Improving impossible differential cryptanalysis

Christina Boura, **Virginie Lallemand**,
María Naya-Plasencia, **Valentin Suder**

ESC Luxembourg, January 16, 2017

UNIVERSITÉ DE VERSAILLES ST-QUENTIN-EN-YVELINES
université PARIS-SACLAY

informatics mathematics
Inria

Impossible differential attacks

- Impossible differential cryptanalysis was introduced independently by Knudsen and Biham et al. in 1998.

Idea

- Find a differential with probability zero.
- Extend this differential by some rounds, possibly in both directions.
- Guess the keys bits that intervene in these rounds.
- If a pair is partially encrypted (decrypted) to the impossible differential, the guessed key bits are wrong.

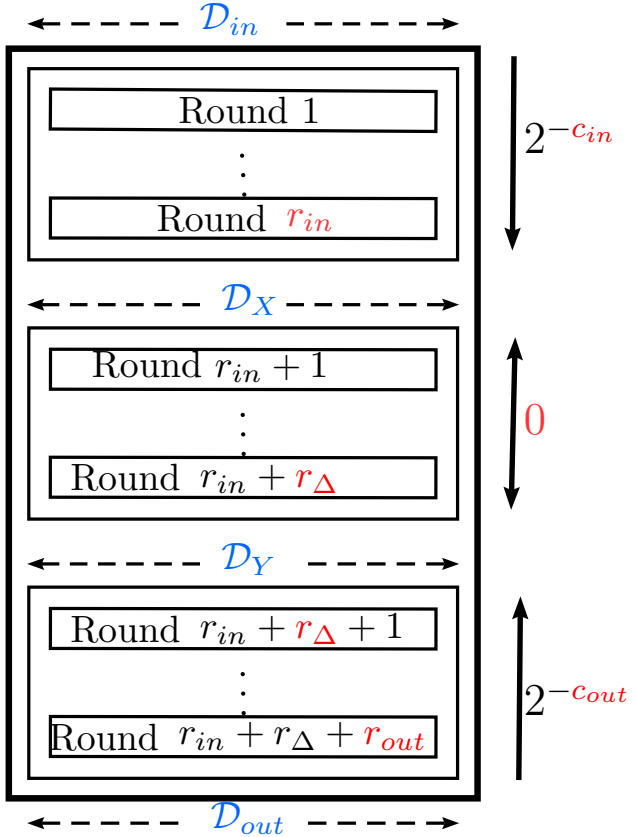
Our contributions

In ASIACRYPT 2014, Boura, Naya-Plasencia and Suder [BN-PS14] introduced a generic complexity analysis of impossible differential attacks.

In this work (to appear in the Journal of Cryptology):

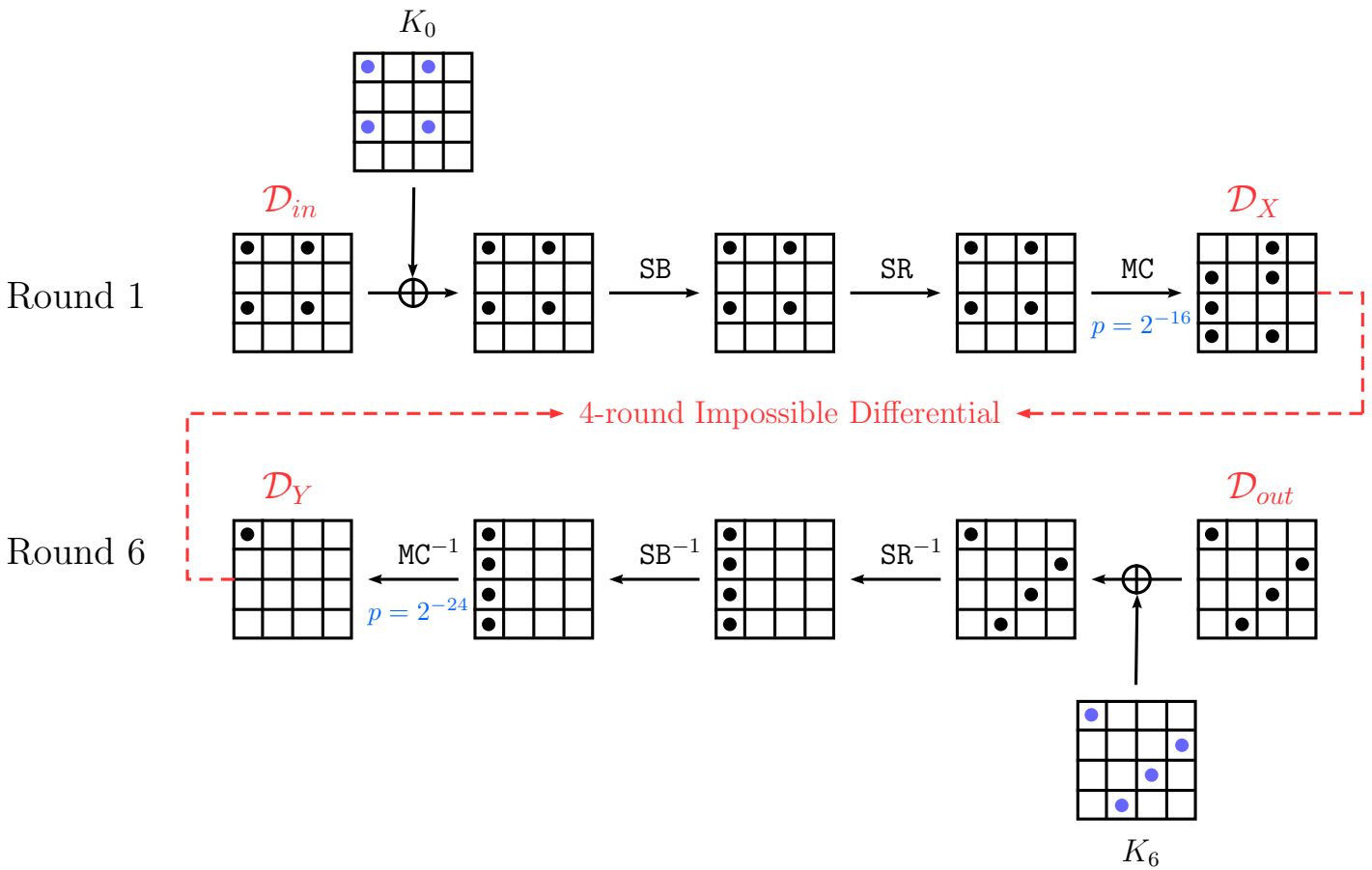
- Extend, correct and complete the analysis of [BN-PS14].
- Correct the time complexity appoximation by taking into account the role of the key schedule.
- New techniques for improving data complexity: Multiple differentials
- Experimental verifications of the introduced techniques.
- Multiple impossible differential vs. simple impossible differentials

Notation



- Δ_V: size in bits of a difference $\mathcal{D}_V$.
- c_{in}, c_{out} : number of bit conditions to be verified.
- k_{in}, k_{out} : number of involved subkey bits.
- $|k_{in} \cup k_{out}|$: key entropy

Example



- $\Delta_{in} = 32, \Delta_{out} = 32$
- $c_{in} = 16, c_{out} = 24$
- $k_{in} = 32, k_{out} = 32.$

How many pairs does an attack require?

By taking N pairs satisfying $(\mathcal{D}_{in}, \mathcal{D}_{out})$, the probability of not discarding a candidate key is

$$P = (1-2^{-(c_{in}+c_{out})})^N$$

How many pairs N are needed for the attack?

- First approach: $(1-2^{-(c_{in}+c_{out})})^N < 2^{-|k_{in} \cup k_{out}|}$
- Better approach: $(1-2^{-(c_{in}+c_{out})})^N < \frac{1}{2}$
- Take at least

$$N_{\mathsf{min}} = 2^{c_{in}+c_{out}}.$$

Memory complexity : N

Cost for finding N pairs

$$C_N = \max\left\{ \min_{\Delta\in\{\Delta_{in},\Delta_{out}\}} \left\{\sqrt{N2^{n-\Delta+1}}\right\}, N2^{n-\Delta_{in}-\Delta_{out}+1} \right\}.$$

Data complexity: C_N

Obviously,

$$C_N < 2^n$$

Time complexity

$$T_{\text{comp}} = C_N +$$

- Encrypt data.

Time complexity

$$T_{\text{comp}} = C_N + \left(N + 2^{|k_{in} \cup k_{out}|} \frac{N}{2^{c_{in}+c_{out}}}\right) C'_E$$

- Encrypt data
- Key sieving

Time complexity

$$T_{\text{comp}} = \left(C_N + \left(N + 2^{|k_{in} \cup k_{out}|} \frac{N}{2^{c_{in}+c_{out}}}\right) C'_E + 2^K P\right) C_E$$

- Encrypt data
- Key sieving
- Test by exhaustive search the remaining keys.

  The last term corresponds to $2^K P = 2^{K-|k_{in} \cup k_{out}|} P 2^{|k_{in} \cup k_{out}|}$.

Outline

The role of the key schedule

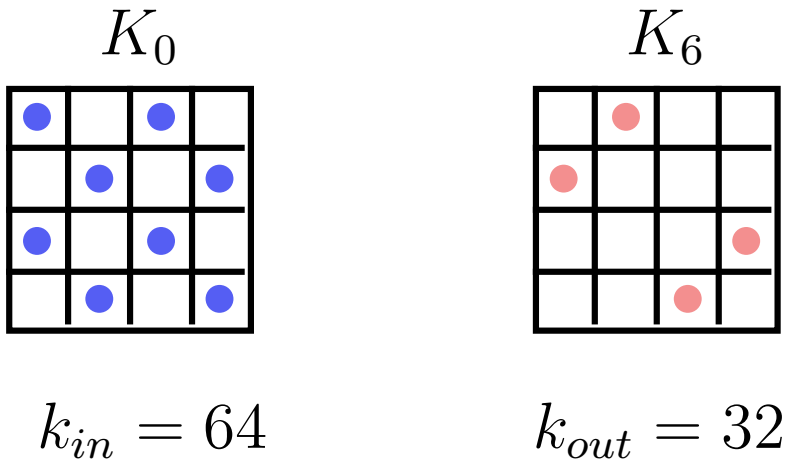
During the key-recovery phase, key bits of different subkeys are guessed.

- How to recover the master key from these guessed bits?

  This depends on the nature of the key-schedule.

- If the key-schedule is (almost) linear, directly translate the k_{in} and k_{out} guessed bits in the same number of bits of the master key.

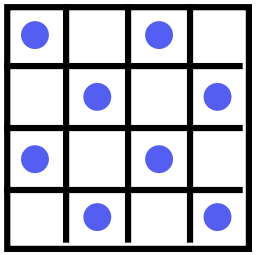
Complex key schedules



If the key-schedule is **complex**, it is not possible to directly translate the information guessed on the subkeys into the same amount of information on the master key.

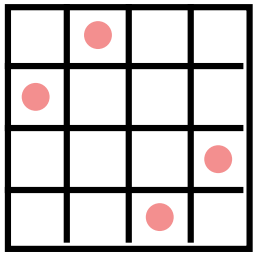
What do we do then?

K_0



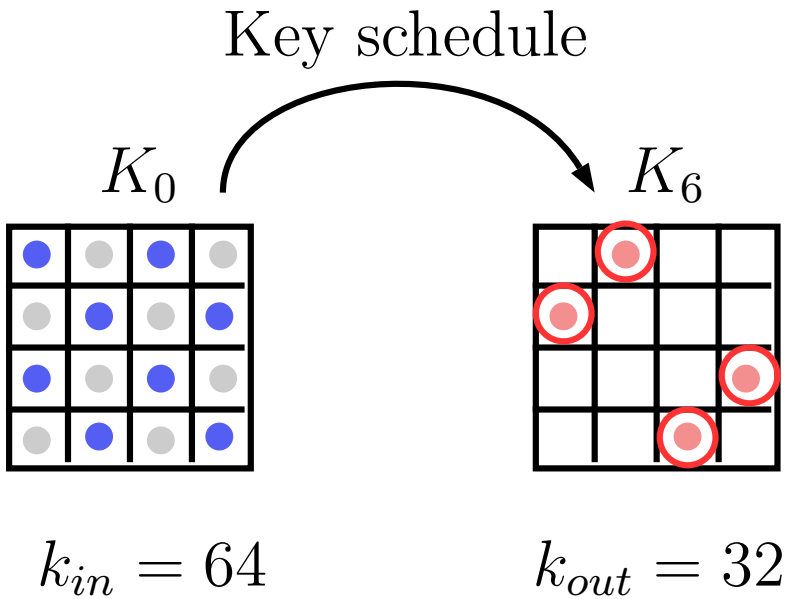
$k_{in} = 64$

K_6



$k_{out} = 32$

What do we do then?



- Complete the missing bits to some of the subkeys.
- Compute through the key schedule.
- Verify if the result matches.

How is the time complexity affected?

- The part of the key schedule connecting the subkeys of the first to the subkeys of the last rounds can be seen as a black box.
- A new term has to be added to the time complexity formula.
- Partition the key bits into two sets k_A and k_B.

$$\min(2^{K-k_A}, 2^{K-k_B}) \cdot P \cdot 2^{k_A+k_B} C_{KS},$$

where C_{KS} is the key schedule cost.

How is the time complexity affected?

- The part of the key schedule connecting the subkeys of the first to the subkeys of the last rounds can be seen as a black box.
- A new term has to be added to the time complexity formula.
- Partition the key bits into two sets k_A and k_B.

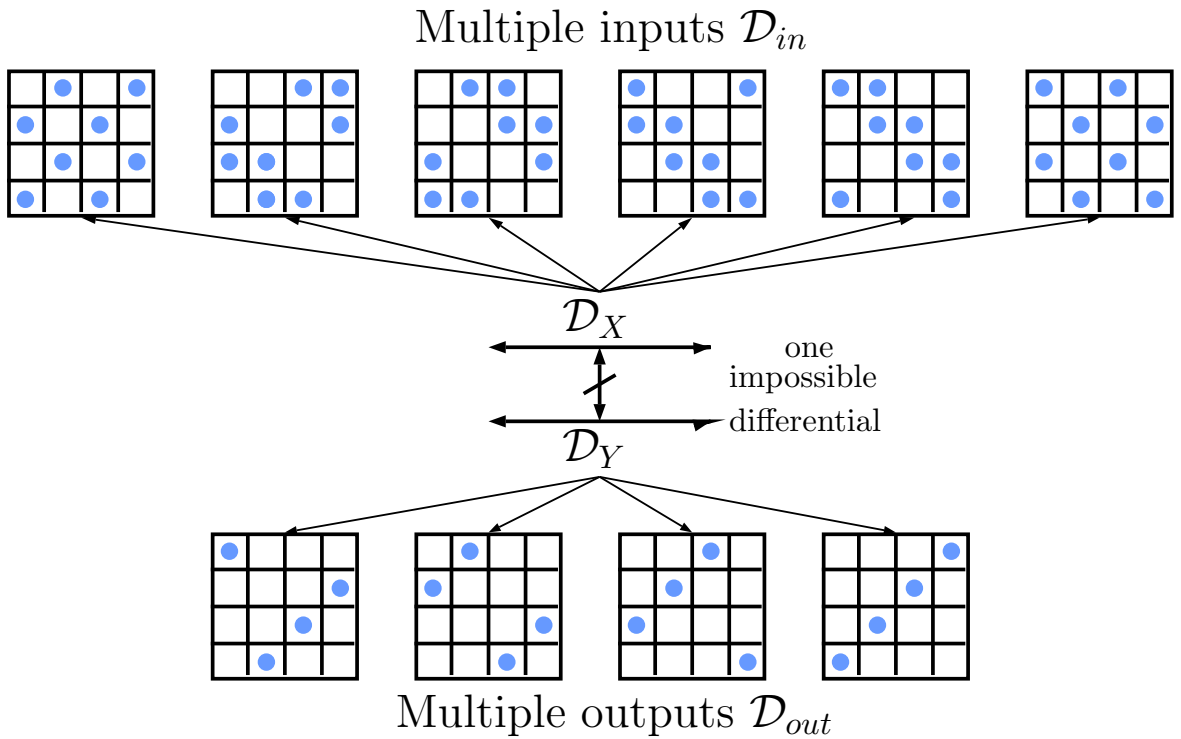
$$\min(2^{K+k_A}, 2^{K+k_B}) \cdot P \cdot C_{KS},$$

where C_{KS} is the key schedule cost.

In previous works, it was wrongly supposed that one guessed word of a subkey could directly be seen as one guessed word of the master key.

Multiple differentials in impossible differential cryptanalysis

- More choices for the input/output patterns of a pair.
- Less data is needed to construct the pairs for the attack $\rightarrow$ reduction in the data complexity

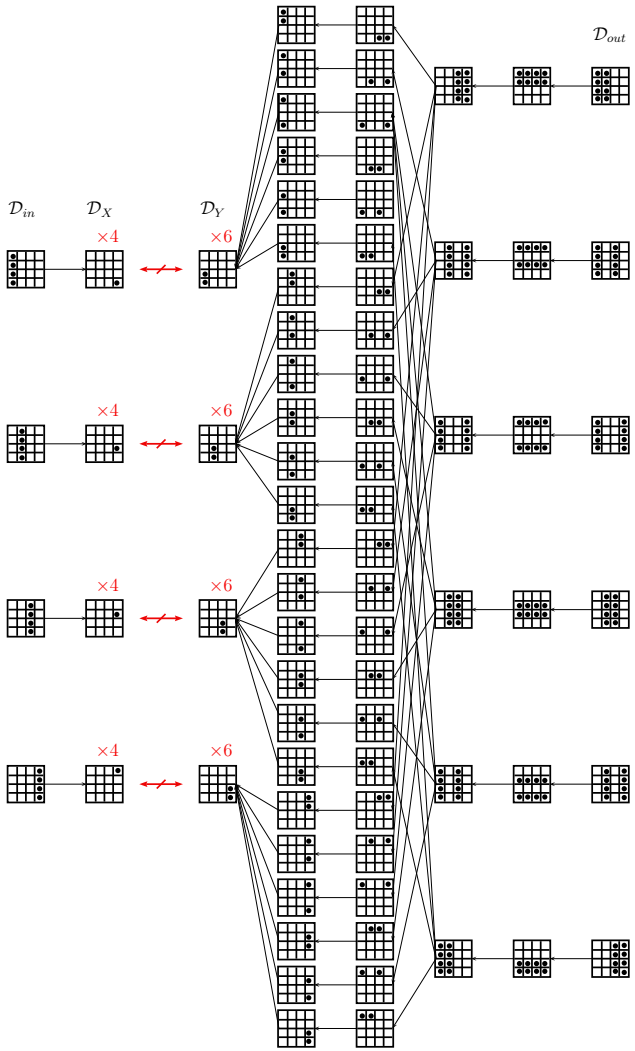


- The $\log_2$ of the data complexity is reduced by

  # input multiples + # output multiples.

Combine multiple impossible diff. with multiple diff.

Use multiple differentials and multiple impossible differentials together to further reduce the amount of data.



New data complexity

- m_{in}: Number of input multiple differentials
- m_{out}: Number of output multiple differentials
- n_{in}: Number of input differences $\mathcal{D}_X$
- n_{out}: Number of output differences $\mathcal{D}_Y$

The new data complexity C'_N

$$C'_N = \frac{C_N}{m_{in} m_{out} n_{in} n_{out}}$$

- In practice, multiple differentials and multiple impossible differentials can be treated in the same way.

Practical verification of the new data complexity formula

Setting

- 6-round, 32-bit Feistel toy cipher
- **Round function**: SPN, with 4-bit PRESENT Sbox, and LED's MDS linear transformation
- 4-round impossible differential, extended one round forwards and one round backwards.

Two cases depending on the probability P of not discarding a key

1. $P = 1/2$
   - Multiple outputs

$\#\ \mathcal{D}_{out}$	1	2	3	4
Theoretical value of $\log_2 C_N$	16.5	15.5	14.9	14.5
Experimental value of $\log_2 C_N$	16.5	15.4	14.8	14.6

1. Multiple inputs
   - The theoretical complexities are matched only if the amount and form of needed pairs allow to optimally exploit the plaintext structures and will be slightly reduced otherwise.
2. P very small
   - Corresponding amount of pairs slightly increased because of the higher number of key bits being involved.

Outline

Results on SPN ciphers

Algorithm	Rounds	Data (CP)	Time	Memory (Blocks)	Tech.	Ref.
AES-128	7	$2^{106.2}$	$2^{110.2}$	$2^{90.2}$	ID	[MDRM 10]
	7	2^{105}	$2^{105} + 2^{99}$	2^{90}	MITM	[DFJ 13]
	7	2^{97}	2^{99}	2^{98}	MITM	[DFJ 13]
	7	2^{105}	$2^{106.88}$	2^{74}	ID	
CRYPTON-128	7	2^{121}	$2^{121} + 2^{116.2}$	2^{119}	ID	[MSD 10]
	7	$2^{114.92}$	$2^{114.92} + 2^{113.7}$	$2^{88.5}$	ID	
ARIA-128	6	2^{113}	$2^{121.6}$	2^{113}	ID	[LSZL 08]
	6	2^{120}	$2^{120} + 2^{96}$	2^{120}	ID	[LS 08]
	6	2^{111}	$2^{111} + 2^{82}$	2^{71}	ID	

Results on Feistel ciphers

Algorithm	Rounds	Data (CP)	Time	Memory (Blocks)	Tech.	Ref.
CLEFIA-128	13	$2^{114.58}$	$2^{116.16}$	$2^{83.16}$	ID	[BM-PS 14]
	13	$2^{114.4}$	$2^{114.4}$	2^{80}	ID	
Camellia-256	13	2^{123}	$2^{251.1}$	2^{203}	ID	[BM-PS 14]
	13	$2^{119.71}$	$2^{225.06}$	$2^{198.71}$	ID	
LBlock	23	2^{59}	$2^{75.36}$	2^{74}	ID	[BN-PS 14]
	23	$2^{63.87}$	$2^{74.30}$	2^{60}	ZC	[BM 14]
	23	$2^{55.5}$	2^{72}	2^{65}	ID	