On the propagation of affine relations through an Sbox

Christina Boura and **Anne Canteaut**

SECRET Project-Team, INRIA Paris-Rocquencourt
Gemalto, France

October 8, 2012

Outline

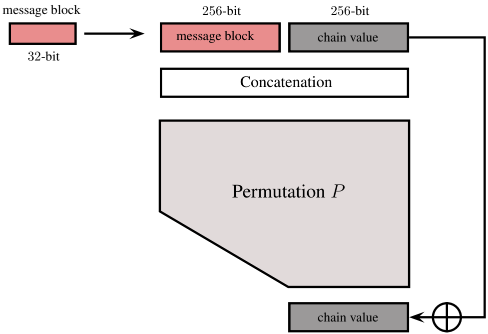
1. Description of Hamsi-256
2. Thomas Fuhr's attack
3. Improvement of the attack
4. (v, w)-linear functions

Outline

Hamsi Hash Function

Designed by Özgül Küçük in 2008 for the SHA-3 competition.
Selected by NIST for the 2nd round (14 candidates).

Compression function of Hamsi-256



Concatenation

State : 4×4 matrix of 32-bit words

s_0	s_1	s_2	s_3
s_4	s_5	s_6	s_7
s_8	s_9	s_{10}	s_{11}
s_{12}	s_{13}	s_{14}	s_{15}

m_0	m_1	c_0	c_1
c_2	c_3	m_2	m_3
m_4	m_5	c_4	c_5
c_6	c_7	m_6	m_7

Permutation P

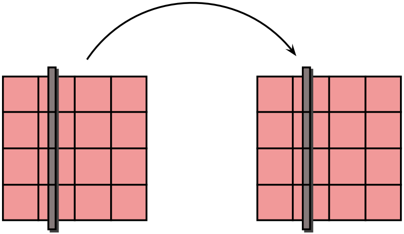
$\mathbf{3}$ rounds of a 512-bit round permutation R

- **XOR** of constants
- **Substitution** by 4×4-bit Sboxes
- **Diffusion** by a linear transformation L

Substitution

128 parallel applications of a 4×4 Sbox S
S is a `Serpent` Sbox

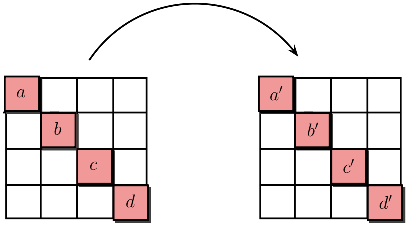
$$S = \{8, 6, 7, 9, 3, 12, 10, 15, 13, 1, 14, 4, 0, 11, 5, 2\}$$



Diffusion

4 parallel applications of a linear function L

$$\begin{aligned} L : \mathbf{F}_2^{128} &\to \mathbf{F}_2^{128} \\ L(a,b,c,d) &= (a',b',c',d'), \end{aligned}$$



- Each bit of a' and c' is the XOR of $\mathbf{7}$ bits of a, b, c, d.
- Each bit of b' and d' is the XOR of $\mathbf{3}$ bits of a, b, c, d.

Outline

First second preimage attack against Hamsi-256 by **Thomas Fuhr** (Asiacrypt 2010)

Idea:

Find some output bits which can be expressed as an **affine function** of some inputs bits when the other input bits are fixed to any arbitrary value.

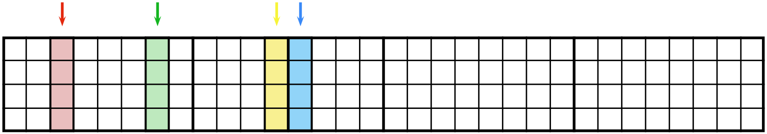
- Build the linear system.
- Solve the system (find preimages for the compression function).
- Use a meet-in-the-middle algorithm to extend these pseudo-preimages to second preimages for the hash function.

Description of the attack in [Fuhr10]

Important property of S

$$S(1, x, 0, \bar{x}) = (1, 0, 0, x) \quad \forall x \in \mathbf{F}_2$$

- Fix N_{var} positions $i = 1, \ldots, N_{var}$ (here $N_{var} = 4$).
- Choose a message block m such that $s_0^i = 1$ (resp. $s_1^i = 1$) and $s_8^i = 0$ (resp. $s_8^i = 1$).
- Consider N_{var} variables x_i, $i = 1, \ldots, N_{var}$.



Description of the attack in [Fuhr10]

Important property of S

$$S(1, x, 0, \bar{x}) = (1, 0, 0, x) \quad \forall x \in \mathbf{F}_2$$

- Fix N_{var} positions $i = 1, \ldots, N_{var}$ (here $N_{var} = 4$).
- Choose a message block m such that $s_0^i = 1$ (resp. $s_1^i = 1$) and $s_8^i = 0$ (resp. $s_8^i = 1$).
- Consider N_{var} variables x_i, $i = 1, \ldots, N_{var}$.

		1				1					1	1																			
		0				0					0	0																			

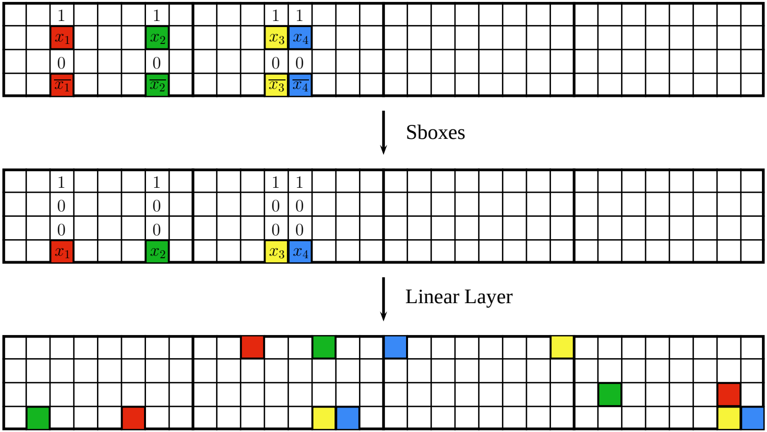
Description of the attack in [Fuhr10]

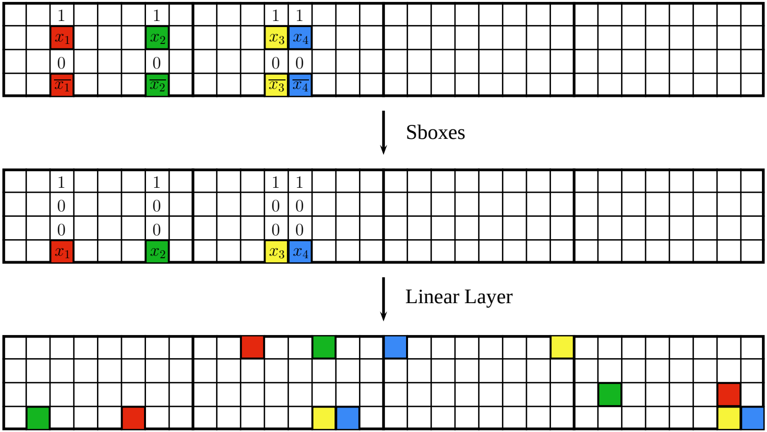
Important property of S

$$S(1, x, 0, \bar{x}) = (1, 0, 0, x) \quad \forall x \in \mathbf{F}_2$$

- Fix N_{var} positions $i = 1, \ldots, N_{var}$ (here $N_{var} = 4$).
- Choose a message block m such that $s_0^i = 1$ (resp. $s_1^i = 1$) and $s_8^i = 0$ (resp. $s_8^i = 1$).
- Consider N_{var} variables x_i, $i = 1, \ldots, N_{var}$.

		1				1					1	1																			
		x_1				x_2					x_3	x_4																			
		0				0					0	0																			
		$\overline{x_1}$				$\overline{x_2}$					$\overline{x_3}$	$\overline{x_4}$																			

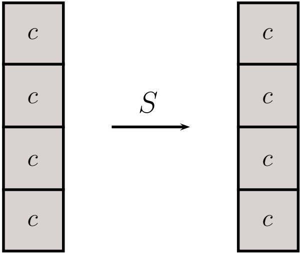

1 1 1 1
x_1 x_2 x_3 x_4
0 0 0 0
$\overline{x_1}$ $\overline{x_2}$ $\overline{x_3}$ $\overline{x_4}$
Sboxes
1 1 1 1
0 0 0 0
0 0 0 0
x_1 x_2 x_3 x_4
Linear Layer



After the first round, the state is linear in the input variables, for any choice of the other constants.

4 different situations

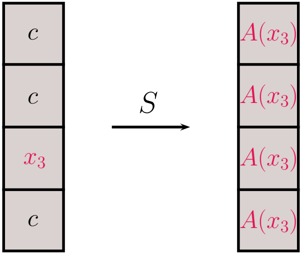
All the input bits are constant.



All output bits are constant.

4 different situations

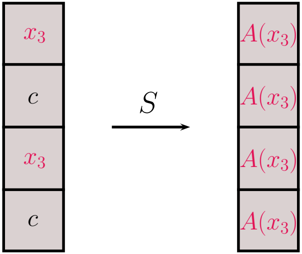
At most one input bit depends on one variable (or a affine combination of variables).



All output bits are an affine combination of this variable.

4 different situations

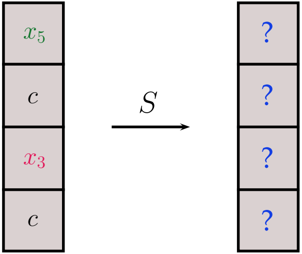
At least two input bits depend on the same variable (or the same affine combination of variables).



All output bits are an affine combination of this variable.

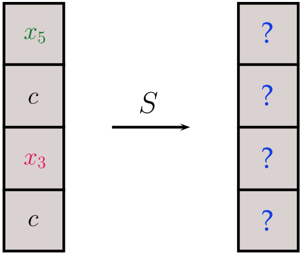
4 different situations

At least two input bits depend on at least two different variables.



4 different situations

At least two input bits depend on at least two different variables.



Are all output bits non-linear?

Two properties of S noticed by Thomas Fuhr

- y_0 is of degree at most 1 if x_0x_2 is of degree at most 1.
- y_3 is of degree at most 1 if x_1x_3 and $x_0x_1x_2$ are of degree at most 1.

$$
\begin{aligned}
y_0 &= x_0x_2 + x_1 + x_2 + x_3 \\
y_1 &= x_0x_1x_2 + x_0x_1x_3 + x_0x_2x_3 + x_1x_2 + x_0x_3 + x_2x_3 + x_0 + x_1 + x_2 \\
y_2 &= x_0x_1x_3 + x_0x_2x_3 + x_1x_2 + x_1x_3 + x_2x_3 + x_0 + x_1 + x_3 \\
y_3 &= x_0x_1x_2 + x_1x_3 + x_0 + x_1 + x_2 + 1.
\end{aligned}
$$

Two properties of S noticed by Thomas Fuhr

- y_0 is of degree at most 1 if x_0x_2 is of degree at most 1.
- y_3 is of degree at most 1 if x_1x_3 and $x_0x_1x_2$ are of degree at most 1.

$$
\begin{aligned}
y_0 &= x_0x_2 + x_1 + x_2 + x_3 \\
y_1 &= x_0x_1x_2 + x_0x_1x_3 + x_0x_2x_3 + x_1x_2 + x_0x_3 + x_2x_3 + x_0 + x_1 + x_2 \\
y_2 &= x_0x_1x_3 + x_0x_2x_3 + x_1x_2 + x_1x_3 + x_2x_3 + x_0 + x_1 + x_3 \\
y_3 &= x_0x_1x_2 + x_1x_3 + x_0 + x_1 + x_2 + 1.
\end{aligned}
$$

Two properties of S noticed by Thomas Fuhr

- y_0 is of degree at most 1 if x_0x_2 is of degree at most 1.
- y_3 is of degree at most 1 if x_1x_3 and $x_0x_1x_2$ are of degree at most 1.

$$
\begin{aligned}
y_0 &= x_0x_2 + x_1 + x_2 + x_3 \\
y_1 &= x_0x_1x_2 + x_0x_1x_3 + x_0x_2x_3 + x_1x_2 + x_0x_3 + x_2x_3 + x_0 + x_1 + x_2 \\
y_2 &= x_0x_1x_3 + x_0x_2x_3 + x_1x_2 + x_1x_3 + x_2x_3 + x_0 + x_1 + x_3 \\
y_3 &= x_0x_1x_2 + x_1x_3 + x_0 + x_1 + x_2 + 1.
\end{aligned}
$$

Results (PhD of T. Fuhr)

$\mathbf{16}$ affine equations on $\mathbf{8}$ variables.
$\mathbf{11}$ affine equations on $\mathbf{9}$ variables.
$\mathbf{9}$ affine equations on $\mathbf{10}$ variables.

Outline

An equivalent notation

y_0 is of degree at most 1 if $x_0 x_2$ is of degree at most 1.

$$\Downarrow$$

y_0 is of degree at most 1 if $x \in V^\perp \subset \mathbf{F}_2^4$ with $V = \langle 1 \rangle$, $V = \langle 4 \rangle$ or $V = \langle 5 \rangle$, or to **any coset** of these hyperplanes.

An equivalent notation

y_0 is of degree at most 1 if x_0x_2 is of degree at most 1.

$$\Downarrow$$

y_0 is of degree at most 1 if $x \in V^\perp \subset \mathbf{F}_2^4$ with $V = \langle 1 \rangle$, $V = \langle 4 \rangle$ or $V = \langle 5 \rangle$, or to **any coset** of these hyperplanes.

y_3 is of degree at most 1 if x_1x_3 and $x_0x_1x_2$ are of degree at most 1.

$$\Downarrow$$

y_3 is of degree at most 1 if x belongs to **any coset** of $V^\perp \subset \mathbf{F}_2^4$ with $V = \langle 1, 2 \rangle$, $V = \langle 2, 4 \rangle$ or $V = \langle 2, 5 \rangle$.

We have identified many such relations for S with $\dim V = 2$

$\langle 1,2\rangle$	$\{1,6,7,8,9,e,f\}$
$\langle 1,4\rangle$	$\{1,e,f\}$
$\langle 1,6\rangle$	$\{1,4,5,a,b,e,f\}$
$\langle 1,8\rangle$	$\{1,e,f\}$
$\langle 1,a\rangle$	$\{1,2,3,c,d,e,f\}$
$\langle 1,c\rangle$	$\{1,e,f\}$
$\langle 1,e\rangle$	$\{1,e,f\}$
$\langle 2,4\rangle$	$\{1,8,9\}$
$\langle 2,5\rangle$	$\{1,8,9\}$
$\langle 2,8\rangle$	$\{e\}$
$\langle 2,9\rangle$	$\{e\}$
$\langle 2,d\rangle$	$\{f\}$
$\langle 3,4\rangle$	$\{1,6,7\}$
$\vdots$	$\vdots$

We have identified many such relations for S with $\dim V = 2$

$\langle 1, 2\rangle$	$\{1, 6, 7, 8, 9, e, f\}$
$\langle 1, 4\rangle$	$\{1, e, f\}$
$\langle 1, 6\rangle$	$\{1, 4, 5, a, b, e, f\}$
$\langle 1, 8\rangle$	$\{1, e, f\}$
$\langle 1, a\rangle$	$\{1, 2, 3, c, d, e, f\}$
$\langle 1, c\rangle$	$\{1, e, f\}$
$\langle 1, e\rangle$	$\{1, e, f\}$
$\langle 2, 4\rangle$	$\{1, 8, 9\}$
$\langle 2, 5\rangle$	$\{1, 8, 9\}$
$\langle 2, 8\rangle$	$\{e\}$
$\langle 2, 9\rangle$	$\{e\}$
$\langle 2, d\rangle$	$\{f\}$
$\langle 3, 4\rangle$	$\{1, 6, 7\}$
$\vdots$	$\vdots$

35 properties in total

Improvement of the attack of [Fuhr10]

1. Use these properties to search for affine propagation of the input variables through the 2nd and the 3rd round.

Improvement of the attack of [Fuhr10]

1. Use these properties to search for affine propagation of the input variables through the 2nd and the 3rd round.

2. Use the following relations of S to go through the 1st round.

$$\begin{array}{rcll} S(1, x, 0, \bar{x}) & = & (1, 0, 0, x) & \forall x \in \mathbf{F}_2 \\ S(1, x, 0, x) & = & (0, x, 1, 0) & \forall x \in \mathbf{F}_2 \end{array}$$

Improvement of the attack of [Fuhr10]

1. Use these properties to search for affine propagation of the input variables through the 2nd and the 3rd round.

2. Use the following relations of S to go through the 1st round.

$$\begin{array}{rcll} S(1,x,0,\bar{x}) & = & (1,0,0,x) & \forall x \in \mathbf{F}_2 \\ S(1,x,0,x) & = & (0,x,1,0) & \forall x \in \mathbf{F}_2 \end{array}$$

3. Track backwards the propagation of the output bits to fix the input variables.

Improvement of the attack of [Fuhr10]

1. Use these properties to search for affine propagation of the input variables through the 2nd and the 3rd round.

2. Use the following relations of S to go through the 1st round.

$$\begin{aligned} S(1,x,0,\bar{x}) &= (1,0,0,x) \quad \forall x \in \mathbf{F}_2 \\ S(1,x,0,x) &= (0,x,1,0) \quad \forall x \in \mathbf{F}_2 \end{aligned}$$

3. Track backwards the propagation of the output bits to fix the input variables.

Results

$\mathbf{13}$ affine equations on $\mathbf{9}$ variables.
$\mathbf{11}$ affine equations on $\mathbf{10}$ variables.

Outline

The notion of (v, w)-linearity

Definition

Let S be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$. Then, S is said to be (v, w)-linear if there exist two subspaces $V \subset \mathbf{F}_2^n$ and $W \subset \mathbf{F}_2^m$ with $\dim V = v$ and $\dim W = w$ such that, for all $\lambda \in W$, S_λ has degree at most 1 on all cosets of V, where S_λ is the Boolean function $x \mapsto \lambda \cdot S(x)$.

We used that the Sbox of Hamsi is $(3, 2)$-linear for **some** (V, W), and that it is (2,2)-linear for **many** (V, W).

Link with the Maiorana-McFarland construction

A function S from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$ is (v, w)-linear if the function S_W that corresponds to all the components S_λ , $\lambda \in W$ can be written as

$$S_W(u, v) = M(u)v + G(u),$$

where $U \times V = \mathbf{F}_2^n$, G is a function from U in F_2^w and $M(u)$ is a $w \times v$ binary matrix.

Link with the Maiorana-McFarland construction

A function S from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$ is (v, w)-linear if the function S_W that corresponds to all the components S_λ , $\lambda \in W$ can be written as

$$S_W(u, v) = M(u)v + G(u),$$

where $U \times V = \mathbf{F}_2^n$, G is a function from U in F_2^w and $M(u)$ is a $w \times v$ binary matrix.

Generalisation of the Maiorana-McFarland construction

The degree of each S_λ is at most $\dim U + 1 = n + 1 - v$.

Boolean functions that are equivalent to the Maiorana-McFarland construction can be characterized by their **second-order derivatives**. (Similar for vectorial functions)

Proposition

Let S be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$. Then, S is (v, w)-linear if and only if there exists a subset of w independent components of S, $S_W = (S_{i_1}, \ldots, S_{i_w})$, and a linear subspace V of dimension v such that all second-order derivatives of S_W, $D_\alpha D_\beta S_W$ with $\alpha, \beta \in V$ vanish.

Boolean functions that are equivalent to the Maiorana-McFarland construction can be characterized by their **second-order derivatives**. (Similar for vectorial functions)

Proposition

Let S be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$. Then, S is (v, w)-linear if and only if there exists a subset of w independent components of S, $S_W = (S_{i_1}, \ldots, S_{i_w})$, and a linear subspace V of dimension v such that all second-order derivatives of S_W, $D_\alpha D_\beta S_W$ with $\alpha, \beta \in V$ vanish.

Easy algorithm for finding all (v, w)-linear subspaces.

Link with non-linearity

Proposition

Let S be a function from $\mathbf{F}_2^n$ into $\mathbf{F}_2^m$. If S is (v, w)-linear, then S has w weakly v-normal coordinates. In particular, $\mathcal{L}(S) \geq 2^v$.

$(n-1,1)$-linear functions

Proposition

Let f be a Boolean function of n variables. Then, f is $(n-1,1)$-linear if and only if $\deg f \leq 2$ and $\mathcal{L}(f) \geq 2^{n-1}$. Moreover, if $\deg(f) = 2$ and $\mathcal{L}(f) \geq 2^{n-1}$, there exist exactly 3 distinct hyperplanes H such that f has degree at most 1 on both H and $\bar{H}$.

$(n-1,1)$-linear functions

Proposition

Let f be a Boolean function of n variables. Then, f is $(n-1,1)$-linear if and only if $\deg f \leq 2$ and $\mathcal{L}(f) \geq 2^{n-1}$. Moreover, if $\deg(f) = 2$ and $\mathcal{L}(f) \geq 2^{n-1}$, there exist exactly 3 distinct hyperplanes H such that f has degree at most 1 on both H and $\bar{H}$.

Remark : The number of subspaces for which S is $(n-1,1)$-linear is determined by the number of the **quadratic components** of S.

Classification of 4×4 Sboxes

A 4×4 Sbox S with **optimal linearity** ($\mathcal{L}(S) = 8$) has $0, 1, 3$, or 7 **quadratic components**.

- Sboxes with 15 **quadratic components** have one linear component.
- Sboxes with 7 **quadratic components** are not optimal against differential cryptanalysis.

Classification of 4×4 Sboxes

A 4×4 Sbox S with **optimal linearity** ($\mathcal{L}(S) = 8$) has $0, 1, 3$, or 7 **quadratic components**.

- Sboxes with 15 **quadratic components** have one linear component.
- Sboxes with 7 **quadratic components** are not optimal against differential cryptanalysis.

Merci pour votre attention !