Message modification, neutral bits and boomerangs

From which round should we start counting in SHA ?

Antoine Joux
DGA
and
University of Versailles St-Quentin-en-Yvelines
France

Joint work with Thomas Peyrin

Differential cryptanalysis of SHA

- Started in 1998 with SHA-0
- Many improvements starting from 2004:
  - Neutral bits technique
  - Multi-block collisions
  - Message modification techniques
  - Non linear differential paths
- In this talk, we focus on:
  - Neutral bits
  - Message modification
  - **Boomerang attack**

Overview of the basic attack

Notations

Notation	Definition
$\mathbb{F}_q$	Finite field with q elements.
$\langle X, Y, \ldots, Z \rangle$	Concatenation of 32-bits words.
$+$	Addition on 32-bits words modulo 2^{32}.
$\oplus$	*Exclusive or* on bits or 32-bits words.
$\vee$	*Inclusive or* on bits or 32-bits words.
$\wedge$	Logical *and* on bits or 32-bits words.
$ROL_\ell(X)$	Rotation by ℓ bits of a 32-bits word.
X_i	The ith bit of 32-bits word X, from the least significant 0 to the most significant 31.

Description of SHA

$\texttt{SHA}$ compression function

Initialization of $\left\langle A^{(0)}, B^{(0)}, C^{(0)}, D^{(0)}, E^{(0)} \right\rangle$

for $i = 0$ to 79

$$A^{(i+1)} =$$
$$ADD\left(W^{(i)}, ROL_5\left(A^{(i)}\right), f^{(i)}\left(B^{(i)}, C^{(i)}, D^{(i)}\right), E^{(i)}, K^{(i)}\right)$$

$$B^{(i+1)} = A^{(i)}$$

$$C^{(i+1)} = ROL_{30}\left(B^{(i)}\right)$$

$$D^{(i+1)} = C^{(i)}$$

$$E^{(i+1)} = D^{(i)}$$

Output

$$\left\langle A^{(0)} + A^{(80)}, B^{(0)} + B^{(80)}, C^{(0)} + C^{(80)}, D^{(0)} + D^{(80)}, E^{(0)} + E^{(80)} \right\rangle$$

Functions $f^{(i)}(X,Y,Z)$, and Constants $K^{(i)}$

Round i	Function $f^{(i)}$		Constant $K^{(i)}$
	Name	Definition	
0 –19	IF	$(X \wedge Y) \vee (\neg X \wedge Z)$	`0x5A827999`
20–39	XOR	$(X \oplus Y \oplus Z)$	`0x6ED9EBA1`
40–59	MAJ	$(X \wedge Y) \vee (X \wedge Z) \vee (Y \wedge Z)$	`0x8F1BBCDC`
60–79	XOR	$(X \oplus Y \oplus Z)$	`0xCA62C1D6`

Expansion of SHA-0

- Input: $\left\langle W^{(0)}, \ldots, W^{(15)} \right\rangle$

$$W^{(i)} = W^{(i-3)} \oplus W^{(i-8)} \oplus W^{(i-14)} \oplus W^{(i-16)} \quad . \tag{1}$$

- Output: $\left\langle W^{(0)}, \ldots, W^{(79)} \right\rangle$

Difference with SHA-1

- Slight difference in the expansion:

$$W^{(i)} = ROL_1\left(W^{(i-3)} \oplus W^{(i-8)} \oplus W^{(i-14)} \oplus W^{(i-16)}\right) \quad . \qquad (2)$$

- $E_0 = (e_0)^{32}$ non-interleaved expansion of SHA-0.

- E_1 interleaved expansion of SHA-1.

Linearized version of SHA

- Replace ADD by XOR.
- Replace f_i by XOR.
- Then, collision can be found with linear algebra

Constructing Differential Collisions

Construction of the Differential Mask

- For `SHA-0`:
  - Find a **disturbance**-vector $\left(m_0^{(0)}, \ldots, m_0^{(79)}\right)$.
  - Apply it on bits 1, in order to obtain perturbative mask $M_0 = \left\langle M_0^{(-5)}, \ldots, M_0^{(79)} \right\rangle$ defined by:

$$\begin{aligned}
\forall i,\ -5 \le i \le -1,\ M_0^{(i)} &= 0 \\
\forall i,\ 0 \le i \le 79,\ M_{0,k}^{(i)} &= 0 \text{ if } k \neq 1; \\
\forall i,\ 0 \le i \le 79,\ M_{0,1}^{(i)} &= m_0^{(i)}\ .
\end{aligned}$$

- For `SHA-1`:
  - Directly find the perturbative mask M_0
  - Use a low weight vector of the expansion E_1
  - Align many bits (not all) on bit 1

Corrective Masks

- From M_0 derive: $M_1, \ldots, M_5$:

$$
\begin{aligned}
\forall i,\ -4 \le i \le 79,\ M_1^{(i)} &= ROL_5\left(M_0^{(i-1)}\right) \;; && (3)\\
\forall i,\ -3 \le i \le 79,\ M_2^{(i)} &= M_0^{(i-2)} \;; && (4)\\
\forall i,\ -2 \le i \le 79,\ M_3^{(i)} &= ROL_{30}\left(M_0^{(i-3)}\right) \;; && (5)\\
\forall i,\ -1 \le i \le 79,\ M_4^{(i)} &= ROL_{30}\left(M_0^{(i-4)}\right) \;; && (6)\\
\forall i,\ 0 \le i \le 79,\ M_5^{(i)} &= ROL_{30}\left(M_0^{(i-5)}\right) \;; && (7)
\end{aligned}
$$

Constraints (basic attack on SHA-0)

- m_0 must be ended by 5 zeroes.
- Differential mask M defined by

$$\forall i,\ 0 \leq i \leq 79,\ M^{(i)} = M_0^{(i)} \oplus M_1^{(i)} \oplus M_2^{(i)} \oplus M_3^{(i)} \oplus M_4^{(i)} \oplus M_5^{(i)} \quad , \tag{8}$$

must be an output of E_0.
Ensured by:

$$M_0^{(i)} = M_0^{(i-3)} \oplus M_0^{(i-8)} \oplus M_0^{(i-14)} \oplus M_0^{(i-16)},\ \forall i,\ 11 \leq i < 80 \quad . \tag{9}$$

Consequence for linearized model

- There exists 64 error vectors m_0 satisfying the constraints.
- There exists 64 masks M: we deduce μ such that $M = E_0(\mu)$.
- For all input $W = \left\langle W^{(0)} \dots W^{(15)} \right\rangle$, $W' = W \oplus \mu$ has same output by the linearized compression function.

- With non-negligible probability, also give attack on real `SHA`

Application to SHA-0

- A few patterns. Best one m_0 with probability $1/2^{61}$:

```
00000 0010001000000000101111
      0110001110000000010100
      0100010010010011101 1
      0011000001111110000000
```

- Complexity goes down to 2^{56} with neutral bits of Biham and Chen

Recent improvements

- Multiblock techniques
- Non linear characteristics
  - Non linearity for a few rounds in the first `SHA-0` collision
  - Non linearity during about 16 rounds in Wang's et al `SHA-1` attack
- Remove a lot of constraints (and improve attacks)

Evaluating the cost of the attack

- Three important phases:
  - Early rounds, where control is possible
  - Late rounds, where behavior is probabilistic
  - Final rounds, where misbehavior can be partially ignored
- Roughly the complexity arises from the probability of success in the late rounds (the final rounds being excepted)
- Evaluated by computing the probability of success of each local collision

Evaluating the cost of a single local collision

- Disturbance insertion: No carry wanted (pr 1/2)
- A correction: Need opposite sign (pr 1)
- B correction: Disturbance propagates with the right sign (pr 1/2)
- C correction: Disturbance propagates (Bit 31, pr 1 or 1/2)
  - Other bits: with the right sign (pr 1/2)
  - Possible dependence on D with `MAJ`
- D correction: Disturbance propagates (Bit 31, pr 1 or 1/2)
  - Other bits: with the right sign (pr 1/2)
- E correction: Need opposite sign (pr 1)

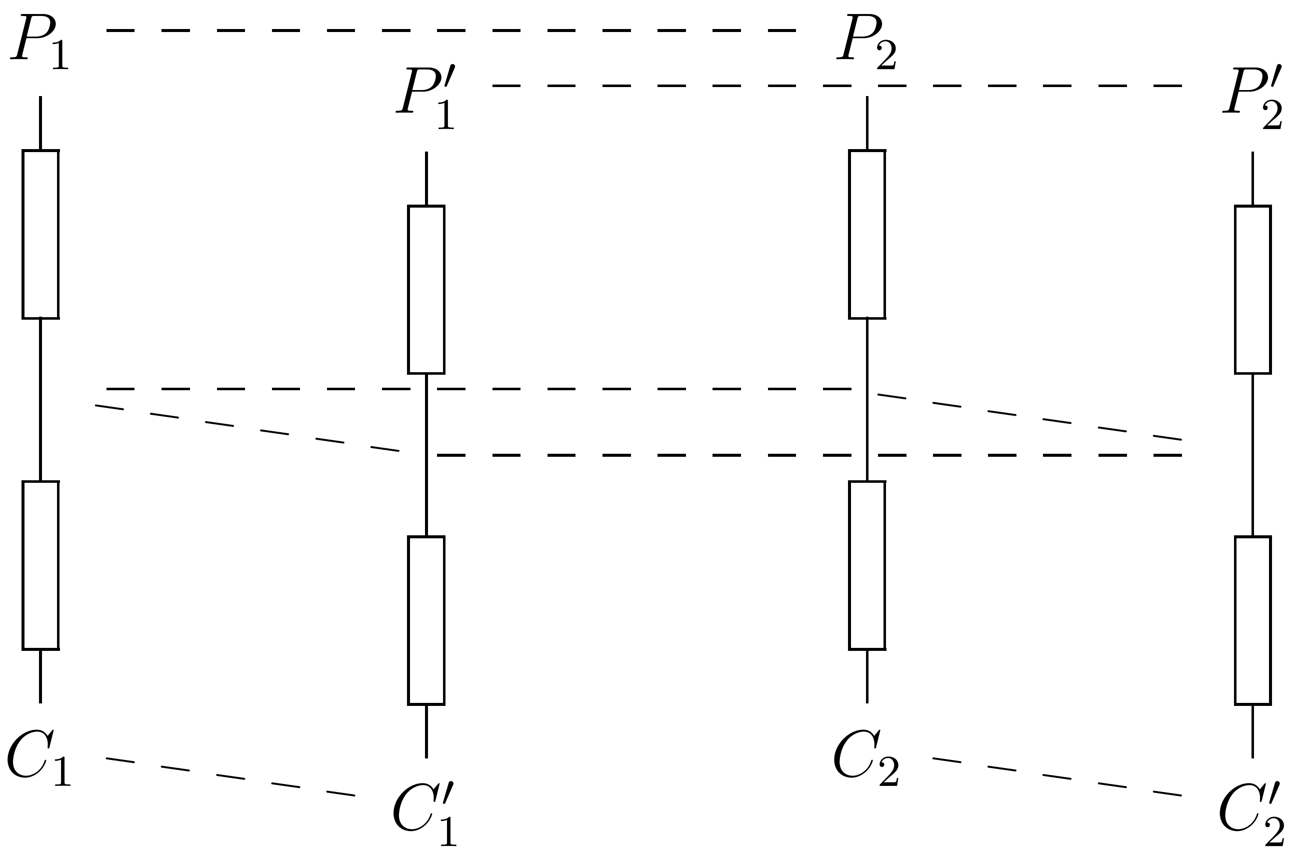
Where do the late rounds start

- In the basic attack, round 16 (or 18 with some care)
- With neutral bits of Biham and Chen, round 23
  - Use the fact that some message "bits" changes do not affect conformance.
  - From one candidate message pair, generates many
- With message modifications of Wang et al., round 26
  - Use ad'hoc message changes to force conformance in early rounds
  - Much fewer pairs to explore, however each pair costs more
  - Wang et al. at first Hash Workshop announced cost $2^{63} + 2^{60}$.
  - Crypto'05 was round 23, cost $2 \cdot 2^{71}$ pairs, 2^{69} SHA computations

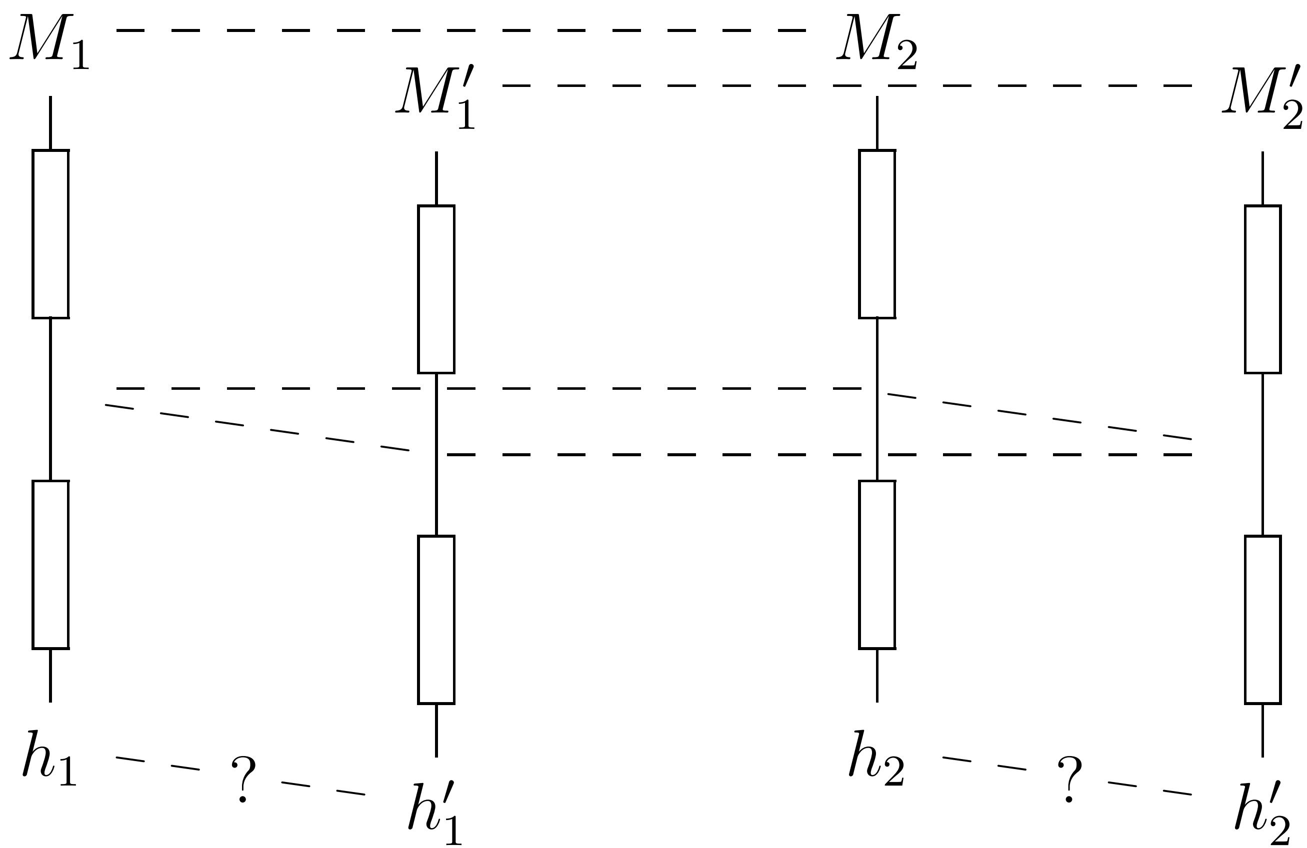
Where do the late rounds start

- Can we do better and improve the overall complexity ?
  - One track is to improve message modification. For example Gröbner approach.
  - The cost per message pair is potentially high
  - Another track is to improve neutral bits.
  - Our approach here: Use a variant of the **boomerang attack**

Boomerang picture for block ciphers



Boomerang picture for hash compression



Boomerang for hash compression

- Each M, M' pair is a partially conformant pair of the main differential
- Both pairs are related by a high probability auxillary differential
- The auxillary differential preserves conformance in the early rounds
- Beyond these rounds, the main differential holds (heuristic)

- Each auxillary differential thus doubles the number of conformant pairs
- Very similar to the neutral bit technique
- Longer range of the conformance preserving property

Construction of auxillary differentials

- A simple technique is to use collisions on pairs at some intermediate round
- First example of auxillary differential (experimentally seen in neutral bits)
  - Insert difference in round 6 at bit i
  - Correct in round 7 at bit $i+7$
  - Correct in round 11 at bit $i-2$
  - Rely on non-linearity for other correction
- With a well-chosen message pair, collision in round 12
- No more (auxillary) difference up to round 19
- Conformance to the main differential continues for a few additional rounds

An auxillary differential with pairwise collision up to round 26

- Found by simple search on bits $i-2$, i and $i+5$
- Contains 5 local collision patterns
- Collision in round 16, no more difference up to round 26

Bit i	0	4	6	8	10
Bit $i+5$	1	5	7	9	11
Bit i					
Bit $i-2$					
Bit $i-2$		8	10		14
Bit $i-2$	5	9	11	13	15

Associated constraints in initial pair

$M_i^{(0)} = a$ $A_i^{(1)} = a$	$M_i^{(4)} = b$ $A_i^{(5)} = b$	$M_i^{(6)} = c$ $A_i^{(7)} = c$	$M_i^{(8)} = d$ $A_i^{(9)} = d$	$M_i^{(10)} = e$ $A_i^{(11)} = e$
$M_{i+5}^{(1)} = \bar{a}$	$M_{i+5}^{(5)} = \bar{b}$	$M_{i+5}^{(7)} = \bar{c}$	$M_{i+5}^{(9)} = \bar{d}$	$M_{i+5}^{(11)} = \bar{e}$
$A_{i+2}^{(0)} = A_{i+2}^{(-1)}$	$A_{i+2}^{(4)} = A_{i+2}^{(3)}$	$A_{i+2}^{(6)} = A_{i+2}^{(5)}$	$A_{i+2}^{(8)} = A_{i+2}^{(7)}$	$A_{i+2}^{(10)} = A_{i+2}^{(9)}$
$A_{i-2}^{(2)} = 0$	$A_{i-2}^{(6)} = 0$	$A_{i-2}^{(8)} = 0$	$A_{i-2}^{(10)} = 0$	$A_{i-2}^{(12)} = 0$
$A_{i-2}^{(3)} = 1$	$A_{i-2}^{(7)} = 0$ $M_{i-2}^{(8)} = \bar{b}$	$A_{i-2}^{(9)} = 0$ $M_{i-2}^{(10)} = \bar{c}$	$A_{i-2}^{(11)} = 1$	$A_{i-2}^{(13)} = 0$ $M_{i-2}^{(14)} = \bar{e}$
$M_{i-2}^{(5)} = \bar{a}$	$M_{i-2}^{(9)} = \bar{b}$	$M_{i-2}^{(11)} = \bar{c}$	$M_{i-2}^{(13)} = \bar{d}$	$M_{i-2}^{(15)} = \bar{e}$

An auxillary differential with pairwise collision up to round 24

- Contains 4 local collision patterns
- Collision in round 14, no more difference up to round 24

Bit i	2	4	6	8
Bit $i+5$	3	5	7	9
Bit $i-2$	5	7	9	
Bit $i-2$	6	8		12
Bit $i-2$	7	9	11	13

Associated constraints in initial pair

$M_i^{(2)} = a$ $A_i^{(3)} = a$	$M_i^{(4)} = b$ $A_i^{(5)} = b$	$M_i^{(6)} = c$ $A_i^{(7)} = c$	$M_i^{(8)} = e$ $A_i^{(9)} = d$
$M_{i+5}^{(3)} = \bar{a}$	$M_{i+5}^{(5)} = \bar{b}$	$M_{i+5}^{(7)} = \bar{c}$	$M_{i+5}^{(9)} = \bar{d}$
$A_{i+2}^{(2)} = A_{i+2}^{(1)}$	$A_{i+2}^{(4)} = A_{i+2}^{(3)}$	$A_{i+2}^{(6)} = A_{i+2}^{(5)}$	$A_{i+2}^{(8)} = A_{i+2}^{(7)}$
$A_{i-2}^{(4)} = 1$	$A_{i-2}^{(6)} = 1$	$A_{i-2}^{(8)} = 1$	$A_{i-2}^{(10)} = 0$
$A_{i-2}^{(5)} = 0$	$A_{i-2}^{(7)} = 0$	$A_{i-2}^{(9)} = 1$	$A_{i-2}^{(11)} = 0$
$M_{i-2}^{(7)} = \bar{a}$	$M_{i-2}^{(9)} = \bar{b}$	$M_{i-2}^{(11)} = \bar{c}$	$M_{i-2}^{(13)} = \bar{d}$

Ongoing work

- Depending on bit position induces conformance up to round 28, 29 or more
- No high message modification cost
- Compatible with the neutral bit technique
- Technical difficulties:
  - Build a non-linear characteristic compatible with enough auxillary characteristics
    * Useful tool: see talk of De Cannière and Rechberger
  - Combine with simple message modification
- Expected result: `SHA-1` weaker today than `SHA-0` in 1998

A safety measure for collision builders

- Sooner or later a `SHA-1` collision will be produced
- This will be an important milestone for hash functions
- Yet it would be nice to minimize bad consequences
- Proposed safety measure:
  - Change the IV while keeping true `SHA-1`
  - For this, prepend a long enough, publicly announced, string
  - Two simple possibilities:
    * Prepend 1Gbyte of zeroes
    * Prepend 1Gbyte of binary expansion of π, e, $\sqrt{2}$, ...

Conclusion
Questions