Differential Cryptanalysis

Maria Eichlseder

Applied Cryptography 2 – ST 2020

Cryptanalysis

…or how to scale your cipher

The best available cryptanalysis (+security margin) indicates the necessary key size, number of rounds, etc. to achieve a certain security level:

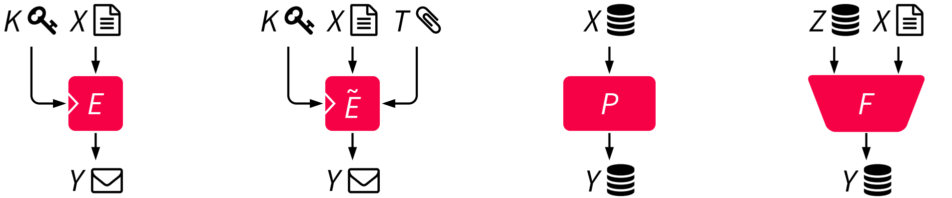
- Asymmetric crypto: Best algorithm to solve hard problem
- Symmetric crypto: Generic and dedicated attack techniques

Outline

- Security Analysis of Symmetric Primitives
- Differential Cryptanalysis
- Exploiting Differentials
- Caveats and Assumptions

Security Analysis of Symmetric Primitives

Reminder: Symmetric Primitives



block cipher
(BC)

tweakable BC
(TBC)

permutation

compression

...

Quantifying Security: The security of a cipher is bounded by…

Generic Attacks: work for any cipher with same interface

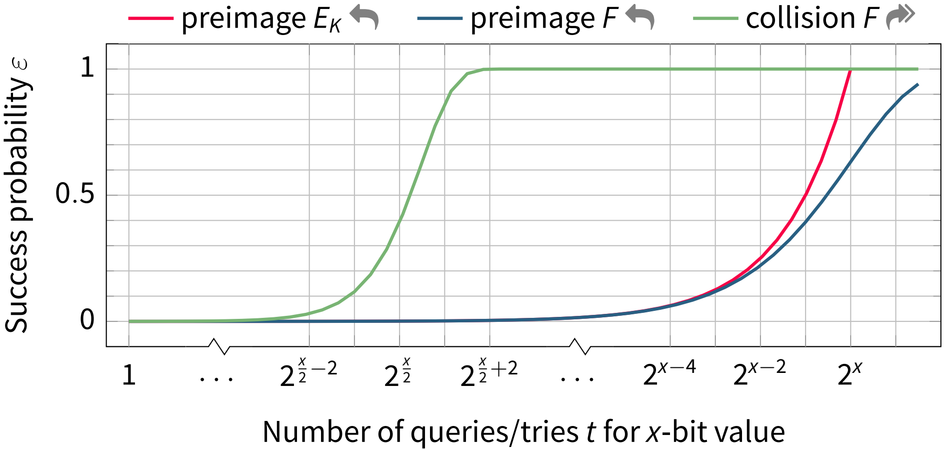
Example: Block cipher with k-bit key and n-bit block:

- Key guessing: Costs about $\approx 2^k$ trial encryptions
- Full codebook: After observing ciphertexts for all 2^n different known plaintexts, attacker knows E_K
- Birthday: After observing ciphertexts for $\approx 2^{n/2}$ different plaintexts, attacker can distinguish from random function (no collisions)

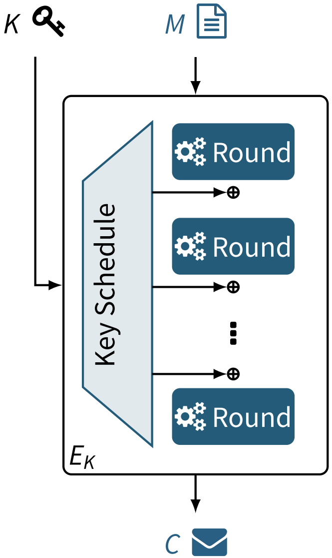
Dedicated Attacks: exploit specific internal details of the cipher

For a good symmetric primitive, we usually want that no dedicated attack is more efficient than the best generic attack (= it is as good as can be expected, given the interface)

Generic Security Levels



The Key-Alternating Construction



2 fundamental ideas:

1. Repeat simple circuit ("round") r times
2. Make the round circuit public but mix input with round key

Important Symmetric Cryptanalysis Techniques

Statistical Analysis

- Differential Cryptanalysis (DC):
  - Predict output difference from input difference
  - Many variants (truncated, impossible, . . .)
- Linear Cryptanalysis (LC):
  - Approximate output as a linear function of input
  - Many links to DC

Other Techniques

- Algebraic Cryptanalysis (many different variants):
  - Describe cipher in equations and solve
  - Derive deterministic properties about output
- . . .

Differential Cryptanalysis

Idea: Tracking Differences

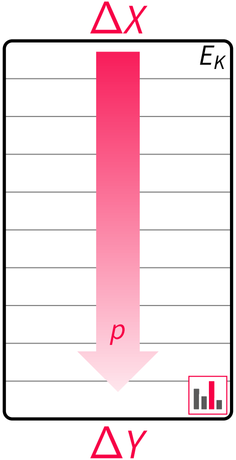
Differential Cryptanalysis – Overview

- Proposed by Biham and Shamir [BS90] for `DES`
- `DES` designers (IBM, NSA) apparently knew about a similar attack before
- Chosen-plaintext attack
- One of the two major statistical attack techniques and design criteria

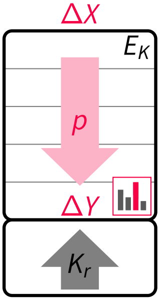
- Main idea:
  1. Predict effect of plaintext difference $\Delta M = M \oplus M^*$ on ciphertext difference $\Delta C = C \oplus C^*$ without knowing K
  2. Use prediction as distinguisher to recover the key

Differential Cryptanalysis – Idea

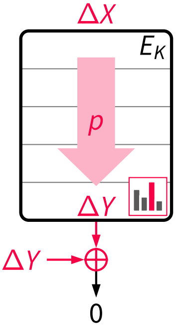
Method



Attack Goals



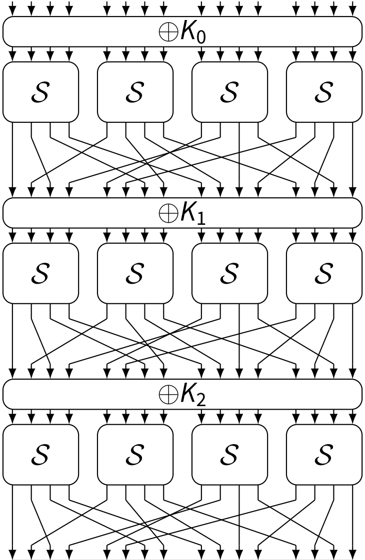
key recovery



collision,
forgery

. . .

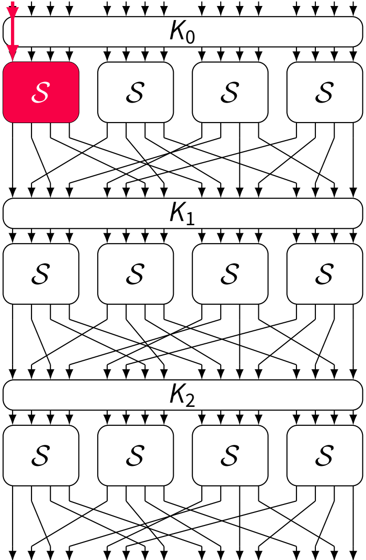
Example: A Toy Block Cipher



x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$\mathcal{S}(x)$	2	0	4	3	9	5	6	7	1	d	e	f	a	8	c	b

Let's Flip a Bit

“active”



Differential Properties of S-boxes (Confusion)

$\Delta\text{in} = 8 \quad \rightarrow \quad \Delta\text{out} = ?$

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$\mathcal{S}(x)$	2	0	4	3	9	5	6	7	1	d	e	f	a	8	c	b

Differential Properties of S-boxes (Confusion)

$\Delta\text{in} = 8 \quad \rightarrow \quad \Delta\text{out} = ?$

$\Delta\texttt{in} = 8$

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$\mathcal{S}(x)$	2	0	4	3	9	5	6	7	1	d	e	f	a	8	c	b

$\Delta\texttt{out} = 3$

Differential Properties of S-boxes (Confusion)

$\Delta \text{in} = 8 \quad \rightarrow \quad \Delta \text{out} = ?$

$\Delta \texttt{in} = 8$

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$\mathcal{S}(x)$	2	0	4	3	9	5	6	7	1	d	e	f	a	8	c	b

$\Delta \texttt{out} = \texttt{d}$

Differential Properties of S-boxes (Confusion)

$\Delta\text{in} = 8 \quad \rightarrow \quad \Delta\text{out} = ?$

$\Delta\texttt{in} = 8$

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$\mathcal{S}(x)$	2	0	4	3	9	5	6	7	1	d	e	f	a	8	c	b

$\Delta\texttt{out} = \texttt{a}$

Differential Properties of S-boxes (Confusion)

$\Delta\text{in} = 8 \quad \to \quad \Delta\text{out} \in \{3, \mathtt{a}, \mathtt{c}, \mathtt{d}\}$

x	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
$\mathcal{S}(x)$	2	0	4	3	9	5	6	7	1	d	e	f	a	8	c	b

- Knowing the value tells us the difference
- Knowing the difference tells us (something about) the value:

$$\text{solutions}(\Delta\text{in}, \Delta\text{out}) := \{x \;:\; \mathcal{S}(x \oplus \Delta\text{in}) \oplus \mathcal{S}(x) = \Delta\text{out}\}$$

Differential Properties of S-boxes – More Formally

We consider pairs of two variables $x, x^* \in \mathbb{F}_2^n$ and evaluate their difference Δx:

$$\Delta x = x^* \oplus x \,.$$

If x and x^* are inputs to two instances of a cryptographic (vectorial Boolean) function $f : \mathbb{F}_2^n \to \mathbb{F}_2^n$, we are interested in the resulting difference Δy of the two outputs y, y^*:

$$\Delta y = y^* \oplus y = f(x \oplus \Delta x) \oplus f(x) \,.$$

For a fixed input difference $\alpha = \Delta x \in \mathbb{F}_2^n$, the output difference Δy depends on the value x. This induces another function on $\mathbb{F}_2^n$, the forward directional derivative by α:

$$\Delta_\alpha f(x) := f(x \oplus \alpha) \oplus f(x) \,.$$

This derivation operator shares many properties with the derivations of differential calculus, such as the “sum rule” and “product rule” (Leibniz’ rule).

These derivatives $\Delta_\alpha f$ may be more amenable to analysis than the initial function f.

Differential Distribution Table (DDT)

Δin \ Δout	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
0	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	-	4	4	-	-	-	-	4	-	-	-	-	4	-	-	-
2	-	-	4	4	-	-	4	-	-	-	-	-	-	-	-	4
3	-	4	-	4	4	-	-	-	-	-	-	-	-	-	4	-
4	-	-	4	-	4	4	-	-	-	-	-	4	-	-	-	-
5	-	-	-	4	-	4	-	4	-	4	-	-	-	-	-	-
6	-	-	-	-	4	-	4	4	-	-	-	-	-	4	-	-
7	-	4	-	-	-	4	4	-	-	-	4	-	-	-	-	-
8	-	-	-	4	-	-	-	-	-	-	4	-	4	4	-	-
9	-	4	-	-	-	-	-	-	-	-	-	4	-	4	-	4
a	-	-	-	-	-	4	-	-	-	-	-	-	4	-	4	4
b	-	-	4	-	-	-	-	-	-	4	-	-	-	4	4	-
c	-	-	-	-	-	-	-	-	16	-	-	-	-	-	-	-
d	-	-	-	-	4	-	-	-	-	4	4	-	-	-	-	4
e	-	-	-	-	-	-	-	4	-	-	4	4	-	-	4	-
f	-	-	-	-	-	-	4	-	-	4	-	4	4	-	-	-

Differential Distribution Table (DDT) – More Formally

We refer to a pair of input difference $\alpha = \Delta x \in \mathbb{F}_2^n$ and output difference $\beta = \Delta y \in \mathbb{F}_2^n$ as a differential $\delta = (\alpha \mapsto \beta)$. The solution set $S(\alpha, \beta)$ of the differential is then

$$S(\alpha, \beta) := \{x \in \mathbb{F}_2^n : \Delta_\alpha f(x) = f(x \oplus \alpha) \oplus f(x) = \beta\} \ .$$

We call the differential impossible if $|S(\alpha, \beta)| = 0$, and possible otherwise.
For example, for $\alpha = 0$, only the trivial differential $(0 \mapsto 0)$ is possible.
Pairs $(x, x \oplus \alpha)$ with $x \in S(\alpha, \beta)$ are called valid.

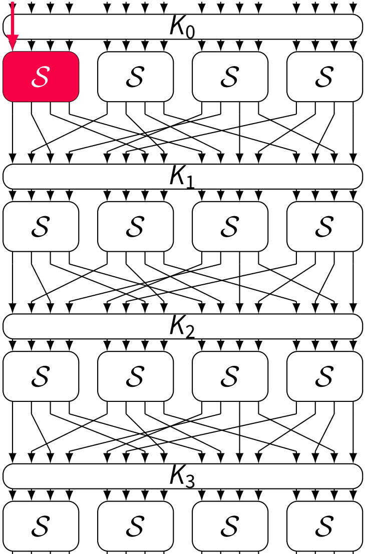
The differential distribution table (DDT) lists the number of solutions $|S(\alpha, \beta)|$ for all α, β.
The multiset of values in this table is referred to as the differential spectrum of f, and its maximum as the differential uniformity $\mathtt{du}_f$ of f:

$$\mathtt{du}_f := \max_{\alpha \neq 0, \beta} |S(\alpha, \beta)|.$$

The probability that f maps $\Delta x = \alpha$ to $\Delta y = \beta$ for uniformly random x is then

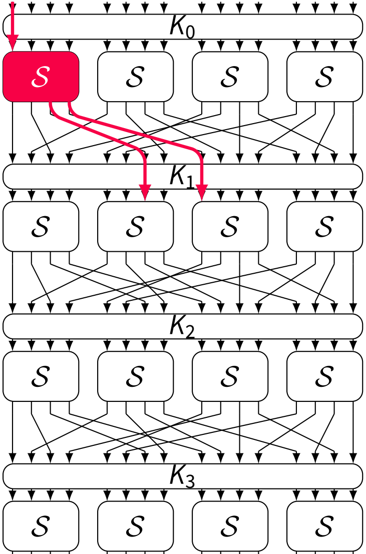
$$\mathtt{dp}(\alpha, \beta) = \mathbb{P}_x[\alpha \overset{f}{\mapsto} \beta] := \mathbb{P}_x\left[f(x \oplus \alpha) \oplus f(x) = \beta\right] = \frac{|S(\alpha, \beta)|}{2^n} \leq \frac{\mathtt{du}_f}{2^n}.$$

Let's Flip a Bit



Δ	p
8000	1

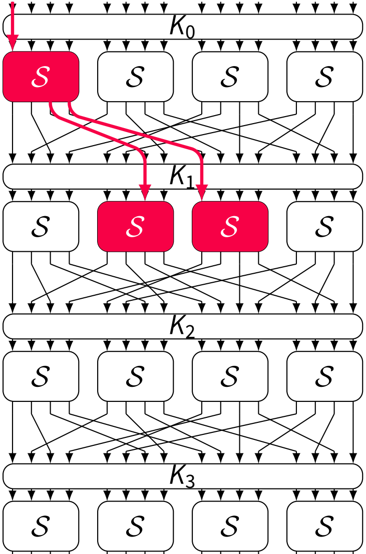
Let's Flip a Bit



Δ	p
8000	1
3000	2^{-2}

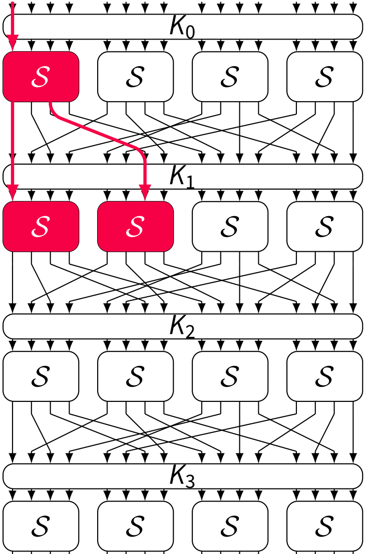
$\cdot 2^{-2}$

Let's Flip a Bit



Δ	p	
8000	1	
		$\cdot 2^{-2}$
3000	2^{-2}	
		$\cdot 1$
0280	2^{-2}	

Let's Flip a Bit

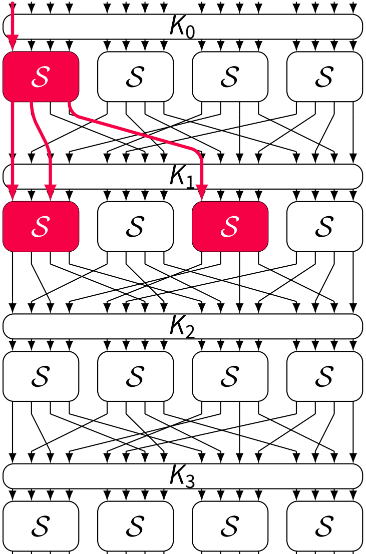


Δ	p
8000	1
a000	2^{-2}
8200	2^{-2}

$\cdot 2^{-2}$

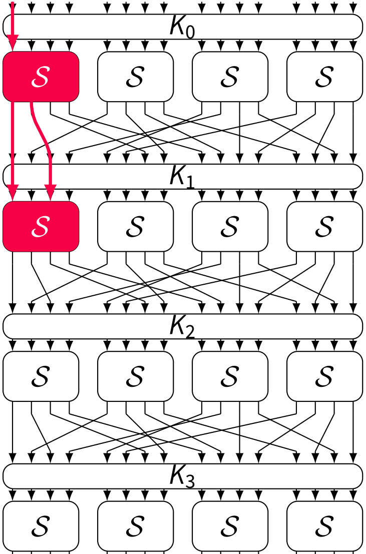
$\cdot 1$

Let's Flip a Bit



Δ	p	
8000	1	
		$\cdot 2^{-2}$
d000	2^{-2}	
		$\cdot 1$
a080	2^{-2}	

Let's Flip a Bit

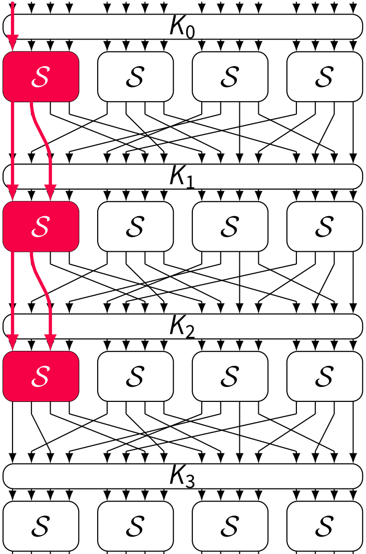


Δ	p
8000	1
c000	2^{-2}
a000	2^{-2}

$\cdot 2^{-2}$

$\cdot 1$

Let's Flip a Bit



Δ	p	
8000	1	
		$\cdot 2^{-2}$
c000	2^{-2}	
		$\cdot 1$
a000	2^{-2}	
		$\cdot 2^{-2}$
c000	2^{-4}	
		$\cdot 1$
a000	2^{-4}	
$\vdots$	$\vdots$	

Differential Properties of Mixing Layers (Diffusion)

If f is an $\mathbb{F}_2$-affine function $f(x) = \ell(x) \oplus c$ with linear part $\ell(x)$ and the input difference is $\alpha = \Delta x$, then the only one value β with non-zero probability is

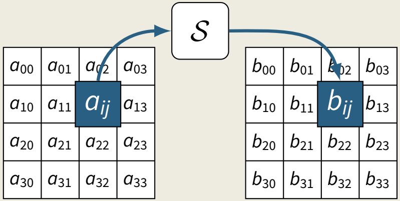
$$\beta = \Delta_\alpha f(x) = \ell(\alpha).$$

When is a linear layer "good"?

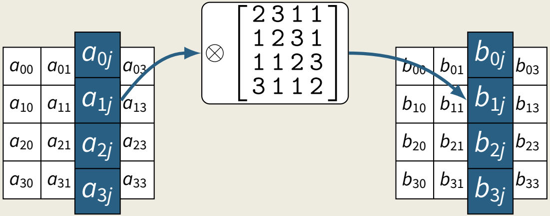
- Branch number $\mathcal{B}$ [Dae95]:
  Min number of active S-boxes in 2 consecutive rounds
- In our toy cipher: $\mathcal{B} = 2$. Can we do better?
- Best case: $\mathcal{B} = 1 +$ number of S-boxes per round
- Requires actual "mixing" (xor), not just bit permutations

Design of AES – Round Function (10 or 12 or 14 Rounds)

1 SubBytes (SB)

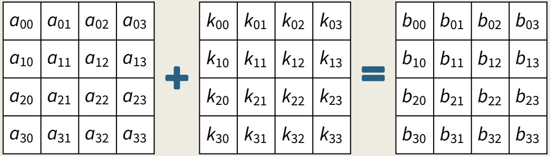


3 MixColumns (MC)



2 ShiftRows (SR)

4 AddRoundKey (AK)



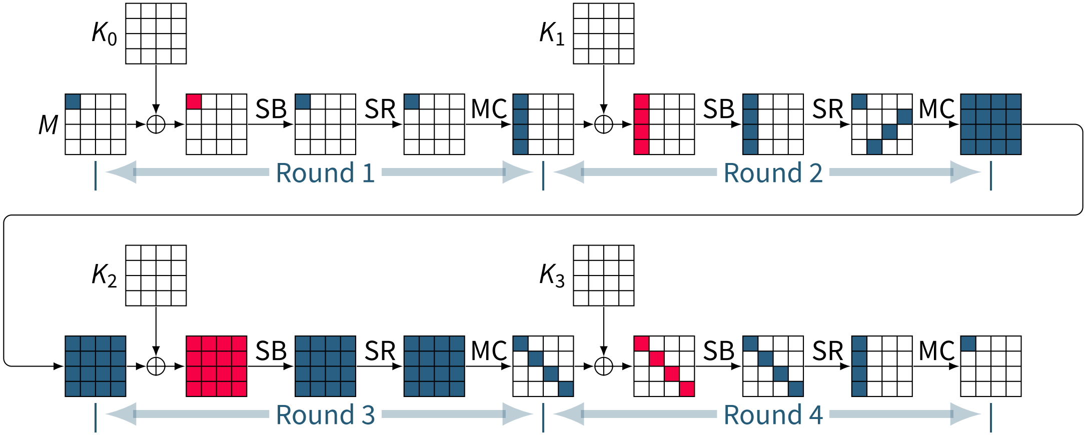
Design of AES – Properties of the Round Function

Let's flip a bit:

$$
\begin{tikzcd}[ampersand replacement=\&]
\begin{bmatrix} 00 & 00 & 00 & 40 \\ 00 & 00 & 00 & 00 \\ 00 & 00 & 00 & 00 \\ 00 & 00 & 00 & 00 \end{bmatrix} \arrow[r, mapsto, "\mathtt{SB}"] \arrow[r, bend right, "2^{-6}"'] \&
\begin{bmatrix} 00 & 00 & 00 & 6a \\ 00 & 00 & 00 & 00 \\ 00 & 00 & 00 & 00 \\ 00 & 00 & 00 & 00 \end{bmatrix} \arrow[r, mapsto, "\mathtt{SR}"] \&
\begin{bmatrix} 00 & 00 & 00 & 6a \\ 00 & 00 & 00 & 00 \\ 00 & 00 & 00 & 00 \\ 00 & 00 & 00 & 00 \end{bmatrix} \arrow[r, mapsto, "\mathtt{MC}\atop\mathtt{AK}"] \&
\begin{bmatrix} 00 & 00 & 00 & d4 \\ 00 & 00 & 00 & 6a \\ 00 & 00 & 00 & 6a \\ 00 & 00 & 00 & be \end{bmatrix} \arrow[r, mapsto, "\mathtt{SB}"] \arrow[r, bend right, "2^{-6\times 4}"'] \&
\begin{bmatrix} 00 & 00 & 00 & 2b \\ 00 & 00 & 00 & 61 \\ 00 & 00 & 00 & 61 \\ 00 & 00 & 00 & cd \end{bmatrix} \arrow[r, mapsto, "\mathtt{SR}"] \&
\begin{bmatrix} 00 & 00 & 00 & 2b \\ 00 & 00 & 61 & 00 \\ 00 & 61 & 00 & 00 \\ cd & 00 & 00 & 00 \end{bmatrix} \arrow[r, mapsto, "\mathtt{SR}\atop\mathtt{MC}"] \&
\begin{bmatrix} cd & 61 & a3 & 56 \\ cd & a3 & c2 & 2b \\ 4c & c2 & 61 & 2b \\ 81 & 61 & 61 & 7d \end{bmatrix}
\end{tikzcd}
$$

- Max differential probability (MDP) of the 8×8 S-box: 2^{-6}
- Mixing layer (based on Maximum Distance Separable code, MDS) with $\mathcal{B} = 5$ (in 2 rounds $\to \geq 5$ active S-boxes)
- Actually, in 4 rounds $\to \geq 25$ active S-boxes $\to p \leq 2^{-6\times 25} = 2^{-150}$ ($\to$ later lecture)

AES – Example for Optimal Pattern with 25 active S-boxes



Automated tools for cryptanalysis

Motivation:

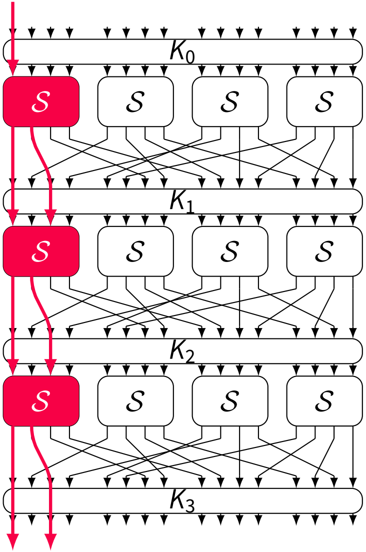
- Finding the best (or very good) characteristics can be very hard
- Necessary to evaluate new primitives

Solvers:

- By hand
- General-purpose solvers:
  - SAT/SMT (Boolean SATisfiability/Sat. Modulo Theories)
  - MILP (Mixed Integer Linear Programming)
  - CP (Constraint Programming)
- Dedicated solvers

Exploiting Differentials

An (Iterative) r-Round Differential Characteristic



Δ $\quad p = 2^{-2 \cdot r}$

8000

$\cdot 2^{-2}$

a000

$\cdot 2^{-2}$

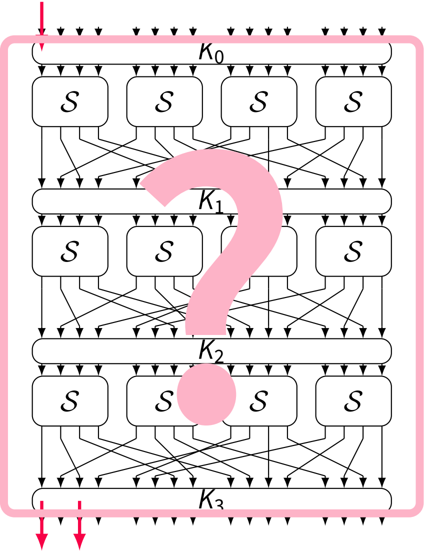
a000

$\cdot 2^{-2}$

a000

$\vdots$

An r-Round Differential



Δ

8000

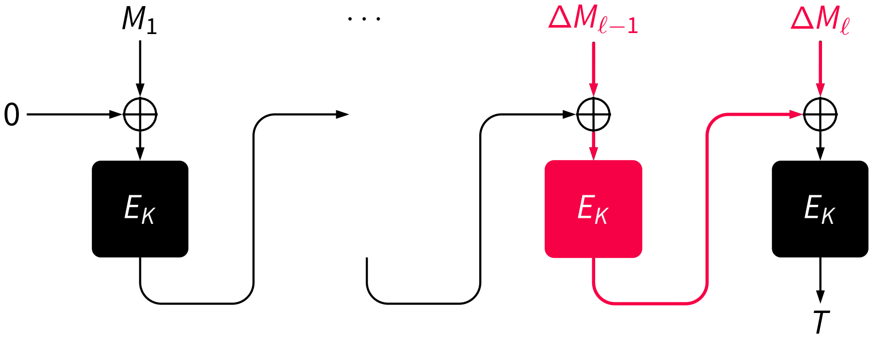
a000

$p \geq 2^{-2 \cdot r}$

$\geq 2^{-2 \cdot r}$

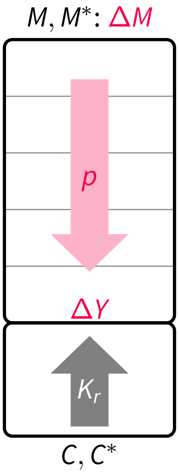
For Forgeries

Example: Forgery with success probability p for `CBC-MAC`



This is useful if $p > 2^{-\text{block size}}$ ($= 2^{-\text{tag size}}$).

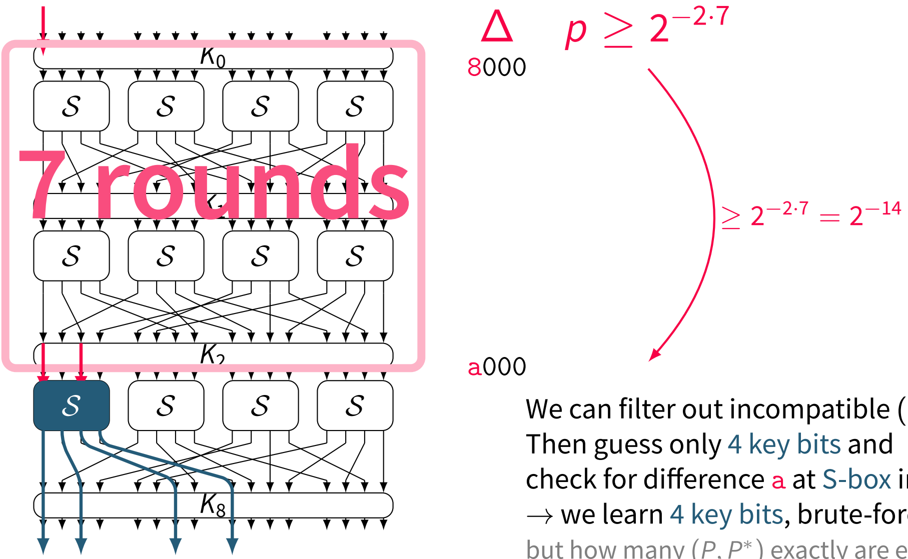
For Key Recovery



- Assume $\Delta M \xrightarrow{r-1 \text{ rounds}} \Delta Y$ has probability $p \gg 2^{-\text{block size}}$
- Query about $1/p$ chosen-plaintext pairs $(M, M^*) \to (C, C^*)$
- Decrypt each pair 1 round with each possible last-round key K_r
- If we get ΔY, upvote candidate K_r 👍

K_r	Upvote counter
0000	👍
0001	👍 👍
0002	👍 👍 👍 👍
0003	👍 👍
...	...

Key Recovery Example: 8-Round Toy Cipher



We can filter out incompatible (C, C^*)
Then guess only 4 key bits and
check for difference a at S-box input
$\rightarrow$ we learn 4 key bits, brute-force the rest
but how many (P, P^*) exactly are enough?

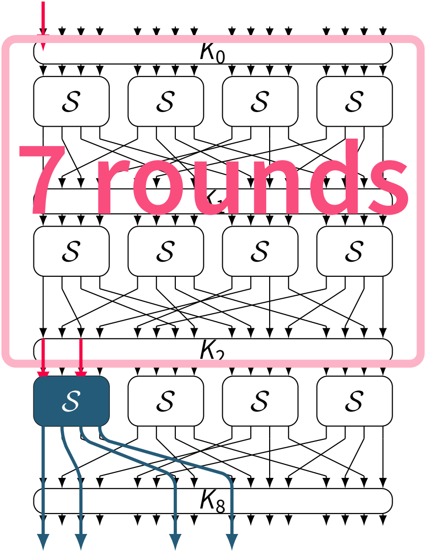
Key Recovery – Details I

$$\text{Signal-to-Noise Ratio SNR} = \frac{N \cdot p_{\text{right}}}{N \cdot p_{\text{wrong}}} = \frac{p}{A \cdot B \cdot 2^{-k}}\,.$$

- p: Expected differential probability for $R - 1$ rounds
- N: Number of queried pairs
- A: Upvoted candidates per pair
- B: Fraction of pairs after filtering ciphertexts
- k: Number of guessed key bits

Need roughly $N \approx 3 \cdot 1/p$ pairs if SNR $\gg 2$, or $N \approx 30 \cdot 1/p$ if $1 <$ SNR ≤ 2. [BS90]

Key Recovery Example: 8-Round Toy Cipher (cont'd)



- p: Expected diff. prob. for $R-1$ rounds
- N: Number of queried pairs
- A: Upvoted candidates per pair
- B: Fraction of pairs after filtering (C, C^*)
- k: Number of guessed key bits

DDT($\mathcal{S}$)	0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
a	-	-	-	-	-	4	-	-	-	-	-	-	4	-	4	4

$$\text{SNR} = \frac{p}{A \cdot B \cdot 2^{-k}} = \frac{2^{-14}}{4 \cdot (2^{-12} \cdot 2^{-2}) \cdot 2^{-4}} = 4$$

About $N \approx 3 \cdot 1/p = 3 \cdot 2^{14}$ pairs (P, P^*) should be ok – but that's $\approx$ the whole codebook!
What if we use a bit less?

Key Recovery – Details II

More precisely, using ranking statistics, to recover the k bits we need about [SB02]:

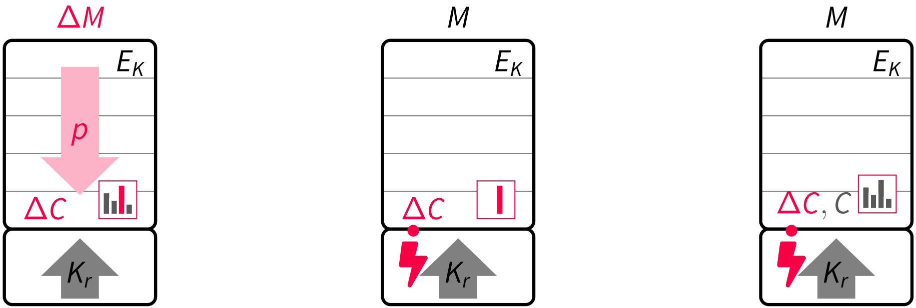
$$N = \frac{\left(\sqrt{\mathrm{SNR}+1}\cdot\Phi^{-1}(\mathbb{P}_s)+\Phi^{-1}(1-2^{-k})\right)^2}{\mathrm{SNR}}\cdot p^{-1}.$$

- p: Expected differential probability for $R-1$ rounds
- N: Number of queried pairs
- k: Number of guessed key bits
- $\mathbb{P}_s$: Target success probability of the attack (= prob. that correct key is ranked first among all guessed keys)
- Φ^{-1}: Quantile function (inverse Cumulative Distribution Function) of the standard normal distribution $\mathcal{N}(0, 1)$

More Tricks

- Clusters: Find multiple characteristics that match the same differential for higher p
  - This is easier if they follow the same pattern of active S-boxes
- Initial Structures: Also append 1 round *before* the characteristic
  - Learn more key material
  - Allow more input differences to generate pairs more efficiently
- For Tweakable Block Ciphers: Put differences in the tweak (“related-tweak” model)
- For unkeyed Permutations, Compression Functions: Use “message modification” to control correct solutions for differences in some steps

"Cheating" with Differences: Changing the Intermediates, not the Input



differential cryptanalysis   differential fault analysis   statistical fault analysis

Caveats and Assumptions

Some Grains of Salt

Some Grains of Salt I

"Markov assumption"

"Expected differential probability (EDP)"

"Hypothesis of stochastic equivalence"

"Wrong key randomization hypothesis"

"Dominant trail assumption"

Some Grains of Salt II

- This "probability" is the average over (all inputs and) all round keys
  - "Expected Differential Probability" (EDP)
  - Ignoring the key schedule's properties
  - Ignoring possible dependencies between rounds: "Markov cipher assumption"
  - Assuming the attacker doesn't know/control intermediate values (hash!)
- The "generic probability" of 2^{-b} is also an average over all $f : \mathbb{F}_2^b \to \mathbb{F}_2^b$
  - For any fixed key, any differential has $p = 0$ or $p \geq 2^{-b+1}$ (DDT!)
  - For a random function and any differential, this p is binomially distributed

Nevertheless, we assume a fixed key behaves $\approx$ like the EDP: this is the

"Hypothesis of stochastic equivalence"

Conclusion

- Differential cryptanalysis is one of the two major statistical attack techniques
  - Attacker tries to find high-probability characteristics
  - Designer tries to show that none exist (but there is no general proof of security)
- It is very versatile
  - many different variants (truncated, impossible, higher-order, ...)
  - many different goals (key, forgery, collision, ...)
- The analysis relies on a number of assumptions & approximations.
  They are usually "reasonably close" to reality, but need to check!

Questions

Questions you should be able to answer

1. Describe the basic idea of differential cryptanalysis. What is the differential distribution table (DDT)?
2. Explain the role of “branch number” and “differential uniformity” in cipher design.
3. Explain an approach to find (optimal) differential characteristics.
4. How is the secret key recovered in differential cryptanalysis?
5. What is a differential characteristic and a differential? How is the probability of a differential computed or approximated? Explain the problems associated with this approximation.
6. Assume you have a new block cipher with 128-bit block size and key size, and you know that the optimal differential characteristic for $r - 1$ (out of r) rounds has a differential probability of $p < 2^{-128}$. Does this guarantee that the cipher is secure against differential cryptanalysis? Discuss why / why not.

Bibliography I

[BS90] Eli Biham and Adi Shamir. **Differential Cryptanalysis of DES-like Cryptosystems.** Advances in Cryptology – CRYPTO 1990. Vol. 537. LNCS. Springer, 1990, pp. 2–21. DOI: 10.1007/3-540-38424-3_1.

[Dae95] Joan Daemen. **Cipher and Hash Function Design. Strategies based on linear and differential cryptanalysis.** PhD thesis. Katholieke Universiteit Leuven, 1995. URL: https://www.esat.kuleuven.be/cosic/publications/thesis-6.pdf.

[LM01] Helger Lipmaa and Shiho Moriai. **Efficient Algorithms for Computing Differential Properties of Addition.** Fast Software Encryption – FSE 2001. Vol. 2355. LNCS. Springer, 2001, pp. 336–350. DOI: 10.1007/3-540-45473-X_28.

[SB02] Ali Aydın Selçuk and Ali Biçak. **On Probability of Success in Linear and Differential Cryptanalysis.** Security in Communication Networks – SCN 2002. Vol. 2576. LNCS. Springer, 2002, pp. 174–185. DOI: 10.1007/3-540-36413-7_13.