MD4-family, SHA-1, SHA-2, SHA-3

Lars R. Knudsen

May 5, 2009

MD4-family

- MD4, MD5, Rivest 1990, 1991
- SHA-0, 1993, US Gov.
- SHA-1, 1995, US Gov.
- SHA-256, SHA-512, 2002, US Gov.
- all hash functions of Davies-Meyer form

$$h(H_{i-1}, M_i) = f(H_{i-1}, M_i) \oplus H_{i-1},$$

  - f is a bijection when M_i is fixed
  - M_i typically much larger than H_{i-1}

Hash functions in real-life

Scheme	Bits in hash code	Compression fct. message bits	Compression fct. state bits	Designer	Year
MD4	128	512	128	Rivest	1990
MD5	128	512	128	Rivest	1991
SHA-1	160	512	160	US Gov.	1995
SHA-256	256	512	256	US Gov.	2002
SHA-512	512	1024	512	US Gov.	2002

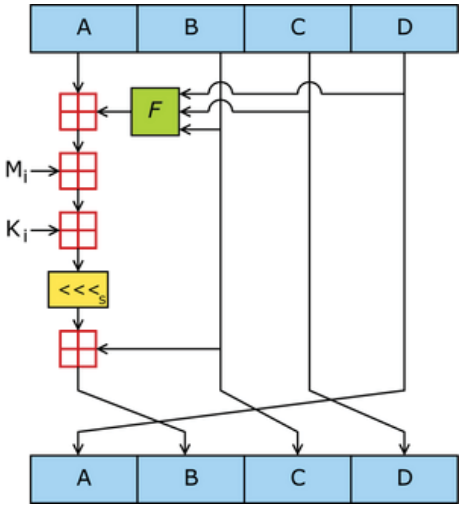
MD: Message Digest
SHA: Secure Hash Algorithm

From MD4 over MD5 to SHA

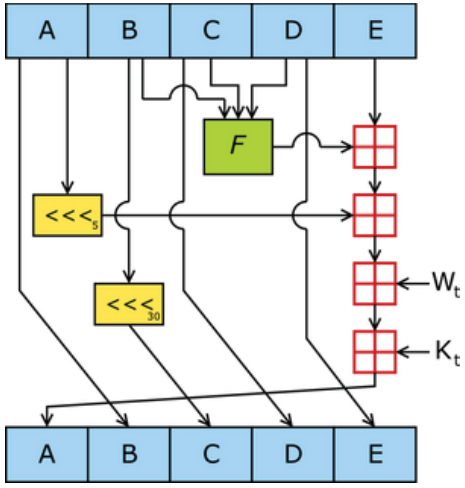
- Iterated hash functions
- Compression functions process message blocks of 512 bits
- Message blocks processed in words of 32 bits
- Message expanded from 512 to $32 \times r$ bits, where r is number of steps of algorithm
-

Algorithm	Steps	# registers of 32 bits
MD4	48	4
MD5	64	4
SHA-1	80	5

MD5 from www.wikimedia.org



SHA-1 from www.wikimedia.org



Hashing with SHA-1

- 80 basic steps in compression function
- Message $W = [W_0 \parallel W_1 \parallel \ldots \parallel W_{15}]$, where W_i are 32 bits
- Expansion:
  $W_i = \text{rot}_1(W_{i-3} \oplus W_{i-8} \oplus W_{i-14} \oplus W_{i-16}), \quad 16 \leq i \leq 79$
- Three functions used:

$$\begin{aligned} f_{if} &= (X \text{ AND } Y) \text{ OR } (\neg X \text{ AND } Z) \\ f_{xor} &= X \oplus Y \oplus Z \\ f_{maj} &= (X \text{ AND } Y) \text{ OR } (X \text{ AND } Z) \text{ OR } (Y \text{ AND } Z) \end{aligned}$$

- SHA-256, SHA-512, etc, follow same design principles

Cryptanalysis - highlights

- 1996: MD4 broken, Dobbertin
- 2004: MD5 broken, Wang
- 2004: SHA-0 broken, Joux et al
- 2005, claim: collisions for SHA-1 in time $\approx 2^{69}$ (Wang)
- 2006, claim: collisions for SHA-1 in time $\approx 2^{63}$ (Wang)
- 2007, claim: collisions for SHA-1 in time $\approx 2^{60}$ (Mendel, Rechberger, Rijmen)
- 2009, claim: collisions for SHA-1 in time $\approx 2^{52}$ (McDonald, Hawkes, Pieprzyk)

Hash function collisions irrelevant ?

- Often heard criticism, collisions are on "random" messages, so not important
- Dobbertin breaks MD4 in 94, after criticsm he shows meaningful collisions on MD4
- Often it requrires only little extra effort to make collisions "meaningful"
- Daum-Lucks, 2005, collision in PostScript
- Lenstra et al, 2005, forging certificate using MD5

Collision in Postscript (Daum-Lucks 2005)

- Notation: $(\mathrm{S1})(\mathrm{S2})\mathrm{eqT1T2ifelse}$
- Meaning: If $S1 = S2$ then $T1$ else $T2$
- Find random messages S1 and S2 which collide under hash function
- Construct PS1 and PS2 for arbitrary T1 and T2
- PS1: $\ldots(\mathrm{S1})(\mathrm{S2})\mathrm{eqT1T2ifelse}\ldots$
- PS2: $\ldots(\mathrm{S2})(\mathrm{S2})\mathrm{eqT1T2ifelse}\ldots$

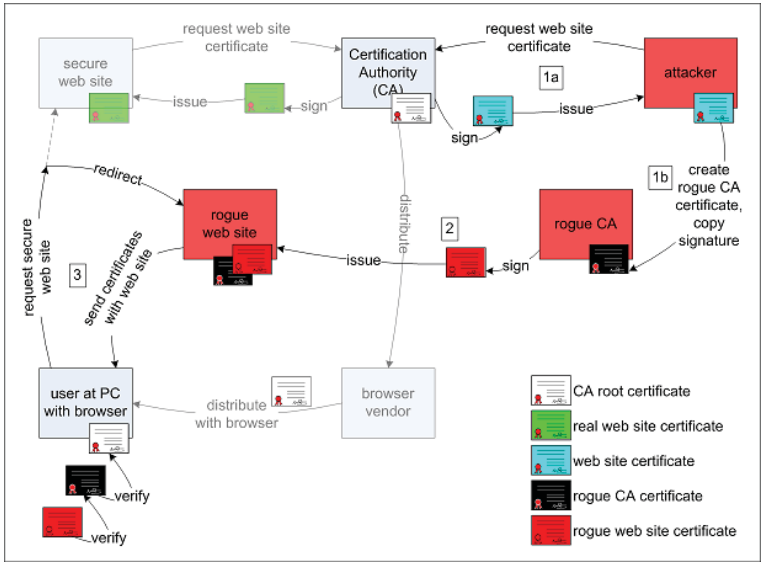
MD5 certificate

Current attacks on MD5 very powerful, lots of freedom for attacker

- 2005, colliding X.509 certificates, but same identities (Lenstra, de Weger, Wang)
- 2007, colliding X.509 certificates, different identities (Stevens, Lenstra, de Weger)
- 2009, Sotirov, Stevens, et al
  - request a legitimate website certificate from commercial CA
  - CA signs certificate which now signs also second certificate
  - second certificate is intermediary CA certificate, can be used to sign arbitrary other website certificates

All above using MD5

MD5 attack



SHA-3

SHA-3 - Call for candidates

- announcement: October 29, 2007
- must provide digests of 224, 256, 384, and 512 bits, not 160.
- available worldwide royalty-free, no IPR
- capable of protecting sensitive information for decades
- should be suitable for
  - digital signatures, FIPS 186-2
  - HMAC, FIPS 198
  - key establishment, SP 800-56A
  - random number generation, SP 800-90
- security strength at least that of the SHA-2s with **significantly** improved efficiency

SHA-3 - Desirable properties

- efficient integral options, e.g., randomized hashing, that "fundamentally improve security"
- parallelizable
- avoid "generic properties" of Damgård/Merkle constructions
- attack on SHA-2 should not lead to attack on SHA-3
- flexible for a wide variety of implementations
- a single family, except if good arguments for more families
- tunable security parameter, e.g., number of rounds, with recommendations

SHA-3 - Security

Message digest of n bits

- Collisions should require $2^{n/2}$ operations
- Preimages should require 2^n operations
- 2nd preimages should require 2^{n-k} for messages shorter than 2^k bits

Higher levels of security against 2nd preimage will be viewed positively

- NIST open to other designs than Damgård/Merkle

SHA-3 - Timeline

- hard submission deadline: 31/10-2008
- documentation and testing like AES
- review is public
- 64 submissions
- 51 candidates selected for round 1
- ≈ 15 selected for round 2 later 2009
- ≈ 5 selected for round 3 late 2010 (?)
- winner selected late 2011 (?)

SHA-3

Two bigger classes

- based on or using AES
- RAX designs, mix of rotations, modular additions and exors

Status:

- 10 of 51 candidates considered broken or withdrawn by designers