Cryptanalysis of MD5 & SHA-1

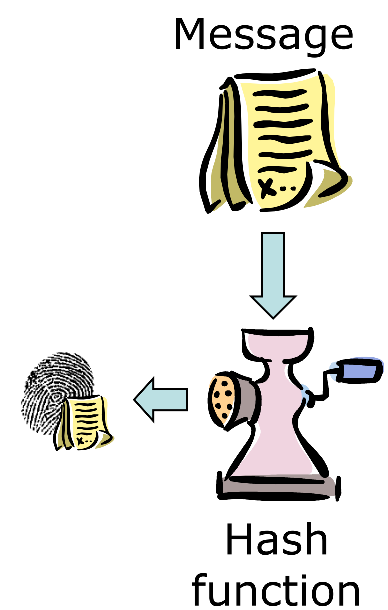
Marc Stevens
marc.stevens@cwi.nl

CWI Amsterdam

- Introduction
  - Cryptographic hash functions
  - Main applications
  - Public hash standards
  - Design of MD5 & SHA-1
- Advances in cryptanalysis of MD5
- Real-world impact of collision attacks
- Recent advances in cryptanalysis of SHA-1

Cryptographic hash functions

- Deterministic algorithm
  - In: message of arbitrary bit-length
  - Out: digital fingerprint of fixed short bit-length



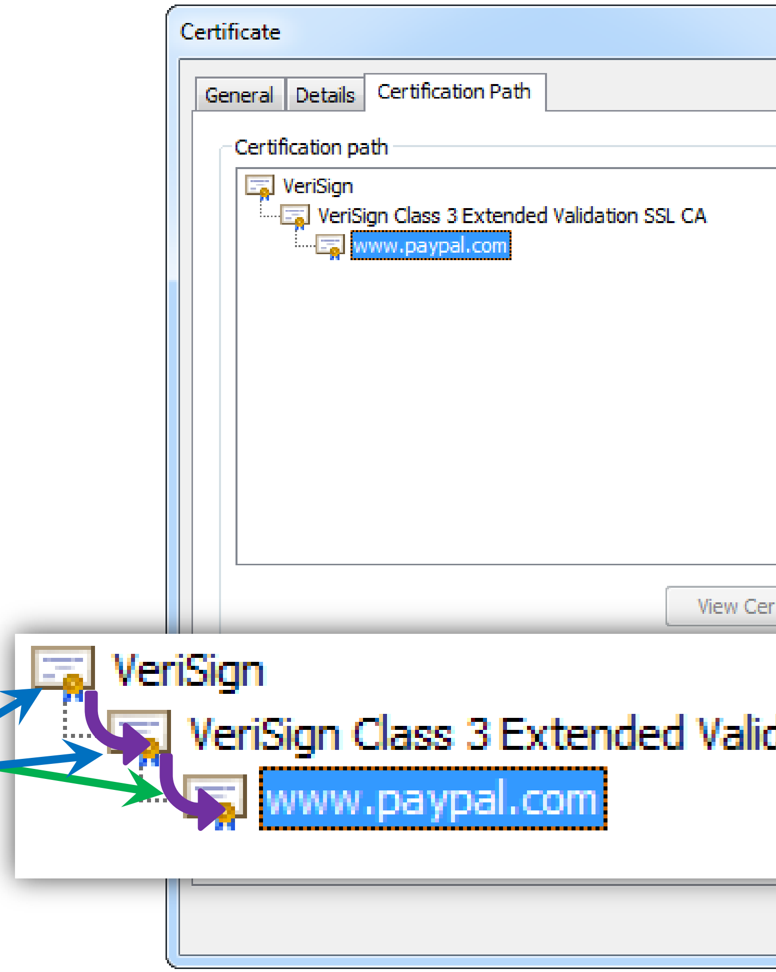
- Security requirement: collision resistance
  - It should be 'hard' to find collisions:

    $a \neq b$ such that $H(a) = H(b)$

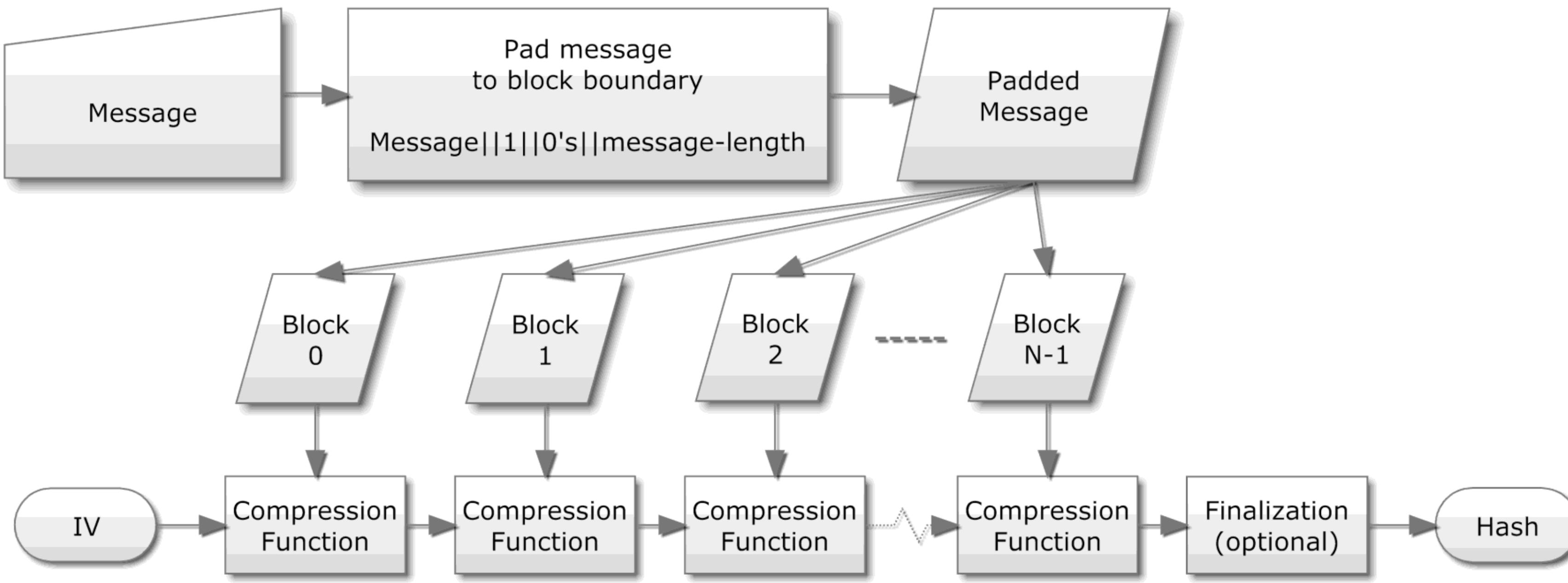
- 'Odd' cryptographic primitive
  - No key
  - No randomness
  - No mathematical definition of collision resistance
    (for fixed non-keyed hash functions)
    Informal definition: there are no *known* attacks better than brute-force

- Digital signatures: hash-then-sign
  - Process message to hash: $h=H(m)$
  - Sign hash with RSA: $s=RSA(sk,h)$
  - If $H(a)=H(b)$ then $Sign(sk,a)=Sign(sk,b)$
  - <u>Requires collision resistant hash function</u>

- Digital certificates
  - Usage: proof of identity in **`https://`**
  - Hierarchy: tree
    - End-node: https server
    - Parent-node: Certification Authority
  - Node certificate signed by parent



- MD5 (‘91, Rivest, 128-bit hash)
  - **broken**: 2^{16} compressions [SSA+09] (~20 ms on 1 core)
  - Still used
- SHA-1 (‘95, NIST, 160-bit hash)
  - **broken**: 2^{61} compressions [MRR07] [S12] (~16,000 years on 1 core)
  - Still widely used
- SHA-2 (‘01, NIST, 224/256/384/512-bit hash)
  - **secure**: attacks up to 41-step SHA-256 & 46-step SHA-512 (of 64/80 steps)
- SHA-3 (‘12, NIST, 224/256/384/512-bit hash) **new!**
  - **secure**: attacks up to 8 rounds (of 12 up to 24 rounds)

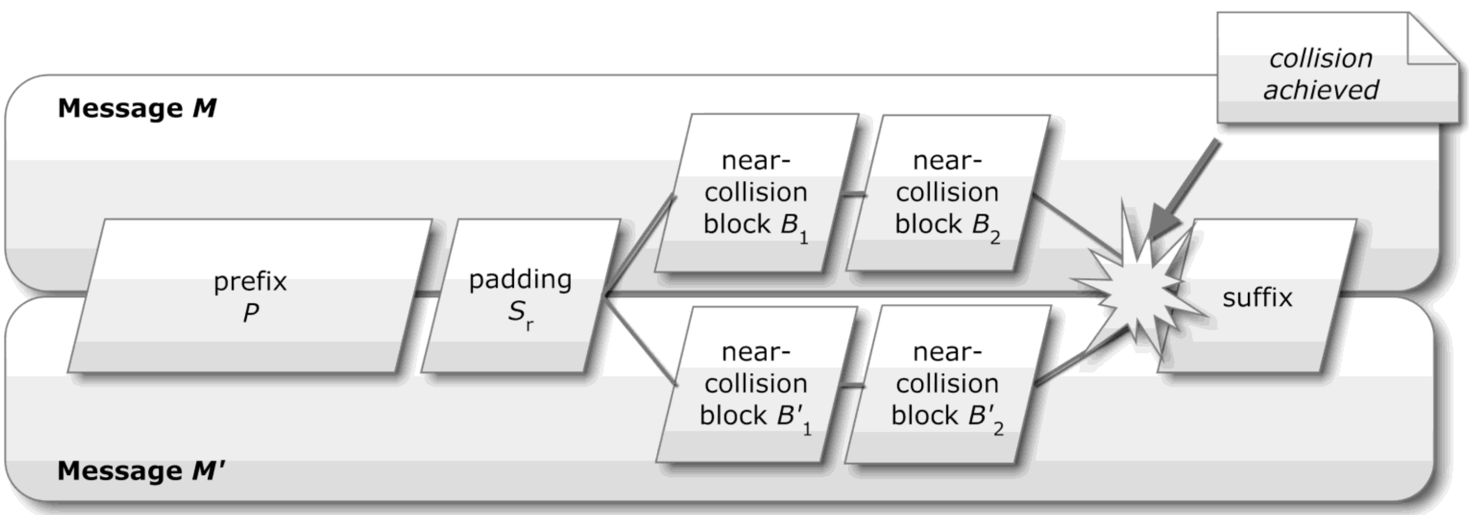

Message
Pad message
to block boundary
Message||1||0's||message-length
Padded
Message
Block
0
Block
1
Block
2
Block
N-1
IV
Compression
Function
Compression
Function
Compression
Function
Compression
Function
Finalization
(optional)
Hash

Overview

- Introduction
- Advances in cryptanalysis of MD5
  - First MD5 collision
  - MD5 chosen-prefix collision attack
  - Update on MD5 collision attacks
- Real-world impact of collision attacks
- Recent advances in cryptanalysis of SHA-1

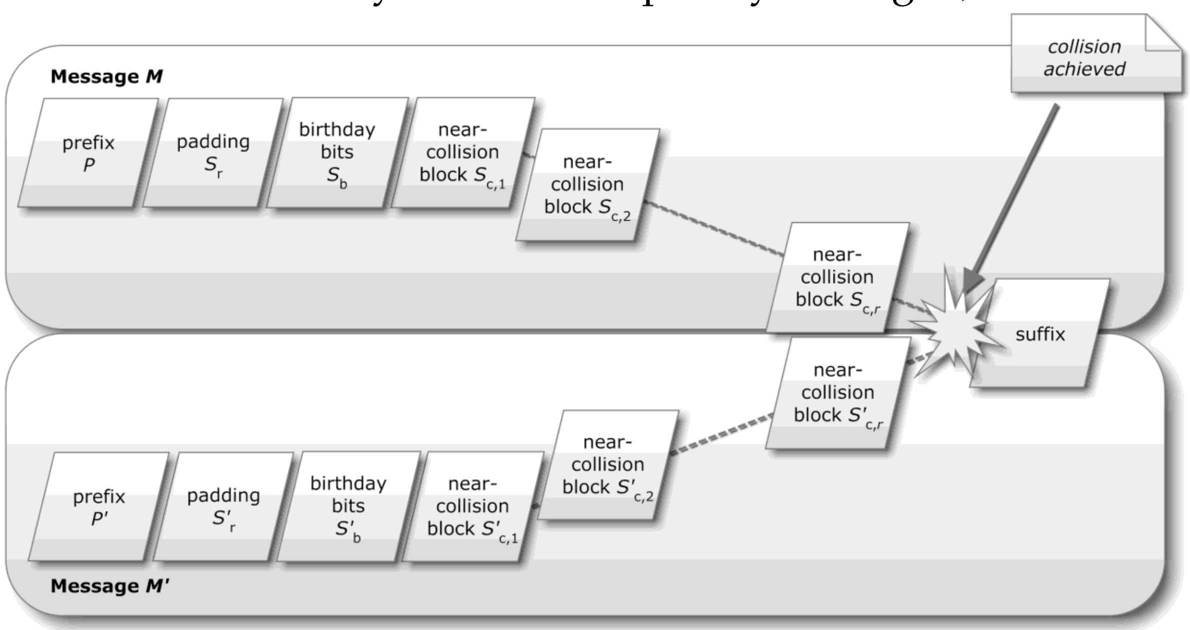
2004 [WY05]

- Breakthrough cryptanalysis `by hand'
- First MD5 collision found: $m \neq m'$ with MD5(m) = MD5(m')
- 2^{40} calls to MD5 (~64 hours on 1 core)
- *Identical-prefix collision* attack
- Skepticism from industry: "no meaningful differences"



2006 [SLdW07]

- Algorithmic cryptanalysis
- *Chosen-prefix collision* attack
  - Create collision from any two messages by appending suffix
  - Allows very meaningful differences
- 2^{49} MD5-calls (~1400 days on 1 core)
- Skepticism from industry: “attack complexity too high”, “no convincing scenario”



2009 [SSA+09]

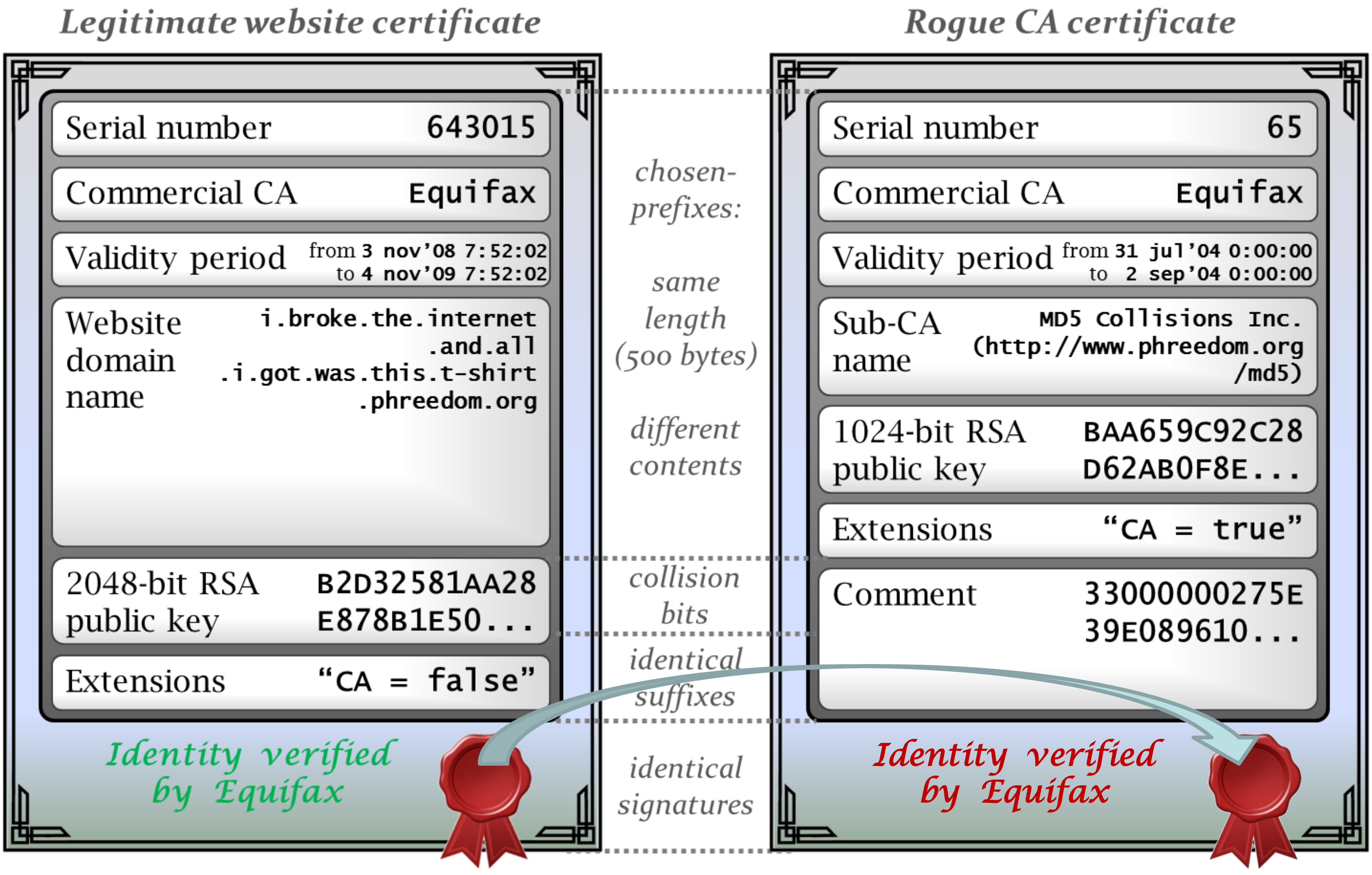
- Speed improvements
- Identical-prefix collision attack
  - New more efficient message differences
  - 2^{16} MD5-calls (~20 ms on 1 core)
- Chosen-prefix collision attack
  - More powerful and flexible birthday search
  - Extended family of differential paths
  - 2^{39} MD5-calls (~32 hours on 1 core)
- Convincing real-world example…

Overview

- Introduction
- Advances in cryptanalysis of MD5
- Real-world impact of collision attacks
  - Rogue Certification Authority
  - Overview colliding certificates
  - Abuse scenario
  - Impact
- Recent advances in cryptanalysis of SHA-1

Real-world impact of collision attacks
Rogue Certification Authority

- Colliding certificates with privilege escalation [SSA+09]
  - Legitimate secure website:
    - e.g., `https://marc-stevens.nl'
  - Illegitimate *sub-C.A.*:
    - ``MD5 Collisions, Inc.''
  - ``MD5 Collisions, Inc.'' trusted by IE9, FireFox, Chrome, ...
  - Successful proof-of-concept construction
    to counter skepticism of real-world danger of MD5 collision attacks



Legitimate website certificate
Rogue CA certificate
Serial number 643015
Commercial CA Equifax
Validity period from 3 nov'08 7:52:02 to 4 nov'09 7:52:02
Website domain name i.broke.the.internet .and.all .i.got.was.this.t-shirt .phreedom.org
2048-bit RSA public key B2D32581AA28 E878B1E50...
Extensions "CA = false"
Identity verified by Equifax
chosen-prefixes:
same length (500 bytes)
different contents
collision bits
identical suffixes
identical signatures
Serial number 65
Commercial CA Equifax
Validity period from 31 jul'04 0:00:00 to 2 sep'04 0:00:00
Sub-CA name MD5 Collisions Inc. (http://www.phreedom.org /md5)
1024-bit RSA public key BAA659C92C28 D62AB0F8E...
Extensions "CA = true"
Comment 33000000275E 39E089610...
Identity verified by Equifax

Real-world impact of collision attacks
Abuse scenario

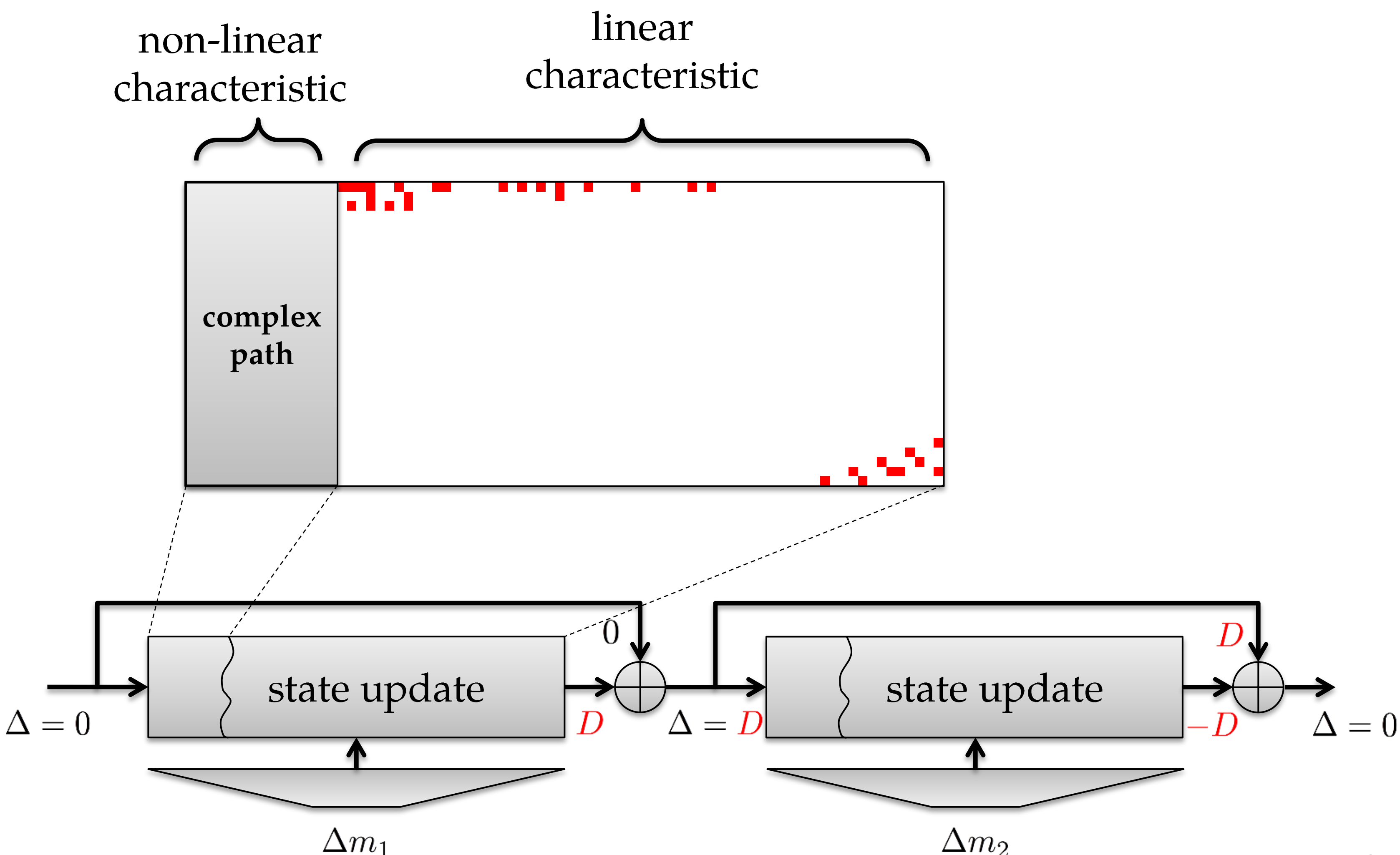
- Very powerful abuse scenario
  - Impersonating *all* secure websites
    - Requires subverting communications
    - Local network access sufficient
    - Man-in-the-middle attack

    - Harvest sensitive private information:
      E.g., usernames, passwords, address, ...
    - Alter queries and responses:
      E.g., financial transactions: account number, amount
  - Demonstrated live at annual Crypto conference

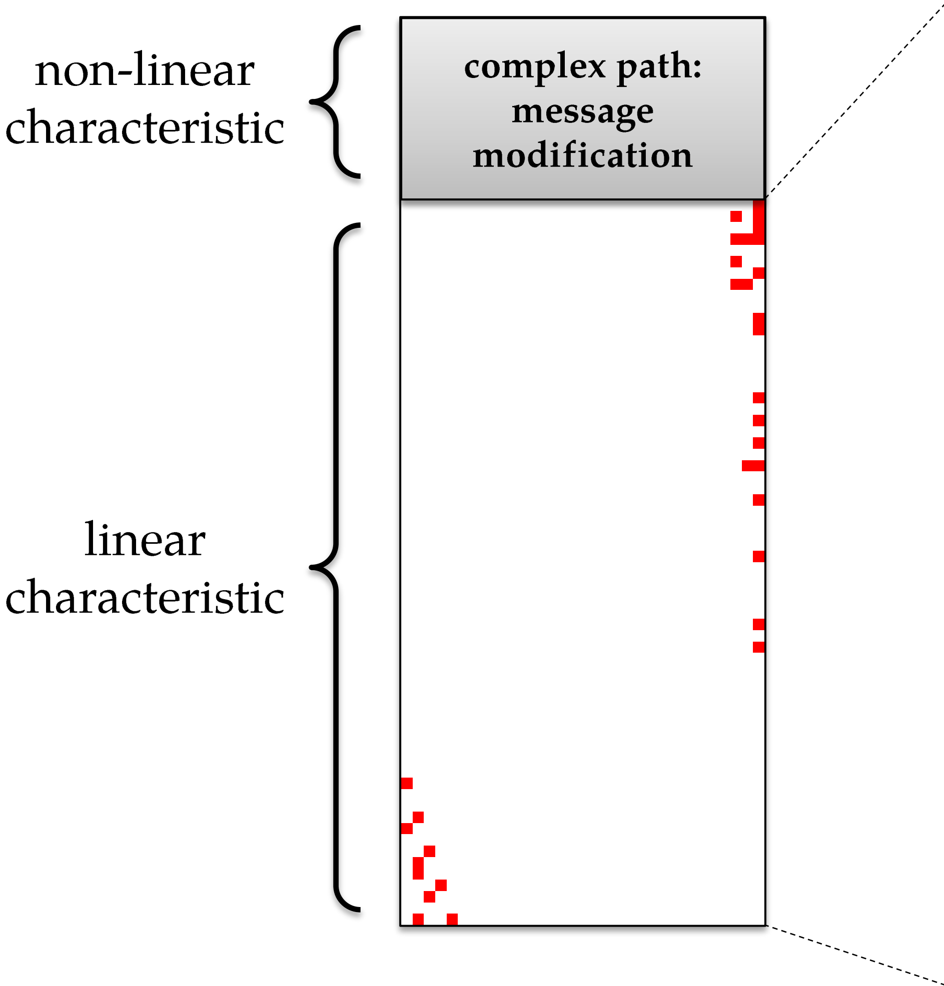
- Impact
  - Collision attacks proven to be very dangerous in practice, not just theoretical
  - Our goal: C.A. abandoned MD5
  - Led to more secure standards for C.A. industry
    - No MD5
    - No SHA-1 after 2012
    - Insert at least N bits of randomness in certificates
    - (RSA public key: at least 2048 bits)
  - New precedent for security researchers
    - Possible legal risk to be silenced
    - Using EFF: Microsoft & Mozilla signed Non-Disclosure Agreement
    - Responsible disclosure through Microsoft & Mozilla

- Introduction
- Advances in cryptanalysis of MD5
- Real-world impact of collision attacks
- Recent advances in cryptanalysis of SHA-1
  - Historic overview
  - Basic attack strategy
  - Novel cryptanalysis
  - New attacks

2005	First SHA-1 collision attack [WYY05a]
	Identical-prefix collision attack: 2^{69} calls (4,000,000 years on 1 core)
2005	Claim: 2^{63} calls [WYY05b] : unpublished
2007	Claim: 2^{61} calls [MRR07] : unpublished
2009	Claim: 2^{52} calls [MHP09] : withdrawn
2011	[PCTH11]: first attack is best *published* attack: 2^{69} calls No actual collision found yet


non-linear characteristic
linear characteristic
complex path
state update
state update
$\Delta = 0$
D
0
$\Delta = D$
D
$-D$
$\Delta = 0$
Δm_1
Δm_2

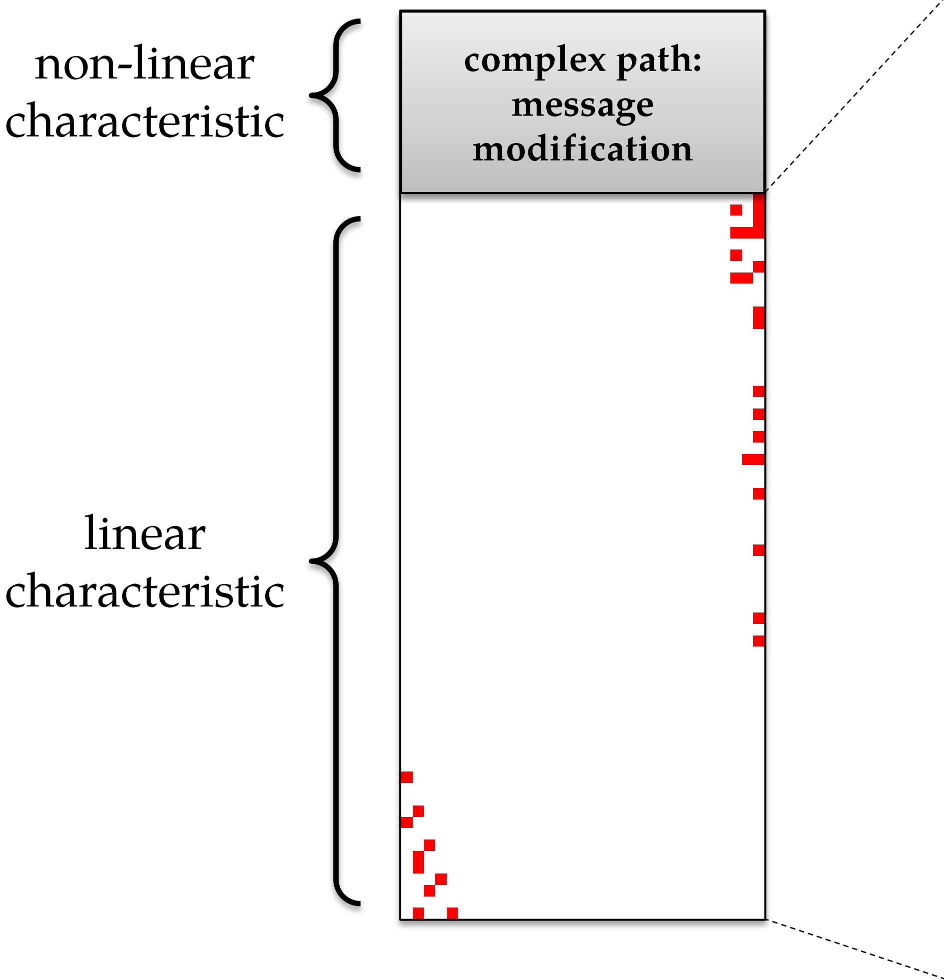
Basic attack strategy



- Linear combination of local collisions
- E.g., last 60 steps

- Most significant factor in total attack complexity

- Study local collision independently
  - Combine probabilities
  - Combine conditions
- Known dependencies
  - Heuristic corrections
  - **Sub-optimal solutions**

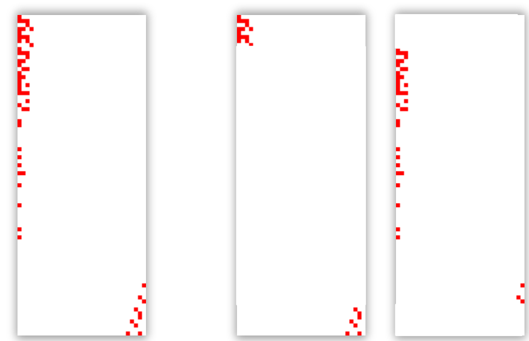


Novel approach [S12]

- Enumerate *all* differential paths
  - With prescribed disturbances
  - Sum exact probabilities
- Automatically captures heuristic techniques & more
  - Carries
  - Compression technique
  - **Dependencies**
- Exact & exhaustive
  $\Rightarrow$ **leads to optimal solution**
- Main problems:
  - Exponential # Δm vectors
  - Exponential # paths

- Problem: *Exponential # Δm vectors*
  - Solution: **message vector classes**
    - Vectors in same class ⇔ same 'characteristics'
    - Only process one vector of each class
    - Deals with major redundancies
- Problem: *Exponential # differential paths*
  - Solution: **differential path reduction**
    - Removes 'independent inner parts'
    - Many paths lead to same reduced path
    - Compute cumulative probabilities removed parts
- Efficient algorithmic solution
  - Iterative process: 1 step, 2 steps, …, 60 steps
  - Simultaneously determines:
    - Reduced paths
    - Cumulative probabilities
    - Message vector classes

$$\begin{aligned} P &= R + S \\ \Pr[P] &= \Pr[R] \cdot \Pr[S] \end{aligned}$$



$$\sum_{P \in \mathcal{P}} \Pr[P] = \sum_{P \in \mathcal{P}} \Pr[R] \cdot \Pr[S] = \sum_{R \in \mathcal{R}} \Pr[R] \cdot \left(\sum_{S \in \mathcal{S}_R} \Pr[S] \right) = \sum_{R \in \mathcal{R}} \Pr[R] \cdot p_R$$

New attacks [S12] based on novel approach:

- New near-collision attack
  - $2^{57.5}$ compressions (~1,400 years on 1 core)
  - First open-source SHA-1 attack
  - Optimal L-part
  - Sub-optimal NL-part & 50+ bits of freedom left
    ⇒ room for improvement
- New identical-prefix collision attack
  - Two near-collisions: >7 times harder
  - 2^{61} compressions (~16,000 years on 1 core)
- New chosen-prefix collision attack
  - Birthday search + near-collision
  - $2^{77.1}$ compressions (~2,000,000,000 years on 1 core)

Conclusion

- Real-world security based on security of hash functions
- Need to understand security of widely used standards
  - Attacks can only get better, not worse
- Yet industry responds slowly to academic results
  - MD5 should be abandoned by now… is it?
  - SHA-1 is currently widely used… while broken for 7 years
- Is the industry waiting till the first SHA-1 collisions?
  - Might not come from Academia
  - Abandoning SHA-1 takes time, see MD5. Why wait?

Thank you for your attention

Questions?

References

[MRR07] *Update on SHA-1*, F. Mendel, C. Rechberger, V. Rijmen, rump session CRYPTO 2007. *(unpublished)*

[MHP09] *Differential path for SHA-1 with complexity $O(2^{52})$*, C. McDonald, P. Hawkes, J. Pieprzyk, Cryptology ePrint Archive, Report 2009/259. *(withdrawn)*

[PCTH11] *Security considerations for the SHA-0 and SHA-1 message digest algorithms*, T. Polk, L. Chen, S. Turner, P. Hoffman, RFC 6194, 2011.

[SLdW07] *Chosen-prefix collisions and colliding X.509 certificates for different identities*, M. Stevens, A.K. Lenstra, B. de Weger, EUROCRYPT 2007, LNCS Vol. 4515, pp. 1-22, Springer, 2007.

[SSA+09] *Short chosen-prefix collisions for MD5 and the creation of a rogue CA certificate*, M. Stevens, A. Sotirov, J. Appelbaum, A.K. Lenstra, D. Molnar, D.A. Osvik, B. de Weger, CRYPTO 2009, LNCS Vol. 5677, pp. 55-69, Springer, 2009.

[S12] *Attacks on hash functions and applications*, Marc Stevens, PhD thesis, Leiden University. (See also the open-source project at: http://code.google.com/p/hashclash/)

[WY04] *How to break MD5 and other hash functions*, X. Wang, H. Yu, EUROCRYPT 2005, LNCS Vol. 3494, pp. 19-35, Springer, 2005.

[WYY05a] *Finding collisions in the full SHA-1*, X. Wang, Y.L. Yin, H. Yu, CRYPTO 2005, LNCS Vol. 3621, pp. 17-36, Springer, 2005.

[WYY05b] *Cryptanalysis on SHA-1*, X. Wang, A.C. Yao, F. Yao, presentation, NIST Cryptographic Hash Workshop, 2005. *(unpublished)*