Hashing

Pradyot Prakash
Web and Coding Club
IIT Bombay

What do you mean?

> “
>
> A hash function is an efficient function mapping binary strings of arbitrary length to binary strings of fixed length (e.g. 128 bits), called the hash-value or digest.
>
> ”

Expectations

- Not unique to crypto; usually many-to-one
- For cryptography, a hash function must be one-way
- Given only a hash, should be computationally infeasible to find a preimage
- Collision:
  - situation where we have two different messages M and M^1 such that $H(M) = H(M^1)$
  - hash function should be collision free

Why do we need it?

- H(M) should be determinable for any M
- (Computational) infeasibility of finding M^1 s.t. $H(M) = H(M^1)$ => can use H(M) rather than M
- Usually digest is smaller than the original data => its use may be more efficient
- Digest as a unique fingerprint of the data

Examples

- MD2, MD4, MD5 (Rivest)
  - Produce 128-bit digests
  - Analysis has uncovered some weaknesses with these
- SHA-1 (Secure Hash Algorithm)
  - Produces 160-bit digests
- SHA-2 family (Secure Hash Algorithm)
  - SHA-224, SHA-256, SHA-384 and SHA-512
  - Digests of sizes 224, 256, 384 and 512 bits respectively

MD5 (Message Digest 5)

- MD5(I love ciphers) = 6663b9bbba0f617947419fd8f9b6f61
- MD5(I love cyphers) = a113fedee88a3335fb4243a680816936

- Convert to bits to see the hodgepodge!
- One actually wants this from a cryptographic hash
- MD5 Broken and not used these days => switch to SHA-256

Practical uses

- Sending passwords over insecure networks
- Storing passwords in databases
- Establishing integrity of data
    - Previous example!
    - Message authentication codes (MAC)
- Digital signatures
- HTTPS protocol
    - Huge importance these days

Breaking them

- Exhaustive search
    - Time confusing
- Rainbow tables
- Forge them
    - Birthday attack
    - Relies on small size of hash

Questions?