COS433/Math 473: Cryptography

Mark Zhandry
Princeton University
Fall 2020
Announcements/Reminders

HW3 due on Oct 20

HW4 will be released today, due Oct 27
Previously on COS 433...
Collision Resistant Hashing
Collision Resistant Hashing

Syntax:
- Key space $K$ (typically $\{0,1\}^\lambda$)
- Domain $D$ (typically $\{0,1\}^l$ or $\{0,1\}^*$)
- Range $R$ (typically $\{0,1\}^n$)
- Function $H: K \times D \rightarrow R$

Correctness: $n \ll l$
Merkle-Damgad

IV (fixed)
Constructing $h$

Common approach: use block cipher

Davies-Meyer

\[ h(x, y) = x \oplus F(y, x) \]
Constructing $h$

Some other possibilities are insecure

$$h(x,y) = F(y,x)$$

$$h(x,y) = F(y,x) \oplus y$$
Constructing $h$

Why do we think Davies-Meyer is reasonable?
• Cannot prove collision resistance just based on $F$ being a secure PRP

Instead, can argue security in “ideal cipher” model
• Pretend $F$, for each key $y$, is a uniform random permutation
Today

• Collision resistance cont.
• Random Oracle Model
• Commitments
We said 128 bit security is usually enough.

However, 128-bit blocks insufficient for compression function. Why?
Birthday Attack

If the range of a hash function is $R$, a collision can be found in time $T = O(|R|^{\frac{1}{2}})$

Attack:
- Given key $k$ for $H$
- For $i=1,\ldots, T$
  - Choose random $x_i$ in $D$
  - Let $t_i \leftarrow H(k,x_i)$
  - Store pair $(x_i, t_i)$
- Look for collision amongst stored pairs
Birthday Attack

Analysis:

Expected number of collisions

\[= \text{Number of pairs} \times \text{Prob each pair is collision}\]
\[\approx (T \text{ choose } 2) \times \frac{1}{|R|}\]

By setting \(T = O(|R|^{\frac{1}{2}})\), expected number of collisions found is at least 1
\[\Rightarrow \text{likely to find a collision}\]
Birthday Attack

Space?

Possible to reduce memory requirements to $O(1)$
Sponge Construction

Absorbing

Squeezing
Sponge Construction

Advantages:
• Round function $\mathbf{f}$ can be public invertible function (i.e. unkeyed SPN network)
• Easily get different input/output lengths
SHA-1,2,3

SHA-1,2 are hash functions built as follows:
• Build block cipher (SHACAL-1, SHACAL-2)
• Convert into compression function using Davies-Meyer
• Extend to arbitrary lengths using Merkle-Damgard

SHA-3 is based on sponge construction
SHA-1,2,3

SHA-1 (1995) is no longer considered secure
• 160-bit outputs, so collisions in time $2^{80}$
• 2017: using some improvements over birthday attack, able to find a collision

SHA-2 (2001)
• Longer output lengths (256-bit, 512-bit)
• Few theoretical weaknesses known

SHA-3 (2015)
• NIST wanted hash function built on different principles
Basing MACs on Hash Functions

Idea: $\text{MAC}(k,m) = H(k \ || \ m)$

Thought: if $H$ is a “good” hash function and $k$ is random, should be hard to predict $H(k \ || \ m)$ without knowing $k$

Unfortunately, cannot prove secure based on just collision resistance of $H$
Random Oracle Model

Pretend $\mathcal{H}$ is a truly random function

Everyone can query $\mathcal{H}$ on inputs of their choice
- Any protocol using $\mathcal{H}$
- The adversary (since he knows the key)

A query to $\mathcal{H}$ has a time cost of 1

Intuitively captures adversaries that simple query $\mathcal{H}$, but don’t take advantage of any structure
MAC in ROM

$$\text{MAC}^H(k,m) = H(k\|m)$$
$$\text{Ver}^H(k,m,\sigma) = (H(k\|m) == \sigma)$$

**Theorem:** $H(k \ || \ m)$ is a CMA-secure MAC in the random oracle model
Meaning

Output 1 iff:

• \( m^* \notin \{m_1, \ldots \} \)
• \( \text{Ver}^H(k, m^*, \sigma^*) = 1 \)
Meaning

Output 1 iff:

- $m^* \notin \{m_1, \ldots\}$
- $H(k||m^*) = \sigma^*$

$k \leftarrow K$

$\sigma_i \leftarrow H(k||m_i)$

$H \leftarrow \text{Funcs}$

$m_i \rightarrow \sigma_i$

$(m^*, \sigma^*)$

$H(x)$

$k||m_i \rightarrow H(x)$

$x \times H(x)$

$m_i \leftarrow H$
The ROM

A random oracle is a good
• PRF: \( F(k, x) = H(k || x) \)

• PRG (assuming \( H \) is expanding):
  • Given a random \( x \), \( H(x) \) is pseudorandom since adv is unlikely to query \( H \) on \( x \)

• CRHF:
  • Given poly-many queries, unlikely for find two that map to same output
The ROM

The ROM is very different from security properties like collision resistant

What does it mean that “Sha-1 behaves like a random oracle”?  
• No satisfactory definition

Therefore, a ROM proof is a heuristic argument for security  
• If insecure, adversary must be taking advantage of structural weaknesses in $H$
When the ROM Fails

$$\text{MAC}^H(k,m) = H(k\|m)$$
$$\text{Ver}^H(k,m,\sigma) = (H(k\|m) == \sigma)$$

Instantiate with Merkle-Damgard (variable length)?
When the ROM Fails

ROM does not apply to regular Merkle-Damgard
• Even if $h$ is an ideal hash function

Takeaway: be careful about using ROM for non-“monolithic” hash functions
• Though still possible to pad MD in a way that makes it an ideal hash function if $h$ is ideal
HMAC
HMAC

Hash

MAC

k

m1

m2

m3

h

h

h

h

h

h

h

h

k

opad

IV

(iv)

Append padding

σ

ipad

⊕

⊕

IV

(fixed)

Hash

MAC
HMAC

ipad, opad?
• Two different (but related) keys for hash and MAC
• ipad makes hash a “secret key” hash function
• Even if not collision resistant, maybe still impossible to find collisions when hash key is secret
• Turned out to be useful after collisions found in MD5
Committments
Anagrams and Astronomy

Galileo and the Rings of Saturn
• 1610: Galileo observed the rings of Saturn, but mistook them for two moons

• Galileo wanted extra time for verification, but not to get scooped

• Circulates anagram
  SMAISMRMILMEPOETALEUMIBUNENUUGTTAUIRAS

• When ready, tell everyone the solution:
  altissimum planetam tergeminum observavi
  ( “I have observed the highest planet tri-form” )
Anagrams and Astronomy

Enter Huygens

• 1656: Realizes Galileo actually saw rings
• Circulates

   AAAAAAA CCCCC D EEEEE G H IIIIIII LLLL MM
     NNNNNNNNNN OOOO PP Q RR S TTTTT UUUUU

• Solution:

   annulo cingitur, tenui, plano, nusquam cohaerente, ad eclipticam inclinato

   ( “it is surrounded by a thin flat ring, nowhere touching, and inclined to the ecliptic” )
Commitment Scheme

Different than encryption

• No need for a decryption procedure
• No secret key

• But still need secrecy ("hiding")
• Should only be one possible opening ("binding")
• (Sometimes other properties needed as well)
Anagrams are Bad Commitments

If too short (e.g. one, two, three words), possible to reconstruct answer
• Even easier if have reasonable guess for answer

If too long, multiple possible solutions
• Kepler tries to solve Galileo’s anagram as

    salue umbistineum geminatum martia proles

    (hail, twin companionship, children of Mars)
(Non-interactive)
Commitment Syntax

Message space $\mathcal{M}$
Ciphertext Space $\mathcal{C}$
(suppressing security parameter)

$\text{Com}(m; r)$: outputs a commitment $c$ to $m$
• Why have $r$?
Commitments with Setup

Message space $\mathcal{M}$
Ciphertext Space $\mathcal{C}$
(suppressing security parameter)

**Setup()**: Outputs a key $k$

**Com($k$, $m$; $r$)**: outputs a commitment $c$ to $m$
Using Commitments

Commit Stage

Reveal Stage

Commit

Reveal

Check that \( c = \text{Com}(m;r) \)
Using Commitments (with setup)

1. **Commit Stage**
   - \( k \leftarrow \text{Setup}() \)
   - \( r \leftarrow R \)
   - \( c \leftarrow \text{Com}(k, m; r) \)

2. **Reveal Stage**
   - \( m, r \)
   - Check that \( c = \text{Com}(k, m; r) \)
Security Properties

Hiding: \( c \) should hide \( m \)

- Perfect hiding: for any \( m_0, m_1, \)

\[
\text{Com}(m_0) \overset{\text{d}}{=} \text{Com}(m_1)
\]

- Statistical hiding: for any \( m_0, m_1, \)

\[
\Delta(\text{Com}(m_0), \text{Com}(m_1)) < \text{negl}
\]

- Computational hiding:

\[
m_0, m_1 \rightarrow c \leftarrow \text{Com}(m_b)
\]
Security Properties (with Setup)

Hiding: \( \mathbf{c} \) should hide \( \mathbf{m} \)

- Perfect hiding: for any \( \mathbf{m}_0, \mathbf{m}_1 \),
  \[
  k, \text{Com}(k, \mathbf{m}_0) \overset{d}{=} k, \text{Com}(k, \mathbf{m}_1)
  \]

- Statistical hiding: for any \( \mathbf{m}_0, \mathbf{m}_1 \),
  \[
  \Delta( [k, \text{Com}(k, \mathbf{m}_0)], [k, \text{Com}(k, \mathbf{m}_1)] ) < \text{negl}
  \]

- Computational hiding:

\[
\begin{align*}
&\text{c} \leftarrow \text{Com}(k, m_b) \\
\end{align*}
\]
Security Properties

Binding: Impossible to change committed value

• Perfect binding: For any $c$, $\exists$ at most a single $m$ such that $c = \text{Com}(m;r)$ for some $r$

• Computational binding: no efficient adversary can find $(m_0,r_0),(m_1,r_1)$ such that:
  $\text{Com}(m_0;r_0) = \text{Com}(m_1;r_1)$
  $m_0 \neq m_1$
Security Properties (with Setup)

Binding: Impossible to change committed value
  • Perfect binding: For any $k,c$, $\exists$ at most a single $m$ such that $c = \text{Com}(k,m;r)$ for some $r$

  • Statistical binding: except with negligible prob over $k$, for any $c$, $\exists$ at most a single $m$ such that $c = \text{Com}(k,m;r)$ for some $r$

  • Computational binding: no PPT adversary, given $k \leftarrow \text{Setup}()$, can find $(m_0,r_0),(m_1,r_1)$ such that $\text{Com}(k,m_0;r_0) = \text{Com}(k,m_1;r_1)$ and $m_0 \neq m_1$
Who Runs \textbf{Setup()}

Alice?
\begin{itemize}
\item Must ensure that Alice cannot devise $k$ for which she can break binding
\end{itemize}

Bob?
\begin{itemize}
\item Must ensure Bob cannot devise $k$ for which he can break hiding
\end{itemize}

Solution: Trusted third party (TTP)
Announcements/Reminders

HW3 due on Oct 20

HW4 will be released today, due Oct 27