The problem about SHA-1 collisions

Marc Stevens

CWI, Amsterdam
MI, Leiden

Overview

- Introduction
  - Hash functions
  - Compression function
  - History of MD5 attacks
  - History of SHA-1 attacks
- Preliminaries
- Differential analysis
- New results
- Future research

Introduction
Hash functions

- *Hash functions*: $\text{SHA-1} : \{0,1\}^* \to \{0,1\}^{160}$
  - MD5 [Rivest, '92], SHA-1 [NSA, '95]

- Collision resistance
  - Hard to find *collision*:

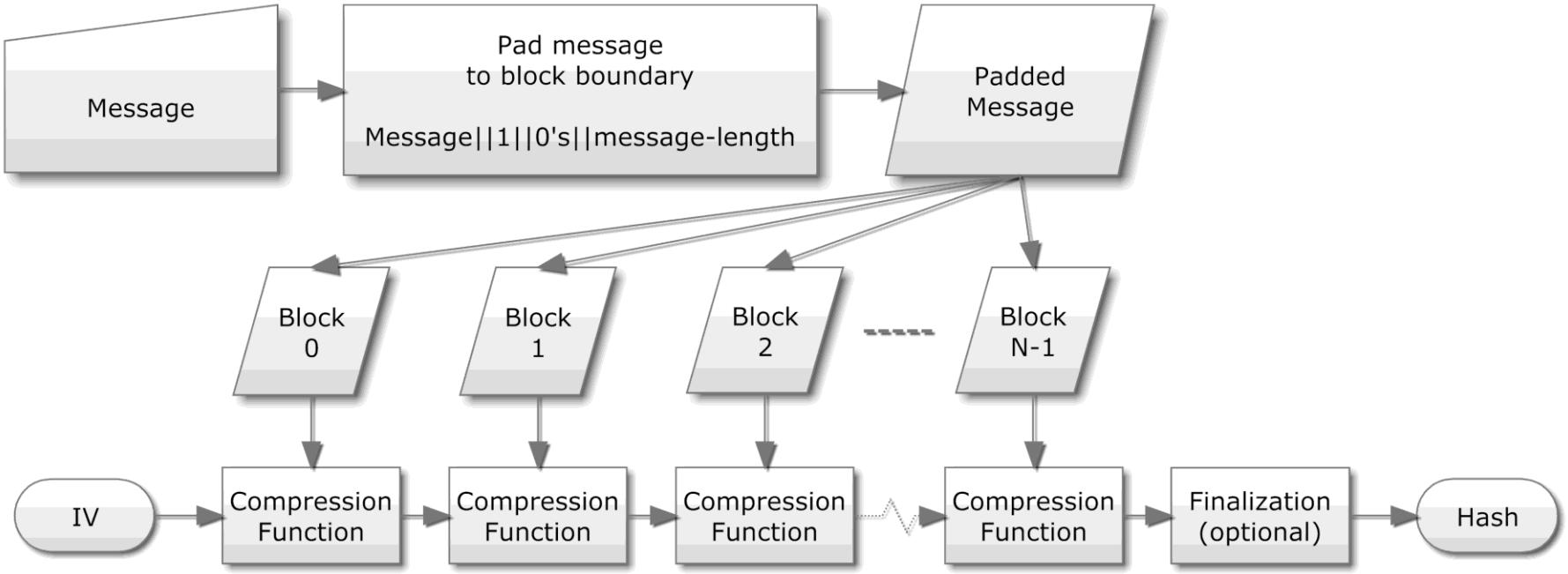
$$M \neq M' : \quad \text{SHA-1}(M) = \text{SHA-1}(M')$$

  - General attack: $\sqrt{2^{160} \cdot \pi/2} \approx 2^{80.33}$ calls to SHA-1

- Used in digital signatures
  - for digital certificates
  - E.g., to secure web: https://

Introduction
Compression function

- Iterative using compression function
- Message M split into pieces $M_0, \ldots, M_{N-1}$
- Internal state: IHV (initialized with IV)



Introduction
History of MD5 attacks

2004 first MD5 collision found [WY]

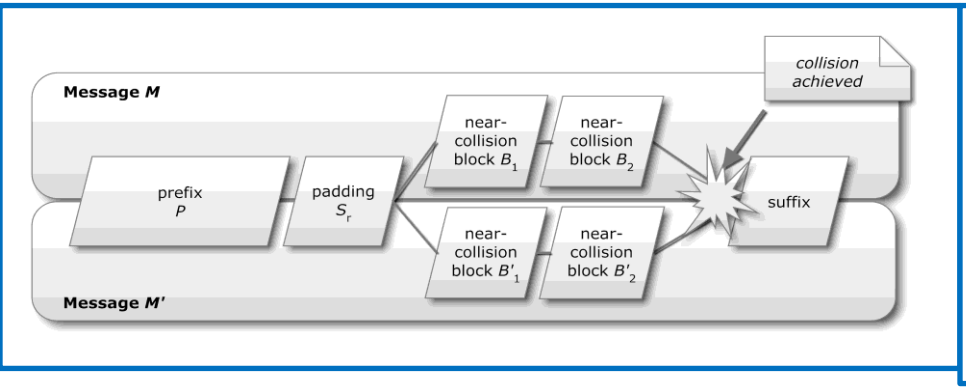
identical-prefix collision: 2^{40} calls

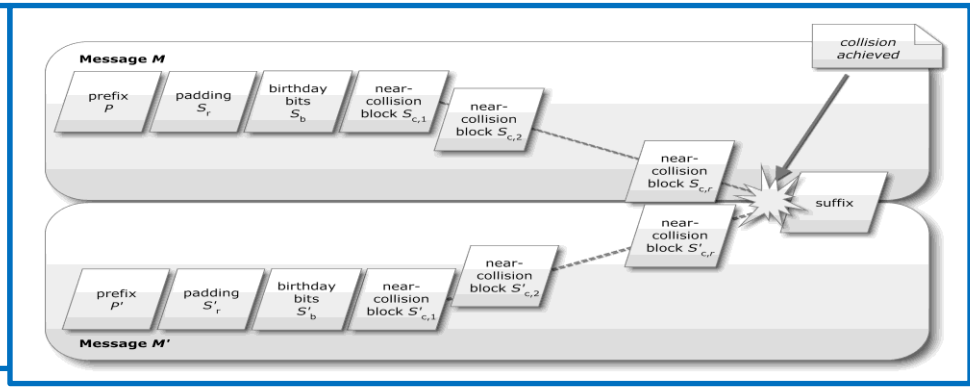
2006 first *chosen-prefix* collision: 2^{49} calls [SLdW]

2009 identical-prefix: 2^{16} calls [SSA+]

chosen-prefix: 2^{39} calls [SSA+]

realistic abuse scenario: rogue CA [SSA+]





Introduction
History of SHA-1 attacks

2005 first SHA-1 collision attack [WYY]
identical-prefix collision: 2^{69} calls
(two near-collision attacks: $2 \cdot 2^{68}$ calls)

2005 claim: 2^{63} calls [WYY] : unpublished

2007 claim: 2^{61} calls [MRR] : unpublished

2009 claim: 2^{52} calls [MHP] : withdrawn

2011 first attack is best attack: 2^{69} calls
WHY?

Overview

- Introduction
- Preliminaries
  - SHA-1 compression function
- Differential analysis
  - Differences
  - Differential path
  - Local collisions
  - Disturbance vector
  - Disturbance vector analysis
- New results
- Future research

Preliminaries
SHA-1 compression function

- Compression function
  - Input: $IHV_{\text{in}} \in \{0,1\}^{160}, \quad B \in \{0,1\}^{512}$
  - Output: $IHV_{\text{out}} \in \{0,1\}^{160}$
- Uses 32-bit words $(x_i)_{x=0}^{31} \in \{0,1\}^{32} \quad \leftrightarrow \quad X \in \mathbb{Z}_{2^{32}} = \mathbb{Z}/2^{32}\mathbb{Z}$
  - bitwise operations: $X \oplus Y,\ X \vee Y,\ X \wedge Y,\ \overline{X},\ X^{\lll n},\ X^{\ggg n}$
  - modular operations: addition, substraction
- Message block
  - split: $B \to m_0, \ldots, m_{15} \in \mathbb{Z}_{2^{32}}$
  - expanded into 80 words:

$$W_t = \begin{cases} m_t & \text{if } 0 \leq t < 16; \\ (W_{t-3} \oplus W_{t-8} \oplus W_{t-14} \oplus W_{t-16})^{\lll 1} & \text{if } 16 \leq t < 80. \end{cases}$$

Preliminaries
SHA-1 compression function

- Working state $(Q_{t+i})_{i=-4}^{0} \in \mathbb{Z}_{2^{32}}^{5}, \quad 0 \le t \le 80$
- Initialization $IHV_{\text{in}} \rightarrow (Q_i)_{i=-4}^{0}$
- 80 steps: step $t = 0, \ldots, 79$
  - boolean function $f_t(X, Y, Z)$, e.g., $X \oplus Y \oplus Z$
  - computes

$$F_t = f_t(Q_{t-1}, Q_{t-2}^{\lll 30}, Q_{t-3}^{\lll 30});$$
$$Q_{t+1} = Q_t^{\lll 5} + F_t + Q_{t-4}^{\lll 30} + W_t + AC_t.$$

- Finalization

$$IHV_{\text{diff}} \leftarrow (Q_i)_{i=76}^{80};$$
$$IHV_{\text{out}} = IHV_{\text{in}} + IHV_{\text{diff}}.$$

Differential analysis
Differences

- Compare two instances of compression function
  - BSDR (binary signed difference representation)
    - $Z \in \{-1, 0, 1\}^{32}$
    - $\sigma : \{-1, 0, 1\}^{32} \to \mathbb{Z}_{2^{32}}, \quad (z_i)_{i=0}^{31} \mapsto \sum_{i=0}^{31} z_i \cdot 2^i$
    - $Z^{\lll n} = (z_{i+n \bmod 32})_{i=0}^{31}$
  - Variable X in one instance has related X′ in other instance
  - Differences between $X, X' \in \mathbb{Z}_{2^{32}}$:

$$\begin{aligned} \delta X &= X' - X && \in \mathbb{Z}_{2^{32}}, \\ \Delta X &= (X'[i] - X[i])_{i=0}^{31} && \in \{-1, 0, 1\}^{32}, \\ \delta X &= \sigma(\Delta X). \end{aligned}$$

Differential analysis
Differential path

- *Differential step* $0 \leq t < 80$ is given by values

$$\delta Q_{t+1}, \Delta Q_t, \Delta Q_{t-1}, \Delta Q_{t-2}, \Delta Q_{t-3}, \delta(Q_{t-4}^{\lll 30}), \Delta F_t, \delta W_t$$

  - such that

$$\delta Q_{t+1} = \sigma((\Delta Q_t)^{\lll 5}) + \sigma(\Delta F_t) + \delta(Q_{t-4}^{\lll 30}) + \delta W_t$$

$$\Delta F_t[b] \in \left\{ \begin{matrix} f_t(Q'_{t-1},(Q'_{t-2})^{\lll 30},(Q'_{t-3})^{\lll 30})[b] \\ -f_t(Q_{t-1},Q_{t-2}^{\lll 30},Q_{t-3}^{\lll 30})[b] \end{matrix} \;\middle|\; Q'_i[j] = Q_i[j] + \Delta Q_i[j] \right\}$$

- *Differential path* $\mathcal{P}$: differential steps $t \in [t_b, t_e]$
  - must match ΔQ_j and $\delta(Q_i^{\lll 30}) = \sigma((\Delta Q_i)^{\lll 30})$
  - success probability $\Pr[\mathcal{P}]$ can efficiently be determined
    - probability that differential steps hold exactly
    - for uniform randomly chosen $(Q_i)_{i=t_b-4}^{t_b}, W_{t_b}, \ldots, W_{t_e}$
    - assuming δW_i and $\delta(Q_{t_b-4}^{\lll 30}), \delta Q_{t_b-3}, \delta Q_{t_b-2}, \delta Q_{t_b-1}, \delta Q_{t_b}$

Differential analysis
Local collisions

- Collision attack based on *local collisions*
  - single disturbance: $\delta W_t = 2^b$
  - 5 corrections: $\delta W_{t+1}, \ldots, \delta W_{t+5}$
- Variations
  - signs
  - carries
- Success probability over all variations under same $(\delta W_i)_{i=t}^{t+5}$ easily determined

$$f_i(Q_{i-1}, Q_{i-2}^{\lll 30}, Q_{i-3}^{\lll 30})$$

i	ΔQ_i	ΔF_i	δW_i
$t-4$	0		
$t-3$	0		
$t-2$	0		
$t-1$	0		
t	0	0	$+2^b$
$t+1$	$+2^b$	0	-2^{b+5}
$t+2$	0	$\pm 2^b$	$\mp 2^b$
$t+3$	0	$\pm 2^{b+30}$	$\mp 2^{b+30}$
$t+4$	0	$\pm 2^{b+30}$	$\mp 2^{b+30}$
$t+5$	0	0	-2^{b+30}
$t+6$	0		

$$\delta Q_{i+1} = \sigma((\Delta Q_i)^{\lll 5}) + \sigma(\Delta F_i) + \delta(Q_{i-4}^{\lll 30}) + \delta W_i$$

Differential analysis
Disturbance vector

- XOR-based message differences $DW_t = W'_t \oplus W_t$
  - message expansion relation holds for $(W_t)_{t=0}^{79}$ and $(W'_t)_{t=0}^{79}$
  - thus must hold for $(DW_t)_{t=0}^{79}$:

$$DW_t = (DW_{t-3} \oplus DW_{t-8} \oplus DW_{t-14} \oplus DW_{t-16})^{\lll 1}$$

$$(\text{for } 16 \leq t < 80)$$

- Combine local collisions
- Disturbance vector $(DV_t)_{t=0}^{79}$
  - follows expansion relation
  - every ‘1’-bit marks start of local collision
  - XOR shifted and rotated D.V. as in local collision

$$DW_t = \bigoplus_{(i,r)\in\mathcal{R}} DV_{t-i}^{\lll r}$$

$$\mathcal{R} = \{(0,0),(1,5),(2,0),(3,30),(4,30),(5,30)\}$$

Differential analysis
Disturbance vector analysis

- Single local collision easily analyzed
- Last 60 steps: most important for attack complexity
  - at least 25 local collisions [JP05]
- Problems analyzing multiple local collisions
  - too many possible $(\delta W_i)_{i=20}^{79}$
  - too many possible $(\Delta Q_t)_{t=20-4}^{79+1}$ and $(\Delta F_t)_{t=20}^{79}$ per $(\delta W_i)_{i=20}^{79}$
  - truncated local collisions
    - non-zero differences $(\delta Q_t)_{t=16}^{20}$ and $(\delta Q_t)_{t=76}^{80}$
- Solution so far: assume independence
  - inaccurate maximum success probabilities [Man11][thesis]
  - leads to bad choices for $(\delta W_i)_{i=20}^{79}$
    - significantly worse collision attack complexity in reality

Overview

- Introduction
- Preliminaries
- Differential analysis
- New results
  - Differential paths
  - Maximum success probability
  - Differential path reduction
  - Differential path extension
  - New disturbance vector cost function
  - Cost function results
  - Near-collision attack construction
- Future research

New results [thesis]
Differential paths

- Limit differences as prescribed by D.V.:

$$\mathcal{Q}_t = \left\{ Y \in \{-1,0,1\}^{32} \;\middle|\; \begin{array}{c} \sigma(Y)=\sum_{i=0}^{31} x_i \cdot DV_{t-1}[i] \cdot 2^i \bmod 2^{32} \\ x_i \in \{-1,1\} \end{array} \right\},$$

$$\Delta Q_t \in \mathcal{Q}_t,$$

$$\delta Q_t \in \sigma(\mathcal{Q}_t) = \{\sigma(Y) \mid Y \in \mathcal{Q}_t\},$$

$$\delta(Q_t^{\lll 30}) \in \sigma(\mathcal{Q}_t^{\lll 30}) = \{\sigma(Y^{\lll 30}) \mid Y \in \mathcal{Q}_t\}.$$

- Limit $\delta W_t \in \mathcal{W}_t = \left\{ \sum_{i=0}^{31} x_i \cdot DW_t[i] \cdot 2^i \;\middle|\; x_i \in \{-1,1\} \right\}$
- Set of all differential paths over steps $[i,j]$

$$\mathcal{D}_{[i,j]} = \left\{ \mathcal{P} \;\middle|\; \begin{array}{r} \delta(Q_{i-4}^{\lll 30}) \in \sigma(\mathcal{Q}_{i-4}^{\lll 30}) \\ \Delta Q_k \in \mathcal{Q}_k, k = i-3, \ldots, j \\ \delta Q_{j+1} \in \sigma(\mathcal{Q}_{j+1}) \\ \delta W_t \in \mathcal{W}_t, t = i, \ldots, j \\ \Delta F_t[b] \in \{f_t(\cdots)[b] - f_t(\cdots)[b] \mid \Delta Q_l\} \end{array} \right\}$$

New results [thesis]
Maximum success probability

- Differential path $\mathcal{P} \in \mathcal{D}_{[20,79]}$
  - message differences (precondition)
    $$w = \Omega(\mathcal{P}) = (\delta W_t)_{t=20}^{79}$$
  - starting differences (precondition)
    $$\Lambda = \Phi(\mathcal{P}) = (\delta(Q_{16}^{\lll 30}), \Delta Q_{17}, \Delta Q_{18}, \Delta Q_{19}, \Delta Q_{20})$$
  - ending differences (postcondition)

$$\delta IHV_{\text{diff}} = \Psi(\mathcal{P}) = (\delta Q_{80}, \delta Q_{79}, \delta(Q_{78}^{\lll 30}), \delta(Q_{77}^{\lll 30}), \delta(Q_{76}^{\lll 30}))$$

- Success probability $p_{w,\Lambda,\delta IHV_{\text{diff}}} = \sum_{\substack{\mathcal{P}\in\mathcal{D}_{[20,79]} \\ \Lambda=\Phi(\mathcal{P}) \\ w=\Omega(\mathcal{P}) \\ \delta IHV_{\text{diff}}=\Psi(\mathcal{P})}} \Pr[\mathcal{P}]$
- Maximum success probability $\max_{\substack{w \\ \Lambda \\ \delta IHV_{\text{diff}}}} p_{w,\Lambda,\delta IHV_{\text{diff}}}$

New results [thesis]
Differential path reduction

- Differential path reduction $\mathrm{Reduce} : \mathcal{P} \mapsto (\mathcal{P}_\mathrm{r}, \mathcal{P}_\mathrm{d})$
  - $\mathcal{P}_r$ reduced differential path
    - differences removed that are not required for `extension'
    - $\Phi(\mathcal{P}) = \Phi(\mathcal{P}_\mathrm{r})$
    - $\Psi(\mathcal{P}) = \Psi(\mathcal{P}_\mathrm{r})$
  - $\mathcal{P}_d$ removed differences differential path
    - $\mathcal{P} = \mathcal{P}_\mathrm{r} + \mathcal{P}_\mathrm{d}$
    - $\Pr[\mathcal{P}] = \Pr[\mathcal{P}_\mathrm{r}] \cdot \Pr[\mathcal{P}_\mathrm{d}]$
    - $\Phi(\mathcal{P}) = \underline{0} \wedge \Psi(\mathcal{P}) = \underline{0} \quad \Rightarrow \quad \mathcal{P}_d = \mathcal{P}$
- Set of reduced paths $\mathcal{R}_{[i,j]} = \{\mathrm{Reduce}(\mathcal{P}) \mid \mathcal{P} \in \mathcal{D}_{[i,j]}\}$
  - success probabilities $p_{w,\mathcal{P}_\mathrm{r}} = \sum_{\substack{\mathcal{P} \in \mathcal{D}_{[i,j]} \\ \mathcal{P}_\mathrm{r} = \mathrm{Reduce}(\mathcal{P}) \\ w = \Omega(\mathcal{P})}} \Pr[\mathcal{P}] / \Pr[\mathcal{P}_\mathrm{r}]$

New results [thesis]
Differential path extension

- Given $\mathcal{R}_{[i,j]}$ and $p_{w,\mathcal{P}_\mathrm{r}}, \quad \forall w \in (\mathcal{W}_t)_{t=i}^{j}, \mathcal{P}_\mathrm{r} \in \mathcal{R}_{[i,j]}$
- We can compute extensions
  - $\mathcal{R}_{[i-1,j]}$ and $p_{w,\mathcal{P}_\mathrm{r}}, \quad \forall w \in (\mathcal{W}_t)_{t=i-1}^{j}, \mathcal{P}_\mathrm{r} \in \mathcal{R}_{[i-1,j]}$
  - $\mathcal{R}_{[i,j+1]}$ and $p_{w,\mathcal{P}_\mathrm{r}}, \quad \forall w \in (\mathcal{W}_t)_{t=i}^{j+1}, \mathcal{P}_\mathrm{r} \in \mathcal{R}_{[i,j+1]}$

- Starting from trivial $\mathcal{R}_{[k,k]}$
  we can compute $\mathcal{R}_{[20,79]}$ and $p_{w,\mathcal{P}_\mathrm{r}}, \ \forall w \in (\mathcal{W}_t)_{t=20}^{79}, \mathcal{P}_\mathrm{r} \in \mathcal{R}_{[20,79]}$

- Thus also $p_{w,\Lambda,\delta IHV_\mathrm{diff}} = \sum_{\substack{\mathcal{P}_\mathrm{r} \in \mathcal{R}_{[20,79]} \\ \Lambda = \Phi(\mathcal{P}_\mathrm{r}) \\ \delta IHV_\mathrm{r} = \Psi(\mathcal{P}_\mathrm{r})}} p_{w,\mathcal{P}_\mathrm{r}} \cdot \Pr[\mathcal{P}_\mathrm{r}]$

New results [thesis]
New disturbance vector cost function

- New disturbance vector cost function

$$FDC((DV_t)_{t=0}^{79}) = \max_{\substack{w \\ \Lambda \\ \delta IHV_{\text{diff}}}} p_{w,\Lambda,\delta IHV_{\text{diff}}} \cdot 2^{w(\Delta Q_{17})+w(\Delta Q_{18})}$$

  - correction due to fulfillment of ΔQ_{17} and ΔQ_{18} before fulfillment of ΔF_{20} in attack implementation

- Comparison cost function

$$FIC((DV_t)_{t=0}^{79}) = \prod_{Y \in \Gamma((DV_t)_{t=0}^{79})} FDC(Y)$$

  where Γ breaks disturbance vector into separate disturbance vectors each containing single local collision

New results [thesis]
Cost function results

- Comparison for selected disturbance vectors

DV	FDC	FIC
$\mathrm{I}(48,0)$	71.4	80.5
$\mathrm{I}(49,0)$	72.2	79.6
$\mathrm{I}(50,0)$	71.9	81.4
$\mathrm{I}(51,0)$	73.3	85.8
$\mathrm{I}(48,2)$	73.8	75.7
$\mathrm{I}(49,2)$	73.8	74.1
$\mathrm{II}(50,0)$	73.0	77.4
$\mathrm{II}(51,0)$	71.9	77.7
$\mathrm{II}(52,0)$	71.8	79.4

  - Results: -$\log_2$
  - Disturbance vectors selected by (near-)optimal FDC
  - Note: maximum success probability only obtained using the optimal message differences

New results [thesis]
Near-collision attack construction

- New analysis allows us to select
  - Λ that lead to optimal success probability $p_{\max}$
    - only a single Λ is used by the attack
  - δIHV_{diff} that (almost) lead to $p_{\max}$
  - w for which $N_w = N_{\max}$
    - where $N_w = (\# : \delta IHV_{\text{diff}} : p_{w,\Lambda,\delta IHV_{\text{diff}}} \geq 0.95 \cdot p_{\max})$

$$N_{\max} = \max_v N_v$$

- This implies
  - (near-)optimal success probability over steps 20-79
  - first near-collision attack has speedup $N_{\max}$

New results [thesis]
Near-collision attack construction

- Preliminary first near-collision attack
  - 192 possible $\delta IHV_{\mathrm{diff}}$
  - $N_{\max}$ is 6
  - runtime complexity of about $2^{57.5}$ calls
  - improves upon 2^{68} by [WYY05]
- Second near-collision attack
  - at least 6 times slower: $2^{60.1}$ calls
  - also more restrictions: slightly more slower
  - very-conservative upper-bound: $2^{64.1}$ calls
- Full collision attack: first+second near-coll. attack
  - estimated lower-bound: $2^{60.3}$ calls
  - very-conservative upper-bound: $2^{64.3}$ calls

Future research

- Current near-collision attack preliminary
- Optimize near-collision attack construction
  - algorithmic (near-)optimal construction
  - (near-)optimal application of speedup techniques
    - tunnels
  - exploit massively-parallel architectures
    - OpenCL, DirectCompute, CUDA
    - higher performance/cost ratio
- Compare attacks using promising D.V.s

Thank you for your attention

Questions?

More information

- Contact: marc@marc-stevens.nl
- Website: http://marc-stevens.nl
- HashClash: http://code.google.com/p/hashclash
  published sources of our implementions
- Information on MD5 attacks:
  http://www.win.tue.nl/hashclash