Outline

MACS: Continued
Collision Resistance
Administrative Details

- Course website: https://courses.grainger.illinois.edu/cs498ac3/fa2020/
- Has syllabus, instructor and TA info, office hours
- IMPORTANT: Join Piazza! piazza.com/illinois/fall2020/ececs498ac/home

I strongly encourage class participation. If you don't understand something in class, please interrupt me and ask questions. Please make abundant use of office hours.

Homework 1 is out. Due in a week.
Message Integrity: MACs
Recap: MACs

MAC = (S, V) is a pair of algorithms that satisfy:
1. **CORRECTNESS.** $V(k, m, \text{tag}) = 1$ when tag = $S(k, m)$
Recap: Security of MACs

- Attacker can demand tags \((t_1, t_2, \ldots, t_n)\) for messages \((m_1, m_2, \ldots, m_n)\).
- Message \(m_{i+1}\) can be chosen adaptively as a function of previous tags \((t_1, \ldots, t_i)\).
- Attacker wins if it outputs \((m', t')\) not in \\{(m_1, t_1), (m_2, t_2), \ldots, (m_n, t_n)\}\ such that \(V(k, m', t') = 1\).
Constructing MACs

• \( S(k,m) = \text{PRF}(k,m) \), \( V(k, m, t) \) = accept if and only if \( t = \text{PRF}(k,m) \)

• What about long messages, larger than the input size of PRF?

• For larger messages, we use:
  • CBC-MAC
  • HMAC
Encrypted CBC-MAC

\[ F(k, \cdot) F(k, \cdot) F(k, \cdot) F(k, \cdot) \]

Block cipher (can be weaker)

Encrypt

\[ F(k_1, \cdot) \]

tag
What happens if we remove encryption step?

Obtain tag for m, then get $t = F(k, m)$
Output forgery $(m', t')$ where $m' = (m || (t \xor m))$ and $t' = t$. 

$t = F(k, m[0])$ m = m[0] ask for tag. Get t.

Generate forgery on $(m[0] || t \oplus m[0])$. Output $(m[0] || t \oplus m[0]), t$
What if message length is not a multiple of block size?

\[ F(k, m[0]) \oplus F(k, m[1]) \oplus F(k, m[3]) \oplus F(k, m[4]) \]

- For message blocks of 128 bits: \( m[4] \) = 0...0
- Or
- For message blocks of 2 bits: \( m[4] \) = 00

\( F(k_{1\cdot}) \)
First idea?
Need invertible padding!

For security, padding must be invertible!

\[ m_0 \neq m_1 \implies m_0 \mid \mid \text{pad}(m_0) \neq m_1 \mid \mid \text{pad}(m_1) \]

Pad with “1000…00”. Add new dummy block if needed.

- The “1” indicates beginning of pad.

\[
\begin{array}{c|c}
\text{m}[0] & \text{m}[1] \\
\end{array}
\quad \longrightarrow \quad
\begin{array}{c|c}
\text{m}[0] & \text{m}[1] \ 100 \\
\end{array}
\]

\[
\begin{array}{c|c}
\text{m’}[0] & \text{m’}[1] \\
\end{array}
\quad \longrightarrow \quad
\begin{array}{c|c}
\text{m’}[0] & \text{m’}[1] \ 1000 \ 000 \\
\end{array}
\]
HMAC

• Hash MAC

• Apply a \textit{hash} function $H$ to your original message

• What properties should $H$ satisfy?
Collision-resistance

Let $H: M \rightarrow T$ be a hash function $(|M| \gg |T|)$

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

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

A function $H$ is **collision resistant** if for all PPT algs. $A$:

$$\text{Adv}_{CR}[A,H] = \Pr[A \text{ outputs collision for } H] = \text{negl}$$

Example: SHA-256 (outputs 256 bits)
MAC from Collision-resistant Hash Functions

Let \((S,V)\) be a MAC for short messages over \((K,M,T)\) (e.g. AES)
Let \(H: M^{\text{big}} \rightarrow M\)

Def: \((S^{\text{big}}, V^{\text{big}})\) over \((K, M^{\text{big}}, T)\) as:

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

Thm: If \((I)\) is a secure MAC and \(H\) is collision resistant
then \(I^{\text{big}}\) is a secure MAC.

Example: \(S(k,m) = \text{AES}_{2\text{-block-cbc}}(k, \text{SHA-256}(m))\) is a secure MAC.

(**Note**: HMAC)
MAC from Collision-resistant Hash Functions

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

Collision resistance is necessary for security:

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

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

\begin{align*}
\text{step 1: adversary asks for } & \quad t \leftarrow S(k, m_0) \\
\text{step 2: output } & \quad (m_1, t) \text{ as forgery}
\end{align*}

If \( H(m_0) = H(m_1) \) then \( \text{tag}_{m_0} = S(k, H(m_0)) = S(k, H(m_1)) = \text{tag}_{m_1} \).
Protecting File Integrity

Software packages:

package name $F_1$

package name $F_2$

package name $F_n$

read-only public space

$H(F_1)$ $H(F_2)$ $H(F_n)$
The birthday attack

Let \( H: M \rightarrow \{0,1\}^n \) be a hash function (\(|M| \gg 2^n\))

Generic alg. to find a collision \textbf{in time} \( O(2^{n/2}) \) hashes

\textbf{Space} \( O(2^{n/2}) \)
The birthday attack

Let \( H: M \rightarrow \{0,1\}^n \) be a hash function ( \(|M| \gg 2^n\) )

Generic alg. to find a collision in time \( O(2^{n/2}) \) hashes

Algorithm:
1. Choose \( 2^{n/2} \) random messages in \( M \): \( m_1, \ldots, m_{2^{n/2}} \) (distinct w.h.p)
2. For \( i = 1, \ldots, 2^{n/2} \) compute \( t_i = H(m_i) \in \{0,1\}^n \)
3. Look for a collision \( (t_i = t_j) \). If not found, go back to step 1.

How well will this work?
The birthday attack

Let \( r_1, \ldots, r_n \in \{1, \ldots, B\} \) be indep. identically distributed integers.

**Thm:** when \( n = 1.2 \times B^{1/2} \) then \( \Pr \left[ \exists i \neq j: r_i = r_j \right] \geq \frac{1}{2} \)

**Proof:** (for uniform indep. \( r_1, \ldots, r_n \))

\[
\Pr \left[ \exists i \neq j: r_i = r_j \right] = 1 - \Pr \left[ \forall i \neq j, r_i \neq r_j \right]
\]
\[
= 1 - \frac{(B-1)}{B} \times \frac{(B-2)}{B} \times \cdots \times \frac{(B-n+1)}{B}
\]

True because
\[
\left( 1 - i \cdot \frac{1}{B} \leq e^{-i/B} \right)
\]
\[
\geq 1 - \prod_{i=1}^{n-1} e^{-i/B} \geq 1 - e^{-n^2/2B} = 0.53 > \frac{1}{2}.
\]
The birthday attack

Take any $m \rightarrow 2^{n/2}$

$H: M \rightarrow \{0,1\}^n$. Collision finding algorithm:

1. Choose $2^{n/2}$ random elements in $M$: $m_1, \ldots, m_{2^{n/2}}$
2. For $i = 1, \ldots, 2^{n/2}$ compute $t_i = H(m_i) \in \{0,1\}^n$
3. Look for a collision ($t_i = t_j$). If not found, got back to step 1.

Expected number of iteration $\approx 2$ (by previous Thm)

Running time: $O(2^{n/2})$ (space $O(2^{n/2})$)

Example: SHA-1 has output size 160 bits. Birthday attack: $2^{80}$. Best attack: $2^{51}$

(Do not use SHA1. Use SHA256 instead!)
Merkle-Damgard

Given \( h : T \times X \rightarrow T \) (compression function)
we obtain \( H : X^{\leq L} \rightarrow T \)

\( H_i \) - chaining variables
PB: padding block

-- If no space for PB add another block
Merkle-Damgird

Theorem: If \( h \) is collision resistant, then so is \( H \).

Proof: collision on \( H \) \( \Rightarrow \) collision on \( h \)

Suppose \( H(M) = H(M') \). We build collision for \( h \).

\[
\begin{align*}
  h(H_t, M_t \parallel PB) &= H_{t+1} = H'_{r+1} = h(H'_r, M'_r \parallel PB') \\
  \Rightarrow h(H_t, M_t \parallel PB) &= h(H'_r, M'_r \parallel PB') \\
  \Rightarrow H_t &= H'_r \quad \text{and} \quad M_t \parallel PB = M'_r \parallel PB'
\end{align*}
\]

BASE CASE:

IV = H_0, H_1, \ldots, H_t, H_{t+1} = H(M)

IV = H_0', H_1', \ldots, H'_r, H'_{r+1} = H(M')

\[
\begin{align*}
  H_{t+1} &= H'_{r+1} \\
  h(H_t, M_t \parallel PB) &= h(H'_r, M'_r \parallel PB') \\
  \Rightarrow h(H_t, M_t \parallel PB) &= h(H'_r, M'_r \parallel PB') \\
  \Rightarrow H_t &= H'_r \quad \text{and} \quad M_t \parallel PB = M'_r \parallel PB'
\end{align*}
\]
Merkle-Damgard

**Theorem:** If h is collision resistant, then so is H.

**Proof:** collision on H $\Rightarrow$ collision on h

Suppose $H(M) = H(M')$. We build collision for h.

$h(H_t, M_t || PB) = H_{t+1} = H'_{r+1} = h(H'_r, M'_r || PB')$

Otherwise suppose $H_t = H'_r$ and $M_t = M'_r$ and $PB = PB'$

Then: $h(H_{t-1}, M_{t-1}) = H_t = H'_t = h(H'_{t-1}, M'_{t-1})$
Merkle-Damgard

Thm: \( h \) collision resistant \( \Rightarrow \) \( H \) collision resistant

Goal: construct compression function \( h: T \times X \rightarrow T \)
Compression function from block cipher

**E**: $K \times \{0,1\}^n \rightarrow \{0,1\}^n$  a block cipher.

The **Davies-Meyer** compression function:  
$$h(H, m) = E(m, H) \oplus H$$

**Thm**: Suppose $E$ is an ideal cipher (collection of $|K|$ random perms.). Finding a collision $h(H,m)=h(H',m')$ takes $O(2^{n/2})$ evaluations of $(E,D)$. 
Goal: find $H, m, H', m'$ s.t. $h(H, m) = h(H', m')$

What about a simpler construction?

Suppose we define $h(H, m) = E(m, H)$ (without XOR step)

Then the resulting $h(., .)$ is not collision resistant:

- to find a collision $(H, m)$ and $(H', m')$
- choose random $(H, m, m')$ and construct $H'$ as follows:

\[ H' = D(m', E(m, H)) \]

\[ E(m', H') = E(m, H) \]

\[ h(H', m') = h(H, m) \]
Standardized Method: HMAC

Most widely used MAC on the Internet.

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

Can we build a MAC directly out of a hash function?

HMAC: $S(k, m) = H(k \oplus opad \ || \ H(k \oplus ipad \ || \ m))$
The HMAC Construction

**Fixed domain**: $h$ : bounded input length

**Variable domain**: $H$ : unbounded input length

The HMAC Construction

$\text{MAC}(k, m)$

Directly build $\text{MAC}$ from $h$ (fixed domain)

Previously: $\begin{align*}
\text{MAC}(k, m) &= h \oplus \text{iv} \\
\text{MAC}(k, m) &= h \oplus \text{iv} \\
\text{MAC}(k, m) &= h \oplus \text{iv}
\end{align*}$ from $h$ (fixed domain)
HMAC: Features

Built from a black-box implementation of SHA-256.

HMAC is assumed to be a secure PRF
- Can be proven under certain PRF assumptions about $h(.,.)$
- Can even be truncated, to say the first 80 bits of output

This is used in TLS
Summary

• Message Authentication Codes (MACs)

• Hash Functions

• HMAC