COS433/Math 473: Cryptography

Mark Zhandry

Princeton University

Spring 2020

Announcements

HW4 Due April 2nd
HW5 Due April 9th (will be posted this afternoon)

PR2 Due April 19th

Previously on COS 433…

Collision Resistant Hashing

Expanding Message Length for MACs

Suppose we have a MAC **(MAC,Ver)** that works for small messages (e.g. 256 bits)

How can I build a MAC that works for large messages?

One approach:

- MAC blockwise + extra steps to insure integrity
- Problem: extremely long tags

Collision Resistant Hashing?

Syntax:

- Domain $\mathbf{D}$ (typically $\{0,1\}^l$ or $\{0,1\}^*$)
- Range $\mathbf{R}$ (typically $\{0,1\}^n$)
- Function $\mathbf{H}: \mathbf{D} \rightarrow \mathbf{R}$

Correctness: $n \ll l$

Theory vs Practice

In practice, the existence of an algorithm with a built in collision isn't much of a concern

- Collisions are hard to find, after all

However, it presents a problem with our definitions

- So theorists change the definition
- Alternate def. will also be useful later

Collision Resistant Hashing

Syntax:

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

Correctness: $n \ll l$

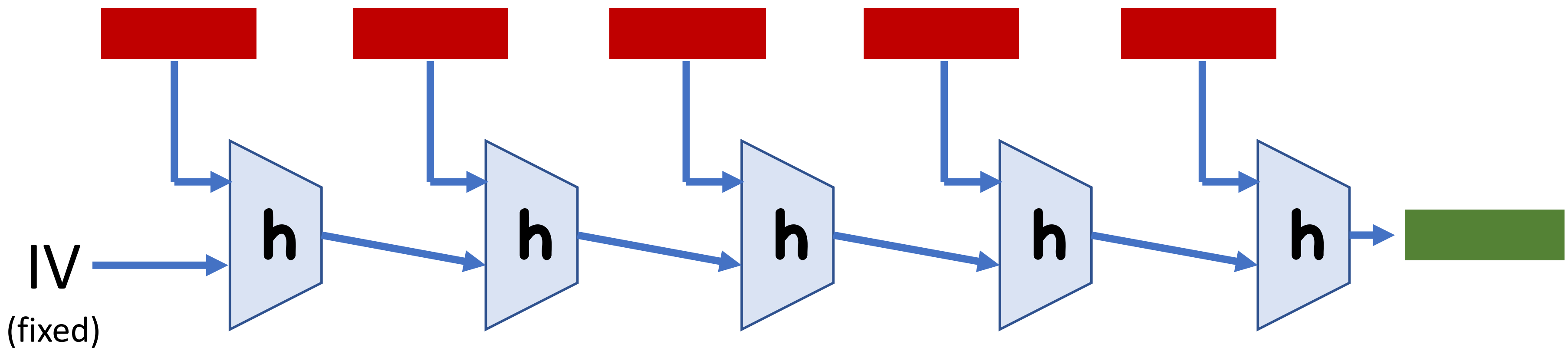
Security

Definition: H is collision resistant if, for all  running in polynomial time, $\exists$ negligible **ε** such that:

$$\Pr[H(k,x_0) = H(k,x_1) \wedge x_0 \neq x_1 :$$

$(x_0,x_1) \leftarrow$  $(k), k \leftarrow K] < \varepsilon(\lambda)$

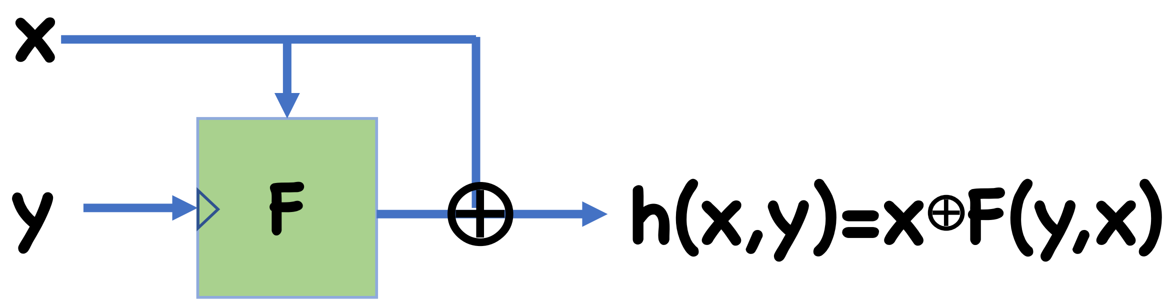
Merkle-Damgard



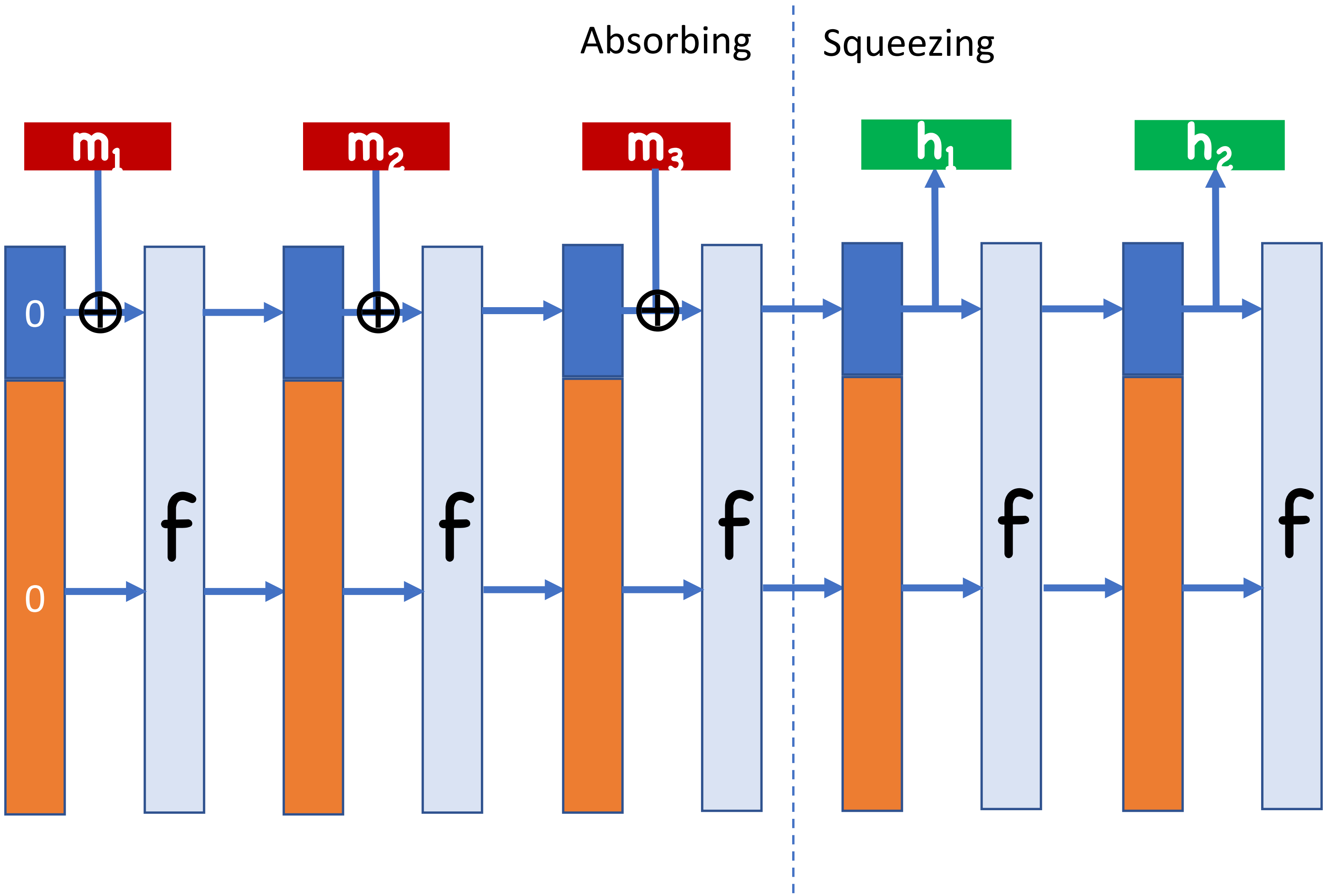
Constructing **h**

Common approach: use block cipher

Davies-Meyer



Sponge Construction



Sponge Construction

Advantages:

- Round function **f** can be public invertible function
  (i.e. unkeyed SPN network)

- Easily get different input/output lengths

Birthday Attack

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

Attack:

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

Birthday Attack

Analysis:

Expected number of collisions

= Number of pairs × Prob each pair is collision

≈ (T choose 2) × 1/|R|

By setting **T=O(|R|^½),** expectend number of collisions found is at least **1**

⇒ likely to find a collision

This time

Commitment Schemes

Anagrams and Astronomy

<u>Galileo and the Rings of Saturn</u>

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

Anagrams and Astronomy

<u>Enter Huygens</u>

- 1656: Realizes Galileo actually saw rings
- Circulates

AAAAAAA CCCCC D EEEEE G H IIIIIII LLLL MM NNNNNNNNN 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 $\mathbf{M}$
Ciphertext Space $\mathbf{C}$
(suppressing security parameter)

Com(m; r): outputs a commitment **c** to **m**

- Why have **r**?

Commitments with Setup

Message space **M**
Ciphertext Space **C**
(suppressing security parameter)

Setup(): Outputs a key **k**
Com(k, m; r): outputs a commitment **c** to **m**

Using Commitments

Commit Stage

$r \leftarrow R$

$c \leftarrow \mathrm{Com}(m;r)$

m

Reveal Stage

m,r

Check that

$c = \mathrm{Com}(m;r)$

Using Commitments (with setup)

k←Setup()

Commit Stage

r←R

c←Com(k,m;r)

m

Reveal Stage

m,r

Check that
c = Com(k,m;r)

Security Properties

Hiding: **c** should hide **m**

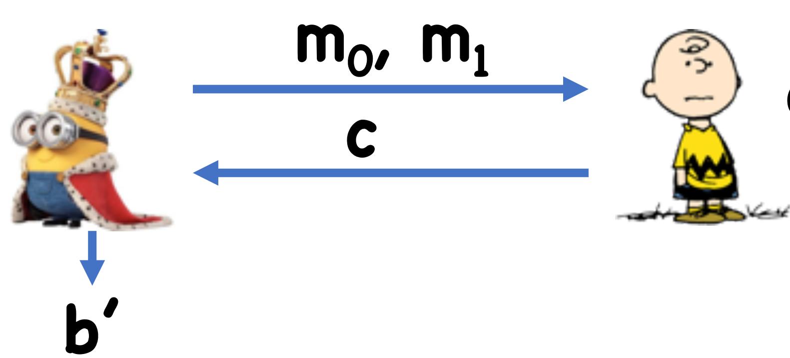
- Perfect hiding: for any m_0, m_1,

$$Com(m_0) \stackrel{d}{=} Com(m_1)$$

- Statistical hiding: for any m_0, m_1,

$$\Delta(\ Com(m_0),\ Com(m_1)\ ) < negl$$

- Computational hiding:



$c \leftarrow Com(m_b)$

Security Properties (with Setup)

Hiding: **c** should hide **m**

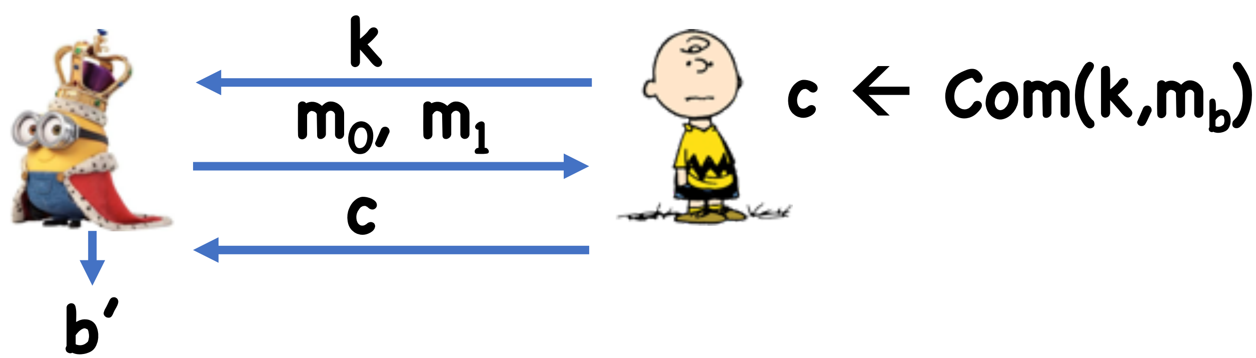
- Perfect hiding: for any $\mathbf{m_0}$, $\mathbf{m_1}$,

$$\mathbf{k,Com(k,m_0)} \stackrel{d}{=} \mathbf{k,Com(k,m_1)}$$

- Statistical hiding: for any $\mathbf{m_0}$, $\mathbf{m_1}$,

$$\Delta(\ [\mathbf{k,Com(k,m_0)}],\ [\mathbf{k,Com(k,m_1)}]\ ) < \mathbf{negl}$$

- Computational hiding:



Security Properties

Binding: Impossible to change committed value

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

- Computational binding: no efficient adversary can find $\mathbf{(m_0,r_0),(m_1,r_1)}$ such that:

$$\mathbf{Com(m_0;r_0)=Com(m_1;r_1)}$$
$$\mathbf{m_0 \neq m_1}$$

Security Properties (with Setup)

Binding: Impossible to change committed value

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

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

- Computational binding: no PPT adversary, given $\mathbf{k \leftarrow Setup()}$, can find $\mathbf{(m_0,r_0),(m_1,r_1)}$ such that

  $\mathbf{Com(k,m_0;r_0)=Com(k,m_1;r_1)}$

  $\mathbf{m_0 \neq m_1}$

Who Runs **Setup()**

Alice?

- Must ensure that Alice cannot devise **k** for which she can break binding

Bob?

- Must ensure Bob cannot devise **k** for which he can break hiding

Solution: Trusted third party (TTP)

Anagrams as Commitment Schemes

Com(m) = sort characters of message

Problems?

- Not hiding: “Jupiter has four moons” vs “Jupiter has five moons”

- Not binding: Kepler decodes Galileo’s anagram to conclude Mars has two moons

Anagrams as Commitment Schemes

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 (perfect) binding, but no hiding

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

- Has (perfect) 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

Com(m;r) = H(m || r)

Theorem: Com(m;r) = H(m||r) has:

- Perfect binding assuming **H** is injective
- Computational binding assuming **H** is collision resistance
- Computational hiding in "random oracle model": **H** is modeled as a random function

“Standard Model” Commitments

Single Bit to Many Bit

Let **(Setup,Com)** be a commitment scheme for single bit messages

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

- $\mathbf{m = (m_1,\ldots,m_t),\ m_i \in \{0,1\}}$
- $\mathbf{r = (r_1,\ldots,r_t),\ r_i}$ are randomness for **Com**

Theorem: If **(Setup,Com)** is statistically/computationally binding, then **(Setup,Com')** is statistically/computationally binding

Theorem: If **(Setup,Com)** is statistically/computationally hiding, then **(Setup,Com')** is statistically/computationally hiding

Binding

Suppose  breaks binding of **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

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

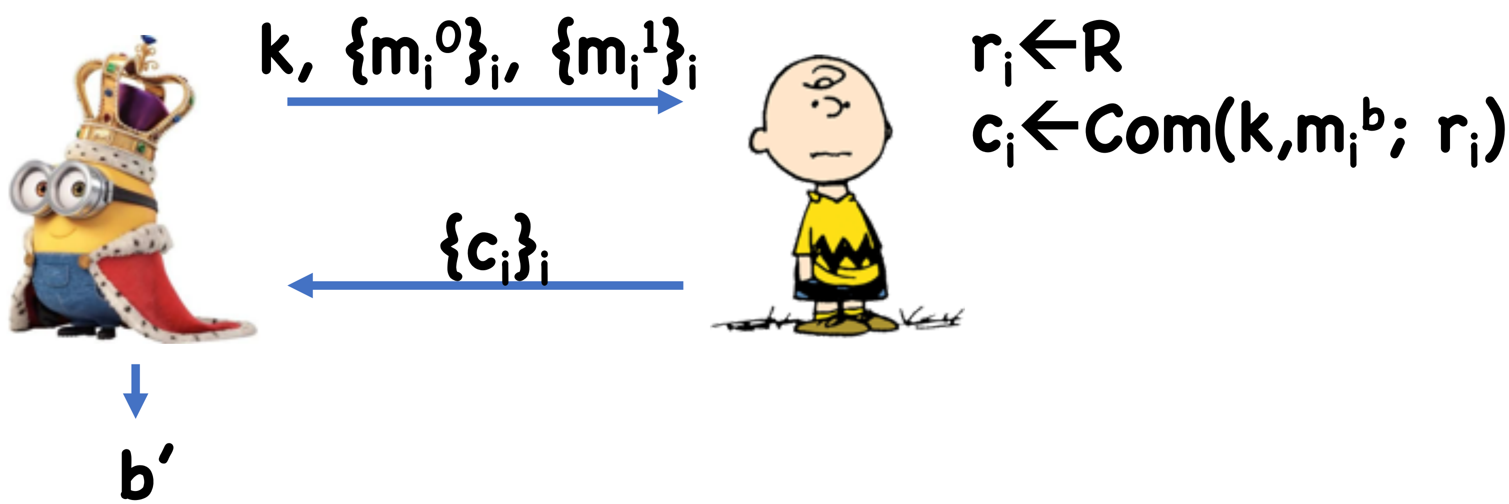
Therefore, $\exists\, i$ such that $m_i^0 \neq m_i^1$ but

$$\mathsf{Com}(k,m_i^0;r_i^0) = \mathsf{Com}(k,m_i^1;r_i^1)$$

$\Rightarrow$ Break binding of **Com**

Hiding

Suppose  breaks (say, computational) hiding



Hiding

Proof by Hybrids

Hybrid $\mathbf{j}$:

- For each $\mathbf{i \leq j}$, $\mathbf{c_i} = \mathbf{Com(k, m_i^1, r_i)}$
- For each $\mathbf{i > j}$, $\mathbf{c_i} = \mathbf{Com(k, m_i^0, r_i)}$

Hybrid $\mathbf{0}$: commit to $\mathbf{\{m_i^0\}_i}$
Hybrid $\mathbf{t}$: commit to $\mathbf{\{m_i^1\}_i}$

$\exists\, \mathbf{j}$ such that  distinguishes Hyb $\mathbf{j\text{-}1}$ from Hyb $\mathbf{j}$
$\Rightarrow$ break hiding of $\mathbf{Com}$

Single Bit to Many Bit

Let **(Setup,Com)** be a commitment scheme for single bit messages

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

- $\mathbf{m = (m_1,\ldots,m_t),\ m_i \in \{0,1\}}$
- $\mathbf{r = (r_1,\ldots,r_t),\ r_i}$ are randomness for **Com**

Therefore, suffices to focus on commitments for single bit messages

Statistically Hiding Commitments?

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

$\mathbf{Setup():}$ $\mathbf{k \leftarrow K}$, output $\mathbf{k}$

$\mathbf{Com(\ k,\ m;\ r) = H(k,\ (m,r))}$

Binding?

Hiding?

Statistically Hiding Commitments

Let **F** be a pairwise independent function family with domain **X={0,1}×R** and range **Y**

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

Setup(): f←F, k←K, output **(f,k)**
Com((f,k), m; r) = H(k, f(m,r))

Theorem: If **|Y|** is "sufficiently large" relative to **|X|** and **H** is collision resistant, then **(Setup,Com)** is computational binding

Theorem: If **|X|** is "sufficiently large" relative to **|Z|**, then **(Setup,Com)** is statistically hiding

Theorem: If **H** is collision resistant and $|X|^2/|Y|$ is negligible, then **(Setup,Com)** is computationally binding

Proof:

- Suppose $|Y| \times \gamma = |X|^2$
- For any $x_0 \neq x_1$, $\Pr[f(x_0)=f(x_1)] < \gamma/(|X|^2)$
- Union bound:

$$\Pr[\exists\ x_0 \neq x_1 \text{ s.t. } f(x_0)=f(x_1)] < \gamma$$

- Therefore, **f** is injective $\Longrightarrow$ any collision for Com must be a collision for **H**

Theorem: If **|X|** is "sufficiently large" relative to **|Z|**, then **(Setup,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 **D** over a set **X**, the min-entropy of **D**, denoted $H_\infty(D)$, is

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

Examples:

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

Leftover Hash Lemma

Lemma: Let **D** be a distribution on **X**, and **F** a family of pairwise independent functions from **X** to **Y**. Then

$$\Delta(\ (f,\ f(D))\ ,\ (f,\ R)\ )\ \leq\ \varepsilon \text{ where}$$

- $f \leftarrow F$
- $R \leftarrow Y$
- $\log |Y| \leq H_\infty(D) + 2 \log \varepsilon$

"Crooked" Leftover Hash Lemma

Lemma: Let **D** be a distribution on **X**, and **F** a family of pairwise independent functions from **X** to **Y**, and **h** be any function from **Y** to **Z**. Then

$$\Delta(\ (f,\ h(f(D)))\ ,\ (f,\ h(R))\ )\ \leq\ \varepsilon \text{ where}$$

- $f \leftarrow F$
- $R \leftarrow Y$
- $\log |Z| \leq H_\infty(D) + 2 \log \varepsilon - 1$

Theorem: If we set $|R|=|Z|^3$ and $|Z|$ is super-poly, then **(Setup,Com)** is statistically hiding

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

Let $D_b=(b,r)$, min-entropy $\log\ |R|$
Set $R\ =|Z|^3$, $\varepsilon\ =\ 2/|Z|$

Then $\log\ |Z| \leq H_\infty(D_b) + 2\ \log\ \varepsilon - 1$

Theorem: If we set $|R|=|Z|^3$ and $|Z|$ is super-poly, then **(Setup,Com)** is statistically hiding

For any **k**,**b**,

$$\Delta(\ (f,\ H(k,\ f(b,r)))\ ,\ (f,\ H(k,\ U))\ )\ \leq\ \varepsilon$$

Thus (for any **k**)

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

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 3λ-bit string **k**

Com(b; r): If **b=0**, output **G(r)**, if **b=1**, output $G(r) \oplus k$

Theorem: **(Setup,Com)** is statistically binding

Theorem: If **G** is a secure PRG, then **(Setup,Com)** is computationally hiding

Theorem: If $\mathbf{G}$ is a secure PRG, then **(Setup,Com)** is computationally hiding

Hybrids:

- Hyb 0: $\mathbf{c = Com(0;r) = G(r)}$ where $\mathbf{r \leftarrow \{0,1\}^\lambda}$
- Hyb 1: $\mathbf{c \leftarrow \{0,1\}^{3\lambda}}$
- Hyb 2: $\mathbf{c = S' \oplus k}$, where $\mathbf{S' \leftarrow \{0,1\}^{3\lambda}}$
- Hyb 3: $\mathbf{c = Com(1;r) = G(r) \oplus k}$ where $\mathbf{r \leftarrow \{0,1\}^\lambda}$

Theorem: (Setup,Com) is statistically binding

Proof:

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

By union bound:

$\Pr[\exists\, r,r'$ such that $Com(k,0)=Com(k,1)]$

$= \Pr[\exists\, r,r'$ such that $G(r) = G(r')\oplus k] < 2^{-\lambda}$

More Problems with Anagrams

Huygens Discovers Saturn's moon Titan

- 1655: Sends the following to Wallis

ADMOVERE OCULIS DISTANTIA SIDERA NOSTRIS,
UUUUUUUCCCRR-HNBQX

(First part meaning "to direct our eyes to distant stars")

Plaintext: **saturno luna sua circunducitur diebus sexdecim horis quatuor**

("Saturn's moon is led around it in sixteen days and four hours")

More Problems with Anagrams

Huygens Discovers Saturn’s moon Titan

- Wallis replies with

AAAAAAAAA B CCCCC DDDD EEEEEEEEE F H
IIIIIIIIIIII LLL MMMMMM NNNNNN OOOOOOO PPPPP
Q RRRRRRRRRRR SSSSSSSSSSSSS TTTTTTTT
UUUUUUUUUUUUUUUUU X

(Contains all of the letters in Huygens’ message, plus some)

More Problems with Anagrams

<u>Huygens Discovers Saturn’s moon Titan</u>

- When Huygens finally reveals his discovery, Wallis responds by giving solution to his anagram:

`saturni comes quasi lunando vehitur. diebus sexdecim circuitu rotatur. novas nuper saturni formas telescopo vidimus primitus. plura speramus`

(“A companion of Saturn is carried in a curve. It is turned by a revolution in sixteen days. We have recently observed new shapes of Saturn with a telescope. We expect more.”)

- Tricked Huygens into thinking British astronomers had already discovered Titan

More Problems with Anagrams

Sometimes, hiding and binding are not enough

For some situations (e.g. claiming priority on discoveries) also want commitments to be "non-malleable"

- Shouldn't be able to cause predictable changes to committed value

Beyond scope of this course

Next Time: Number Theory

Handout on course website with basic number theory primer

Reminders

HW4 Due April 2nd
HW5 Due April 9th

PR2 Due April 19th