Streebog compression function as PRF in secret-key settings

Vitaly Kiryukhin

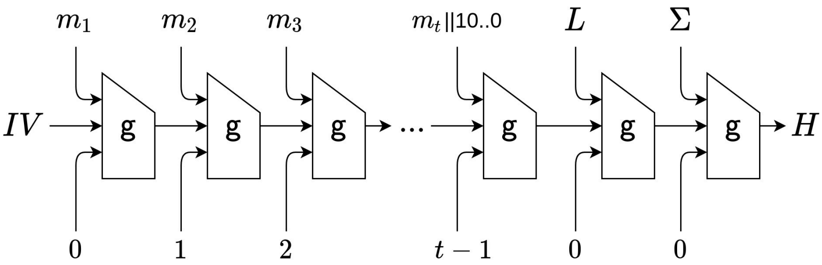
JSC «InfoTeCS», LLC «SFB Lab»

CTCrypt 2021

June 3, 2021

vitaly.kiryukhin@infotecs.ru

GOST R 34.11-2012 – «Streebog»

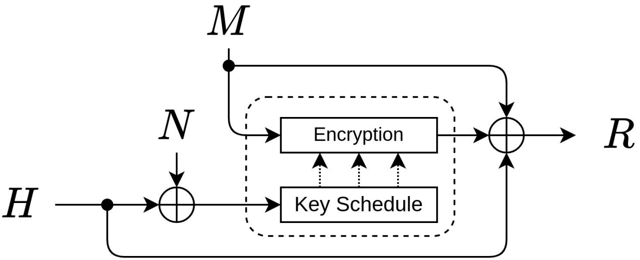


- Slightly modified Merkle-Damgård structure
- 512-bit compression function g : $V^{512} \times V^{512} \times V^{512} \to V^{512}$
- Finalization with message bit-length L and checksum Σ

Compression function

$g_N(H, M)$ – AES-like XSPL-cipher E in the Miyaguchi-Preenel mode

$$g_N(H, M) = E(H \oplus N, M) \oplus H \oplus M = R$$



H – the previous state of the hash function

M – the message block

N – is the number of previously hashed bits

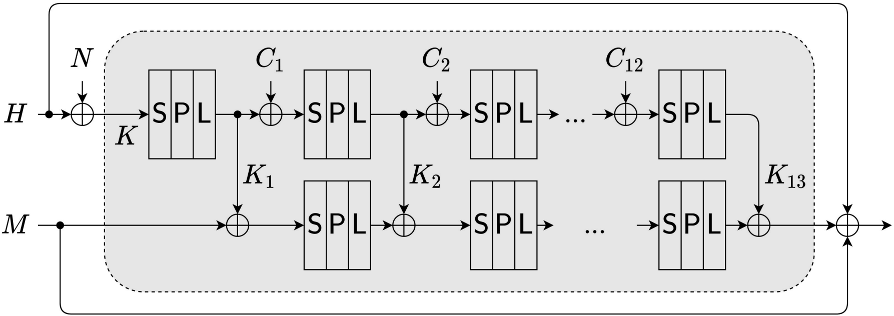
R – the output (the next state)

Block cipher

- 12 rounds (13 keys)
- $v \times v = 8 \times 8$ bytes state ($n = 512$ bits)

$$\mathsf{E}(K = H \oplus N, M) = \mathsf{X}[K_{13}]\mathsf{LPSX}[K_{12}]\ldots\mathsf{LPSX}[K_2]\mathsf{LPSX}[K_1](M)$$

$$K_1 = \mathsf{LPS}(K),\ \ K_{i+1} = \mathsf{LPS}(K_i \oplus C_i),\ \ i = 1, 2, \ldots, 12$$



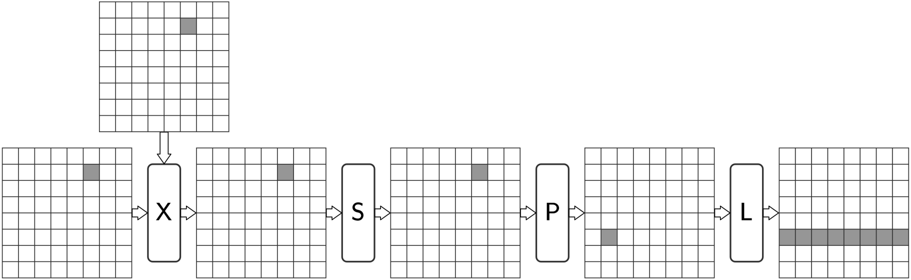
Round

X – modulo 2 addition with a round key

S – parallel application of substitution to each byte

P – transposition

L – parallel application of the linear transformation to each row



Main security properties of a keyless hash function

We expect the keyless hash-functions and the compression function to have three properties:

- preimage resistance: $H = \mathrm{Hash}(M) \Rightarrow M$
- second preimage resistance: $M \Rightarrow M' \neq M$, $\mathrm{Hash}(M) = \mathrm{Hash}(M')$
- collision resistance: (M, M'), $\mathrm{Hash}(M) = \mathrm{Hash}(M')$

Main security properties of a keyless hash function

We expect the keyless hash-functions and the compression function to have three properties:

- preimage resistance: $H = \text{Hash}(M) \Rightarrow M$
- second preimage resistance: $M \Rightarrow M' \neq M$, $\text{Hash}(M) = \text{Hash}(M')$
- collision resistance: (M, M'), $\text{Hash}(M) = \text{Hash}(M')$

Many papers devoted to the preimage, the second preimage, various types of the collisions, «known-key» and «chosen-key» distinguishers of Streebog (as well as its compression function and block cipher).

Secret-key settings

Keyless hash function is often used as part of the **secret-key** cryptoalgorithms:

1. HMAC, NMAC, secret-IV MAC etc.
2. Key trees, key derivation functions

The security of such algorithms depends significantly on the fact that the **compression function** is a **PRF**.

Secret-key settings

PRF: compression function $g_K(M)$ with the secret-key K must be **indistinguishable** from the random function ρ under adaptively chosen message attacks

Secret-key settings

PRF: compression function $g_K(M)$ with the secret-key K must be **indistinguishable** from the random function ρ under adaptively chosen message attacks

$$\mathrm{Adv}_{\mathsf{g}}^{PRF}(\mathcal{A}) = \Big| \Pr\left(K \xleftarrow{\$} V^n : \mathcal{A}^{g_K(\cdot)} \Rightarrow 1\right) - \\ - \Pr\left(\rho \xleftarrow{\$} \mathrm{Func}(V^n, V^n) : \mathcal{A}^{\rho(\cdot)} \Rightarrow 1\right) \Big|$$

Secret-key settings

We have two cases, as a secret key can be used:

1. the previous state H
2. the message block M

1) The previous state H as a secret key

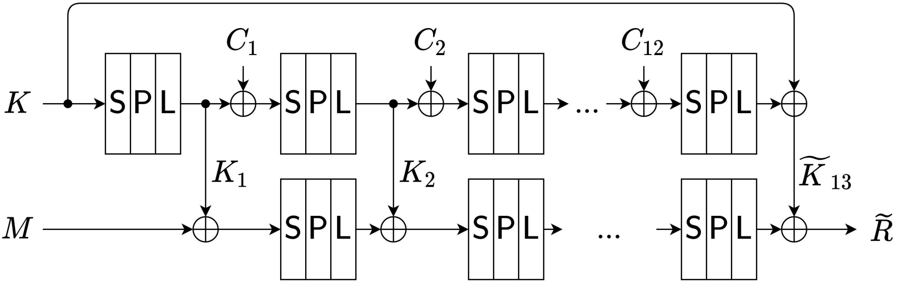
H as a secret key

The analysis is reduced to the block cipher

$$\mathsf{E}(H, M) \oplus H = R \oplus M = \widetilde{R},$$

$$\mathsf{E}(H, M) = \mathsf{X}[K_{r+1} \oplus H]\mathsf{LPSX}[K_r] \ldots \mathsf{LPSX}[K_1](M),$$

where the last round key is $\widetilde{K}_{r+1} = K_{r+1} \oplus H$.



Generic attacks

Secure as the underlying block cipher (up to the birthday-paradox):

$$\mathrm{Adv}^{PRF}_{\mathsf{g}(K,\cdot)}(t, q) \leq \mathrm{Adv}^{PRP}_{\widetilde{\mathsf{E}}}(t, q) + \frac{q^2}{2^{n+1}}.$$

1. Key guessing: time-complexity $t \approx 2^n$ operations
2. Birthday-paradox distinguisher: data-complexity $q \approx 2^{n/2}$ queries

Previously known results

Rounds	Time	Memory	Data	Description
6.75	$2^{399.5}$	2^{349}	2^{483}	[AAY15]
6.75	$2^{261.5}$	2^{205}	$2^{495.5}$	[AAY15]
12	2^{256}	2^{256}	2^{256}	birthday-paradox
12	2^{512}	$\sim$	2	key guessing

[AAY15] Abdelkhalek A., AlTawy R., Youssef A. M. –
Impossible Differential Properties of Reduced Round Streebog – 2015

$q \gg 2^{n/2} \Rightarrow$ the attack is built only against the PRP-property

Previously known results

We can use a lot of results about AES-128.

The most effective of them are:

- Meet-ih-the-Middle ($t \approx q \approx 2^{99}$ against 7-rounds)
- Impossible Differentials ($t \approx q \approx 2^{112}$ against 7-rounds)

And again $q \gg 2^{n/2}$.

New method against Streebog compression function

We propose key-recovery algorithm with $q \ll 2^{n/2}$
for 7-round Streebog compression function.

The proposed method based on *polytopic* approach.
[Tiessen T. – *Polytopic Cryptanalysis* – EUROCRYPT 2016]

Impossible Polytopic (multidimensional differential)

Differential method

- pair of blocks B_0 and B_1
- difference $\Delta B = B_0 \oplus B_1$

Impossible Polytopic (multidimensional differential)

Differential method

- pair of blocks B_0 and B_1
- difference $\Delta B = B_0 \oplus B_1$

Polytopic (multidimensional differential) method

- vector of $(d+1)$ blocks B_0, B_1, B_2, ..., B_d
- d-difference $\boldsymbol{\delta B} = (B_0 \oplus B_1, B_0 \oplus B_2, \ldots, B_0 \oplus B_d)$
- B_0 is an «anchor» or «reference point»

[Tiessen T. – *Polytopic Cryptanalysis* – EUROCRYPT 2016]

Impossible Polytopic (multidimensional differential)

Difference ΔB and d-difference $\boldsymbol{\delta B}$ are propagated in a similar way:

Impossible Polytopic (multidimensional differential)

Difference ΔB and d-difference $\boldsymbol{\delta B}$ are propagated in a similar way:

- X — not change

Impossible Polytopic (multidimensional differential)

Difference ΔB and d-difference $\boldsymbol{\delta B}$ are propagated in a similar way:

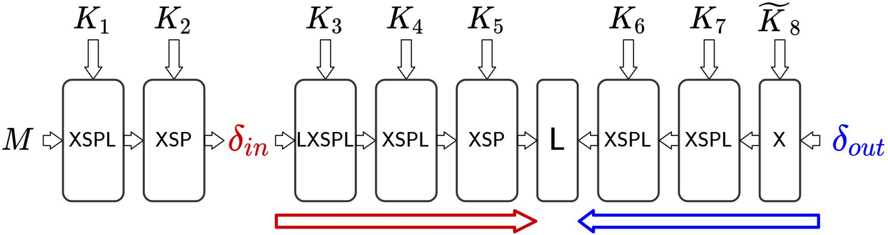
- X — not change
- P — bijective
- L — bijective

Impossible Polytopic (multidimensional differential)

Difference ΔB and d-difference $\boldsymbol{\delta B}$ are propagated in a similar way:

- X — not change
- P — bijective
- L — bijective
- S — **non-bijective**
    - if «anchor» B_0 is known then the propagation is also **bijective**

New method



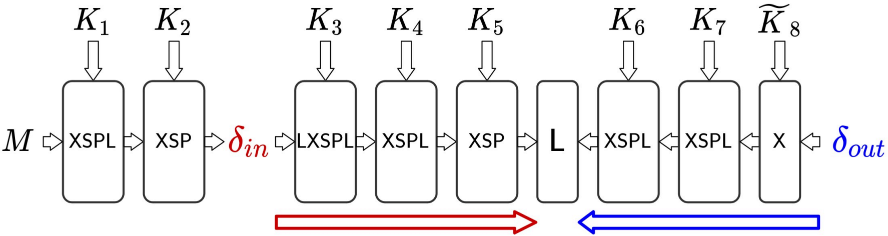
1) Choose structure of 2^{64} messages

2) Guess 64 bits of K_1. Partially encrypt all messages

3) Choose $d = 2^7$ blocks (of 2^{64}) and

obtain d-difference δ_{in} with only one active S-box

New method



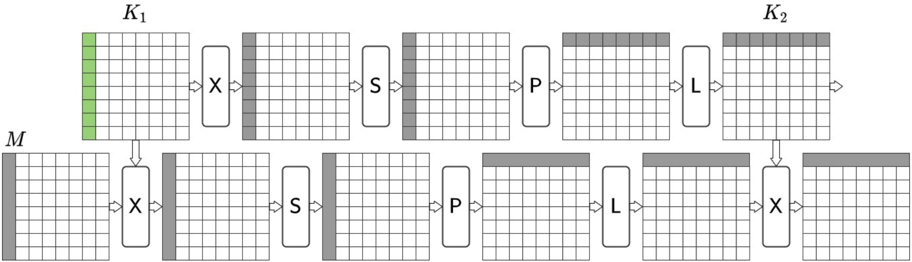
4) Propagate δ_{in} forward by guessing 136 bits

5) Propagate δ_{out} backward by guessing 72 bits eight times

6) Check by naive algorithm for «generalized birthday problem» that δ_{in} and δ_{out} are compatible

- failed $\Rightarrow$ go to step 2 and try another bits of K_1
- passed $\Rightarrow$ the key bits and the state bits are guessed correctly

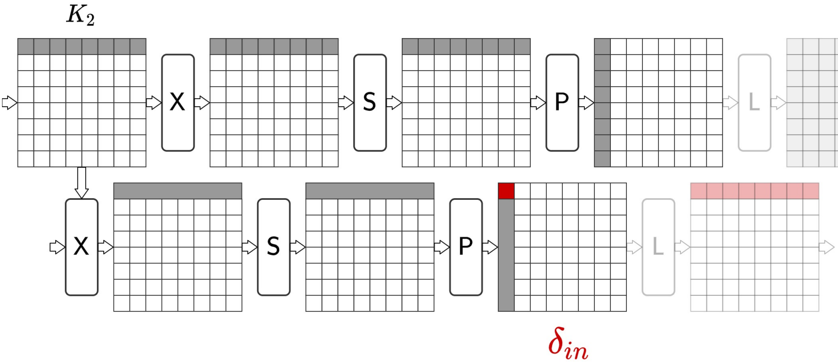
New method – steps 1-2



Choose structure of 2^{64} messages

Guess 64 bits of K_1. Partially encrypt all messages

New method – step 3



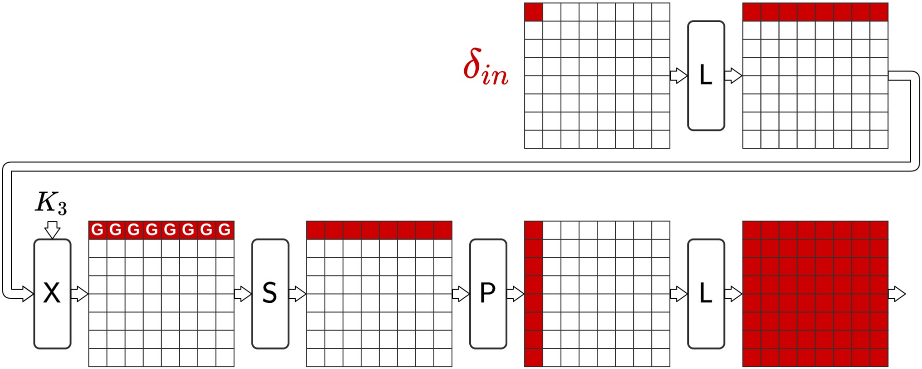
Choose $d = 2^7$ blocks (of 2^{64}) and

obtain d-difference δ_{in} with only one active S-box

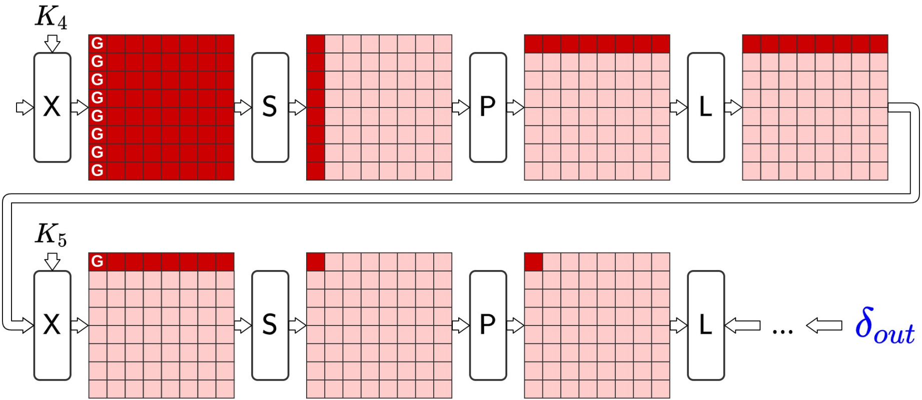
New method – step 4

Propagate δ_{in} forward by guessing $8 \cdot (8 + 8 + 1) = 136$ bits

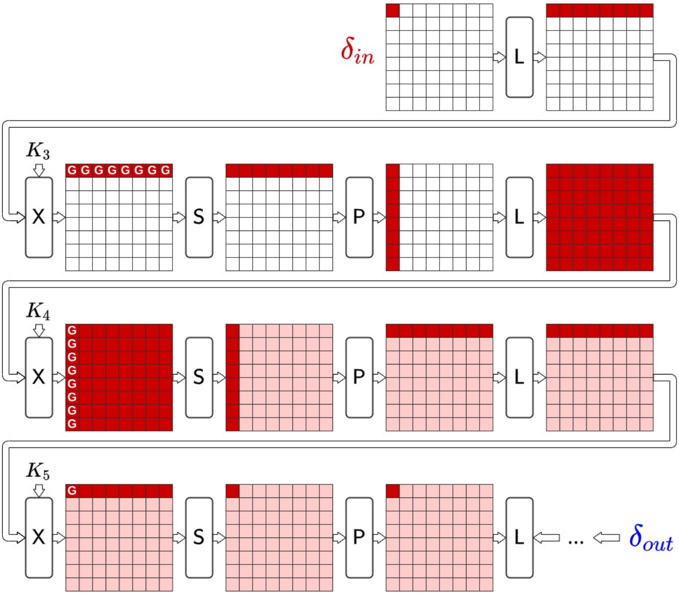
New method – step 4



New method – step 4



New method – step 4

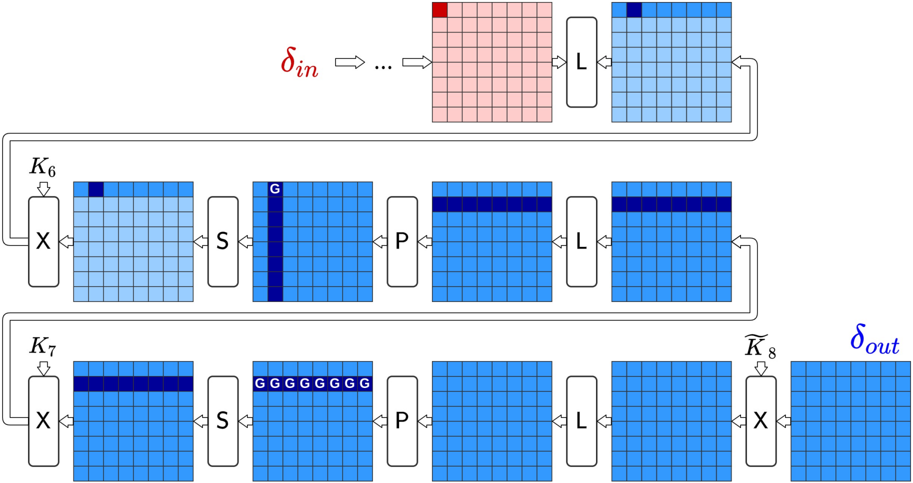


New method – step 5

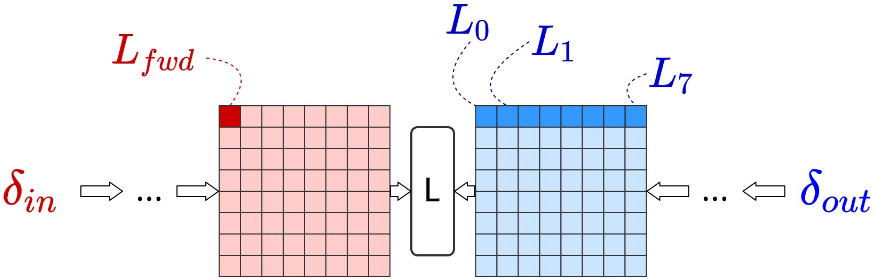
Propagate $\boldsymbol{\delta}_{out}$ backward by

guessing $8 \cdot (8+1) = 72$ bits independently eight times

New method – step 5

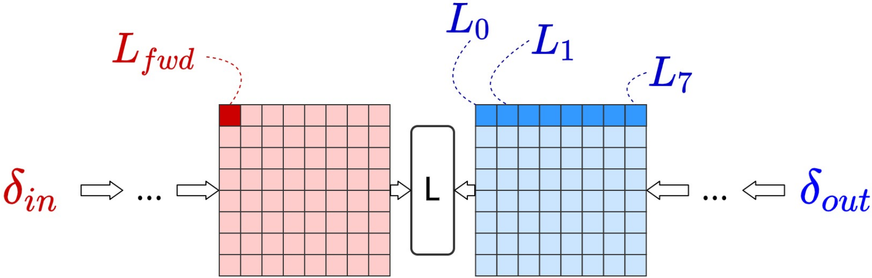


New method – step 6



- $\mathcal{L}_{\text{fwd}}$ – array of «forward» *d*-differences, $|\mathcal{L}_{\text{fwd}}| = 2^{136}$
- $\mathcal{L}_0$, $\mathcal{L}_1$, ..., $\mathcal{L}_7$ – arrays of «backward» *d*-differences, $|\mathcal{L}_j| = 2^{72}$

New method – step 6



- $\mathbb{L} \in \mathbb{F}_{2^8}^{8\times 8}$ is the matrix of the linear transformation
- $c_0, c_1, \ldots, c_7 \in \mathbb{F}_{2^8}$ are the coefficients from the column of $\mathbb{L}^{-1}$

$$\mathcal{L}_{\text{fwd}}[i_{\text{fwd}}] = c_0 \cdot \mathcal{L}_0[i_0] \oplus c_1 \cdot \mathcal{L}_1[i_1] \oplus \ldots \oplus c_7 \cdot \mathcal{L}_7[i_7]$$

New method – step 6 – «generalized birthday problem»

We obtain a «generalized birthday problem»

$$\mathcal{L}_{\mathrm{fwd}}[i_{\mathrm{fwd}}] = c_0 \cdot \mathcal{L}_0[i_0] \oplus c_1 \cdot \mathcal{L}_1[i_1] \oplus \ldots \oplus c_7 \cdot \mathcal{L}_7[i_7]$$

but we have no task to find at least some «collision».

Our goal is **one unique correct** solution

$$(i_{\mathrm{fwd}},\ i_0, i_1, i_2, \ldots i_7).$$

New method – step 6 – «generalized birthday problem»

$$\mathcal{L}_{\mathrm{fwd}}[i_{\mathrm{fwd}}] = c_0 \cdot \mathcal{L}_0[i_0] \oplus c_1 \cdot \mathcal{L}_1[i_1] \oplus \ldots \oplus c_7 \cdot \mathcal{L}_7[i_7]$$

Rearrange the components:

$$\underbrace{\mathcal{L}_{\mathrm{fwd}}[i_{\mathrm{fwd}}] \oplus c_0 \cdot \mathcal{L}_0[i_0] \oplus c_1 \cdot \mathcal{L}_1[i_1] \oplus c_2 \cdot \mathcal{L}_2[i_2]}_{\mathrm{left}} = \underbrace{c_3 \cdot \mathcal{L}_3[i_3] \oplus \ldots \oplus c_7 \cdot \mathcal{L}_7[i_7]}_{\mathrm{right}}$$

Combine all lists:

$$\mathcal{L}_{\mathrm{left}}[i_{\mathrm{left}}] = \mathcal{L}_{\mathrm{right}}[i_{\mathrm{right}}]$$

$\mathcal{L}_{\mathrm{left}}$ is stored in memory, $|\mathcal{L}_{\mathrm{left}}| = 2^{136} \cdot \left(2^{72}\right)^3 = 2^{352}$

$\mathcal{L}_{\mathrm{right}}$ is iterated dynamically, $|\mathcal{L}_{\mathrm{right}}| = \left(2^{72}\right)^5 = 2^{360}$

New method – step 6 – «generalized birthday problem»

If solution $(i_{\text{left}}, i_{\text{right}})$ of $\mathcal{L}_{\text{left}}[i_{\text{left}}] = \mathcal{L}_{\text{right}}[i_{\text{right}}]$ is found then

- d-difference trail $\delta_{in} \to \delta_{out}$ exists
- all key and state bits are correctly guessed
- $2^{64} \cdot 2^{352} \cdot 2^{360} \cdot 2^{-d \cdot 8} = 2^{-240} \approx 0$ false solutions

else

- try another 64 bits of K_1

Complexity

7-round attack

$$t \approx \underbrace{2^{64}}_{K_1} \cdot d \cdot \left(\underbrace{2^{136}}_{\rightarrow} + \underbrace{8 \cdot 2^{72}}_{\leftarrow} + \underbrace{2^{352}}_{\mathcal{L}_{\text{left}}} + \underbrace{2^{360}}_{\mathcal{L}_{\text{right}}} \right)$$

- $t \approx 2^{431}$ table lookups $\Rightarrow$ about $t = 2^{431} \cdot 2^{-10} = 2^{421}$ computations
- 2^{354} (n-bit states) of memory
- $q = 2^{64}$ chosen pairs (M, R)
- the success probability is equal to one

Application to AES-128

The ideas of the proposed method can be applied to 6 rounds of AES-128:

- $t = 2^{120}$ memory access operations
- small amount of the chosen plaintexts $q = d + 1 < 2^5$

Application to AES-128

The ideas of the proposed method can be applied to 6 rounds of AES-128:

- $t = 2^{120}$ memory access operations
- small amount of the chosen plaintexts $q = d + 1 < 2^5$

«Meet-ih-the-Middle» approach:

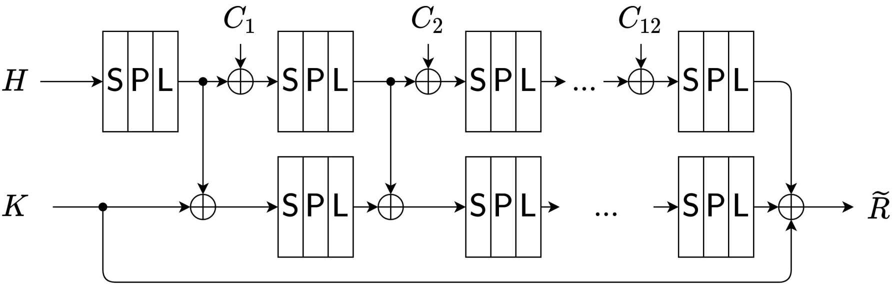
- $t_{MitM} = 2^{106} < t$
- $q_{MitM} = 2^8 > q$

[Derbez P., Fouque P.-A. Exhausting Demirci-Selcuk Meet-in-the-Middle Attacks against Reduced-Round AES – 2015]

2) The message block M as a secret key

M as a secret key

An adversary has a full control over H and the round keys



Generic attacks

$\mathrm{g}(\cdot, K)$ is a secure PRF in the ideal cipher model
(i.e. if E is a family of random permutations)

$$\mathrm{Adv}^{PRF}_{\mathrm{g}(\cdot,K)}(t, q) \leq \frac{t}{2^{n-1}}$$

1. Key guessing: time-complexity $t \approx 2^n$ operations
2. ~~Birthday-paradox distinguisher: data-complexity $q \approx 2^{n/2}$ queries~~
   In this case, there is NO simple birthday-paradox distinguisher

Previously known results

Rounds	Time	Memory	Data	Description
12	2^{512}	$\sim$	2	key guessing

As far as we know,

the non-trivial results in this model have not been published.

New method

We propose the algorithm against seven rounds.

New method

We propose the algorithm against seven rounds.

«Offline» stage

- rebound approach
- 2^{112} pairs (H, H') are formed
- each pair generates a truncated differential trail

  «8 – 1 – 8 – 64 – 16 – 16 – 64 – 64»

New method

We propose the algorithm against seven rounds.

«Offline» stage

- rebound approach
- 2^{112} pairs (H, H') are formed
- each pair generates a truncated differential trail

  «8 – 1 – 8 – 64 – 16 – 16 – 64 – 64»

«Online» stage

- the truncated related-key trail with a probability of at least 2^{-112}

  «8 – 0 – 8 – 0 – 16 – 16 – 64 – 64»
- for each attempt about 2^{128} possible values of the unknown state
- if trail was realized then

  among the constructed solutions there will be a true one

«Offline» stage

Construct the suitable round keys for the block cipher.

Rebound approach:

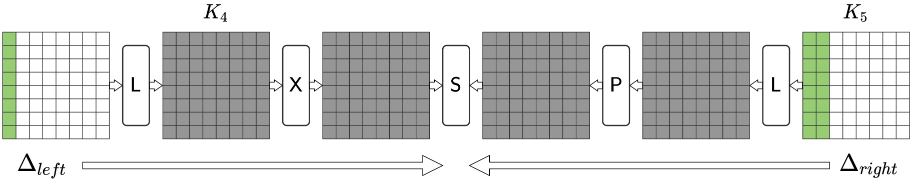
$$\Delta K_4 \Rightarrow \cdot \Leftarrow \Delta K_5$$

$$\Delta K_1 \Leftarrow \Delta K_2 \Leftarrow \Delta K_3 \Leftarrow \Delta K_4 \Leftarrow \cdot \Rightarrow \Delta K_5 \Rightarrow \Delta K_6 \Rightarrow \Delta K_7 \Rightarrow \Delta K_8$$

$$8 \Leftarrow \quad 1 \Leftarrow \quad 8 \Leftarrow \quad 64 \Leftarrow \cdot \Rightarrow 16 \quad \Rightarrow 16 \quad \Rightarrow 64 \quad \Rightarrow 64$$

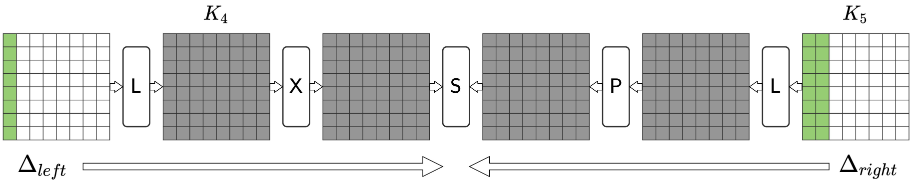
«Offline» stage

Rebound approach. «Inbound»



«Offline» stage

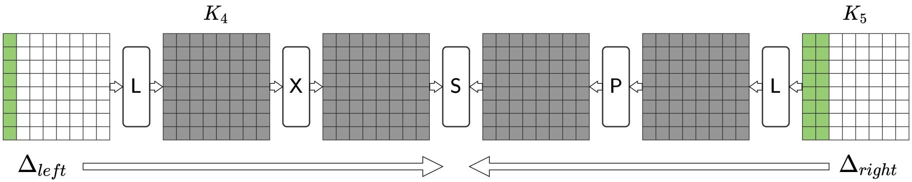
Rebound approach. «Inbound»



- one active column from the left Δ_{left}
- two active columns from the right Δ_{right}

«Offline» stage

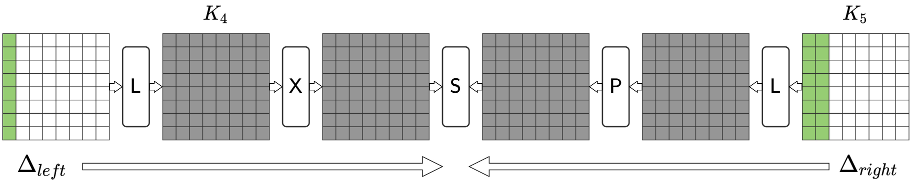
Rebound approach. «Inbound»



- one active column from the left Δ_{left}
- two active columns from the right Δ_{right}
- $\approx \left(2^8\right)^{8+16} = 2^{192}$ pairs $(\Delta_{left}, \Delta_{right})$

«Offline» stage

Rebound approach. «Inbound»

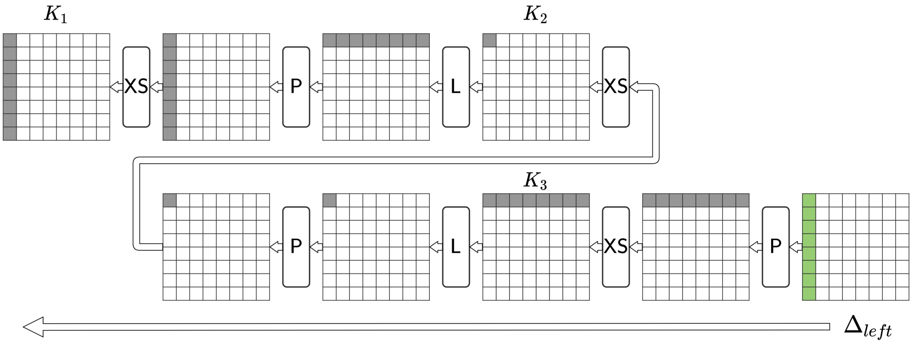


- one active column from the left Δ_{left}
- two active columns from the right Δ_{right}
- $\approx (2^8)^{8+16} = 2^{192}$ pairs $(\Delta_{left}, \Delta_{right})$
- 2^{192} solutions

$$\mathsf{S}(x \oplus \mathsf{L}(\Delta_{left})) \oplus \mathsf{S}(x) = \mathsf{P}^{-1}\mathsf{L}^{-1}(\Delta_{right})$$

«Offline» stage

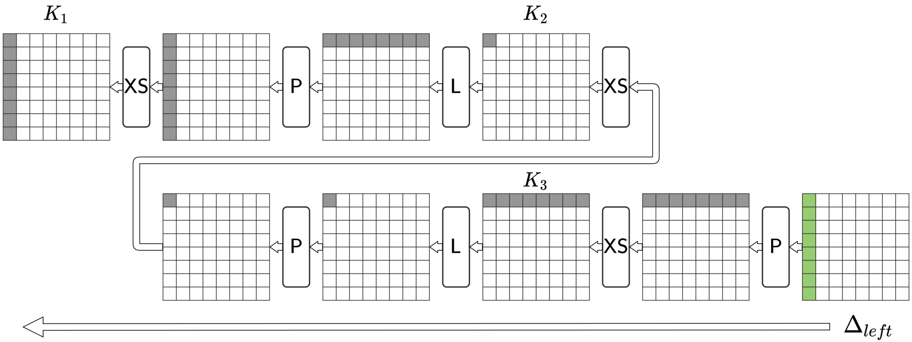
Rebound approach. «Outbound». Left side.



- one transition «$1 \leftarrow 8$»

«Offline» stage

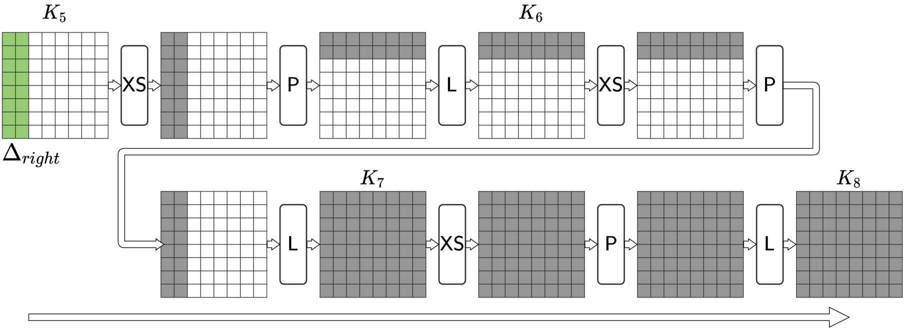
Rebound approach. «Outbound». Left side.



- one transition «$1 \leftarrow 8$»
- there are only $\approx 2^{136} = 2^{192} \cdot 2^{-56}$ solutions remain

«Offline» stage

Rebound approach. «Outbound». Right side.



- Almost all 2^{136} solutions remain

«Online» stage

Truncated related-key differential trail

- About 2^{136} pairs (H, H')

«Online» stage

Truncated related-key differential trail

- About 2^{136} pairs (H, H')
- Input $H \Rightarrow K_1 \Rightarrow \ldots \Rightarrow K_8$
- Input $H' \Rightarrow K'_1 \Rightarrow \ldots \Rightarrow K'_8$

«Online» stage

Truncated related-key differential trail

- About 2^{136} pairs (H, H')
- Input $H \Rightarrow K_1 \Rightarrow \ldots \Rightarrow K_8$
- Input $H' \Rightarrow K'_1 \Rightarrow \ldots \Rightarrow K'_8$
- Differential trail $\Delta H \Rightarrow \Delta K_1 \Rightarrow \ldots \Rightarrow \Delta K_8$ over key schedule

«Online» stage

Truncated related-key differential trail

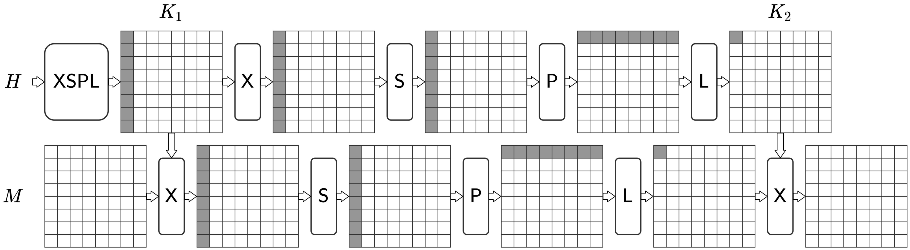
- About 2^{136} pairs (H, H')
- Input $H \Rightarrow K_1 \Rightarrow \ldots \Rightarrow K_8$
- Input $H' \Rightarrow K'_1 \Rightarrow \ldots \Rightarrow K'_8$
- Differential trail $\Delta H \Rightarrow \Delta K_1 \Rightarrow \ldots \Rightarrow \Delta K_8$ over key schedule
- Secret M «encrypted» under $H \Rightarrow$ output R
- Secret M «encrypted» under $H' \Rightarrow$ output R'

«Online» stage

Truncated related-key differential trail

- About 2^{136} pairs (H, H')
- Input $H \Rightarrow K_1 \Rightarrow \ldots \Rightarrow K_8$
- Input $H' \Rightarrow K'_1 \Rightarrow \ldots \Rightarrow K'_8$
- Differential trail $\Delta H \Rightarrow \Delta K_1 \Rightarrow \ldots \Rightarrow \Delta K_8$ over key schedule
- Secret M «encrypted» under $H \Rightarrow$ output R
- Secret M «encrypted» under $H' \Rightarrow$ output R'
- Related-key differential trail $\Delta M_1 \Rightarrow \ldots \Rightarrow \Delta M_8$ over «encryption»

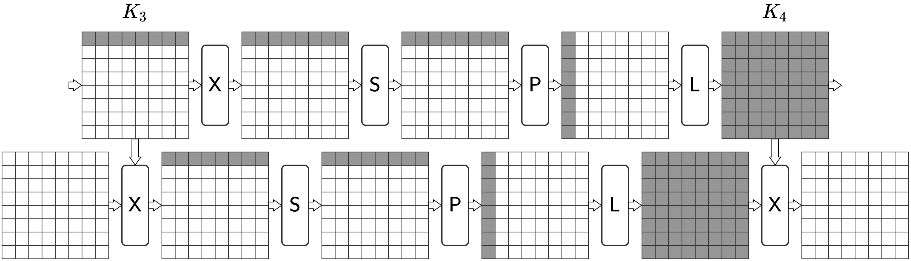
«Online» stage



Both transitions through S are the same:

$$\Pr \geq \left(\frac{2}{256}\right)^8 = 2^{-56}$$

«Online» stage



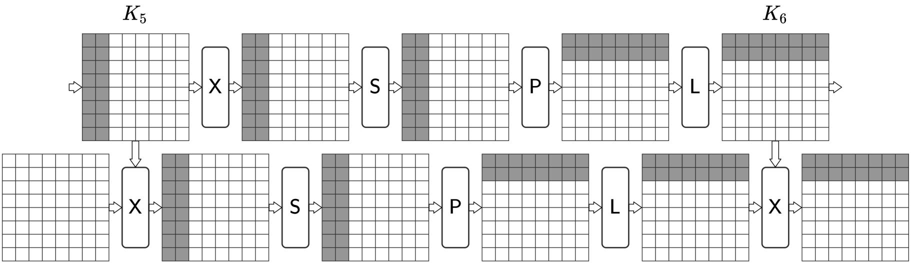
Both transitions through S are the same:

$$\Pr \geq \left(\frac{2}{256}\right)^8 = 2^{-56}$$

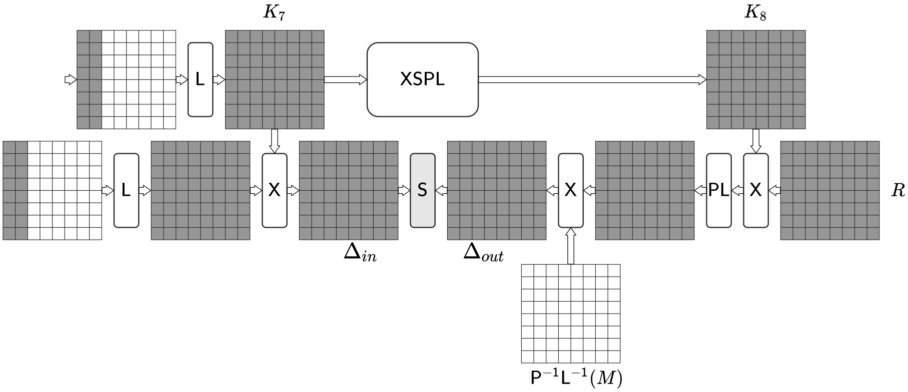
$\Rightarrow$ the probability of the related-key differential trail $\geq 2^{-56} \cdot 2^{-56} = 2^{-112}$

«Online» stage

Two more rounds...

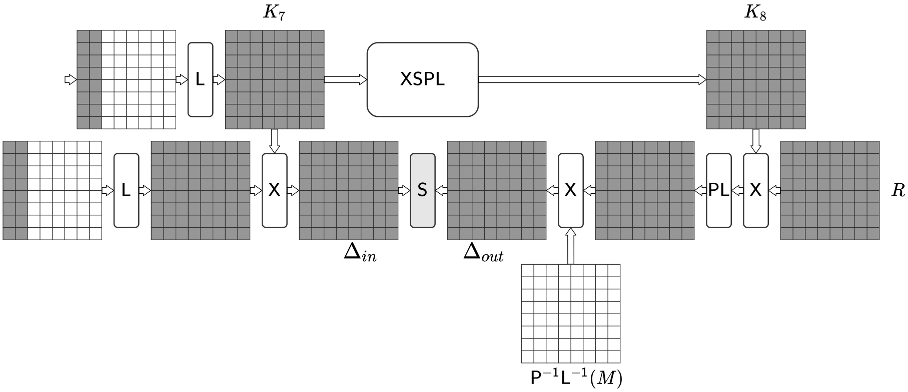


«Online» stage



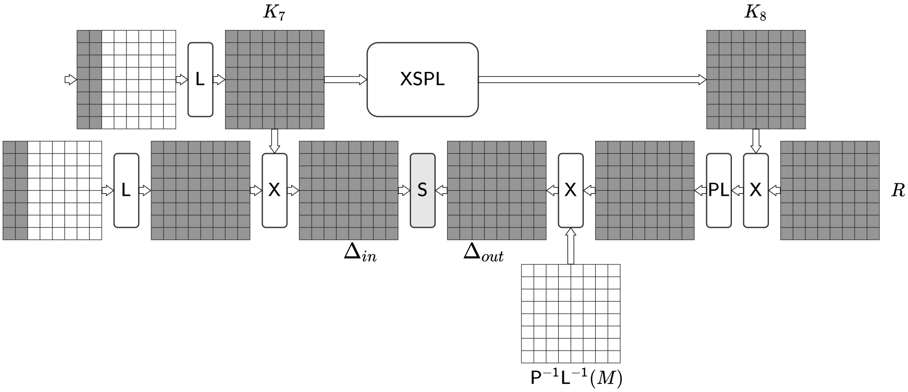
- We know K_8, K_8', R, R'

«Online» stage



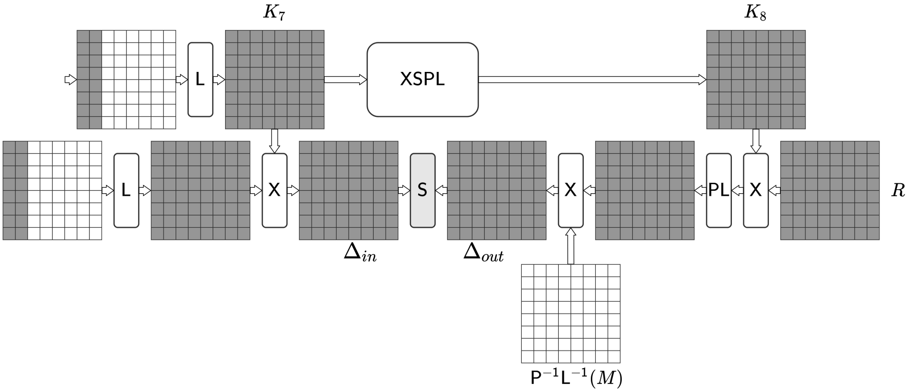
- We know K_8, K_8', R, R'
- Partially decrypt to the last S $\Rightarrow$ we know Δ_{out}

«Online» stage



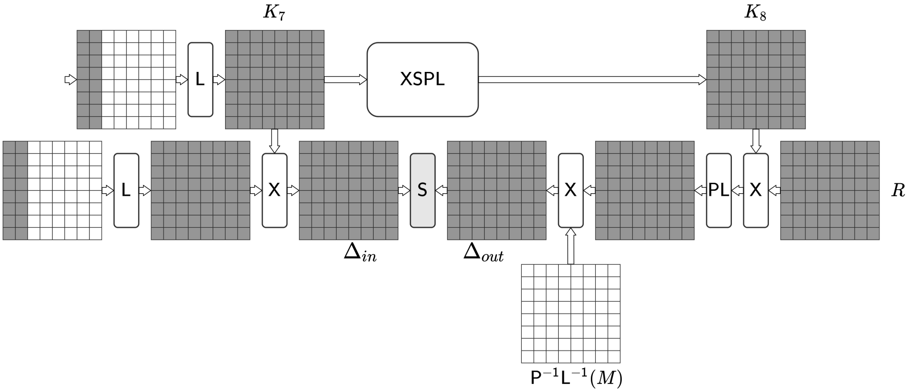
- We know K_8, K'_8, R, R'
- Partially decrypt to the last S $\Rightarrow$ we know Δ_{out}
- The trail is realized $\Rightarrow$ the rows of Δ_{in} belong to the 2^{16}-element set

«Online» stage



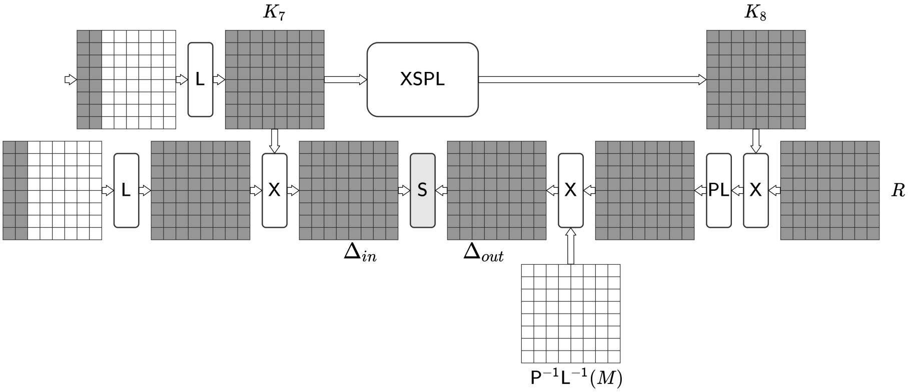
- Solve equation $\mathsf{S}(x \oplus \Delta_{in}) \oplus \mathsf{S}(x) = \Delta_{out}$ row-by-row for all Δ_{in}

«Online» stage



- Solve equation $\mathsf{S}(x \oplus \Delta_{in}) \oplus \mathsf{S}(x) = \Delta_{out}$ row-by-row for all Δ_{in}
- About $\left(2^{16}\right)^8 = 2^{128}$ solutions for the full secret state

«Online» stage



- Solve equation $\mathsf{S}(x \oplus \Delta_{in}) \oplus \mathsf{S}(x) = \Delta_{out}$ row-by-row for all Δ_{in}
- About $\left(2^{16}\right)^8 = 2^{128}$ solutions for the full secret state
- The truth of each M is checked on an arbitrary input-output pair

Complexity

7-round attack

- $t = \underbrace{2^{128} \cdot 2^{64}}_{\text{"offline"}} + \underbrace{2^{112} \cdot 2^{128}}_{\text{"online"}} \approx 2^{240}$ operations
- $q = 2^{113}$ chosen pairs (H, R)
- «Offline» and «Online» stages can be performed simultaneously (negligible memory)
- success probability $1 - (1 - 2^{-112})^{q/2} \approx 1 - e^{-1} \approx 0.63$

Conclusion

We examine Streebog compression function as preudo-random function.

Each of the two inputs (the previous state and the message block) can be used as a secret parameter.

We present two key-recovery algorithms for 7 rounds (of 12).

Setting	Rounds	Time	Memory	Data	Method
secret state	7	2^{421}	2^{354}	2^{64}	impossible polytopic
secret message	7	2^{240}	2^{20}	2^{113}	truncated differentials

Thank you for attention!

Questions?