Structural Evaluation of AES and Chosen-Key Distinguisher of 9-round AES-128

Thomas Peyrin

joint work with Pierre-Alain Fouque and Jérémy Jean (CRYPTO 2013)

NTU - Singapore

ISCAS Seminar

Beijing, China - October 23, 2013

NANYANG TECHNOLOGICAL UNIVERSITY

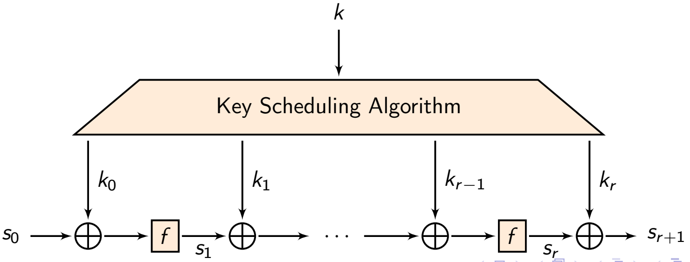
Outline

Outline

Block Ciphers

Iterated SPN Block Ciphers

- Internal Permutation : f
- Number of Iterations : r
- SPN : $f = \text{P} \circ \text{S}$ applies Substitution (S) and Permutation (P).
- Secret Key : k
- Key Scheduling Algorithm : $k \rightarrow (k_0, \ldots, k_r)$
- Ex : AES, PRESENT, SQUARE, Serpent, etc.

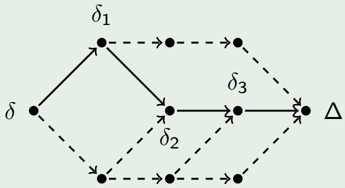


Differentials and Differential Characteristics

Differential (Characteristics)

- Used in differential cryptanalysis
- Sequence of differences at each round for an iterated primitive.
- A differential is a collection of characteristics.

Examples



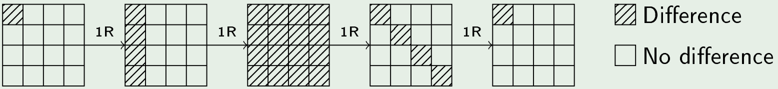
- $\delta \rightarrow \Delta$ is a differential.
- $\delta \rightarrow \delta_1 \rightarrow \delta_2 \rightarrow \delta_3 \rightarrow \Delta$ is a differential characteristic.
- $\mathbb{P}(\delta \rightarrow \delta_1 \rightarrow \delta_2 \rightarrow \delta_3 \rightarrow \Delta)$ is its differential probability.

Differentials and Differential Characteristics

Differential Characteristics

- Differential characteristics are easier to handle than differentials
  $\Longrightarrow$ We usually focus on characteristics
- Designers' goal : upper-bound the differential probability of characteristics.

Example : 4-round AES



- 4-round characteristic with 25 active S-Boxes (minimal).
- AES S-Box : $p_{max} = 2^{-6}$.
- Differential probability : $p \leq 2^{-6 \times 25} = 2^{-150}$.

AES

Design of the AES

- AES Permutation : structurally bounded diffusion for any rounds
- Provably resistant to Single-Key differential attacks
- Very easy get the bounds by hand (just using the fact that the MixColumns matrix is MDS)

Minimal Number of Active S-Boxes for AES in the SK model

Rounds	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
min	1	5	9	25	26	30	34	50	51	55

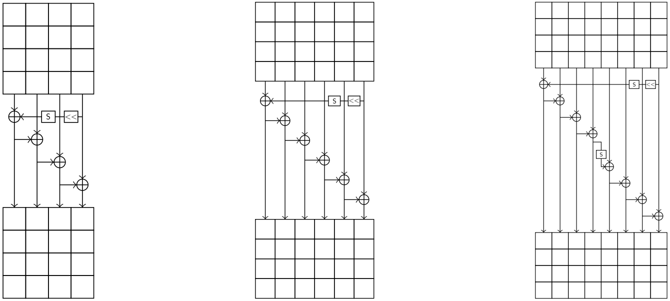
Question

What would this table look like for the AES structure in the RK model ?

AES key schedule

Design of the AES key schedule

- Ad-hoc key schedule
  $\Longrightarrow$ RK Attacks for AES-192/256 [BKN-C09], [BK-A09], [BN-E10].
- hard to analyze, so far no simple proof/analysis exist, except the computer-based ones.



(a) AES−128. (b) AES−192. (c) AES−256.

Related-key attacks

Why studying related-keys attacks ?

- some protocols might use simple updates to generate new keys
- RK analysis helps to understand hash functions
- in the ideal case, the cipher shouldn't have any structural flaw, so we can even extend the SK/RK model to known-key/chosen-key analysis

Our current knowledge for building key schedules/message expansion is sparse

- AES has a rather efficient key schedule (about 25% to 40% of the internal permutation part), but no clue about its security
- in order to get simple provable confidence in the key schedule, designers proposed inefficient solutions :
  - Whirlpool has a very strong message expansion, but then one round is not efficient
  - LED has no key schedule, but requires more rounds to resist RK

Our Contributions

Main contribution

We propose an algorithm finding all the "smallest" RK characteristics :

- runs in time linear in the number of rounds, exponential in the state size (previous algorithms are exponential in both)
- for AES−128, requires a few hours on a single PC instead of several days previously
- for AES−128, depending on the output required, memory usually ranges from 0.5GB to 60GB (100 GB in the worst case where one wants **all** the best characteristics)

Side results for AES−128

- we provide the first chosen-key distinguisher for 9-round AES−128
- AES−128 can not be proven secure against RK attacks with structural arguments only
- best RK characteristic for 5 rounds AES−128 has probability 2^{-105} (not 2^{-102} as previously believed)

Outline

Outline

Existing Algorithms (1/2)

Matsui's Algorithm (e.g. DES)

- Works by **induction** :
  derive best n-round char. from best chars. on $1, \ldots, n-1$ rounds
- Compute best char. for 1R
- Traverse a **tree** of depth 2 for 2R
- Pruning possible (A^* optim.)

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \to \Delta_j)$$

Δ_1

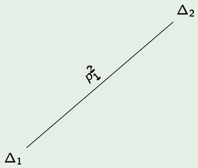
Existing Algorithms (1/2)

Matsui's Algorithm (e.g. DES)

- Works by **induction** :
  derive best n-round char. from best chars. on $1, \ldots, n-1$ rounds
- Compute best char. for 1R
- Traverse a **tree** of depth 2 for 2R
- Pruning possible (A^* optim.)

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \rightarrow \Delta_j)$$



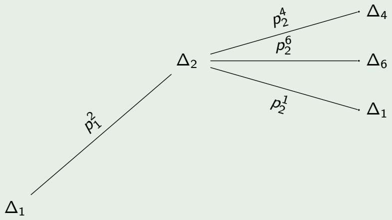
Existing Algorithms (1/2)

Matsui's Algorithm (e.g. DES)

- Works by **induction** : derive best n-round char. from best chars. on $1, \dots, n-1$ rounds
- Compute best char. for 1R
- Traverse a **tree** of depth 2 for 2R
- Pruning possible (A^* optim.)

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \to \Delta_j)$$



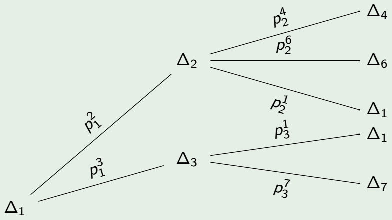
Existing Algorithms (1/2)

Matsui's Algorithm (e.g. DES)

- Works by **induction** : derive best n-round char. from best chars. on $1, \ldots, n-1$ rounds
- Compute best char. for 1R
- Traverse a **tree** of depth 2 for 2R
- Pruning possible (A^* optim.)

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \to \Delta_j)$$



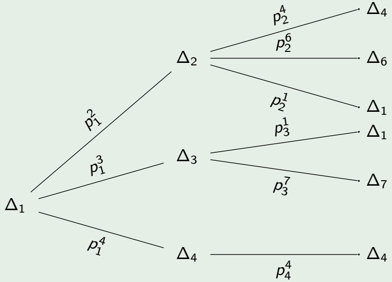
Existing Algorithms (1/2)

Matsui's Algorithm (e.g. DES)

- Works by **induction** :
  derive best n-round char. from best chars. on $1, \ldots, n-1$ rounds
- Compute best char. for 1R
- Traverse a **tree** of depth 2 for 2R
- Pruning possible (A^* optim.)

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \to \Delta_j)$$



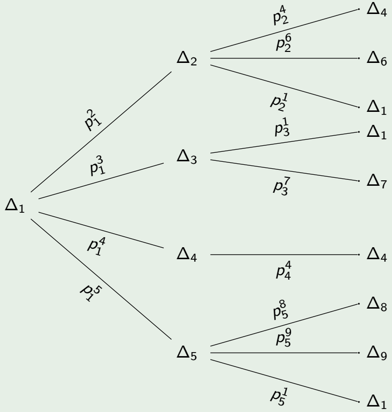
Existing Algorithms (1/2)

Matsui's Algorithm (e.g. DES)

- Works by **induction** : derive best n-round char. from best chars. on $1, \ldots, n-1$ rounds
- Compute best char. for 1R
- Traverse a **tree** of depth 2 for 2R
- Pruning possible (A^* optim.)

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \to \Delta_j)$$



Existing Algorithms (1/2)

Matsui's Algorithm (e.g. DES)

- Works by **induction** : derive best n-round char. from best chars. on $1, \ldots, n-1$ rounds
- Compute best char. for 1R
- Traverse a **tree** of depth 2 for 2R
- Pruning possible (A^* optim.)

Pros

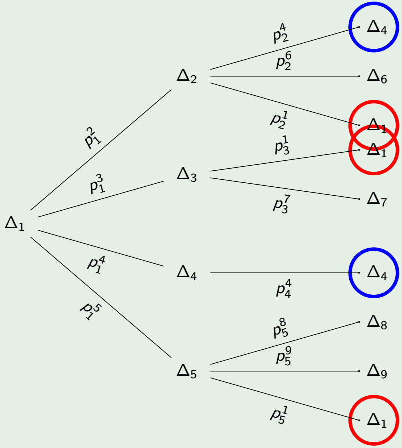
- works on DES in single-key

Drawbacks

- Rely on non-equivalent differential probabilities : needs dominant characteristic(s)
- Poor performances for AES
- Differences visited several times

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \to \Delta_j)$$



Existing Algorithms (2/2)

Biryukov-Nikolic [BN-E10]

- Adapt Matsui's algorithm
- Different algos for several KS

Pros

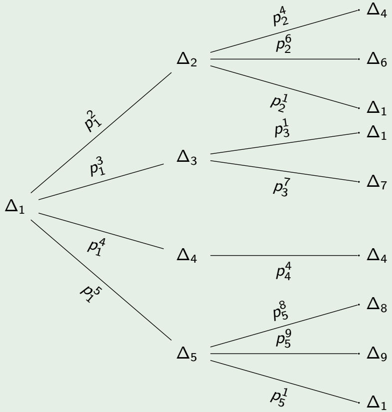
- Switch to truncated differences
  $\Longrightarrow$ less edges
- Representation of trunc. differences
  $\Longrightarrow$ handle branching in the KS
- Works on AES

Cons

- Not that fast because AES−128 has no predominant char.
- Differences visited several times
- Nodes visited exponential in the number of rounds

Tree Example

$$p_i^j \stackrel{\text{def}}{=} \mathbb{P}(\Delta_i \rightarrow \Delta_j)$$

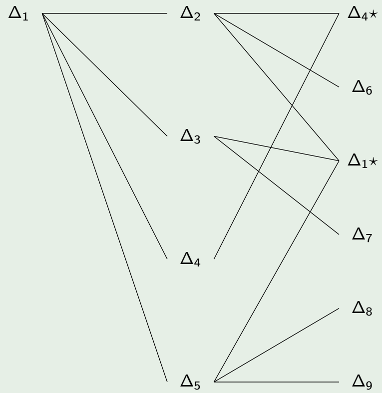


Our Algorithm

Algorithm

- Switch to a graph representation
- Merge equal diff. of the same round
- Graph traversal similar as Dijkstra
- Path search seen as Markov process

Graph Example

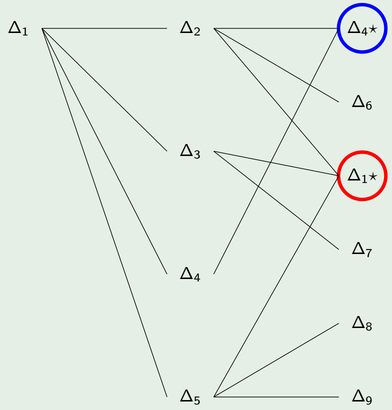


Our Algorithm

Algorithm

- Switch to a graph representation
- Merge equal diff. of the same round
- Graph traversal similar as Dijkstra
- Path search seen as Markov process

Graph Example

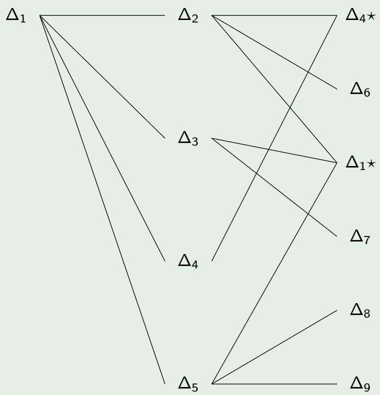


Our Algorithm

Algorithm

- Switch to a graph representation
- Merge equal diff. of the same round
- Graph traversal similar as Dijkstra
- Path search seen as Markov process

Graph Example



Our Algorithm

Algorithm

- Switch to a graph representation
- Merge equal diff. of the same round
- Graph traversal similar as Dijkstra
- Path search seen as Markov process

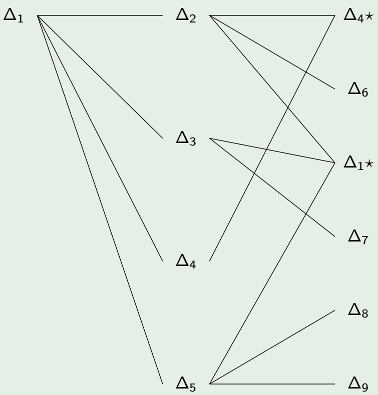
Pros

- Each difference in each round is visited only once
- Numbers of nodes and edges are linear in the number of rounds
- A^* optimization still applies

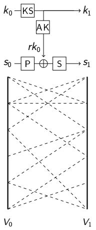
Notes

- Only partial information propagated
- Need to adapt the Markov process

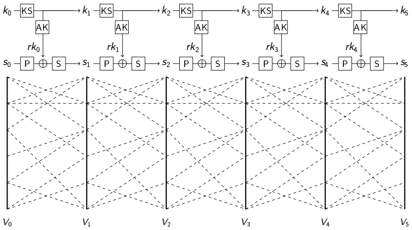
Graph Example



The graph G

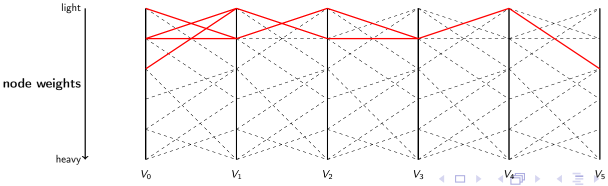


(d) Graph G.



(e) Graph G_5.

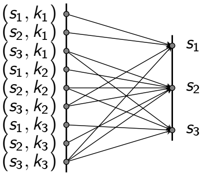
G is a bipartite directed acyclic graph, with the weight on the nodes



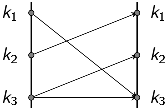
Implementation tricks

Implementation tricks

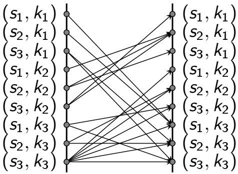
- we store only the graph G for one round, the entire graph is obtained by repeating G.
- instead of storing a huge graph G of all the best differential transitions for one round, we store separate graphs G_{BC} and G_{KS}. Then, G can be obtained by making the product of G_{BC} and G_{KS}.



(f) Graph G_{BC}.



(g) Graph G_{KS}.



(h) Graph G.

Outline

Outline

1. Motivations
2. Algorithms
3. Application to AES−128
   - Truncated differences
   - Actual differences
4. Distinguishing 9R AES−128
5. The End

Outline

Application to the Structure of AES−128

Structural Analysis

- We ignore the semantic definition of the S-Box and the MDS matrix
- We count the number of active S-Boxes (truncated differences)
- Do not apply to AES−128 with the instantiated S and P
- Give an estimation of the structural quality of the AES family

Related-Key Model (XOR difference of the keys)

Rounds	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
min	0	1	3	9	11	13	15	21	23	25

Hash Function Setting (KS considered independently)

Rounds	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
minmax	0	1	3	6	7	9	11	14	15	17

Examples of best truncated differential characteristics

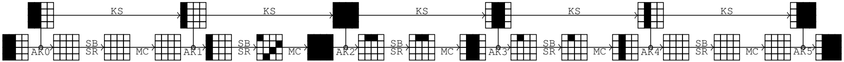


Figure: Best truncated differential characteristics for AES−128 when $r = 5$ rounds with 11 active Sboxes.

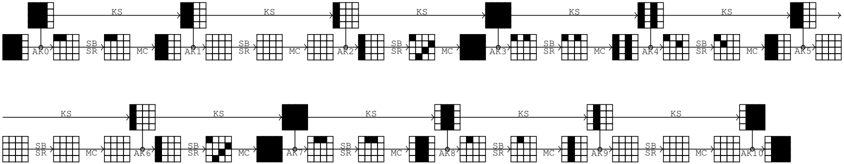


Figure: Best truncated differential characteristics for AES−128 when $r = 10$ rounds with 25 active Sboxes.

Impossibility Results for the Structure of AES−128

There exists a characteristic on 10 rounds with only 25 active S-Boxes
$\Longrightarrow$ best RK differential attack in p_{max}^{-25} computations.

Result 1

It is impossible to prove the security of the full AES−128 against **related-key differential attacks** without considering the differential property of the S-Box.

Notes

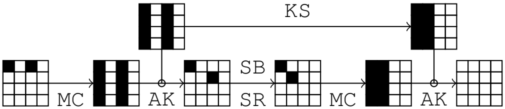
- With a random S-Box, p_{max}^{-25} might be smaller than 2^{128}
  $\Longrightarrow$ when $p_{max} \geq 2^{-5}$
- AES structure on its own not enough for RK security
- For a specified S-Box with bounded $p_{max} \leq 2^{-6}$
  $\Longrightarrow$ security against RK attacks

Outline

Markov process and filtering

Example of linear incompatibility in the case of AES−128 :

The linearity of the key schedule imposes all the active columns $[a, b, c, d]^T$ to be equal, which contradicts the first key addition (AK) $\mathbf{M} \cdot [x, 0, 0, 0]^{\mathrm{T}} \oplus [x', 0, 0, 0]^{\mathrm{T}} = \mathbf{M} \cdot [y, 0, 0, 0]^{\mathrm{T}} \oplus [0, y', 0, 0]^{\mathrm{T}}$.



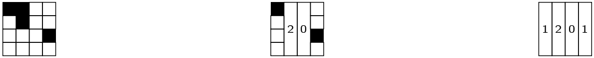
Post-filtering

The problem with Markov process is that we loose all information from the past (how did I get to this difference ?) ... which is exactly what we need to detect the incompatibilities.
We can still apply a filter on the output of the diff. characteristic search algorithm : test all the paths one by one and try to instantiate them.

State compression

State compression

Example of compressed truncated state and semi-compressed truncated state from a truncated state



(a) Truncated state. (b) Semi-compressed state. (c) Compressed state.

Dilemma

- if we compress the state too much, there will be too many inconsistent path, the filtering process will be too long
- if we don't compress enough, the differential characteristic search will be too long (or require too much memory)

Related-Key attacks on AES−128

RK attacks against AES−128

- After 6 rounds, there is no RK characteristic for AES−128 with a probability greater than 2^{-128}.
- For $1, \ldots, 5$ rounds, our algorithm has found the best characteristics
- Same truncated characteristics as [BN-E10]
- Best instantiations of differences : maximal probabilities.

Best bounds on RK attacks for AES−128

Rounds	**1**	**2**	**3**	**4**	**5**
#S-Boxes	0	1	5	13	17
[BN-E10]	0	-6	-30	-78	-102
max $\log_2(p)$	0	-6	-31	-81	-105

Outline

Outline

Distinguishing model [KR-A07, BKN-C09]

Solve Open-Problem

We can use the best 5-round characteristic to construct a chosen-key distinguisher for 9-round AES−128.

Let E_k be the 9-round AES−128 block cipher using key k.

Limited Birthday Problem [GP-FSE10]

Given

- a fully instantiated difference δ in the key,
- a partially instantiated difference Δ_{IN} in the plaintext,
- a partially instantiated difference Δ_{OUT} in the ciphertext,

find

- a key k,
- a pair of messages (m, m'),

such that :

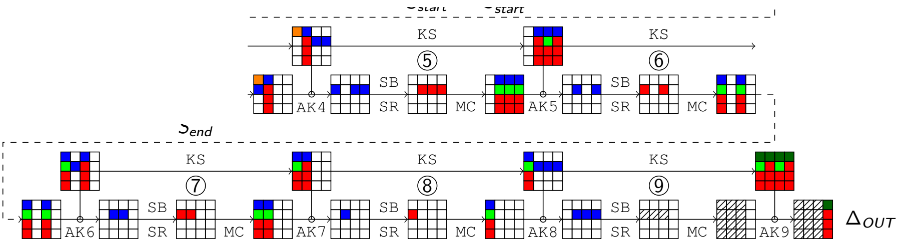
$$m \oplus m' \in \Delta_{IN}$$

$$\text{and : } E_k(m) \oplus E_{k \oplus \delta}(m') \in \Delta_{OUT}.$$

9-Round characteristic for AES−128

Construction of the characteristic

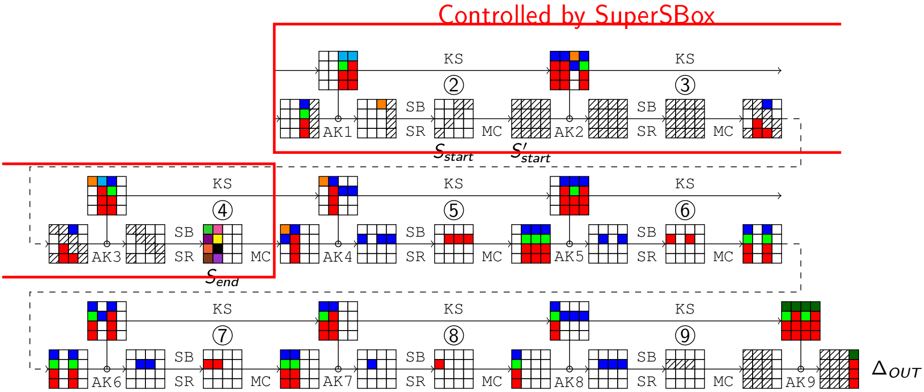
Take the best 5-round characteristic for AES−128 we have found.



9-Round characteristic for AES−128

Construction of the characteristic

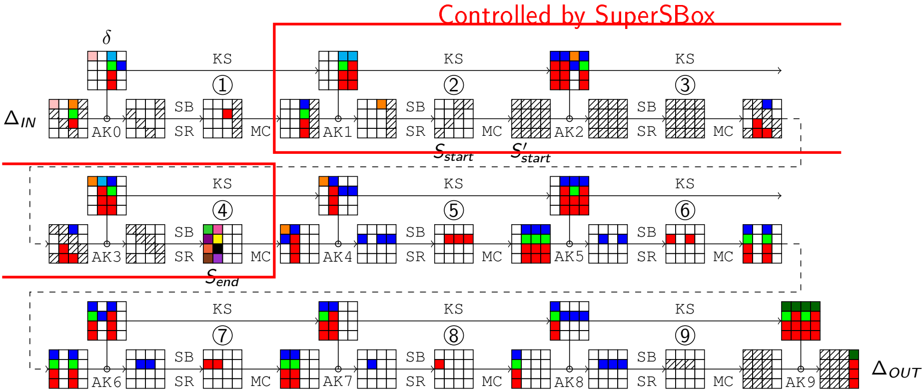
Prepend three rounds to be controlled by the SuperSBox technique.



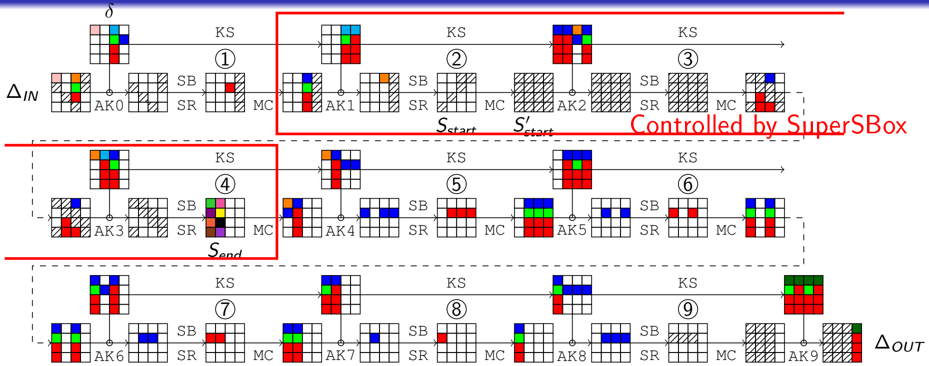
9-Round characteristic for AES−128

Construction of the characteristic

Prepend one other round, as inactive as possible.



9-Round CK Distinguisher for AES−128



Distinguishing algorithm

- Generate 2^{15} valid pairs of keys (about 2^{27} of them exist, since $\mathbb{P}_{KS} = 2^{-101}$)
  - Store the ith SuperSBox from S'_{start} to S_{end} in T_i (costs 2^{32})
  - For all 5 differences at S_{start} (costs 2^{40}), check the tables and :
    - Check backward direction : $p = 2^{-7}$ (a single S-Box)
    - Check forward direction : $p = 2^{-6\times 8} = 2^{-48}$ (8 S-Boxes)

Time complexity

Complexity of the distinguishing algorithm

- Check probability : $2^{-7-48} = 2^{-55}$
- Time complexity :

$$2^{15} \times (2^{32} + 2^{40}) \approx 2^{55} \text{ computations}$$

- For 2^{15} different pairs of keys :
  - Construct the SuperSBoxes in 2^{32} operations
  - Try all values for the 5 byte-differences in 2^{40} operations

Generic time complexity

- Limited-Birthday Problem [GP-FSE10]
- Input space (Δ_{IN}) of size $4 \times 8 + 7 = 39$ bits
- Output space (Δ_{OUT}) of size $3 \times 7 = 21$ bits
- Time complexity : 2^{68} encryptions

Outline

Outline

Conclusion

- New differential characteristics finding algorithm for SPN ciphers
  - Graph-based approach : Dijkstra and A^* optimization
  - Search the best truncated differential characteristics
  - Time complexity linear in the number of rounds considered
- Applications to the structure of AES−128 :
  - Impossibility results for related-key attacks
  - Impossibility results for the hash function setting
  - Exact probabilities for the best differential characteristics (eg. 2^{-105} for 5 rounds)
- Chosen-key distinguisher for 9-round AES−128
  - Solve open problem
  - Time Complexity : 2^{55} encryptions
  - Generic Complexity : 2^{68} encryptions
- More details in the paper and its extended version (ePrint/2013/366)

Thank you for your attention !

We are looking for good PhD students
in symmetric key crypto.

If interested, please contact me at :
thomas.peyrin@ntu.edu.sg

NANYANG TECHNOLOGICAL UNIVERSITY