Collision Resistance

Let $H: \mathcal{M} \to \mathcal{T}$ be a hash function

$(|\mathcal{T}| \ll |\mathcal{M}|)$

A collision for H is a pair $m_0, m_1 \in \mathcal{M}$

s.t. $m_0 \neq m_1$ and $H(m_0) = H(m_1)$

$|\mathcal{T}| < |\mathcal{M}| \Rightarrow$ collisions must exist!

Def. A function $H: \mathcal{M} \to \mathcal{T}$ is collision resistant (CRH)

if for all explicit eff. algos A:

$\text{CRadv}[A, H] := \Pr[A \text{ outputs collision for } H]$

is negligible

$\Rightarrow$ hard to find an explicit collision

Std. examples

2001: SHA256, SHA384, SHA512 $\Leftarrow$ most widely used, Intel HW instructions (since 2019, AMD since 2017)

2014: SHA3-256, SHA3-384, SHA3-512

Immediate approach

small MAC $\Rightarrow$ big MAC

→ (S, V) secure MAC over $(\mathcal{K}, \mathcal{M}, \mathcal{T})$ for short msgs.

→ $H: \mathcal{M}^{big} \to \mathcal{M}$ a CRH

→ Def (S, V) a MAC over $(\mathcal{K}, \mathcal{M}^{big}, \mathcal{T})$

where $\begin{cases} S'(k, m) := S(k, H(m)) \\ V'(k, m, t) := V(k, H(m), t) \end{cases}$

Thm: (S, V) a secure MAC, H a CRH

Then (S', V') is a secure MAC

Why CRH needed?

Suppose adv has $m_0 \neq m_1 \in \mathcal{M}^{big}$ s.t. $H(m_0) = H(m_1)$

attack on (S', V'):

→ req tag on m_0, get t

→ output forgery (m_1, t)

$\Rightarrow$ valid forgery b/c $V(k, H(m_1), t)$ succeeds

CRH generic attacks

General attack: "birthday attack"

Bday paradox:

Let $r_0 \dots r_n \xleftarrow{R} \{1, B\}$ be ind. uniform RVs

Thm: When $n \geq 1.2\sqrt{B}$ then $Pr[\exists i \neq j : r_i = r_j] \geq \frac{1}{2}$

Proof:

$$Pr[\exists i \neq j : r_i = r_j] = 1 - Pr[\forall i \neq j : r_i \neq r_j]$$

$$= 1 - (1 - \tfrac{1}{B})(1 - \tfrac{2}{B})(1 - \tfrac{3}{B}) \dots (1 - \tfrac{n}{B})$$

$$= 1 - \prod_{i=1}^{n} (1 - \tfrac{i}{B}) \geq 1 - \prod_{i=1}^{n} e^{-i/B}$$

$$\geq 1 - e^{-n^2/2B} \geq 1 - e^{-1/2 \cdot 1.2^2} \approx 0.53$$

Bday attack:

1. Choose random $m_0 \dots m_{2^{\ell/2}} \in M$
2. Compute $H(m_0) \dots H(m_{2^{\ell/2}})$
3. Look for collision
4. If no collision, goto 1.

after exp. 2 iters, will find collision

time $= O(\sqrt{|T|})$

So: 128-bit hash: collision time 2^{64} (bad)

256-bit hash 2^{128} ✓

Generic attack on SHA256 takes ~ same time as on AES128

Naively: memory $O(2^{\ell/2})$

Puzzle: can find collision in time $O(2^{\ell/2})$ using $O(1)$ space

Quantum

Some evidence that collisions can be found in time $O(2^{\ell/3})$, but still open

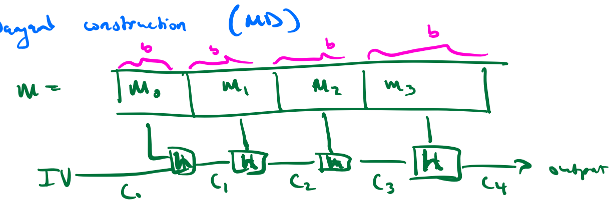
(lots of space too)

Constructing a CRH

1. Merkle-Damgard construction (MD)

msg:

H_{MD}



Terminology: ① $h: \{0,1\}^b \times T \to T$ compression func

② $c_0 \dots c_n$ chaining vars

③ IV: fixed initial value

SHA256: input size 512 bits (32 bytes)

pad: ensures msglen is a multiple of b

pad = $[100\dots0 \,|\, \underbrace{\text{msg-len}}_{64\text{ b}}]$

if no space for pad - dummy block.

Thm. h is CRH → H_{MD} is CRH.

Proof. Suppose $HMD(m) = HMD(m')$

Goal. Find collision h.

m: $IV = c_0, c_1, \dots c_t$

m': $IV = c'_0, c'_1, \dots c'_r$

$HMD(m) = HMD(m') \Rightarrow c_t = c'_r$

$\Rightarrow h(m[t-1], c_{t-1}) = h(m'[r-1], c'_{r-1})$

Suppose $(m[t-1], c_{t-1}) \neq (m'[r-1], c'_{r-1})$

$\Rightarrow$ found collision! ✓

if not: $m[t-1] = m'[r-1]$ ← last block has msg len

$c_{t-1} = c'_{r-1}$

$\Rightarrow r = t$

$m[t-1] = m'[t-1]$

$c_{t-1} = c'_{t-1}$

$\Rightarrow h(m[t-2], c_{t-2}) = h(m'[t-2], c_{t-2})$

if $h(m[t-2], c_{t-2}) \neq h(m'[t-2], c_{t-2})$

found collision for h! ✓

If not: repeat to beginning,

either found collision

or $m = m'$ ← contradiction

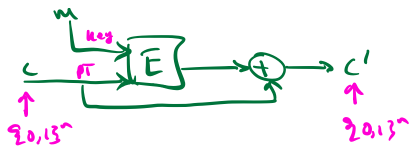
Constructing compression functions h

Davies-Meyers

Let $E(K, X)$ be block cipher over (K, X)
where $X = \{0,1\}^n$

key, PT

$h(m, c) := E(m, c) \oplus c$



Thm. if E is "ideal cipher" (random collection of permutation)
then finding collision on $h(m, c)$ takes time $\geq 2^{n/2}$

Best possible b/2 Bday attack finds coll. in $\approx 2^{n/2}$

SHA 256 uses cipher called SHACA2.

Applications

→ Software integrity

Files $F_1 \ldots F_n$ ← read/write no security

need: public read-only space

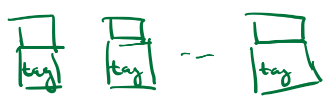
write: $H(F_1) \ldots H(F_n)$

CRH: attacker can't find $F' \neq F_i$
s.t. $H(F') = H(F_i)$

⇒ If downloaded F_i has correct hash, F_i is an authentic file

Two approaches to data integrity

1. read/write large data files
   + read-only small storage
2. MAC: read/write large data r/o
   owner + recipient have shared secret key



Building a MAC from a hash fn

How to build a PRF F and MAC from hash fn $H: M \to T$?

Attempt 1: $F(k,m) := H(k \| m)$

bad idea for MD H! (extension attack)

adv. can ask for MAC on M and obtain MAC on $M \| m$.

Given $Y = F(k,m)$ anyone can compute

$$Y' = F(k,\ m \| pad \| x)$$

for all one-block msgs x.

Standard method: HMAC

$$F_{HMAC}(k,m) := H\big((k \oplus \underbrace{opad}_{\text{outer pad}}) \,\|\, H(k \oplus \underbrace{ipad}_{\text{inner pad}} \| m)\big)$$

fixed values

Thm. if compr fn $h(m,c)$ is a secure PRF (w/ either inp as key) then HMAC is a secure PRF.