Accurate Estimate of the Advantage of Impossible Differential Attacks

Céline Blondeau

presented by Christina Boura

March 7, 2018

FSE 2018, Bruges

Outline

Outline

Introduction

Multivariate Distribution

Key-Recovery Attacks

Impossible Differential Cryptanalysis

- Defined at the end of the 90's as a generalization of differential cryptanalysis
- Given a cipher $E = E_0 \circ E' \circ E_1$
- A differential $(\delta_{\mathrm{x}}, \delta_{\mathrm{y}})$ over E' is impossible if

$$\forall k \in \mathcal{K} \;\; P_x[\; E'(x \oplus \delta_x) \oplus E'(x) = \delta_y \;] = 0$$

- Usually a set of differentials $(\delta_x, \delta_y) \in \Delta_{\mathrm{x}} \times \Delta_{\mathrm{y}}$ fulfill this property
- From this distinguisher on E', we can mount a key-recovery attack on E

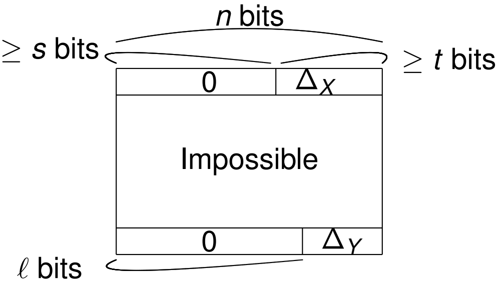
Complexity

The 3 phases of the key-recovery attack:

- Data generation: Generating pairs from a set of plaintexts
- Key sieving: Partial inversion with a selected number of potential candidate
- Exhaustive key search

- Recent publications: [BN-PS14], [Der16] Analyzing and minimizing the time complexity of the attack, with maximal focus on:
  - the data generation phase
  - and the key sieving phase
- This work: Providing a statistical analysis of the relation between the data complexity and the time complexity of:
  - the exhaustive key-search phase

Distinguishing Attack



- Δ_X and Δ_Y are linear (or affine) spaces
- $|\Delta_Y| = 2^{n-\ell}$
- A structure: subset of 2^t elements in Δ_X
- From a data complexity $N = 2^{s+t}$, we can generate $N_s = 2^{s+t}(2^t - 1)$ pairs

Classical Model: Binomial Distribution

- Statistical modeling similar to classical differential attacks
- Statistically a pair is a sample
- $[\ T\ =\ i\ ]$: the event that the differential(s) appears i times
- Given $p = |\Delta_Y| 2^{-n} = 2^{-\ell}$

Assuming a binomial distribution:

- For a random permutation,

$$
\begin{aligned}
P[\ T_{\mathcal{B}}\ =\ 0\ ] &= \binom{N_S}{0} p^0 (1-p)^{N_S} \\
&= (1-p)^{N_S} \approx \exp[\ -N_S\, p\ ]
\end{aligned}
$$

Advantage of an ID Distinguisher

- Advantage of a key recovery attack: number of won key-bits
- False alarm error probability: ratio of random permutations for which the differential(s) is impossible

Wrong key randomization hypothesis:

- Advantage of a distinguishing attack: $a = \log_2(P[\ T = 0\ ])$

Binomial distribution and its approximation

- Advantage estimate

$$\tilde{a}_{\mathcal{B}} = \frac{N_S}{\ln(2)} \left(2^{-\ell}\right)$$

Motivation and Contribution

- Experiments on 12-bit random permutations
- The data complexity is 2^{s+t}
- $\hat{a}$: the experimental advantage
- $a_{\mathcal{B}}$: classical advantage
- $a_{\mathcal{MH}}$: Advantage obtained with the theory developed in this paper

ℓ	s	t	$s+t$	$\hat{a}$	$a_{\mathcal{B}}$	$a_{\mathcal{MH}}$
7	2	3	5	1.25	1.27	1.25
7	4	3	7	4.99	5.07	4.99
9	4	3	7	1.11	1.26	1.11
9	6	3	9	4.44	5.05	4.44
9	8	3	11	17.73	20.22	17.77

Outline

Counting the Number of Pairs

To derive the new model: we do not manipulate pairs

$$E'(x \oplus \delta_x) \oplus E'(x) \in \Delta_y \Leftrightarrow \lfloor (E'(x \oplus \delta_x) \oplus E'(x) \rfloor_\ell = 0$$

- "Algorithm", inside a structure:
  - Create a vector L of ℓ bits
  - For all x, increment $L[\lfloor E'(x) \rfloor_\ell]$
  - Number of pairs is $S_j = \sum_{i=0}^{2^\ell - 1} \left(L[i](L[i]-1)\right)/2$
- Total number of pairs: Sum of S_j for each structure

Remarks:

- Inside a structure, the counting is similar to the counting in the multidimensional linear context
- Already used to show the relation between truncated differential and multidimensional linear attacks

Focussing on ONE Structure

- Focusing on one structure:

$$S_j = \sum_{i=0}^{2^\ell - 1} \left(L[i](L[i] - 1) \right) / 2$$

- Impossible differential attacks: No pairs $\Leftrightarrow$ each $L[i]$ should be equal to 0 or 1

- L follows a multivariate hypergeometric distribution

- If the structure has 2^t plaintexts, L should have:
  - $2^\ell - 2^t$ items equal to "0"
  - and 2^t items equal to "1"

$$P[\text{ No pairs }] = \frac{\binom{2^\ell}{2^t}}{\binom{2^n}{2^t}} (2^{n-\ell})^{2^t}$$

Multiple Structures

- Classical attacks: More than one structure
- If we assume independence between the structures we can derive the following estimate:

$$\tilde{a}_{MH} = \frac{N_S}{\ln(2)}\left(2^{-\ell} - 2^{-n}\right)$$

- To compare with the classical estimate

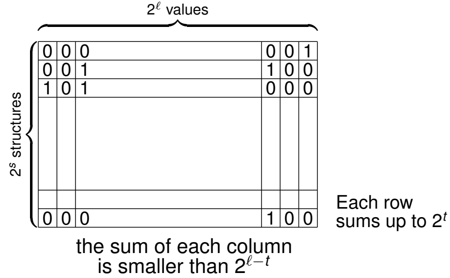
$$\tilde{a}_{B} = \frac{N_S}{\ln(2)}\left(2^{-\ell}\right)$$

- In general, the independence assumption is accurate as long as

$$N = 2^{s+t} \ll 2^{\ell}$$

- If we do not make this assumption, the model is more complicated and is based on the bi-multivariate hypergeometric distribution

Bi-Multivariate Hypergeometric Distribution: Maximal Advantage



- The maximal advantage of an impossible differential distinguisher (δ_X, δ_Y) is

$$a_{\max} = \frac{(2^{n-\ell}-1)(2^t-1)}{2\ln(2)}\left(1+\mathcal{O}(2^{-min(n,\ell+t)})\right)$$

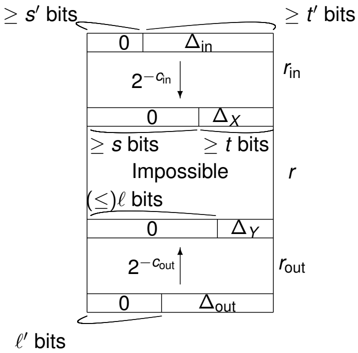
Outline

Introduction

Multivariate Distribution

Key-Recovery Attacks

Key-Recovery Attacks



- Attack on $r_{in} + r + r_{out}$ rounds
- Δ_{in} and Δ_{out} as the sets of all possible input respectively output differences
- $N^A = 2^{s'+t'}$: Data complexity of the key-recovery attack

- When more than one structure is involved, the data complexity of a distinguishing and a key recovery attack is the same

Experiments on a 16-bit Feistel

- Key-recovery attack on a 16-bit Feistel with 4 branches
- Taking an impossible differential distinguisher with $|\Delta_X| = |\Delta_Y| = 2^4$
- $\hat{a}$: experimental advantage
- $\bar{a}$: obtained in the paper
- $\tilde{a}_{\mathcal{B}}$: classical advantage

$\log(N)$	s'	t'	$\hat{a}$	$\bar{a}$	$\tilde{a}_{\mathcal{B}}$
10	0	10	0.51	0.51	0.68
11	0	11	2.53	2.54	2.70
12	0	12	10.14	10.14	10.82

$\log(N)$	s'	t'	$\hat{a}$	$\bar{a}$	$\tilde{a}_{\mathcal{B}}$
10	2	8	2.53	2.54	2.71
11	3	8	5.01	5.07	5.41
12	4	8	9.81	10.14*	10.82

- Left: 2 rounds before the distinguisher, 2 rounds after
- Right: 1 round before the distinguisher, 2 rounds after

*:non-accuracy: due to the non-independence of the structures

Only ONE Differential

- In the case of a single input differential ($t = 1$)
- $N_S = 2^{s+t-1}(2^t - 1)$ can not be estimated as $N_S \approx 2^{s+2t-1}$
- Maximal advantage: advantage using the full codebook
- Without the approximation, the maximal advantage is 0.72
- This advantage was previously estimated as 1.42
- The time complexity of the recent impossible differential attacks on SIMON is larger than estimated.

LBlock and CRYPTON

- Key-recovery attack on 23 rounds of LBlock of Boura *et al*
  - Data complexity of $2^{55.5}$
  - The time complexity has been computed for an advantage of 30.6 bits
  - Corrected advantage: 28.69 bits
- Key-recovery on 7 rounds of CRYPTON of Boura *et al*
  - Data complexity: $2^{114.9}$ known plaintexts
  - The time complexity has been computed for an advantage of 148.44 bits
  - Corrected advantage: 145.45 bits
- This result does not influence the overall time complexity since it is not dominated by the exhaustive key-search

Conclusion

- We analyze the advantage of impossible differential attacks
- We corrected it from

$$\frac{N_S}{\ln(2)}\left(2^{-\ell}\right)$$

to

$$\frac{N_S}{\ln(2)}\left(2^{-\ell}-2^{-n}\right)$$

- This result has an impact on the complexity of the exhaustive key search when ℓ is close to n
- We partially solve the problem of asymmetry between chosen plaintext and chosen ciphertext impossible differential attacks