CSIT 495/595 - Introduction to Cryptography
Hash Functions

Bharath K. Samanthula
Department of Computer Science
Montclair State University

Outline

- Hash Functions: Motivation and Definition
- HMAC
- Generic Attacks
- Applications

Hash Functions: Motivation

- Hash Function: a function that takes inputs of long length and compress them into short, fixed-length outputs, called *digests* or *hash values*
- Key requirement: avoid collisions for two different inputs that map to the same digest
- Classic Example: hash tables that enable $O(1)$ lookup time

Collision Resistance

Let $H: M \to T$ be a hash function ($|M| >> |T|$)

A **collision** for H is a pair $m_0, m_1 \in M$ such that:

$$H(m_0) = H(m_1) \quad \text{and} \quad m_0 \neq m_1$$

A function H is **collision resistant** if for all (explicit) "eff" algs. A:

$$\mathbf{Adv_{CR}[A,H] = Pr[\ A \text{ outputs collision for } H]}$$

is "neg".

Example: SHA-256 (outputs 256 bits)

some slides are adopted from *Collision Resistance* by Dan Boneh

MACs from Collision Resistance (1)

Let $I = (S,V)$ be a MAC for short messages over (K,M,T) (e.g. AES)

Let $H: M^{big} \rightarrow M$

Def: $I^{big} = (S^{big}, V^{big})$ over (K, M^{big}, T) as:

$S^{big}(k,m) = S(k,H(m))$; $V^{big}(k,m,t) = V(k,H(m),t)$

<u>**Thm**</u>: If I is a secure MAC and H is collision resistant

then I^{big} is a secure MAC.

MACs from Collision Resistance (2)

$$S^{big}(k, m) = S(k, H(m)) \quad ; \quad V^{big}(k, m, t) = V(k, H(m), t)$$

Collision resistance is necessary for security:

Suppose adversary can find $m_0 \neq m_1$ s.t. $H(m_0) = H(m_1)$.

Then: S^{big} is insecure under a 1-chosen msg attack

step 1: adversary asks for $t \longleftarrow S(k, m_0)$

step 2: output (m_1, t) as forgery

Protecting File Integrity: Example

Software packages:

package name	package name		package name
F_1	F_2	…	F_n

read-only public space

$H(F_1)$ $H(F_2)$ $H(F_n)$

When user downloads package, can verify that contents are valid

H collision resistant $\Rightarrow$

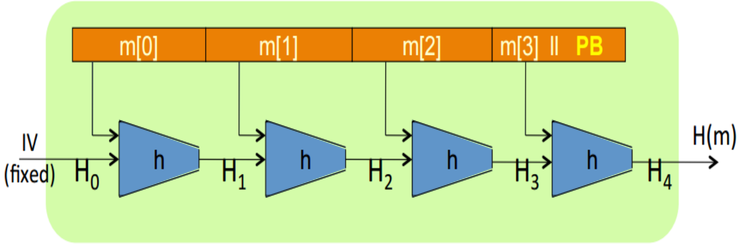
attacker cannot modify package without detection

no key needed (public verifiability), but requires read-only space

Domain Extension: The Merkle-Damgard Transform

- Given collision-resistant compression function with fixed-length inputs and output, domain extension is used to handle arbitrary-length inputs
- Merkle-Damgard Transform: a common approach used for extending a compression function to a full-fledged hash function, while still maintaining the collision property
- Extensively used in practice, for example, MD5 and SHA family of hash functions

The Merkel-Damgard Iterative Construct



Given $h: T \times X \longrightarrow T$ (compression function)

we obtain $H: X^{\leq L} \longrightarrow T$. H_i - chaining variables

PB: padding block

If no space for PB add another block

Dan Boneh

MAC construction from Hash Functions

Can we use $H(\cdot)$ to directly build a MAC?

$H: X^{\leq L} \longrightarrow T$ a C.R. Merkle-Damgard Hash Function

Attempt #1: $S(k, m) = H(k \parallel m)$

This MAC is insecure because:

- Given $H(k \parallel m)$ can compute $H(w \parallel k \parallel m \parallel PB)$ for any w.
- Given $H(k \parallel m)$ can compute $H(k \parallel m \parallel w)$ for any w.
- Given $H(k \parallel m)$ can compute $H(k \parallel m \parallel PB \parallel w)$ for any w.
- Anyone can compute $H(k \parallel m)$ for any m.

Standardized Method: Hash-MAC (HMAC)

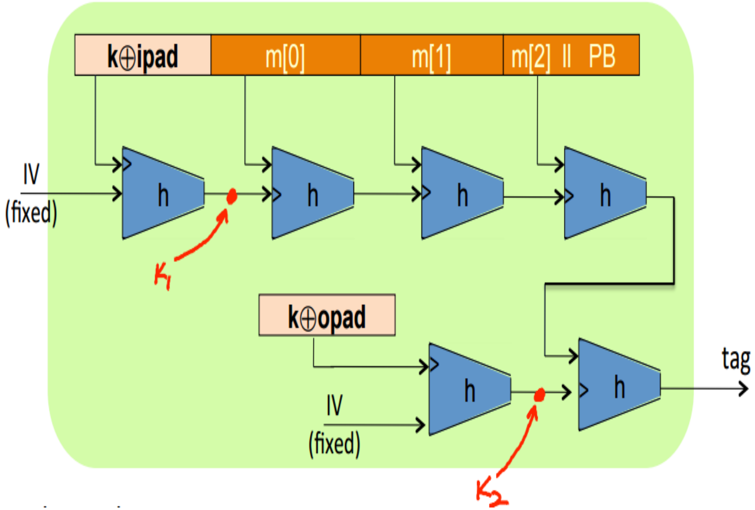
Most widely used MAC on the Internet.

H:	hash function.
	example: SHA-256 ; output is 256 bits

Building a MAC out of a hash function:

$$\text{HMAC:} \quad S(k, m) = H\big(\, k \oplus opad \,\|\, \mathbf{H(k \oplus ipad \,\|\, m)} \,\big)$$

HMAC: Graphical Interpretation



Hash Functions: Generic Attacks

Birthday Attack

- Given the length of the output is ℓ, a trivial collision-finding attack can be run in $O(2^\ell)$
- Can the attacker do better? YES
- Given q distinct inputs $x_1, \ldots, x_q$ and their hash values, what is the probability for the attacker to find a collision?
- This problem is analogous to Birthday Problem - Given q people in a room, what is the probability that two of the have the same Birthday

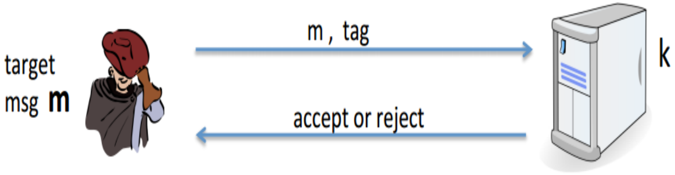
Hash Functions: Generic Attacks

Birthday Attack

- For the Birthday problem, when $q = \Theta(N^{1/2})$, the probability for two of them have the same birthday is greater than 1/2 (where $N = 365$)
- For hash functions, q should be at least $\Theta(2^{\ell/2})$ to achieve collision probability of roughly 1/2
- Example: to make finding hash collisions as difficult as an exhaustive search over 128-bit keys, the output length should be at least 256 bits

Hash Functions: Generic Attacks

MAC Timing attacks



Timing attack: to compute tag for target message m do:

Step 1: Query server with random tag

Step 2: Loop over all possible first bytes and query server.

stop when verification takes a little longer than in step 1

Step 3: repeat for all tag bytes until valid tag found

Hash Functions: Generic Attacks

MAC Timing attacks

- Possible Solution: code comparison operation to take same amount of time for every verification step
- Other solutions exist

Some Applications of Hash Functions

- Fingerprinting and Deduplication
  - Virus Fingerprinting
  - Deduplication
  - (Peer-to-peer) P2P file sharing
- Merkle Trees
- Password Hashing
- ... and many others

Summary

- Hash Functions: Motivation and Definition
- HMAC
- Generic Attacks
- Applications

Useful References

- Chapter 5, Introduction to Modern Cryptography by Jonathan Katz and Yehuda Lindell, 2nd Edition, CRC Press, 2015.
- http://research.microsoft.com/pubs/64588/hash_survey.pdf
- https://cseweb.ucsd.edu/~mihir/papers/hmac.html
- https://cseweb.ucsd.edu/~mihir/papers/hmac-cb.pdf