Two Attacks on a White-Box AES Implementation

Tancrède Lepoint[1,2], Matthieu Rivain[1],
Yoni De Mulder[3], Peter Roelse[4], and Bart Preneel[3]

[1] CryptoExperts, France
{tancrede.lepoint,matthieu.rivain}@cryptoexperts.com

[2] École Normale Supérieure, France

[3] KU Leuven and iMinds, Belgium
{yoni.demulder,bart.preneel}@esat.kuleuven.be

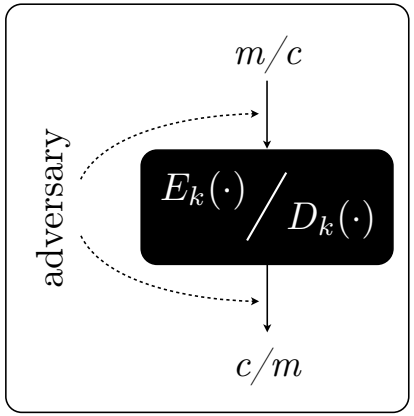
[4] Irdeto B.V., The Netherlands
peter.roelse@irdeto.com

SAC 2013, Vancouver, Canada *August 15th, 2013*

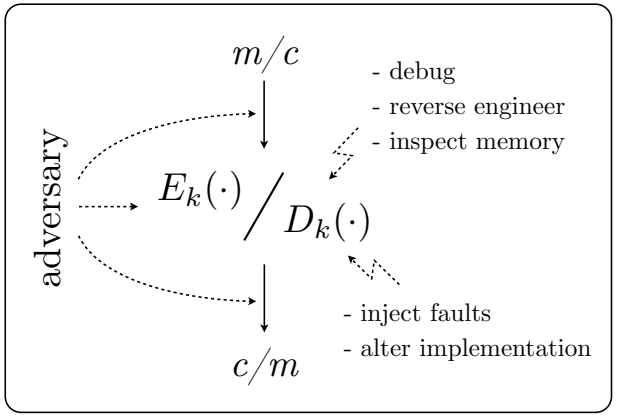
What is White-Box Cryptography ?

White-Box Cryptography

- focuses on the software implementation of cryptographic primitives executed in an **untrusted** environment.
- aims at protecting the embedded secret cryptographic key; it has the objective that the white-box implementation behaves as a **"virtual black box"**:
  - a white-box adversary may not have any advantage over a black-box attacker, i.e., he is unable to extract any more key information than he could extract under a black-box attack (oracle access to the WB implementation).



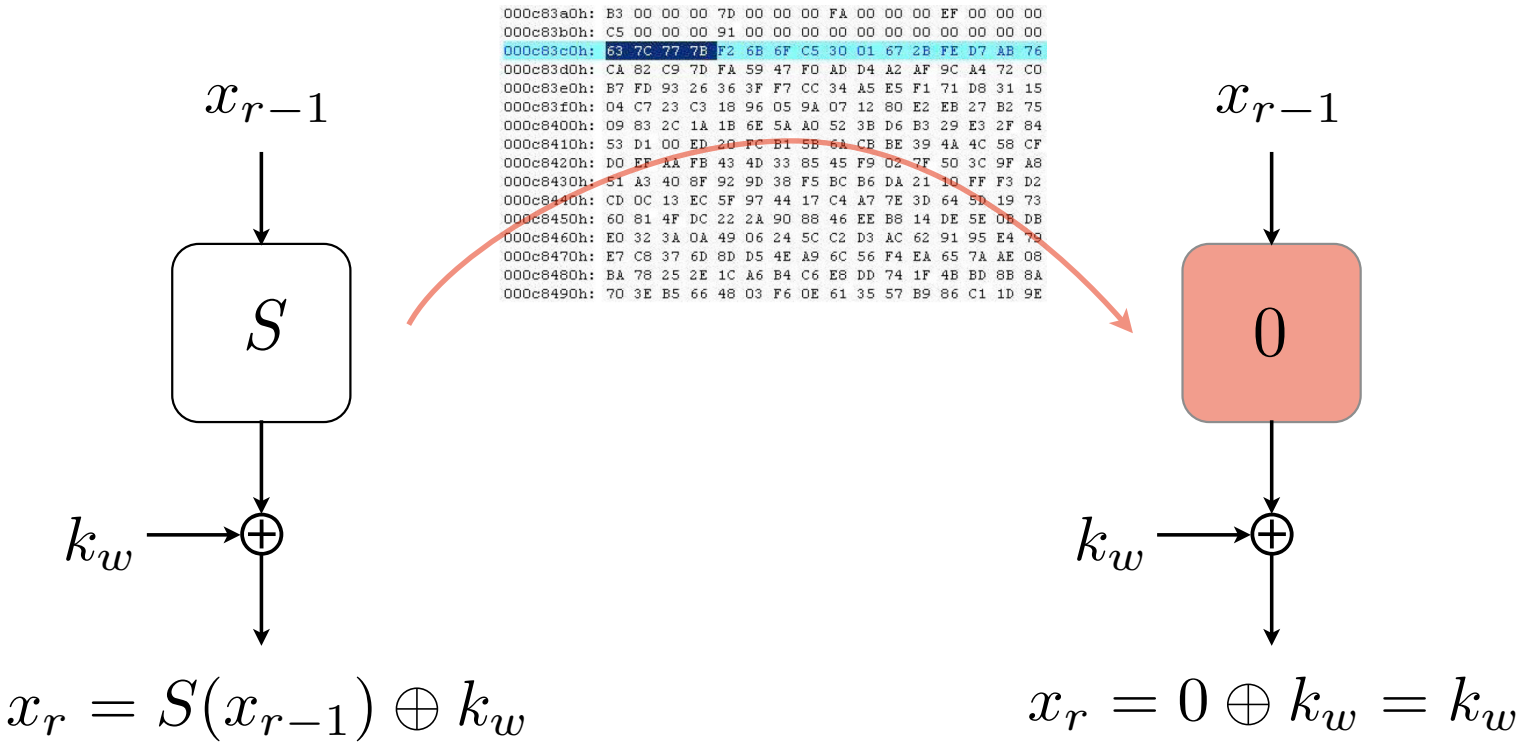
- Black-box attacker:
  - only has access to the input/output behavior of the cryptographic algorithm.
  - has no visibility into its execution.



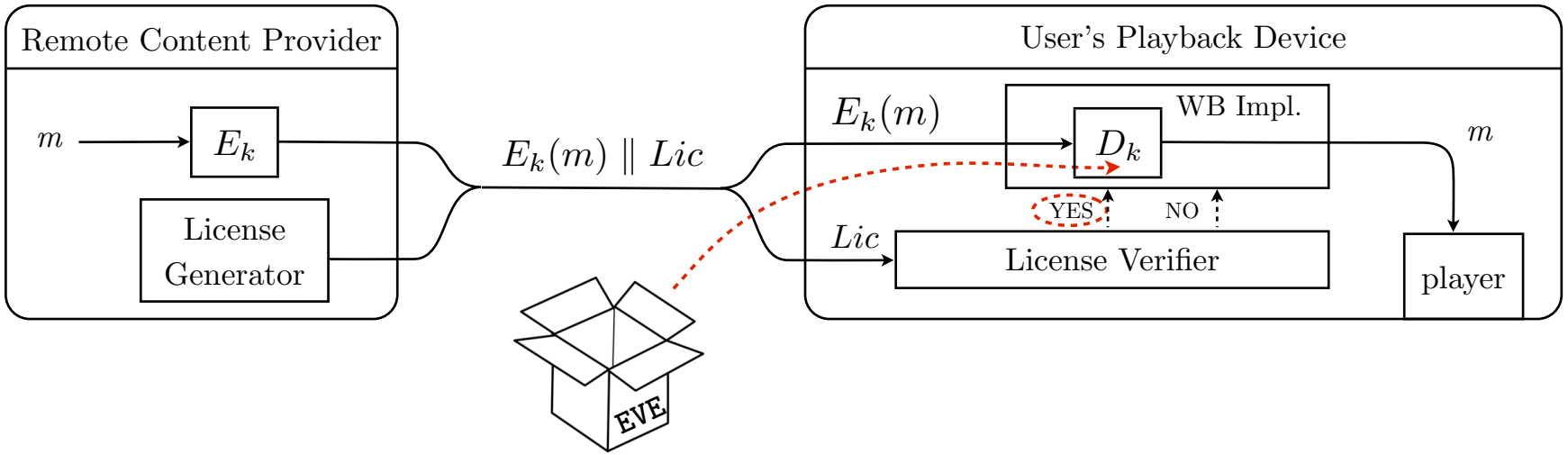
- White-box attacker:
  - has <u>full access</u> to the software implementation of the cryptographic algorithm.
  - has <u>full control</u> over its execution environment.
  - has the goal to extract the secret cryptographic key (**key recovery**).

When the attacker has knowledge of the internal structure of a cryptographic primitive, *the way how it is implemented is the sole remaining line of defense*.

- S-box Blanking Attack [Kerins and Kursawe, 2006]



Use case: a (very) simplified DRM model



- The **trusted** digital media player (containing the WB implementation) is deployed in an **untrusted** environment (the end-user's playback device).
- The goal of a malicious behaving end-user is to extract the secret decryption key out of the decryption routine in order to:
  - decrypt the encrypted content while circumventing the License Verification
  - distribute the key to non-authorized end-users

White-Box AES Implementation

State-of-the-Art

White-box AES Implementation
Chow, Eisen, Johnson, van Oorschot [2002]

Billet, Gilbert and Ech-Chatbi [2004]

Generic Class

Michiels, Gorissen and Hollmann [2008]

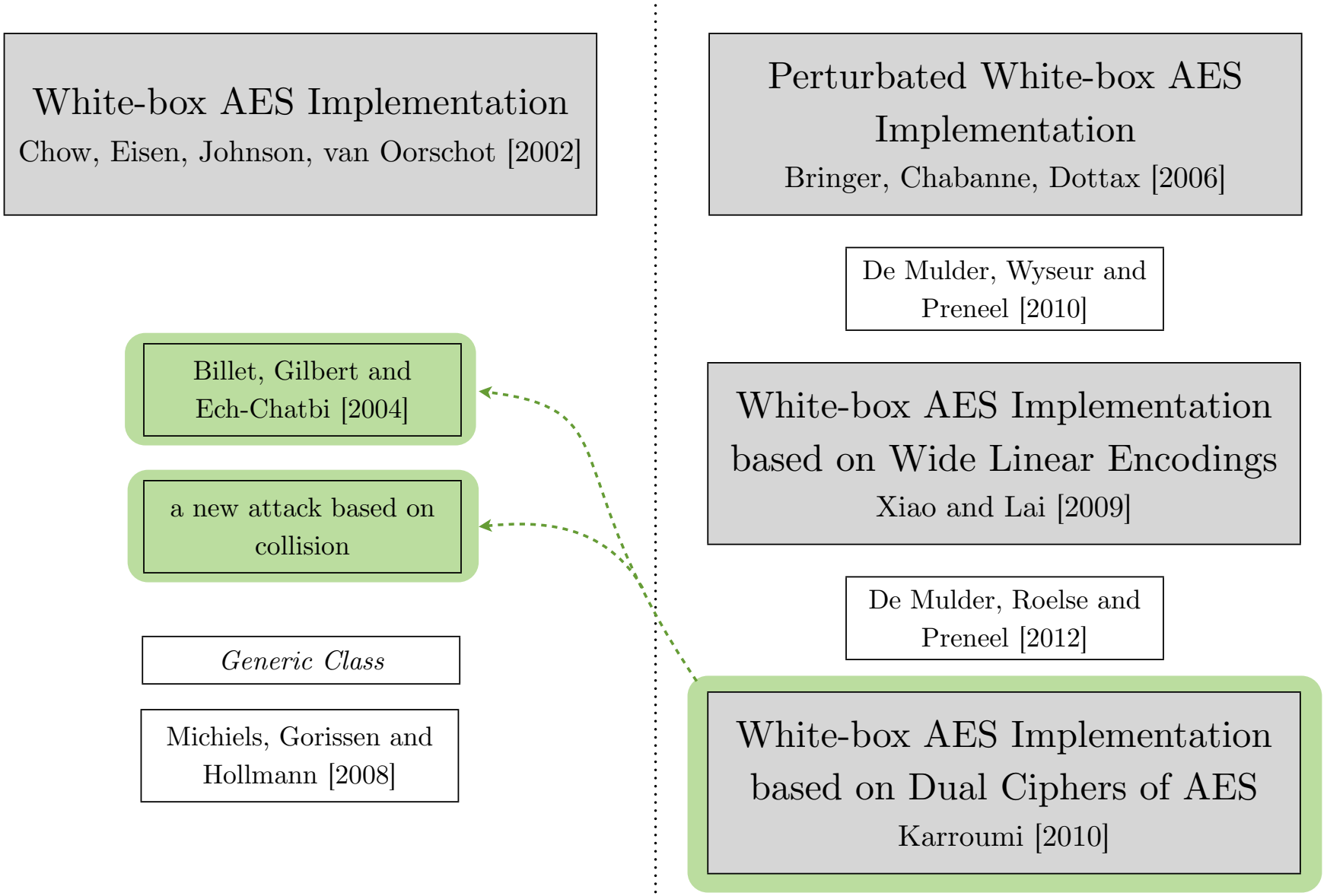
Perturbated White-box AES Implementation
Bringer, Chabanne, Dottax [2006]

De Mulder, Wyseur and Preneel [2010]

White-box AES Implementation based on Wide Linear Encodings
Xiao and Lai [2009]

De Mulder, Roelse and Preneel [2012]

White-box AES Implementation based on Dual Ciphers of AES
Karroumi [2010]


White-box AES Implementation
Chow, Eisen, Johnson, van Oorschot [2002]
Perturbated White-box AES Implementation
Bringer, Chabanne, Dottax [2006]
De Mulder, Wyseur and Preneel [2010]
Billet, Gilbert and Ech-Chatbi [2004]
White-box AES Implementation based on Wide Linear Encodings
Xiao and Lai [2009]
a new attack based on collision
De Mulder, Roelse and Preneel [2012]
Generic Class
White-box AES Implementation based on Dual Ciphers of AES
Karroumi [2010]
Michiels, Gorissen and Hollmann [2008]

Aspects of Chow's White-Box AES Implementation

Descriptions of AES-128

Conventional way:

1. AddRoundKey ($K^{(1)}$);
2. for r from 1 to 9: (Round 1-9)
   (a) SubBytes;
   (b) ShiftRows;
   (c) MixColumns;
   (d) AddRoundKey ($K^{(r+1)}$);
3. SubBytes; (Round 10)
4. ShiftRows;
5. AddRoundKey ($K^{(11)}$).

Used for WB AES:

1. for r from 1 to 9: (Round 1-9)
   (a) ShiftRows;
   (b) AddRoundKey ($\hat{K}^{(r)}$);
   (c) SubBytes;
   (d) MixColumns;
2. ShiftRows; (Round 10)
3. AddRoundKey ($\hat{K}^{(10)}$);
4. SubBytes;
5. AddRoundKey ($K^{(11)}$).

AES Subround

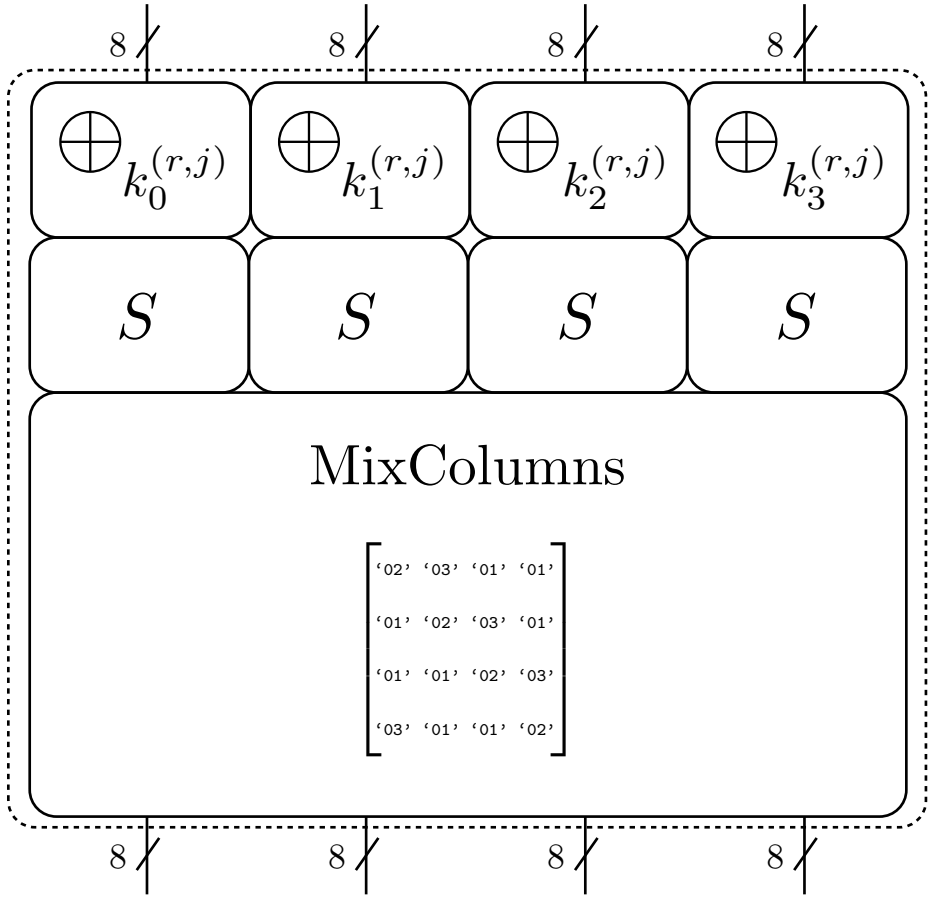
for r from 1 to 9:

(a) ShiftRows;

(b) AddRoundKey ($K^{(r)}$);

(c) SubBytes;

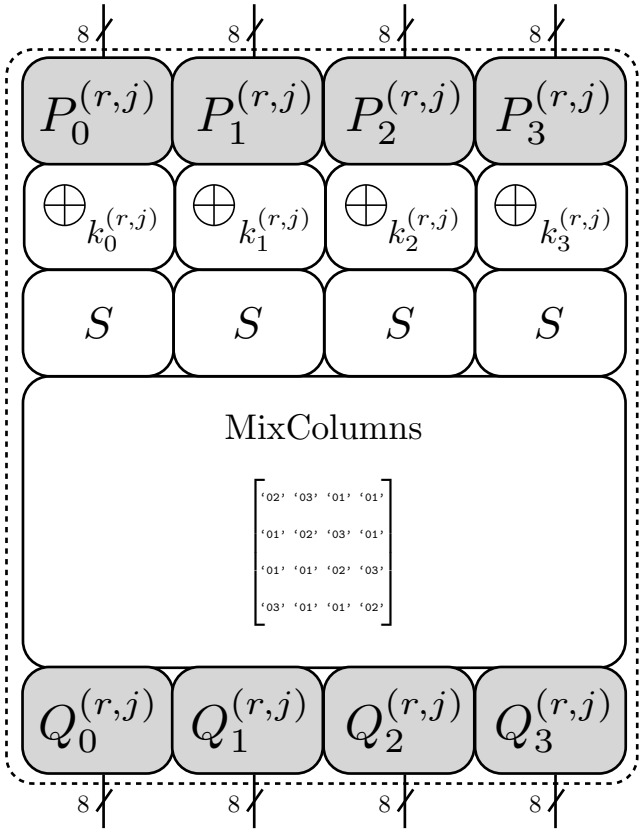
(d) MixColumns;



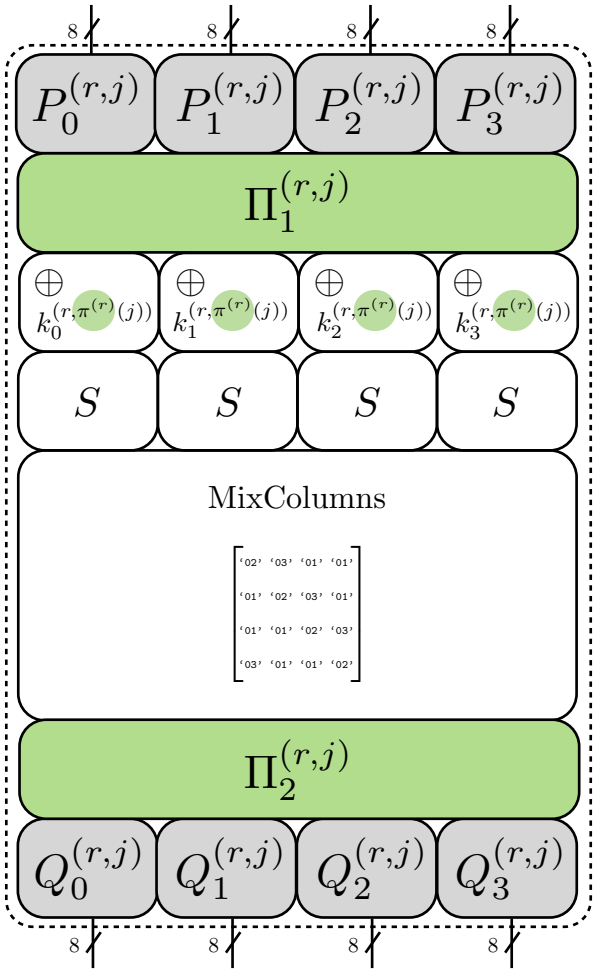
for $0 \leq j \leq 3$ and $1 \leq r \leq 9$

Encoded AES Subround

the naive version



the generic version



new attack (collisions)

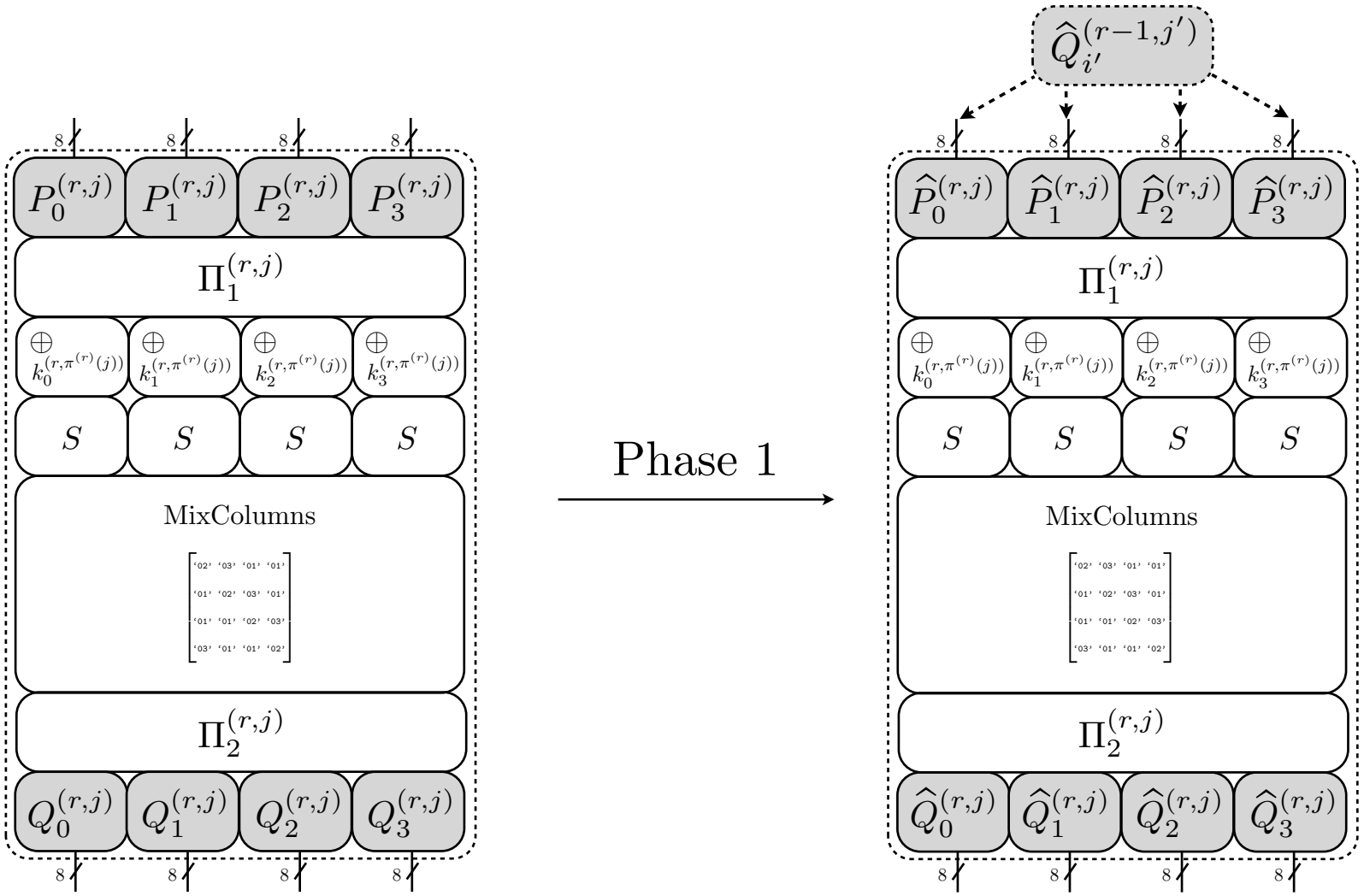
BGE attack

Revisiting the BGE Attack

BGE Attack

Phase 1

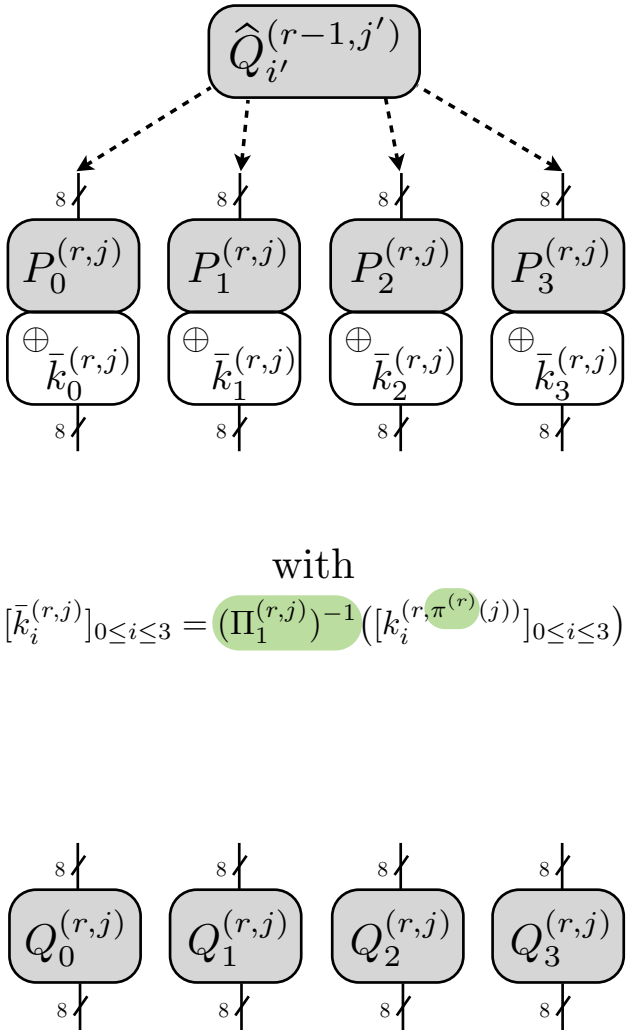
- obtain the output encodings up to an affine part
- the same for the input through round *r-1*



BGE Attack

Phase 2

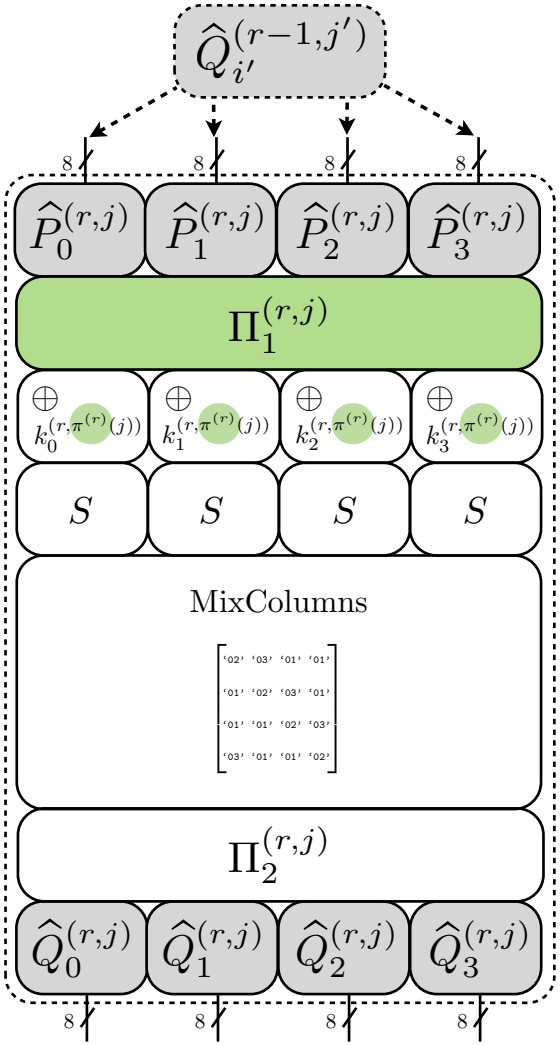
- fully recover the affine output encodings and the key-dependent affine input encodings
- fully recover the affine input encodings through round *r-1*



with

$$[\bar{k}_i^{(r,j)}]_{0\le i\le 3} = (\Pi_1^{(r,j)})^{-1}\left([k_i^{(r,\pi^{(r)}(j))}]_{0\le i\le 3}\right)$$

Phase 2

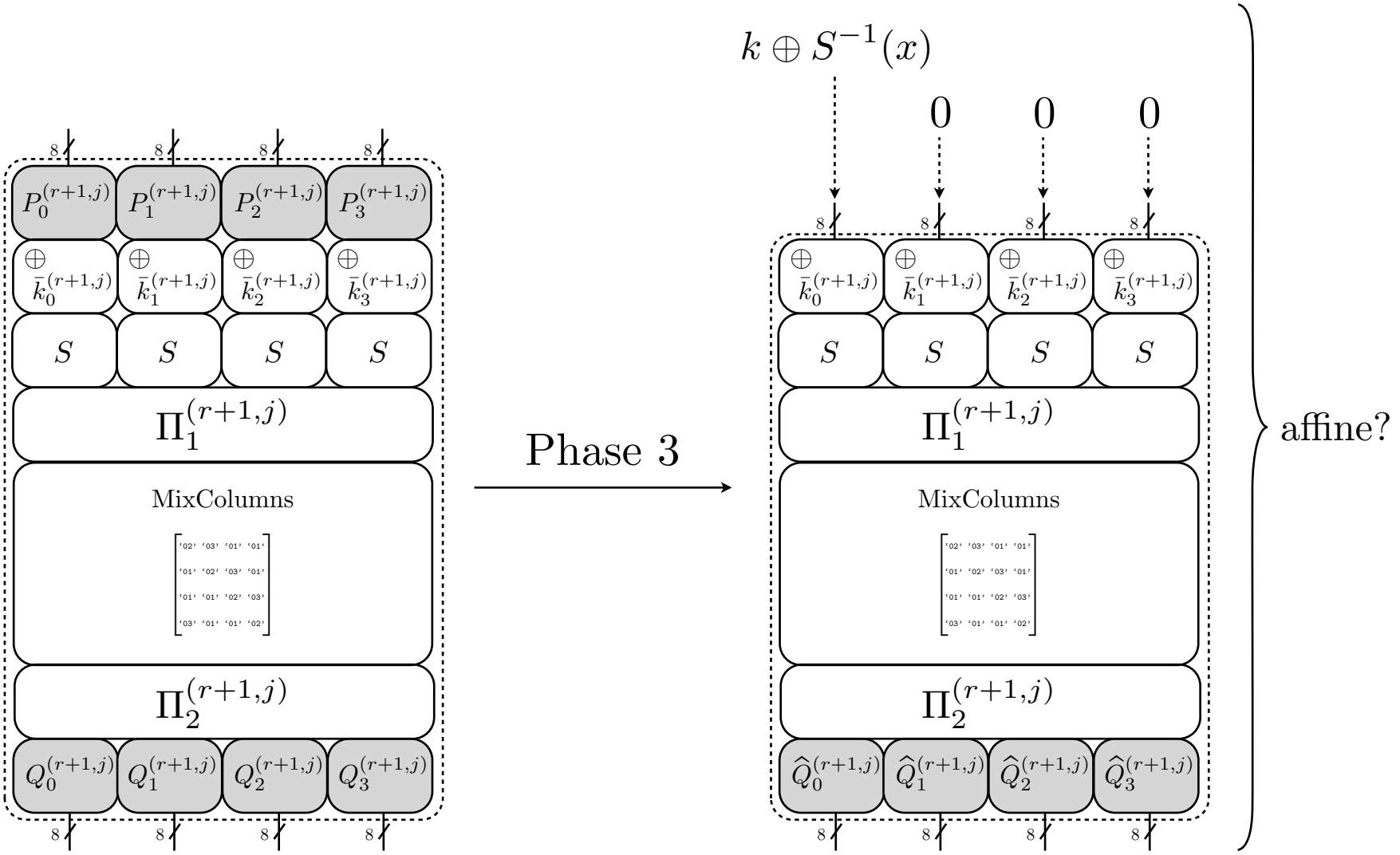


Original BGE Attack	Improved BGE Attack
Phase 1: 2^{30}	[Tolhuizen, 2012] ⇢ 2^{19}
Phase 2: 2^{29}	⇢ 2^{22}
Phase 3: no given work factor	
▸ apply Phases 1 and 2 to rounds r and $r+1$ to obtain the round keys ▸ relate both round keys via: 1. the white-box implementation 2. the AES key scheduling algorithm	▸ apply Phases 1 and 2 only to round r to obtain the r^{th} round key ▸ a new method to obtain the $(r+1)^{\text{th}}$ round key 2^{18} ▸ an efficient method to determine the correct order of the round keys 2^{13}
total work factor: 2^{30}	total work factor: 2^{22}

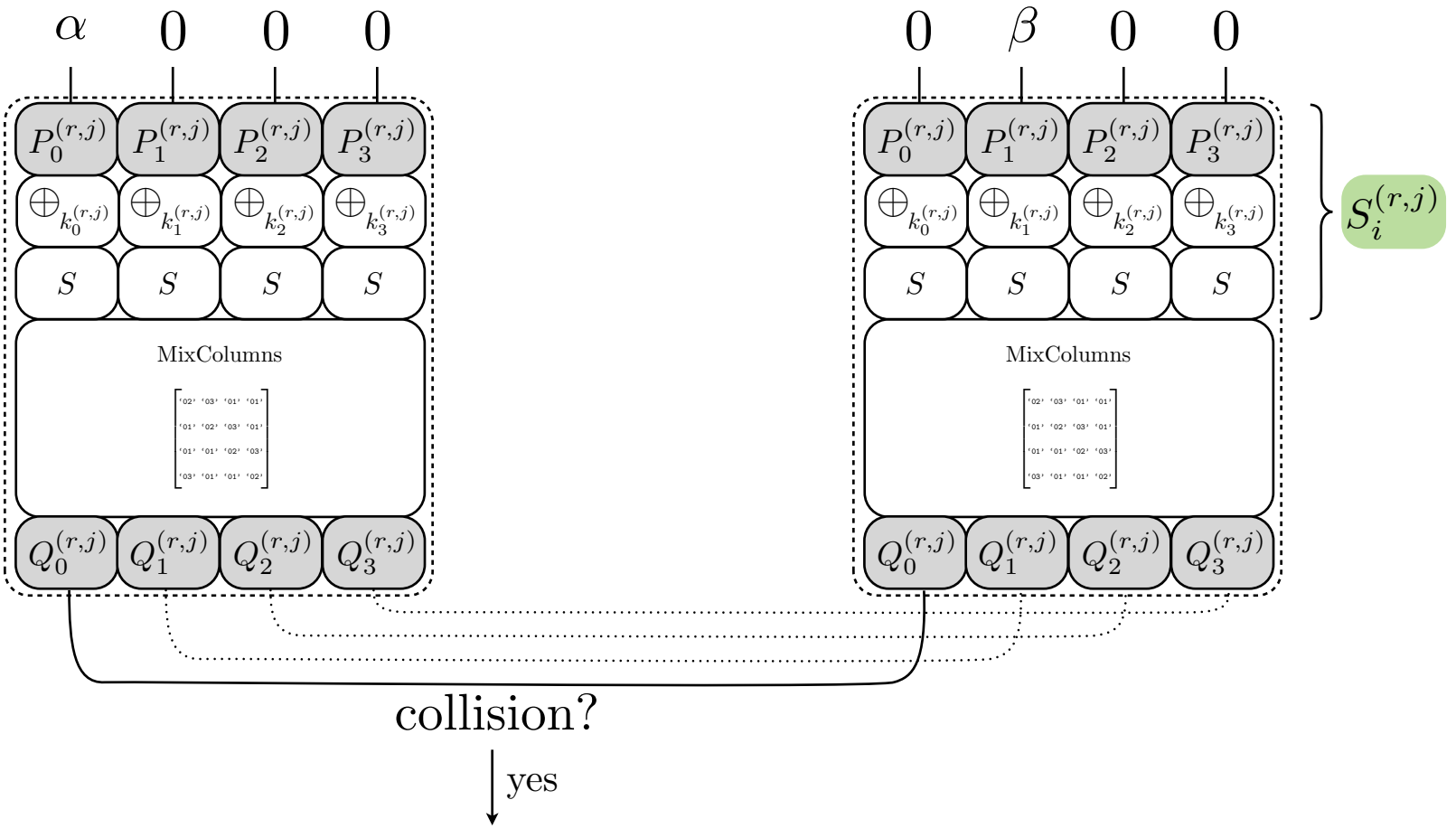
BGE Attack

Phase 3

- obtain the $(r{+}1)^{\text{th}}$ round key
- correctness: $S\big(c \oplus S^{-1}(x)\big)$ is non-affine for all non-zero values of c



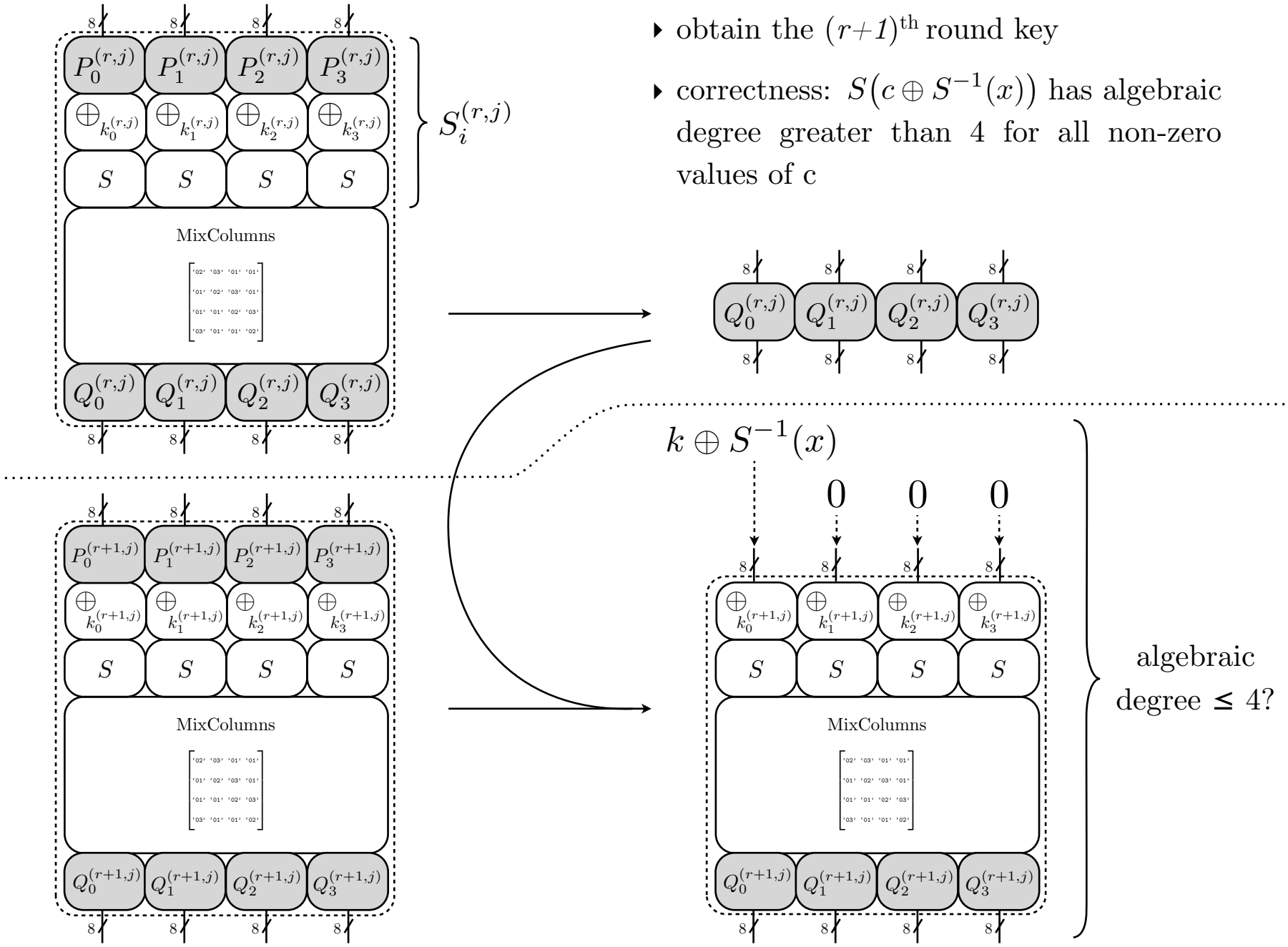
New Attack based on Collisions



$$02 \otimes S_0^{(r,j)}(\alpha) \oplus 03 \otimes S_1^{(r,j)}(0) = 02 \otimes S_0^{(r,j)}(0) \oplus 03 \otimes S_1^{(r,j)}(\beta)$$

256 pairs (α,β) with the trivial solution $(\alpha,\beta)=(0,0)$

2^{22}



- obtain the $(r+1)^{\text{th}}$ round key
- correctness: $S\big(c \oplus S^{-1}(x)\big)$ has algebraic degree greater than 4 for all non-zero values of c

Cryptanalysis of Karroumi's White-Box AES Implementation

Dual AES Ciphers

[Barkan and Biham, 2002] [Biryukov, De Cannière, Braeken, and Preneel, 2003]

$$C = \mathtt{AES}_k(P)$$

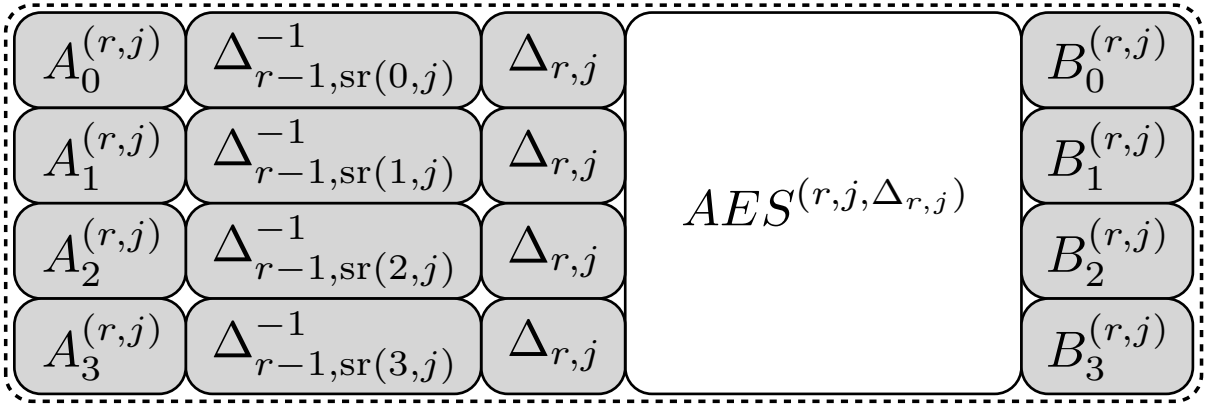
$$\Big\downarrow (\Delta \in \mathcal{T}) \qquad \text{with} \quad |\mathcal{T}| = 61.200$$

$$\Delta(C) = \mathtt{AES}_k^{\Delta}\big(\Delta(P)\big)$$

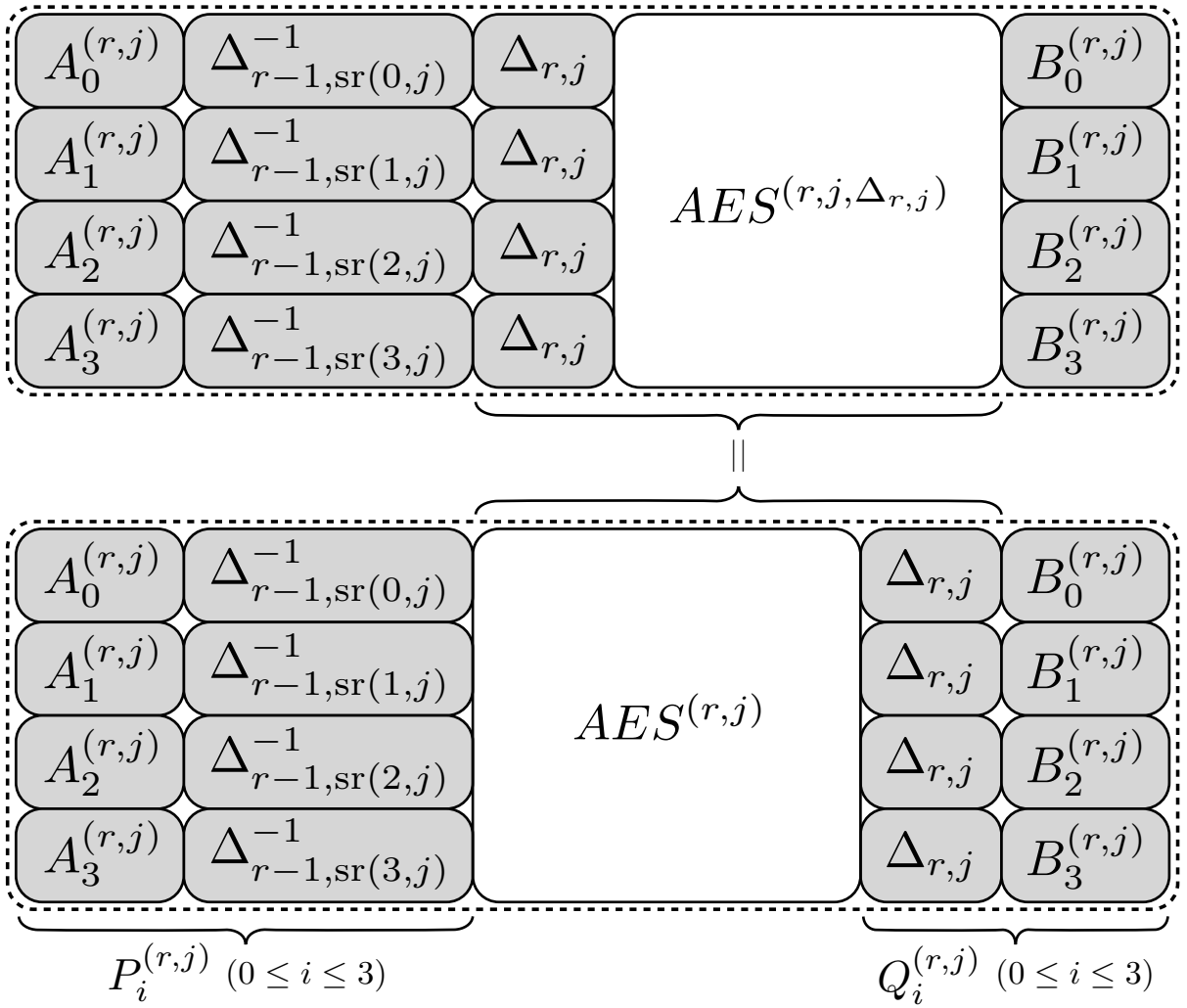
$$\mathcal{T} = \{R_l \circ m_\alpha \circ f_t \mid 1 \le l \le 30, \alpha \in \mathbf{F}_{256}^* \text{ and } 0 \le t \le 7\}$$

- squaring operations $f_t(x) = x^{2^t}$
- multiplication with a non-zero constant $m_\alpha(x) = \alpha \otimes x$
- isomorphisms between the AES polynomial representation of $\mathbf{F}_{256}$ and one of the 30 polynomial representations of $\mathbf{F}_{256}$

An encoded *dual* AES subround ...



... is in fact an encoded AES subround



Conclusions

- reduced the work factor of the BGE attack from 2^{30} to 2^{22}
  - non-affine encodings and permutations on the round key bytes have a negligible contribution to the overall work factor of the improved BGE attack
- a new attack based on internal collision with work factor 2^{22}
- insecurity of Karroumi's white-box AES implementation

Open problem: new WB-AES designs?

- Research for new secure WBAES designs: fixed key vs. dynamic key
- WBC part of bigger program: additional layers of security by obfuscation techniques
- Companies: "security through obscurity"

irdeto

NAGRA KUDELSKI

whiteCryption
White-Box Cryptography Solutions

Questions?