ (best bias given by Matsui on 8 rounds: 5.95×10^{-4}).
- experimentations:
  - $|\mathcal{L}_i|\leq 100$: favourable case!
  - running time $\approx$ one day.
  - many approximations found on different linear combinations of ciphertext bits. Example in Table 1.

Bias			Linear Combination				
-2.49	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,31]$	$\oplus$	$K[4,9,13,31,33,41,44,52,54]$
4.86	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,31]$	$\oplus$	$K[4,9,13,31,33,41,44,47,52,54]$
-4.68	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,28]$	$\oplus$	$K[4,9,15,31,33,41,44,52,54]$
4.81	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,28]$	$\oplus$	$K[4,9,15,31,33,41,44,47,52,54]$
-2.18	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,28,29,31]$	$\oplus$	$K[9,13,15,31,33,41,44,47,52,54]$
-3.67	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,28,31]$	$\oplus$	$K[4,9,13,15,31,33,41,44,47,52,54]$
-4.59	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,30]$	$\oplus$	$K[4,9,30,31,33,41,44,52,54]$
2.63	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,30]$	$\oplus$	$K[4,9,30,31,33,41,44,47,52,54]$
2.3	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,29,30,31]$	$\oplus$	$K[9,13,30,31,33,41,44,52,54]$
2.69	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,29,30,31]$	$\oplus$	$K[9,13,30,31,33,41,44,47,52,54]$
3.77	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,30,31]$	$\oplus$	$K[4,9,13,30,31,33,41,44,52,54]$
3.23	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,30,31]$	$\oplus$	$K[4,9,13,30,31,33,41,44,47,52,54]$
2.43	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,27,28,29,30]$	$\oplus$	$K[9,15,30,31,33,41,44,47,52,54]$
-3.33	$\times$	10^{-4}	$P_H[15]$	$\oplus$	$P_L[0,7,18,24,28,30]$	$\oplus$	$K[4,9,15,30,31,33,41,44,52,54]$

Table 1: Ciphertext bits combination: $C_L[12,16]\oplus C_H[7,18,24]$

4 Soft decoding to reconstruct the key

- given relations $\langle C,\gamma_i\rangle+\langle P,\pi_i\rangle+\langle K,\kappa_i\rangle=0$ with probability $p_i=1/2+\varepsilon_i$, $i=1,\ldots,l$,
- determine vectors $\Delta_0,\Delta_1,\ldots,\Delta_t$ such that $\kappa_1,\ldots,\kappa_l\in\Delta_0+Vect(\Delta_1,\ldots,\Delta_t)$ (Table 1: $\Delta_0=K[9,31,33,41,44,52]$, $\Delta_1=K[4]$, $\Delta_2=K[13]$, $\Delta_3=K[15]$, $\Delta_4=K[30]$, $\Delta_5=K[47]$ and $\Delta_6=K[54]$).
- for $X=(X_1,\ldots,X_t)\in\{0,1\}^t$: $\Delta\cdot X\stackrel{def}{=}\Delta_0+\Delta_1X_1+\cdots+\Delta_tX_t$.
- *Goal:* given a sample of size s of plaintext-ciphertext pairs, associated with the key $\bar{K}$,determine $A(X)=\langle\bar{K},\Delta\cdot X\rangle\in RM(1,t)$. This will lead to the knowledge of the key bits $\bar{K}[\Delta_i]$, $i=0,\ldots,t$.
  - construct a codeword $Y(X)\in\mathbb{Z}$ using *log-probability* associated to the biases ε_i of the given relations: each κ_i corresponds to one position x_i ($Y(x_i)=0$ if no corresponding κ_i).
  - decode Y by maximizing $\sum_{X\in\{0,1\}^t}(-1)^{A(X)}Y(X)$
- Figure 1 : experimental success rate of the attack, depending on the size of the sample (between 500000 and 2000000). We also show the success rate of the attack if we set $y(x_i)=\pm1$ instead of the log-probability.

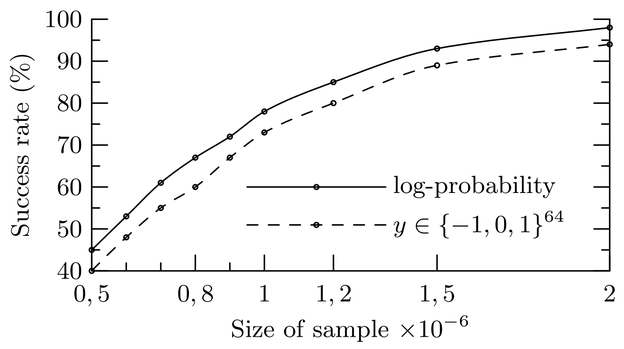

Figure 1: Experimental success rate of the attack.

5 Applications: Side channel attack

Cipher	Model	rounds	# linear equ.	# key bits	# Plaintexts	Pr(Success)
DES	HW	1	349	30	2^{10}	0.79
DES	HW	1	349	48	2^{12}	0.99
DES	HW	2	728	6	2^{9}	0.97
DES	HW	2	728	48	2^{12}	0.95
DES	HW	3	164	12	2^{17}	0.96
DES	HW	3	164	27	2^{20}	0.99
DES	HD	2	27	16	2^{14}	0.71
DES	HD	2	27	16	2^{16}	0.99
AES	HW	Last	1410	128	2^{10}	0.80
AES	HW	Last	1410	128	2^{11}	0.99

Simulation Results [14]

6 Conclusion

- The algorithm we designed enabled us to find many multiple linear approximations of 8 rounds of DES, or 5 rounds of Weight(DES_3(X)).
- These approximations can be used in improving the efficiency of linear cryptanalysis.
- The problem of finding more ciphertext masks, leading to linear approximations with good biases, remains open.
- original algorithm based on soft decoding that permits to reconstruct efficiently key bits.

Acknowledgement

This work was done in collaboration with the Pr. Ilya Dumer,Rafael Fourquet, Pr. Grigory Kabatiansky, Pierre Loidreau and Thomas Roche.

References

[1] A. Biryukov, C. De Cannière, and M. Quisquater. On multiple linear approximations. In M. Franklin, editor, *Advances in Cryptology – CRYPTO 2004*, volume 3512 of *Lecture Notes in Computer Science*, pages 1–22. Springer, 2004.

[2] B. Collard, F. X. Standaert, and J.-J. Quisquater. Experiments on the multiple linear cryptanalysis of serpent. In *Fast Software Encryption, FSE 2008*, 2008.

[3] B. Gérard and J.-P. Tillich. On linear cryptanalysis with many linear approximations. Technical report, INRIA, 2007. preprint.

[4] O. Goldreich and L. A. Levin. A hard core predicate for all one-way functions. In *Proceedings of the 21-st ACM Symposium on Theory of Computing*, pages 25–32, May 1989.

[5] O. Goldreich, R. Rubinfeld, and M. Sudan. Learning polynomials with queries: the highly noisy case. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pages 294–303, 1995. Extended version: http://people.csail.mit.edu/madhu/papers.html.

[6] T. Helleseth, T. Kløve, and V. Levenshtein. Bounds on the error-correcting capability of codes beyond half the minimum distance. In D. Augot, P. Charpin, and G. Kabatianski, editors, *Proceedings of the 3rd International Workshop on Coding and Cryptography, WCC 2003*, pages 243–251, 2003.

[7] T. Johansson and F. Jönsson. Fast correlation attacks through reconstruction of linear polynomials. In M. Bellare, editor, *Advances in Cryptology – CRYPTO 2000*, volume 1880 of *Lecture Notes in Computer Science*, pages 300–315. Springer, 2000.

[8] G. Kabatiansky and C. Tavernier. List decoding of Reed-Muller codes. In *Ninth International Workshop on Algebraic and Combinatorial Coding Theory, ACCT'2004*, pages 230–235, June 2004. http://ced.tavernier.free.fr/Balgaria.pdf.

[9] G. Kabatiansky and C. Tavernier. List decoding of first order Reed-Muller codes II. In *Tenth International Workshop on Algebraic and Combinatorial Coding Theory, ACCT'2006*, pages 131–134, September 2006. http://ced.tavernier.free.fr/Kabat.pdf.

[10] B. Kaliski and M. Robshaw. Linear cryptanalysis using multiple linear approximations. In Y. Desmedt, editor, *Advances in Cryptology – CRYPTO'94*, Lecture Notes in Computer Science, pages 26–39. Springer, 1994.

[11] M. Matsui. Linear cryptanalysis method for the DES cipher. In *Advanced in cryptology - EUROCRYPT'93*, volume 765 of *Lecture Notes in Computer Science*, pages 386–397. Springer, 1993.

[12] M. Matsui. The first experimental cryptanalysis of the Data Encryption Standard. In Y. Desmedt, editor, *Advances in Cryptology – CRYPTO'94*, Lecture Notes in Computer Science, pages 1–11. Springer, 1994.

[13] S. Murphy. The independance of linear approximations in symmetric cryptanalysis. *IEEE Transactions on Information Theory*, 52(12):5510–5518, December 2006.

[14] T. Roche and C. Tavernier. Multi-linear cryptanalysis in power analysis mlpa. arXiv:0906.0237.

[15] L. Trevisan. Some applications of coding theory in computational complexity. In *Electronic Colloquium on Computational Complexity*, number 43, 2004.