6.857 Rivest
L2.1 2/9/09

Administrivia: homework groups
term project ideas: slashdot ("security")
www.iacr.org/eprint

Outline: Hash functions

- ☐ Intro
- ☐ Random Oracle Model
- ☐ Desirable Properties
- ☐ Applications
- ☐ Construction (MD5)

Intro: What is a hash fn?

Maps arbitrary strings of data to fixed-length output in deterministic, public, "random" manner.

$h: \{0,1\}^* \rightarrow \{0,1\}^d$

$\{0,1\}^d$: strings of length d

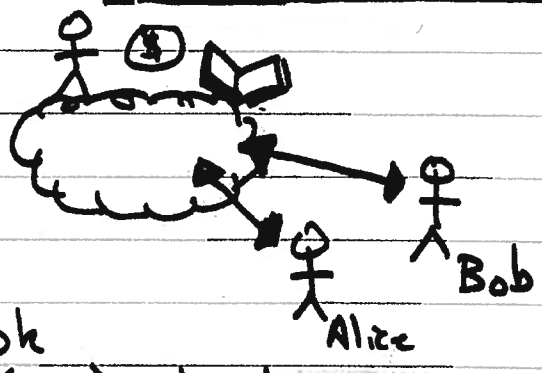
$\{0,1\}^*$: strings of arbitrary length ≥ 0

Also called "message digest" function.
Typical output lengths are 128, 160, 256, 512.
No secret key. All operations public. Anyone can compute h.
Examples: MD4, MD5, SHA-1, SHA-256, SHA-512

	MD4, MD5	SHA-1	SHA-256	SHA-512
d	128	160	256	512
broken (CR):	✓	✓?		

Should be easy to compute (poly-time)

Ideal: Random Oracle
(not achievable in practice)



Oracle: on input $x \in \{0,1\}^*$
  if x not in book
    - flip coin d times to determine h(x)
    - record (x, h(x)) in book
  else: return y where $(x,y) \in$ book.

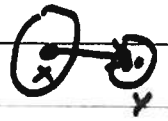
Gives random answer every time, except as req'd for consistency with previous answers. (h must be deterministic...)
("randomness on demand, with memory")
Many cryptographic schemes are proved secure in ROM (random oracle model); assumes we have RO.

In practice, $\nexists$ RO, so we use something "pseudorandom".

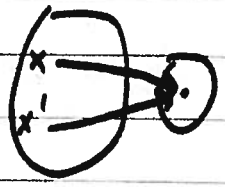
Desirable Properties

6.857 Rivest
L2.3 2/9/09

OW (1) "One-way" (pre-image resistance)
Infeasible, given $y \in_R \{0,1\}^d$, to find any x s.t. $h(x)=y$ (a "pre-image" of y)



CR (2) Collision-resistance (strong collision resistance)
Infeasible to find x, x' s.t. $x \neq x'$ and $h(x)=h(x')$ (a "collision")



TCR (3) Weak Collision Resistance (target collision resistance, 2nd pre-image resistance)
In-feasible, given x, to find $x' \neq x$ s.t. $h(x)=h(x')$

PRF (4) Pseudo-randomness
Behavior indistinguishable from RO

NM (5) Non-malleability
Infeasible, given $h(x)$, to produce $h(x')$ where x and x' are "related". (e.g. $x'=x+1$)

These are informal definitions, & don't really work as given, since h is fixed & public.
Need for good theory, to have family of hash functions...

Facts: h is CR $\Rightarrow$ h is TCR. (But not reverse)
h is OW $\not\Leftrightarrow$ h is CR (neither implication holds)

Fact: Collisions can be found in time $O(2^{d/2})$ [Birthday paradox]

Fact: Inversion can be done in time $O(2^d)$.

Applications

6.857 Rivest
L2.4 2/9/09

① Password storage

- Store $h(PW)$, not PW, on computer
- Disclosure of $h(PW)$ should not reveal PW (orequiv.)
- Need OW

② File modification detector

- For each file F, store $h(F)$ securely (on DVD)
- Check if F modified by recomputing $h(F)$
- need WCR (adversary wants to change F but not $h(F)$)
- (hashes of downloadable software equivalent problem)

③ Digital signatures

PK_A = Alice's PK
SK_A = Alice's SK

Signing: $\sigma = sign(SK_A, M)$ [Alice's sig on M]

Verify: $Verify(M, \sigma, PK_A) = true/false$

Adversary wants to forge a signature that verifies.

For large M, easier to sign $h(M)$

$\sigma = sign(SK_A, h(M))$ ["hash & sign"]

Need CR (Alice gets Bob to sign x, then claims he signed x', if $h(x)=h(x')$)

Don't need OW.

Applications (cont)

6.857 Rivest
L2.5 2/9/09

(4) Commitments

- Alice has value x (e.g. auction bid)
- Alice computes $C(x)$ & submits it as her bid
  ↑ "commitment to x"
- $C(x)$ is her "sealed bid"
- When bidding is over, Alice should be able to "open" $C(x)$ to "reveal" x

Binding ☐ Alice should not be able to open $C(x)$ in more than one way!

Secrecy ☐ Auctioneer (or anyone) seeing $C(x)$ should not learn anything about x.

NM ☐ Given $C(x)$ shouldn't be possible to produce $C(x-1)$

Need: NM, CR, OW (really need more for secrecy...)

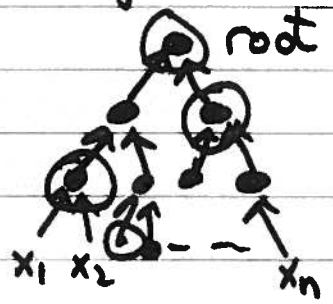
How: $C(x) = h(r \| x)$ $\quad r \in_R \{0,1\}^{256}$

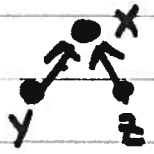
to open, reveal r & x

randomized

(5) To authenticate n objects (e.g. time-stamping)

Merkle tree:





value at x
$= h$(value at y || value at z)

root is authenticator for all n values (put in NYT)

Show leaf & ancestors & their siblings to prove leaf is in tree

need CR

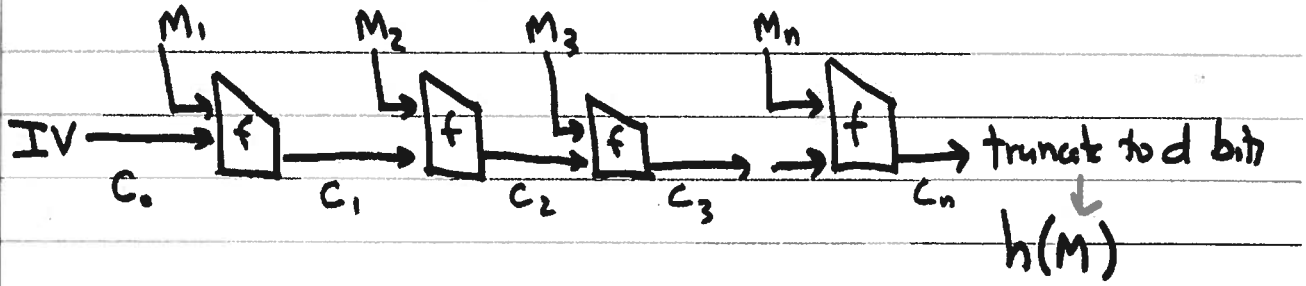
Construction ("Merkle-Damgard" style)

6.857 Rivest
L2.6 2/9/09

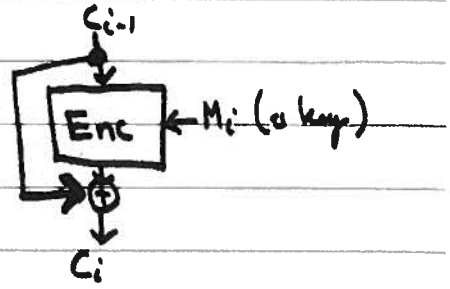
- Choose output size d (e.g. d = 160)
- Choose chaining variable size c (e.g. c = 160) (better if c ≥ 2d !)
- Choose block size b for message
- Design "compression fn" f (OW, CR, PR, NM, TCR, ...)

$$f : \{0,1\}^c \times \{0,1\}^b \to \{0,1\}^c$$

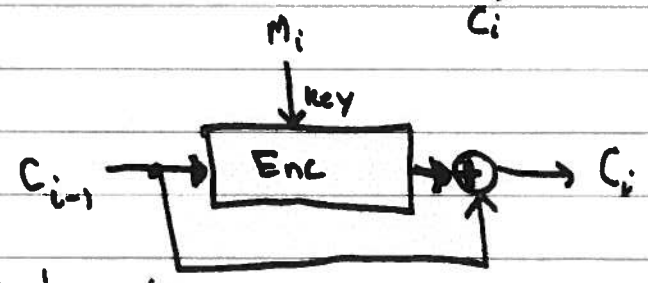
- Choose c-bit IV (initialization vector)
- Given message, add both 10*-bits & "length of m" [illegible] |m|
  so that m's new length is multiple of b bits
  now $M = M_1 M_2 \cdots M_n$ (n b-bit blocks)



- Like "mode of operation" for encryption algorithm.
- IV is arbitrary, but fixed.
- Thm: If f is CR, then so is h.
  Pf: Work backwards through chain from h-collision to find f-collision.
- Thm: Same for OW.
- Common pattern: M_i, C_{i-1} → f → C_i =



or



But AES etc. are hard to change keys in.