The Trail Backtracking Attack

Guido Bertoni, Joan Daemen,
Michaël Peeters* and Gilles Van Assche
*NXP Semiconductors STMicroelectronics

Outline

- Introduction
- Trail backtracking
- Panama
- RadioGatún
- Grindahl
- Conclusion

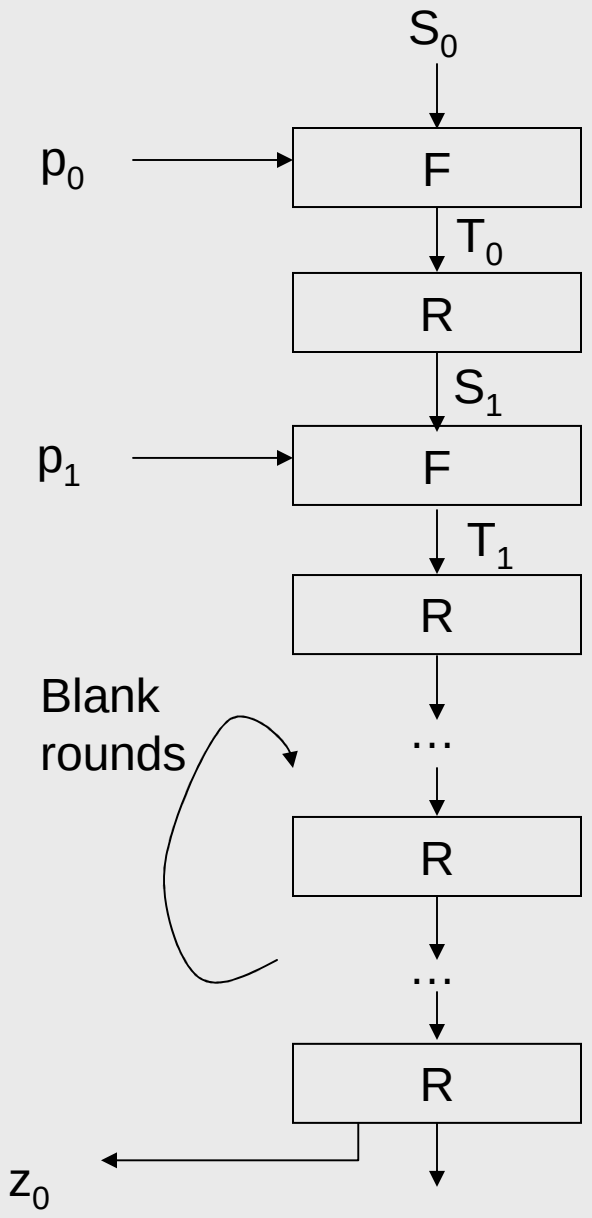
Context

- Trial backtracking is an attack for searching collision in hash functions
- It consists in following a differential trail and estimate the cost of stay in it
- Mainly for stream-oriented hash function

Names, Symbols etc

- H iterated hash function
- p_i blocks of the input message of l bits
- R round function
- F “adds” the input block to the state S
- S_i state at round i

  $S_{i+1} = R(T_i)$
- Size of the state larger than l

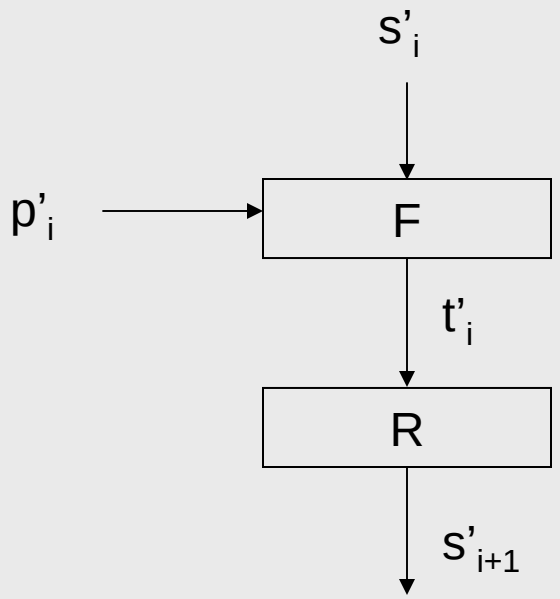


Round Differential

- Differential over round i (t'_i, s'_{i+1})
- Differential Probability DP as the proportion of the states pairs $T_i, T_i + t'_i$ such that

$$R(T_i) \oplus R(T_i \oplus t'_i) = s'_{i+1}$$

- A differential over a round is possible if the $DP > 0$



Round Differential

- A differential (t'_i, s'_{i+1}) imposes a set of conditions over the non linear mapping
- A right pair satisfies these conditions
- The probability of a single round can be expressed as

$$DP \approx 2^{-w_r(t'_i, s'_{i+1})}$$

- w_r is the restriction weight

Collision Trails

- A trail Q is defined as the sequence of differential

$$Q : ((s'_0, p'_0, t'_0), (s'_1, p'_1, t'_1) \ldots s'_r)$$

$$DP(Q) \approx 2^{-w_r(Q)} = 2^{-\sum_{i=0}^{r-1} w_r(t'_i, s'_{i+1})}$$

- A differential trail that starts and ends with zero differences is a collision trail

Trail backtracking

- Fix a differential trail of *r* rounds
- Take N random input pairs entering first round
- Every pair gives $s'_1 = R(t'_0)$ if s'_1 is in the trail, the pair is a right pair
- Number of pairs in output of first round is:

$$N 2^{-w_r(t'_0, s'_1)}$$

Increase/Decrease in the number of pairs

- Depending on the relation between the input l and the weight of the round
  - $l<w_r$ the number of pairs decrease
  - $l>w_r$ the number of pairs increase

Excess Weight

- The number of right pairs entering round g is

$$N2^{gl-\sum_{i=0}^{g-1} w_r(t'_i, s'_{i+1})}$$

- We define W_e as the excess weight

$$N2^{-W_e(g-1)}$$

$$W_e(g) = \sum_{j=0}^{g} w_r(t'_j, s'_{j+1}) - l$$

Trail Backtracking

- There are two interesting rounds
  - The crowded: the round where the number of pairs is maximum, thus the W_e is minimum
  - The lonesome: minimum number of pairs, W_e reaches the maximum
- From these two rounds we can derive:
  - The amount of pairs N
  - The workload

Number of Pairs and Workload

- In every round the pairs coming out should be at least one

$$N \geq \max_{h} 2^{l+W_e(h)} = 2^{l+\max_h W_e(h)}$$

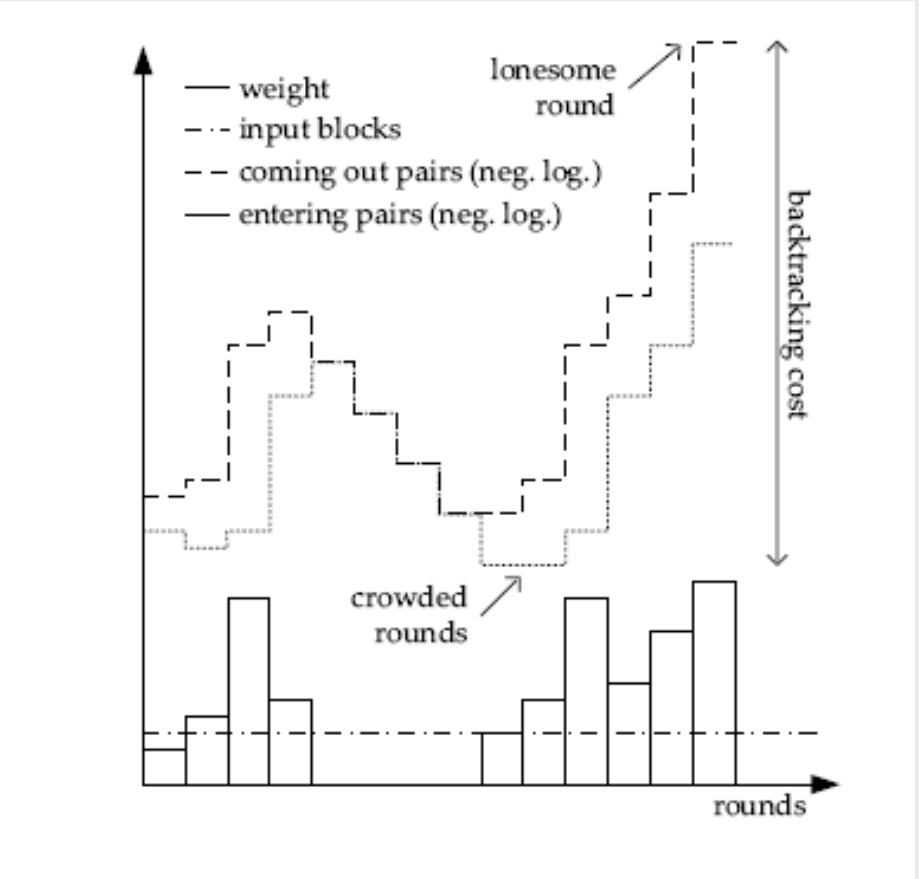
- Work load can be approx with the number of pairs entering the crowded round input

$$\max_{g} N 2^{-W_e(g-1)} = N 2^{-\min_g W_e(g-1)}$$

Backtracking cost

Work factor:

backtracking cost of a trail



$$2^{\max_{0,g,h \le g \le h < r} (W_e(h) - W_e(g-1)) + l}$$

$$\max_{0,g,h \le g \le h < r} (W_e(h) - W_e(g-1)) + l$$

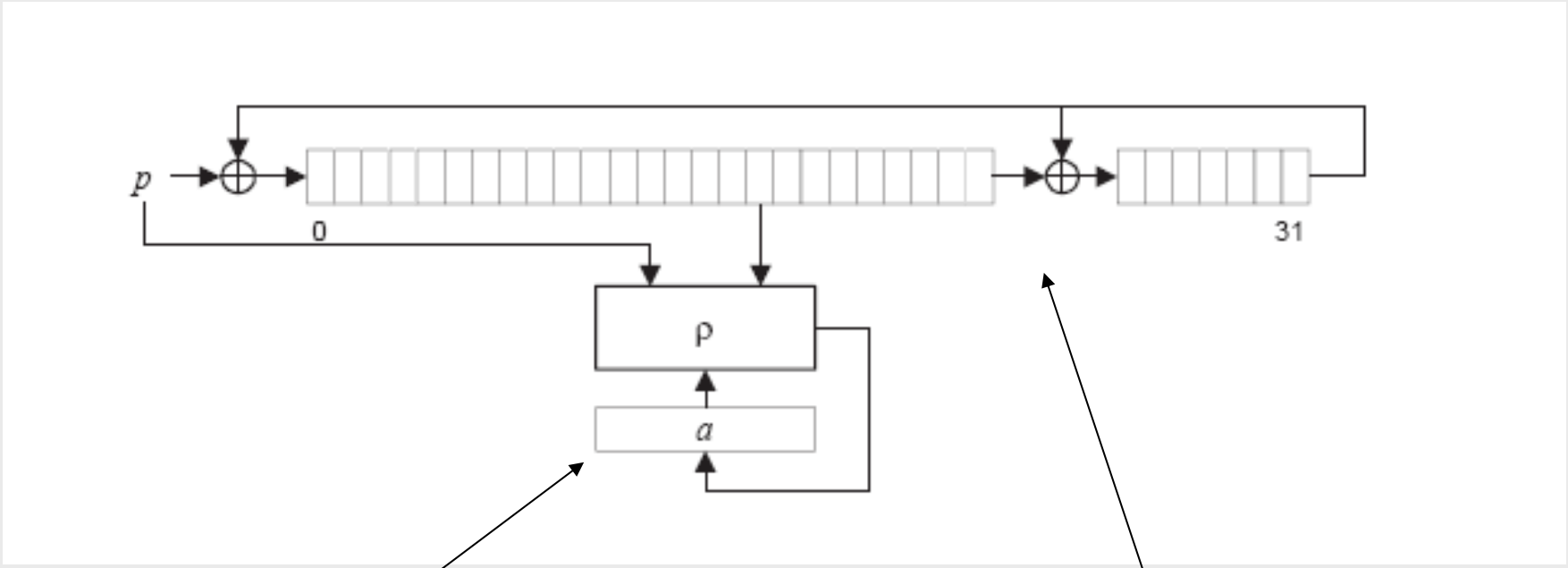
Outline

- ✓ Introduction
- ✓ Trail backtracking
- ❑ **Panama**
- ❑ **RadioGatún**
- ❑ **Grindahl**
- ❑ **Conclusion**

Panama

- Hash/stream cipher presented in 98 by Daemen and Clapp
- Collisions
  - First attack Rijmen et al. 2^{82}
  - broken last year by Daemen and Van Assche

Panama Structure

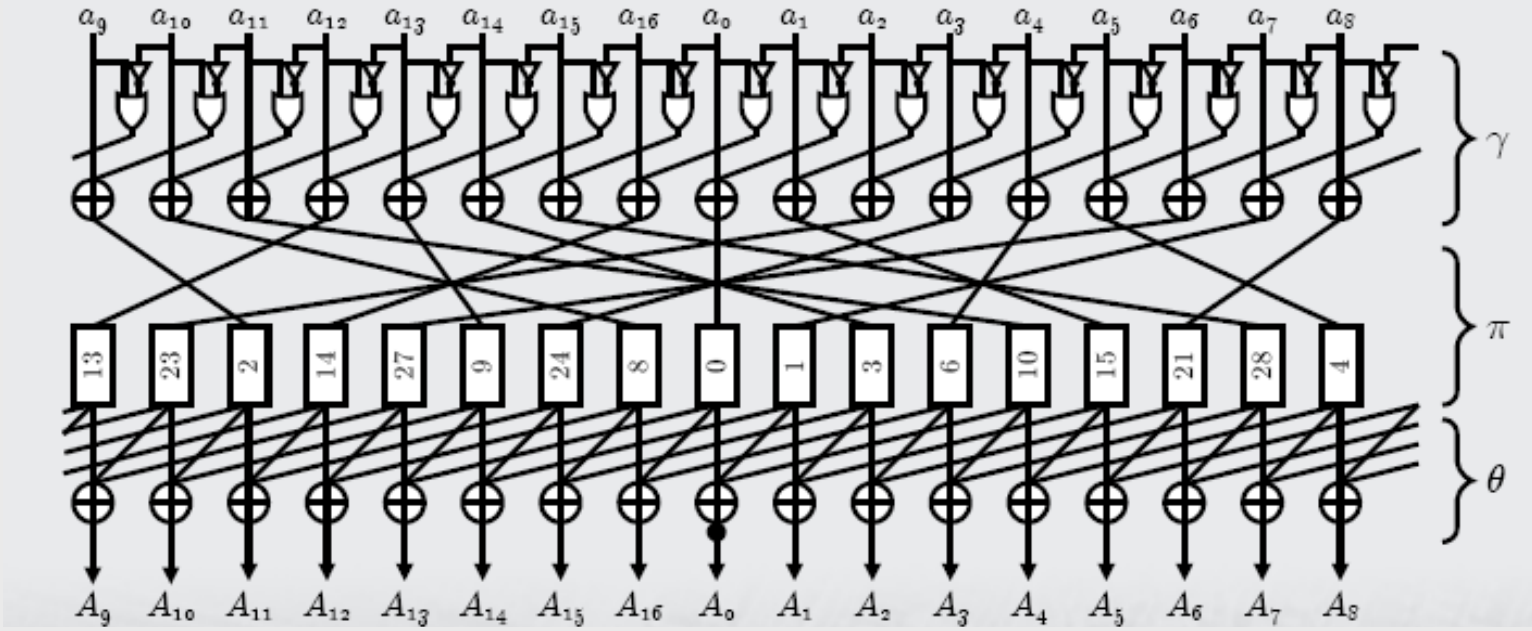


state

buffer

rho

- The rho function is composed by three transformations



Trails for Panama

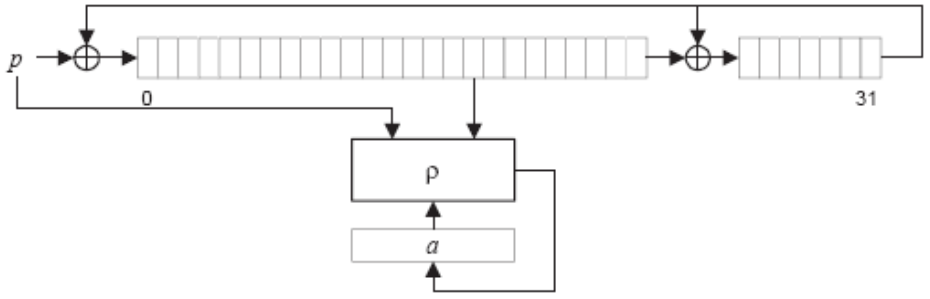
- Once a difference is injected, it spreads fast in the state and move slowly in the buffer
- Buffer is linear, a difference sequence

  $dp^{(0)}, r(dp)^{(7)}, dp^{(32)}$

  will cause a collision in the buffer
- And any combination of this shifted
- For attacking Panama a combination of three of these differences are used

Subcollision in the state

- Difference injected in the state can be canceled in the next two rounds
- A difference is injected:
  - From the input
  - Or from the buffer



Subcollisions

- 5 subcollisions needs to be fixed
- 3 are the inputs, and two are given by the contribution of the buffer to the state
- In the case of Rijmen et al attack only part of the conditions are algebraically moved to inputs, other are satisfied by trials
- In the case of Daemen and Van Assche all are moved to input blocks

Solving subcollision

- The State is composed by 17 words
- 8 are directly controllable through the input
  - Immediate satisfaction
- Others conditions can be moved to previous round or the round before

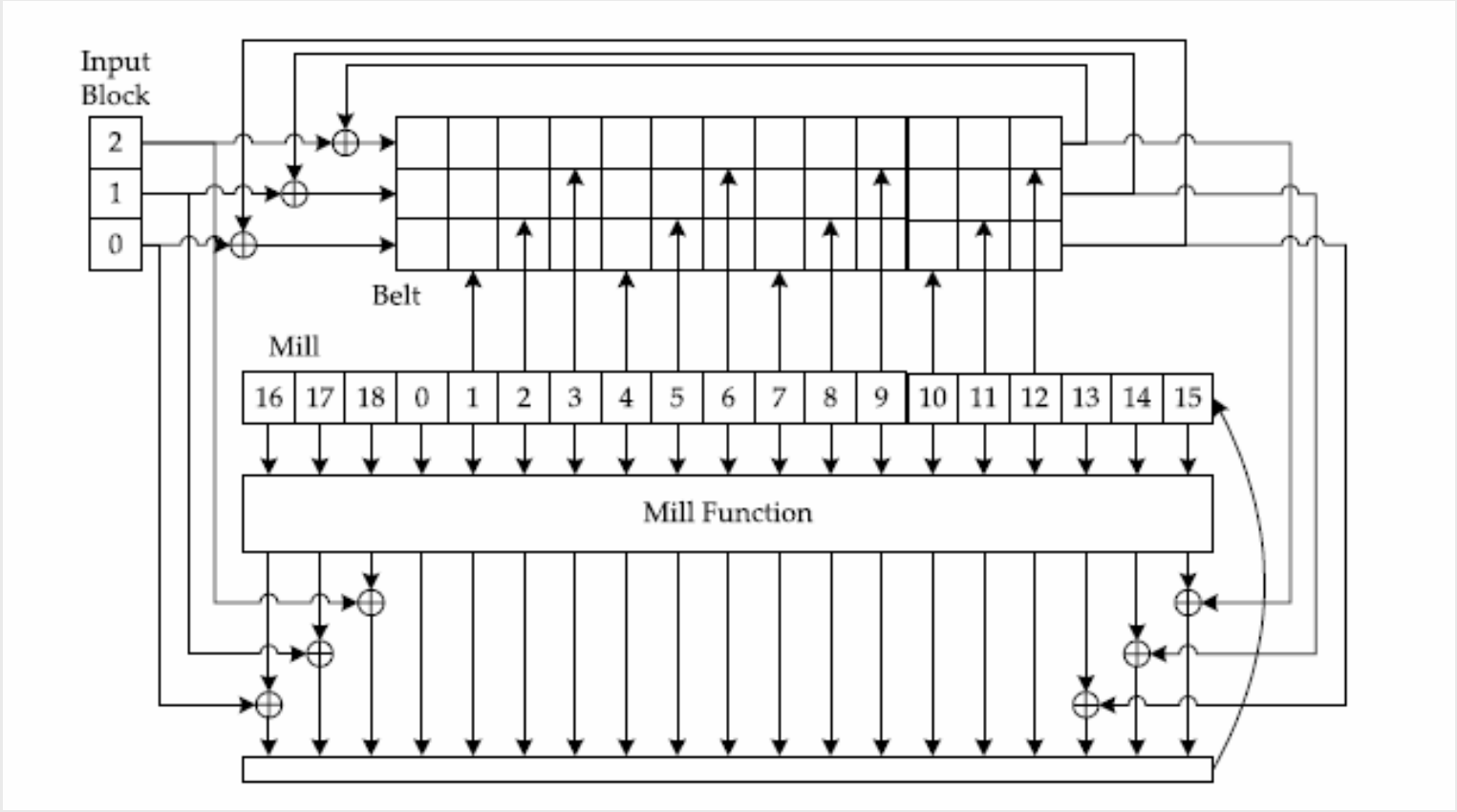
Outline

- ✓ Introduction
- ✓ Trail backtracking
- ✓ Panama
- ❑ **RadioGatún**
- ❑ **Grindahl**
- ❑ **Conclusion**

RadioGatún

- Inspired from Panama
- Two parts:
  - Belt is like the Buffer of Panama
  - Mill is like the state, a similar rho function
- 3 input words instead of 8
- Feedback from the Mill to the Belt

RadioGatún


Input
Block
2
1
0
Belt
Mill
16 17 18 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Mill Function

Design principles

- Trail backtracking cost is influenced by the lower value of the input (from 8 to 3)
  - Less degree of freedom
  - Increase w_r-l, thus the backtracking cost
  - Increase the depth of the backtracking

RadioGatún

- Use of trail backtracking
  - Searching for trails, and evaluate costs
- Best trail found RadioGatún[1]
  - backtracking cost 46
- Different alternatives have been explored
  - Number of feedbacks from mill to belt
  - Size of the belt
  - Size of the mill

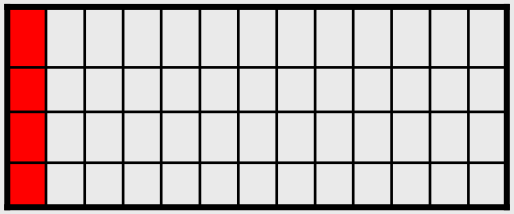
Outline

- ✓ Introduction
- ✓ Trail backtracking
- ✓ Panama
- ✓ RadioGatún
- ❑ **Grindahl**
- ❑ **Conclusion**

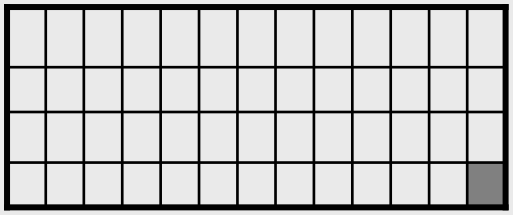
Grindahl

- New hash function proposed at FSE2007 by Knudsen, Rechberger and Thomsen
- Concatenate-permute-truncate
- The permutation is based on Rijndael building blocks
- Two variants 256 and 512
  - Difference: size internal state, shifts and digest

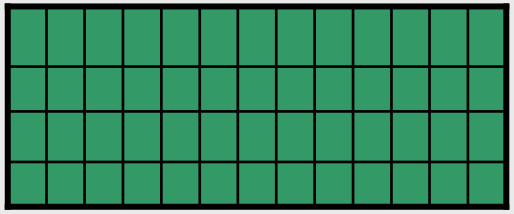
Grindahl iteration



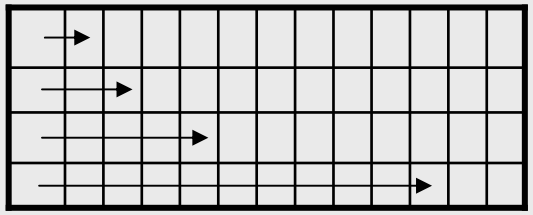
Concatenate



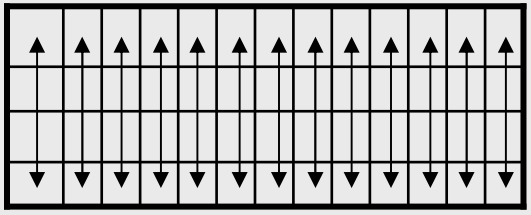
AddConstant



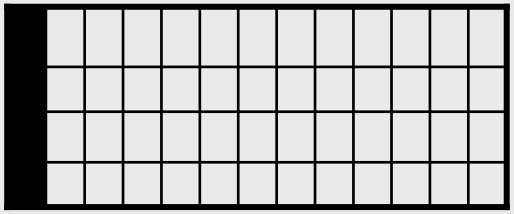
SubBytes



ShiftRows



MixColumns



Truncate

Use of truncated differences

- Presented by Peyrin ASIACRYPT2007
- Use a bit to represent if there is a difference or not in a byte of the state
- SubBytes and AddConstant can be not considered
- MixColumns is the fundamental part

Differential of MixColumns

- #Input Difference + #output difference ≥ 5
- Approx Probabilities ($\log_2$):

Output difference

Input difference \ Output difference	0	1	2	3	4
0	0	-	-	-	-
1	-	-	-	-	0
2	-	-	-	-8	0
3	-	-	-16	-8	0
4	-	-24	-16	-8	0

Spreading of Differences in 3 Rounds

Differences before the application of MixColumns

	0	1	2	3	4	5	6	7	8	9	10	11	12
0		X											
1			X										
2					X								
3											X		

	0	1	2	3	4	5	6	7	8	9	10	11	12
0			X	X		X						X	
1				X	X		X						X
2		X				X	X		X				
3		X						X				X	X

	0	1	2	3	4	5	6	7	8	9	10	11	12
0	X		X	X	X	X	X	X	X	X			X
1	X	X		X	X	X	X	X	X	X	X		
2			X	X		X	X	X	X	X	X	X	X
3	X	X	X	X	X	X			X	X		X	X

The All-difference state

- Grindahl has an all-difference state that is a kind of attractor
- It is very easy to “fall in” and difficult to exit
- But how difficult?
- Is it possible to avoid it?

The Attack from Peyrin

- Peyrin shows two trails from all-difference to zero-difference, one of 4 rounds with $p=2^{-312}$ and 8 rounds with $p=2^{-440}$
- But using the degree of freedom it turns out that the latter is better and a collision can be found in 2^{112}
- Degree of freedom is given by the control bytes

Open points

- Not demonstrated if it is possible to avoid the all-difference state and build a collision trail with reasonable cost
- In the case of 8 rounds, there are degree of freedom not used
- The number of differential trails grows rapidly
  - not easy to search

Conclusions

- Trail Backtracking can be used for bit-sliced algorithms but even with byte/word oriented algorithms using truncated differences
- Conditions can be imposed and satisfied algebraically or trying pairs

references

- All papers are on the net:
- http://radiogatun.noekeon.org/
- http://radiogatun.noekeon.org/panama/index.html
- http://www.cosic.esat.kuleuven.be/publications/article-81.ps
- http://www.ramkilde.com/grindahl
- http://tpeyrin.no-ip.org/