COS433/Math 473: Cryptography

Mark Zhandry
Princeton University
Spring 2017
Previously...

Encryption
  +
Authentication
  =
Authenticated Encryption

Collision Resistance
Security Notions for Hashing

Collision resistance as a game:

\[ k \xrightarrow{\, x_0, x_1 \,} 1 \text{ iff } x_0 \neq x_1 \]
\[ H(k, x_0) = H(k, x_1) \]
Security Notions for Hashing

2nd Preimage Resistance (or target collision resistance):

\[ H(k, x_0) = H(k, x_1) \iff x_0 \neq x_1 \]
Security Notions for Hashing

2-Universal:

\[ x_0 \quad x_1 \quad k \]

1 iff \( x_0 \neq x_1 \)

\[ H(k, x_0) = H(k, x_1) \]
Security Notions for Hashing

One-wayness (or pre-image resistance):

\[ k, y \quad \leftrightarrow \quad x_1 \quad \leftrightarrow \quad x_0 \leftarrow D \]

\[ y \leftarrow H(k, x_0) \]

1 iff \( y = H(k, x_1) \)
Implications

Collision Resistance

\[ \rightarrow \]

2\textsuperscript{nd} Pre-image Resistance

\[ \rightarrow \]

One-wayness
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
Today

Commitment Schemes

Start: number-theoretic constructions of symmetric key primitives
Remember Galileo

• 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

\[ \text{SMAISRMILMEOETALEUMIBUNENUGTTAUIRAS} \]

• When ready, tell everyone the solution:

\[ \text{altissimum planetam tergeminum observavi} \]

(“I have observed the highest planet tri-form”)
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
(Non-interactive)
Commitment Syntax

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

$\text{Com}(m; r)$: outputs a commitment $c$ to $m$
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

\[ m \leftarrow \mathcal{R} \]

\[ c \leftarrow \text{Com}(m;r) \]

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

Commit Stage

- \( k \leftarrow \text{Setup}() \)
- \( r \leftarrow R \)
- \( c \leftarrow \text{Com}(k, m; r) \)

Reveal Stage

- \( m \)
- \( m, r \)

Check that \( c = \text{Com}(k, m; r) \)
Security Properties

Hiding: \( c \) should hide \( m \)
- Perfect hiding: for any \( m_0, m_1 \),
  \[
  \text{Com}(m_0) \equiv \text{Com}(m_1)
  \]
- Statistical hiding: for any \( m_0, m_1 \),
  \[
  \Delta( \text{Com}(m_0), \text{Com}(m_1) ) < \text{negl}
  \]
- Computational hiding:
  
\[
\text{b'} \overset{\text{m_0, m_1}}{\longrightarrow} c \overset{c}{\longleftarrow} \text{Com}(m_b)
\]
Security Properties (with Setup)

Hiding: \( c \) should hide \( m \)
- Perfect hiding: for any \( m_0, m_1, \)
  \[
k,\text{Com}(k,m_0) \overset{d}{=} k,\text{Com}(k,m_1)\]
- Statistical hiding: for any \( m_0, m_1, \)
  \[
  \Delta( [k,\text{Com}(k,m_0)], [k,\text{Com}(k,m_1)] ) < \text{negl}
  \]
- Computational hiding:

\[ b' \]

\[ k \]

\[ m_0, m_1 \]

\[ c \]

\[ c \leftarrow \text{Com}(k,m_b) \]
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 PPT adversary can find $(m_0, r_0), (m_1, r_1)$ such that $\text{Com}(m_0; r_0) = \text{Com}(m_1; r_1)$
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 $\text{Com}(k, m_0; r_0) = \text{Com}(k, m_1; r_1)$
Who Runs Setup()

Trusted third party (TTP)?

Alice?
• Must ensure that Alice cannot devise $k$ for which she can break binding
• If binding holds, can actually devise scheme Com’ without setup

Bob?
• Must ensure Bob cannot devise $k$ for which he can break hiding
Honest-but Curious vs Malicious

Honest-but Curious receiver: runs \textit{Setup} as expected, tries to learn committed message

Malicious receiver: can generate $k$ however he wants, tries to learn message
Anagrams as Commitment Schemes

\( \text{Com}(m) = \text{sort characters of message} \)

Problems?
• Not hiding: “Jupiter has four moons” vs “Jupiter has five moons”

• Not binding: Kepler recodes Galileo’s anagram to conclude Mars has two moons
Anagrams as Commitment Schemes

\( \text{Com}(m) = \) add random superfluous text, then sort characters of message

Might still not be hiding

• Need to guarantee, for example that expected number of each letter in output is independent of input string

Still not binding...
Other Bad Commitments

\( \text{Com}(m) = m \)
- Has binding, but no hiding

\( \text{Com}(m;r) = m \oplus r \)
- Has hiding, but no binding
Can a commitment scheme be both statistically hiding and statistically binding?
A Simple Commitment Scheme

Let $H$ be a hash function

$\text{Com}(m;r) = H(m \ || \ r)$

Binding?

Hiding?
Theorem: $\text{Com}(m;r) = H(m||r)$ has:

- Perfect binding assuming $H$ is injective
- Computational binding assuming $H$ is collision resistance (implied by RO)
- Computational hiding in the Random Oracle Model
Hiding

\[ x \xrightarrow{H(X)} H \]

\[ m_0, m_1 \]

\[ c \]

\[ r \xleftarrow{R} c \xleftarrow{H(m_b \|l r)} \]

\[ b' \]
Proof of Hiding

Suppose never queries $H$ on $m_b||r$

Then all query answers and commitment $c$ seen by are independent uniform strings
• has no chance of determining $b$

Probability queries on $m_b||r$?
• At most $q/|R|$ = negligible
“Standard Model” Commitments?

Random oracle model proof is heuristic argument for security

Can we prove it under assumptions such as collision resistance, etc?
Single Bit to Many Bit

Let \((\text{Setup}, \text{Com})\) be a commitment scheme for single bit messages

Let \(\text{Com}'(k, m; r) = (\text{Com}(k, m_1; r_1), \ldots, \text{Com}(k, m_t; r_t))\)

- \(m = (m_1, \ldots, m_t), \ m_i \in \{0, 1\}\)
- \(r = (r_1, \ldots, r_t), \ r_i \) are randomness for \(\text{Com}\)
Theorem: If \((\text{Setup, Com})\) is perfectly/statistically/computationally binding, then so is \((\text{Setup, Com'})\)

Theorem: If \((\text{Setup, Com})\) is perfectly/statistically/computationally, semi-honest/malicious hiding, then so is \((\text{Setup, Com'})\)
Binding

Suppose a break (say comp) binding of \(\text{Com}'\)

Given \(k\), produces \((m_1^0, r_1^0), \ldots, (m_t^0, r_t^0), (m_1^1, r_1^1), \ldots, (m_t^1, r_t^1)\) such that

- \((m_1^0, \ldots, m_t^0) \neq (m_1^1, \ldots, m_t^1)\)
- \(\text{Com}(k, m_i^0; r_i^0) = \text{Com}(k, m_i^1; r_i^1)\) for all \(i\)

Therefore, \(\exists i\) such that \(m_i^0 \neq m_i^1\) but

\[\text{Com}(k, m_i^0; r_i^0) = \text{Com}(k, m_i^1; r_i^1)\]

\[\Rightarrow\] Break binding of \(\text{Com}\)
Suppose breaks (say, computational malicious) hiding
Hiding

Proof by Hybrids

Hybrid $j$:
• For each $i \leq j$, $c_i = \text{Com}(k, m^1_i, r_i)$
• For each $i > j$, $c_i = \text{Com}(k, m^0_i, r_i)$

Hybrid $0$: commit to $\{m^0_i\}_i$
Hybrid $1$: commit to $\{m^1_i\}_i$

$\exists j$ such that distinguishes Hyb $j-1$ from Hyb $j$
$\Rightarrow$ break hiding of $\text{Com}$
Single Bit to Many Bit

Let \((\text{Setup,Com})\) be a commitment scheme for single bit messages.

Let \(\text{Com}'(k,m; r) = (\text{Com}(k,m_1;r_1),\ldots,\text{Com}(k,m_t;r_t))\)
- \(m = (m_1,\ldots,m_t)\), \(m_i \in \{0,1\}\)
- \(r = (r_1,\ldots,r_t)\), \(r_i\) are randomness for \(\text{Com}\)

Therefore, suffices to focus on commitments for single bit messages.
Statistically Hiding Commitments

Let $H$ be a collision resistant hash function with domain $X = \{0,1\} \times R$ and range $Z$.

Setup(): $k \leftarrow K$, output $k$
Com( $k$, $m$; $r$) = $H(k, (m,r))$

Binding?

Hiding?
Statistically Hiding Commitments

Let $F$ be a pairwise independent function family with domain $X=\{0,1\} \times \mathbb{R}$ and range $Y$

Let $H$ be a collision resistant hash function with domain $Y$ and range $Z$

**Setup():** $f \leftarrow F$, $k \leftarrow K$, output $(f,k)$

$Com((f,k), m; r) = H(k, f(m,r))$
Theorem: If $|Y|/|X|$ is “sufficiently large” and $H$ is collision resistant, then $(\text{Setup,Com})$ has computational binding

Theorem: If $|X|$ is “sufficiently large”, then $(\text{Setup,Com})$ has statistical hiding
Theorem: If $|Y|/|X|$ is “sufficiently large” and H is collision resistant, then $(\text{Setup,Com})$ has computational binding

Proof:
• Suppose $|Y| > |X|^2 \times 2^\lambda$
• For any $x_0 \neq x_1$, $\Pr[f(x_0) = f(x_1)] < 1/(|X|^2 \times 2^\lambda)$
• Union bound:
  \[ \Pr[\exists \ x_0 \neq x_1 \text{ s.t. } f(x_0) = f(x_1)] < 1/2^\lambda \]
Theorem: If $|X|$ is “sufficiently large”, then $(\text{Setup}, \text{Com})$ has statistical hiding

Goal: show $(f, k, H(k, f(0,r)))$ is statistically close to $(f, k, H(k, f(1,r)))$
Min-entropy

**Definition:** Given a distribution $\mathcal{D}$ over a set $\mathbb{X}$, the min-entropy of $\mathcal{D}$, denoted $H_\infty(\mathcal{D})$, is

$$- \min_x \log_2( \Pr[x \leftarrow \mathcal{D}] )$$

**Examples:**
- $H_\infty( \{0,1\}^n ) = n$
- $H_\infty( \text{random n bit string with parity 0} ) = ?$
- $H_\infty( \text{random } i > 0 \text{ where } \Pr[i] = 2^{-i} ) = ?$
Leftover Hash Lemma

**Lemma:** Let $\mathcal{D}$ be a distribution on $\mathbf{X}$, and $\mathcal{F}$ a family of pairwise independent functions from $\mathbf{X}$ to $\mathbf{Y}$. Then

$$\Delta( (f, f(D)) , (f, R) ) \leq \varepsilon$$

where

- $f \leftarrow \mathcal{F}$
- $R \leftarrow \mathbf{Y}$
- $\log |\mathbf{Y}| \leq H_\infty(\mathcal{D}) + 2 \log \varepsilon$
Lemma: Let $\mathbf{D}$ be a distribution on $\mathbf{X}$, and $\mathbf{F}$ a family of pairwise independent functions from $\mathbf{X}$ to $\mathbf{Y}$, and $h$ be any function from $\mathbf{Y}$ to $\mathbf{Z}$. Then

$$\Delta((f, h(f(D)))) , (f, h(R)) \leq \varepsilon$$

where

- $f \leftarrow \mathbf{F}$
- $R \leftarrow \mathbf{Y}$
- $\log |Z| \leq H_\infty(D) + 2 \log \varepsilon - 1$
Goal: show \((f, k, H(k, f(0,r)))\) is statistically close to \((f, k, H(k, f(1,r)))\)

Suppose \(|Z| = 2^{\lambda}\)

\((0,r)\) has min-entropy \(\log |R|\)

Set \(R = \{0,1\}^{3\lambda}, \varepsilon = 2 \times 2^{-\lambda}\)

Then \(\log |Z| \leq H_\infty(D) + 2 \log \varepsilon - 1\)
Theorem: If $|X|$ is “sufficiently large”, then $(\text{Setup,Com})$ has statistical hiding

For any $k$,
$$\Delta( (f, H(k, f(0,r))) , (f, H(k, U)) ) \leq \varepsilon$$

Thus
$$\Delta( (f, H(k, f(0,r))) , (f, H(k, f(1,r))) ) \leq 2\varepsilon$$

Therefore
$$\Delta( (f, k, H(k, f(0,r))) , (f, k, H(k, f(1,r))) ) \leq 2\varepsilon$$
Statistically Binding Commitments

Let $G$ be a PRG with domain $\{0,1\}^\lambda$, range $\{0,1\}^{3\lambda}$

**Setup():** choose and output a random $3\lambda$-bit string $k$

**Com(b; r):** If $b=0$, output $G(r)$, if $b=1$, output $G(r) \oplus k$
Theorem: \((\text{Setup,Com})\) is statistically binding

Theorem: If \(G\) is a secure PRG, then \((\text{Setup,Com})\) has computational hiding
Theorem: If $G$ is a secure PRG, then $(\text{Setup}, \text{Com})$ has computational hiding

Hybrids:

- **Hyb 0:** $S = \text{Com}(0; r) = G(r)$ where $r \leftarrow \{0,1\}^\lambda$
- **Hyb 1:** $S \leftarrow \{0,1\}^{3\lambda}$
- **Hyb 2:** $S = S' \oplus k$, where $S' \leftarrow \{0,1\}^{3\lambda}$
- **Hyb 3:** $S = \text{Com}(1; r) = G(r) \oplus k$ where $r \leftarrow \{0,1\}^\lambda$
Theorem: \((\text{Setup,Com})\) is statistically binding

Proof:

For any \(r,r',\) \(\Pr[G(r) = G(r') \oplus k] = 2^{-3\lambda}\)

By union bound:

\[
\Pr[\exists r,r' \text{ such that } \text{Com}(k,0) = \text{Com}(k,1)] \\
= \Pr[\exists r,r' \text{ such that } G(r) = G(r') \oplus k] < 2^{-\lambda}
\]
Number-theoretic Constructions
So Far...

Two ways to construct cryptographic schemes:
• Use others as building blocks
  • PRGs $\rightarrow$ Stream ciphers
  • PRFs $\rightarrow$ PRPs
  • PRFs/PRPs $\rightarrow$ CPA-secure Encryption
  • ...

• From scratch
  • RC4, DES, AES, etc

In either case, ultimately scheme or some building block built from scratch
Cryptographic Assumptions

Security of schemes built from scratch relies solely on our inability to break them
• No security proof
• Perhaps arguments for security

We gain confidence in security over time if we see that nobody can break scheme
Number-theory Constructions

Goal: base security on hard problems of interest to mathematicians
• Wider set of people trying to solve problem
• Longer history
Integer Factorization

Given an integer $N$, factor $N$ into its prime factors

Studied for centuries, presumed computationally difficult
• Grade school algorithm: $O(N^{1/2})$
• Much better algorithms: 
  $$\exp(C (\log n)^{1/3} (\log \log n)^{2/3})$$
• However, all require super-polynomial time
Factoring Assumption: Let $p, q$ be two random $\lambda$-bit primes, and $N = pq$. Then any PPT algorithm, given $N$, has at best a negligible probability of recovering $p$ and $q$. 


One-way Functions From Factoring

\( P_\lambda = \{\lambda\text{-bit primes}\} \)

\( F: P_\lambda^2 \rightarrow \{0,1\}^{2\lambda} \)
\( F(p,q) = p \times q \)

**Trivial Theorem:** If factoring assumption holds, then \( F \) is one-way
Sampling Random Primes

**Prime Number Theorem:** A random $\lambda$-bit number is prime with probability $\approx 1/\lambda$

**Primality Testing:** It is possible in polynomial time to decide if an integer is prime

Fermat Primality Test (randomized, some false positives):
- Choose a random integer $a \in \{0,\ldots,N-1\}$
- Test if $a^N = a \mod N$
- Repeat many times
Discrete Log

Let $p$ be a large integer (maybe prime)

Given $g \in \mathbb{Z}_p^*$, $a \in \mathbb{Z}$, easy to compute $g^a \mod p$

However, no known efficient ways to recover $a$ from $g$ and $g^a \mod p$
Discrete Log Assumption: Let $p$ be a $\lambda$-bit integer.

Then the function $(g, a) \rightarrow (g, g^a \mod p)$ is one-way, where

- $g \in \mathbb{Z}_p^*$
- $a \in \mathbb{Z}_{\Phi(p)}$
Generalizing Discrete Log

Let $G_\lambda$ be multiplicative groups of size $n_\lambda$

**Definition:** The discrete log assumption holds on \{G_\lambda\} if the function $F:G_\lambda \times \{0, \ldots, n_\lambda - 1\} \rightarrow G_\lambda^2$ is one-way, where

$$F(g,a) = (g,g^a)$$

Examples:
- $G = \mathbb{Z}_p^*$ for a prime $p$, $n = p-1$
- $G = $ subgroup of $\mathbb{Z}_p^*$ of order $q$, where $q \mid p-1$
- $G = $ "elliptic curve groups"
Hardness of Discrete Log

Brute force search: $O(n)$

Better generic algorithm: $O(n^{1/2})$
• Known to be optimal for generic algorithms

Much better algorithms are known for $\mathbb{Z}_p^*$
• Similar running times to integer factorization
• Still super-polynomial