Cryptanalysis of the cryptographic standard SHA-1

Marc Stevens
Cryptology Group
CWI Amsterdam

joint work with Pierre Karpman (Inria, NTU)
& Thomas Peyrin (NTU)
Background
Background

- **Cryptographic hash functions**
  \[ H: \{0,1\}^* \rightarrow \{0,1\}^N \]

- **Collision resistance (informal)**
  Infeasible to find \( x \neq y \) with \( \text{SHA-1}(x) = \text{SHA-1}(y) \)
  (Generic attack: \( O(2^{N/2}) \))

- Weak: MD5 [Riv92], SHA-1 [NIST1995]
- Secure: SHA-2 [NIST2001], SHA-3 [NIST2015]
**Merkle-Damgård Construction**
- Splits message into 512-bit blocks
- Processes them iteratively using compression function

**Security proof**
- \( (H \text{ collision} \Rightarrow C.F. \text{ collision}) \)
- \( \Rightarrow (C.F. \text{ collision resistant} \Rightarrow H \text{ collision resistant}) \)
- \( \Rightarrow (C.F. \text{ collision} \Rightarrow ? (\text{no security proof})) \)
SHA-1 Compression function

- Linearly expand 16 words (32-bits) of message to 80 words
- Non-linear step function on 5 state words & 1 message word
- Davies-Meyer feedforward of Chaining Value
Digital signature standards based on

Hash - Then - Sign

Widely-used standards: (MD5-RSA,) SHA-1-RSA, SHA-2-RSA
Security depends on collision resistance of hash function
[Wang et al. 2004]

- Breakthrough cryptanalytic attacks

- Theoretical & practical break of hash function MD5

- Limited impact: identical-prefix collisions
Background

[2007&2009 Stevens et al.]

• more versatile: chosen-prefix collision attacks

• Practical: realistic abuse scenario with high impact
# Background

<table>
<thead>
<tr>
<th></th>
<th>MD5</th>
<th></th>
<th>SHA-1</th>
<th></th>
<th>SHA-256</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Id.Pr.</td>
<td>Ch.Pr.</td>
<td>Id.Pr.</td>
<td>Ch.Pr.</td>
<td>Id.Pr.</td>
</tr>
<tr>
<td>Birthday</td>
<td>$2^{64.3}$</td>
<td>$2^{64.8}$</td>
<td>$2^{80.3}$</td>
<td>$2^{80.8}$</td>
<td>$2^{128.3}$</td>
</tr>
<tr>
<td>2004</td>
<td>$2^{40}$</td>
<td></td>
<td>$2^{69}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>$2^{37}$</td>
<td></td>
<td>(2$^{63}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td>$2^{32}$</td>
<td>$2^{49}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td>$2^{25}$</td>
<td>$2^{42}$</td>
<td>(2$^{61}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>$2^{21}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>$2^{16}$</td>
<td>$2^{39}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td>$2^{61}$</td>
<td>$2^{77}$</td>
<td></td>
</tr>
<tr>
<td>today</td>
<td>$2^{16}$</td>
<td>$2^{39}$</td>
<td>$2^{61}$</td>
<td>$2^{77}$</td>
<td>$2^{128.3}$</td>
</tr>
</tbody>
</table>

Published collision attacks on MD5 & SHA-1
Background

- [NIST2011] Special Publication 800-131A

<table>
<thead>
<tr>
<th>Hash Function</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHA-1</td>
<td>Digital signature generation</td>
</tr>
<tr>
<td></td>
<td>Acceptable: -2010</td>
</tr>
<tr>
<td></td>
<td>Disclosed: 2011-2013</td>
</tr>
<tr>
<td></td>
<td>Disclosed: 2014-</td>
</tr>
<tr>
<td></td>
<td>Digital signature verification</td>
</tr>
<tr>
<td></td>
<td>Acceptable: -2010</td>
</tr>
<tr>
<td></td>
<td>Legacy-use: 2011-</td>
</tr>
<tr>
<td></td>
<td>Other applications</td>
</tr>
<tr>
<td></td>
<td>Acceptable</td>
</tr>
</tbody>
</table>

- [Schneier2012]: Projected costs of SHA-1 collisions
  - $2.77M in 2012
  - $700K by 2015
  - $173K by 2018
  - $43K by 2021
  (based on [Stevens12], Amazon EC2 rates & Moore’s Law)

- Actual CA/Browser Forum policy:
  - SHA-1 digital signature verification up to 1 Jan. 2017
Prior work

- *(Identical-prefix) collision attacks on full SHA-1*
  - Birthday search: $2^{80}$
  - [WYY05]: $2^{69}$
  - Wang, Yao, Yao 2005: $2^{63}$ (no publication, partially verified)
  - [SKI06]: ?? ($2^{52}$ symbolic message modifications $\times 2^{23}$?)
  - Mendel et al. 2007: $2^{60.3}$ (no publication)
  - [MHP09]: $2^{52}$ (withdrawn)
  - Chen 2011: $2^{58}$ (not peer-reviewed, too optimistic by factor $2^{3.5}$)
  - [Stevens13]: $2^{61}$

- *Example reduced-round SHA-1 collisions*
  - [DR06]: $2^{35}$ (64 out of 80 steps)
  - [DMR07]: $2^{44}$ (70 out of 80 steps)
  - [Gre10]: $2^{50.7}$ (73 out of 80 steps)
  - [GA11]: $2^{57.7}$ (75 out of 80 steps) (10,000 GPU-days, 1 GPU $\approx$ 40 cores)
Our work
Our work

- Example SHA-1 collisions thought to be imminent since 2005
- Previous works show analysis more complicated & too high cost
- Our research directions
  1. Precise analysis
     ⇒ optimal complexity & degrees of freedom
  2. Use massively-parallel architectures: graphic cards (GPUs)
     ⇒ more cost efficient
  3. Collisions on (reduced-round) SHA-1’s Compression Function
     ≡ freestart collision attack on (reduced-round) SHA-1
- Our results: freestart collision attacks on SHA-1
  - [KPS15] : $2^{50.3}$ (76 out of 80 steps) (5 GPU-days, 1GPU≈140cores)
  - [SKP16] : $2^{57}$ (80 out of 80 steps) (640 GPU-days, 1GPU≈140cores)
  - First practical attack on full SHA-1!
  - More efficient GPU implementation (prev: 1GPU≈40cores)
  - Estimations for cost of collision attack on full SHA-1
    : $2^{61}$ (SHA-1 collision) (40,000 GPU-days, EC2 ≈ $100k$)
SHA-1 cryptanalysis

Differential path

- Precise description of how differences propagate through compression function
- Last 60 steps determine most of attack’s complexity
- [Stevens13] precise methods to determine optimal differential paths
  [KPS15,SKP16] improvements
  (very technical, omitted here)
- Translate differential path into system of equations to solve
SHA-1 cryptanalysis

System of equations

- Simple equations on expanded message bits
  ⇒ linear equations on input message bits
- Simple equations on state bits
- First 16 steps easily solved
  ⇒ all message bit equations fulfilled
  ⇒ determines remaining 64 steps
- Make predictable small changes to solve up to step 24
  (amortizes cost of earlier steps)
  ⇒ only control about 30% of SHA-1
- Find many solutions up to step 24 to probabilistically fulfill remaining steps
Freestart collision attack

- Start from the middle
  - Advantage: higher probability diff. path \(\Rightarrow\) lower complexity
  - Disadvantage: cannot control input CV
  - \(\Rightarrow\) collision for C.F.

- Motivation
  - Invalidates security proof
  - Intermediate results
  - To perfect cryptanalysis tools
  - Testbed for GPU implementation
○ Nvidia GTX-970
○ Recent, high-end, good price/performance
○ 13 x 128 = 1664 cores @ 1.2 GHz
○ High-level programming with CUDA
○ Throughput for 32-bit arithmetic: all 1/cycle/core (except rotl/rotr)
○ ≈ € 350

○ Single Instruction Multiple Threads
  Execution is bundled in **warps** of 32 threads
  Control-flow divergence is serialized ⇒ minimize branching
○ Hide latency by running more threads than cores
  Transparent scheduling of actionable warps to cores
○ Be careful: incoherent memory reads/writes are slow
• [KPS15, SKP16] GPU tree search framework
  1. Store partial solutions up to some step in shared buffers
  2. Every thread of a warp loads one solution
  3. ... tries all degrees of freedom for this step
  4. ... stores successful larger partial solution in next step buffer
• Depth-first search: always process last queue with enough work
GPU vs CPU

○ Freestart 76-step SHA-1 [KPS15]
  ○ Initial GPU implementation: only easy speed-up tricks
  ○ On one GPU, the attack takes $\approx 4.2$ days
  ○ On one CPU core @ 3.2 GHz, the attack takes $\approx 606$ days
  ○ $\Rightarrow$ One GPU $\equiv 140$ CPUcores
  ○ (To compare with $\equiv 40$ [GA11])
  ○ For raw SHA-1 computations, ratio is 320
  ○ $\Rightarrow$ Relative loss of only $\times 2.3$ due to branching
    (better than expected for a highly branching tree search!)
○ Freestart full SHA-1 (80-steps) [SKP16]
  ○ Second generation implementation: also advanced speed-up tricks
  ○ Complexity: $2^{57}$
  ○ $\approx 10$ days on 64 GPUs (16 desktops with 4 GTX970 each)
  ○ First practical attack on full SHA-1

<table>
<thead>
<tr>
<th></th>
<th>Message 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IV_1$</td>
<td>50 6b 01 78 ff 6d 18 90 20 22 91 fd 3a de 38 71 b2 c6 65 ea</td>
</tr>
<tr>
<td>$M_1$</td>
<td>9d 44 38 28 a5 ea 3d f0 86 ea a0 fa 77 83 a7 36</td>
</tr>
<tr>
<td></td>
<td>33 24 48 4d af 70 2a aa a3 da b6 79 d8 a6 9e 2d</td>
</tr>
<tr>
<td></td>
<td>54 38 20 ed a7 ff fb 52 d3 ff 49 3f c3 ff 55 1e</td>
</tr>
<tr>
<td></td>
<td>fb ff d9 7f 55 fe ee f2 08 5a f3 12 08 86 88 a9</td>
</tr>
<tr>
<td>Compr($IV_1, M_1$)</td>
<td>f0 20 48 6f 07 1b f1 10 53 54 7a 86 f4 a7 15 3b 3c 95 0f 4b</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Message 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$IV_2$</td>
<td>50 6b 01 78 ff 6d 18 91 a0 22 91 fd 3a de 38 71 b2 c6 65 ea</td>
</tr>
<tr>
<td>$M_2$</td>
<td>3f 44 38 38 81 ea 3d ec a0 ea a0 ee 51 83 a7 2c</td>
</tr>
<tr>
<td></td>
<td>33 24 48 5d ab 70 2a b6 6f da b6 6d d4 a6 9e 2f</td>
</tr>
<tr>
<td></td>
<td>94 38 20 fd 13 ff fb 4e ef ff 49 3b 7f ff 55 04</td>
</tr>
<tr>
<td></td>
<td>db ff d9 6f 71 fe ee ee e4 5a f3 06 04 86 88 ab</td>
</tr>
<tr>
<td>Compr($IV_2, M_2$)</td>
<td>f0 20 48 6f 07 1b f1 10 53 54 7a 86 f4 a7 15 3b 3c 95 0f 4b</td>
</tr>
</tbody>
</table>
● Predictions for cost of collisions for full SHA-1
  ● Complexity: $2^{61}$ [Stevens13]
  ● $\approx 40,000$ GPU days (Amazon EC2: older GPUs)
  ● $\approx$ $100k$ renting fee on Amazon EC2 (spot-prices)
  ● $\times 7$ lower cost in 2015 than predicted earlier by Schneier
Impact & Conclusion
Industry Impact

- CA/Browser Forum: Ballot 152
  - Extend issuance SHA-1 certificates up to 1 Jan. 2017 (before: 1 Jan. 2016)
  - (unaltered: deprecate SHA-1 certificates after 1 Jan. 2017)
  - Proposed/endorsed by Entrust, Microsoft, Trend Micro
  - Seemingly enough support to pass
  - Our recommendations on 8 Oct. ensured Ballot did not pass on 16 Oct.

- Certification Authorities have found loop-hole
  - Withdraw older CA certificate from Browser root-CA-stores
  - $\Rightarrow$ not encumbered by CA/Browser regulations $\Rightarrow$ can sign SHA-1

- Mozilla, Microsoft & Google:
  - Possibly deprecate SHA-1 certificates per 1 July 2016

- TLS 1.3 draft 9
  - Deprecated all uses of SHA-1 digital signatures
Conclusion

- Improved cryptanalysis of SHA-1 using
  - Precise analysis methods (omitted here)
  - More efficient GPU tree search framework
- Freestart collision attacks on
  - 76-step SHA-1
  - Full SHA-1 ! $\Rightarrow$ first practical attack on full SHA-1
  - $\Rightarrow$ invalidates SHA-1’s collision resistance proof
- Work-in-progress
  - Collision attack on full SHA-1
- Industry is deprecating SHA-1 painstakingly slow
  - SHA-1 has been used ubiquitously as de facto industry standard
    $\Rightarrow$ very hard and costly to deprecate everywhere
  - CA/Browser forum is at the frontier, but deprecating per 1 Jan. 2017
  - $\Rightarrow$ Need practical examples to speed-up deprecation
- Note: counter-cryptanalysis [Stevens13b]
  - Detect digital signature forgeries constructed using collision attack
  - Practical & real-time: only $\times$ 2 as slow as plain SHA-1
Thank you!
SHA-1 cryptanalysis

Attacks on SHA-1 based on near-collision attacks

Identical-prefix collision attack

Chosen-prefix collision attack
SHA-1 cryptanalysis

- Attacks on SHA-1 based on near-collision attacks
- Near-collision attack on compression function:
  - Given input chaining value pair
  - Compute message block pair
  - To achieve ‘desired’ difference between output chaining values

Based on a differential path