Previously on COS 433...
Perfect Security for Multiple Messages

Definition: A stateless scheme \((\text{Enc},\text{Dec})\) has perfect secrecy for \(n\) messages if, for any two sequences of messages \((m_0^{(i)})_{i \in [d]}, (m_1^{(i)})_{i \in [d]} \in \mathcal{M}^d\)

\[
(\text{Enc}(K, m_0^{(i)}))_{i \in [d]} \overset{d}{=} (\text{Enc}(K, m_1^{(i)}))_{i \in [d]}
\]

Notation: \((f(i))_{i \in [d]} = (f(1), f(2), ..., f(n))\)
Theorem: No stateless deterministic encryption scheme can have perfect security for multiple messages.
Randomized Encryption

Syntax:
• Key space $K$ (usually $\{0,1\}^\lambda$)
• Message space $M$ (usually $\{0,1\}^n$)
• Ciphertext space $C$ (usually $\{0,1\}^m$)
• $Enc: K \times M \rightarrow C$ (potentially probabilistic)
• $Dec: K \times C \rightarrow M$ (usually deterministic)

Correctness:
• For all $k \in K$, $m \in M$,
  \[ \Pr[ Dec(k, Enc(k,m) ) = m ] = 1 \]
Theorem: No stateless randomized encryption scheme can have perfect security for multiple messages
What do we do now?

Tolerate tiny probability of distinguishing
• If $\Pr[c^{(0)} = c^{(1)}] = 2^{-128}$, in reality never going to happen

How small is ok?
• Usually $2^{-80}$, $2^{-128}$, or maybe $2^{-256}$

Next time: formalize weaker notion of secrecy to allow for small probability of detection
Statistical Distance

Given two distributions $D_1$, $D_2$ over a set $X$, define

$$\Delta(D_1,D_2) = \frac{1}{2} \sum_x | \Pr[D_1=x] - \Pr[D_2=x] |$$

Observations:

$$0 \leq \Delta(D_1,D_2) \leq 1$$

$$\Delta(D_1,D_2) = 0 \iff D_1 \equiv D_2$$

$$\Delta(D_1,D_2) \leq \Delta(D_1,D_3) + \Delta(D_3,D_2)$$

($\Delta$ is a metric)
Another View of Statistical Distance

Theorem: $\Delta(D_1, D_2) \geq \varepsilon$ iff $\exists A$ s.t.

$\left| \Pr[A(D_1) = 1] - \Pr[A(D_2) = 1] \right| \geq \varepsilon$

Terminology: for any $A$,

$\left| \Pr[A(D_1) = 1] - \Pr[A(D_2) = 1] \right|$

is called the “advantage” of $A$ in distinguishing $D_1$ and $D_2$
Another View of Statistical Distance

Theorem: $\Delta(D_1, D_2) \geq \varepsilon$ iff $\exists A$ s.t. $\left| \Pr[A(D_1) = 1] - \Pr[A(D_2) = 1] \right| \geq \varepsilon$

To lower bound $\Delta$, just need to show adversary $A$ with that advantage
Examples

\( D_1 = \) Uniform distribution over \( \{0,1\}^n \)
- \( \Pr[D_1=x] = 2^{-n} \)

\( D_2 = \) Uniform subject to even parity
- \( \Pr[D_2=x] = 2^{-(n-1)} \) if \( x \) has even parity, 0 otherwise

\[
\Delta(D_1,D_2) = \frac{1}{2} \sum_{x \text{ even}} |2^{-n} - 2^{-(n-1)}| + \frac{1}{2} \sum_{x \text{ odd}} |2^{-n} - 0|
\]

\[
= \frac{1}{2} \sum_{x \text{ even}} 2^{-n} + \frac{1}{2} \sum_{x \text{ odd}} 2^{-n}
\]

\[= \frac{1}{2} \]
Examples

\[ D_1 = \text{Uniform over } \{1, \ldots, n\} \]
\[ D_2 = \text{Uniform over } \{1, \ldots, n+1\} \]

\[ \Delta(D_1, D_2) = \frac{1}{2} \sum_{x=1}^{n} |1/n - 1/(n+1)| + \frac{1}{2} |0 - 1/(n+1)| \]

\[ = \frac{1}{2} \sum_{x=1}^{n} 1/n(n+1) + \frac{1}{2} 1/(n+1) \]

\[ = \frac{1}{2} 1/(n+1) + \frac{1}{2} 1/(n+1) = 1/(n+1) \]
Statistical Security

**Definition:** A scheme \((\text{Enc}, \text{Dec})\) has \(\varepsilon\)-statistical secrecy for \(d\) messages if \(\forall\) two sequences of messages \((m_0^{(i)})_{i \in [d]}, (m_1^{(i)})_{i \in [d]} \in M^d\)

\[
\Delta\left[ (\text{Enc}(K, m_0^{(i)}))_{i \in [d]}, (\text{Enc}(K, m_1^{(i)}))_{i \in [d]} \right] < \varepsilon
\]

We will call such a scheme \((d, \varepsilon)\)-secure
Statistical Security

We will consider a scheme “secure” for $d$ messages if it is $(d, \varepsilon)$-secure for very small $\varepsilon$

- E.g. $2^{-80}$, $2^{-128}$, etc

For comparison: chance of
- Being struck by lightning twice: $2^{-23}$
- Winning the lottery: $2^{-26}$
- World-ending asteroid while on this slide: $2^{-46}$
Stateless Encryption with Multiple Messages

Ex:

\[ M = C = \mathbb{Z}_p \quad (p \text{ a prime of size } 2^{-128}) \]
\[ K = \mathbb{Z}_p^* \times \mathbb{Z}_p \]
\[ \text{Enc( (a,b), m) = (am + b) mod p} \]
\[ \text{Dec( (a,b), c) = (c-b)/a mod p} \]

Q: Is this statistically secure for two messages?
Example

Ex:

\[ M = \mathbb{Z}_p \text{ (} p \text{ a prime of size } 2^{-128}) \]
\[ C = \mathbb{Z}_p^2 \]
\[ K = \mathbb{Z}_p^2 \]

\[ \text{Enc}( (a,b), m) = (r, (ar+b) + m) \]
\[ \text{Dec}( (a,b), (r,c) ) = c - (ar+b) \]

Q: Is this statistically secure for two messages?
Proof of Example

Let $D_b$ be distribution of $(\text{Enc}(k,m_b^{(i)}))_I$

Let $D_b'$ be $D_b$, but conditioned on $r_0 \neq r_1$

Fix $r_0 \neq r_1, m_0, m_1, c_0, c_1$

\[
\Pr[ar_0+b+m_0=c_0, ar_1+b+m_1=c_1] = \frac{1}{p^2}
\]

So $D_0' \overset{d}{=} D_1'$ ( $\Delta(D_0', D_1') = 0$ )
Proof of Example

Lemma: \( \Delta(D_1, D_2) \leq \Pr[\text{bad}|D_1] + \Pr[\text{bad}|D_2] \\
+ \Delta(D_1', D_2') \)

Where:
- “\text{bad}” is some event
- \( \Pr[\text{bad}|D_b] \) is probability “\text{bad}” when sampling from \( D_b \)
- \( D_b' \) is \( D_b \), but conditioned on not “\text{bad}”
Proof of Lemma

\[ \Delta(D_1,D_2) = \sum_x \left| \Pr[D_1=x] - \Pr[D_2=x] \right| \]

\[ = \sum_{x:\text{bad}} \left| \Pr[D_1=x] - \Pr[D_2=x] \right| \]

\[ + \sum_{x:\text{good}} \left| \Pr[D_1=x] - \Pr[D_2=x] \right| \]

\[ \leq \sum_{x:\text{bad}} \left| \Pr[D_1=x] \right| + \sum_{x:\text{bad}} \left| \Pr[D_2=x] \right| \]

\[ + \sum_{x:\text{good}} \left| \Pr[D_1=x] - \Pr[D_2=x] \right| \]

\[ \leq \Pr[\text{bad}|D_1] + \Pr[\text{bad}|D_2] + \Delta(D_1,\text{good},D_2,\text{good}) \]
Proof of Example

Let $D_b$ be distribution of $(\operatorname{Enc}(k, m_b^{(i)}) )_I$
Let $\textbf{bad}$ be when $r_0 = r_1$
Let $D_b'$ be $D_b$, but conditioned on $\textbf{not bad}$

$\Pr[\text{bad}|D_b] = 1/p$
$\Delta(D_0', D_1') = 0$

Therefore, $\Delta(D_0, D_1) \leq 2/p$
Summary so Far

Stateless encryption for multiple messages ✓

But, key length grows with number of messages ✗

And, key length grows with length of message ✗
Limits of Statistical Security

**Theorem:** Suppose \((\text{Enc}, \text{Dec})\) has plaintext space \(M = \{0,1\}^n\) and key space \(K = \{0,1\}^t\). Moreover, assume it is \((d, \frac{1}{3})\)-secure. Then:

\[
t \geq d \cdot n
\]

In other words, the key must be at least as long as the total length of all messages encrypted.
Proof Idea

Use an encryption protocol to build a compression protocol

\[ m \xrightarrow{\text{Comp}(m)} m' \]

\[ m' \xleftarrow{\text{Decomp}(m')} m \]

Goal: \(|m'| < |m|\)
For Now: Easier Goal

\[ \text{Goal: } |m'| < |m| \]
The Protocol

Let $m_0$ be some message in $M$

Setup():
• Choose random $k_0 \leftarrow K$
• Let $c_1 \leftarrow \text{Enc}(k_0, m_0), \ldots, c_d \leftarrow \text{Enc}(k_0, m_0)$
• Output $(c_1, \ldots, c_d)$

Comp( $(c_1, \ldots, c_d), (m_1, \ldots, m_d)$ ):
• Find $k, r_1, \ldots, r_d$ such that $c_i = \text{Enc}(k, m_i; r_i) \ \forall i$
• If no such values exist, abort
• Output $k$
The Protocol

Let \( m_0 \) be some message in \( M \)

\[ \text{Comp}( (c_1, \ldots, c_d), (m_1, \ldots, m_d) ) : \]
- Find \( k, r_1, \ldots, r_d \) such that \( c_i = \text{Enc}(k, m_i; r_i) \ \forall i \)
- If no such values exist, abort
- Output \( k \)

\[ \text{Decomp}((c_1, \ldots, c_d), k) : \]
- Compute \( m_i = \text{Dec}(k, c_i) \)
- Output \((m_1, \ldots, m_d)\)
Analysis of Protocol

If \textbf{Comp} succeeds, \textbf{Decomp} must succeed by correctness
• Since $c_i = \text{Enc}(k, m_i; r_i)$, \text{Dec}(k, c_i) must give $m_i$

Therefore, must figure out when \textbf{Comp} succeeds

\textbf{Claim:} For any sequence of messages $m_1, \ldots, m_d$, \textbf{Comp} succeeds with probability at least $1 - \varepsilon$

(Probability over the randomness used by \textbf{Setup}())
Claim: For any sequence of messages $m_1, \ldots, m_d$, $\text{Comp}$ succeeds with probability at least $1-\varepsilon$

Proof:
• Suppose $\text{Comp}$ succeeds with probability $1-p$ for messages $m_1, \ldots, m_d$
• Let $A(c_1, \ldots, c_d)$ be the algorithm that runs $\text{Comp}((c_1, \ldots, c_d), (m_1, \ldots, m_d))$ and outputs 1 if $\text{Comp}$ succeeds
• If $c_i = \text{Enc}(k_0, m_i)$, then $\Pr[A(c_1, \ldots, c_d) = 1] = 1$
• If $c_i = \text{Enc}(k_0, m_0)$, then $\Pr[A(c_1, \ldots, c_d) = 1] = 1-p$
• By $(d, \varepsilon)$ statistical security of $\text{Enc}$, $p$ must be $\leq \varepsilon$
Claim: For any sequence of messages $m_1, \ldots, m_d$, Comp succeeds with probability at least $1-\varepsilon$.

Claim: For a random sequence of messages $m_1, \ldots, m_d$, Comp succeeds with probability at least $1-\varepsilon$.

(Probability over the randomness used by Setup() and the random choices of $m_1, \ldots, m_d$.)
Next step: Removing Setup

We know:

$$\Pr[\text{Comp succeeds: } (c_1, \ldots, c_d) \leftarrow \text{Setup}(), \ m_i \leftarrow M] \geq 1 - \varepsilon$$

Therefore, there must exist some $$(c_1^*, \ldots, c_d^*)$$ such that

$$\Pr[\text{Comp succeeds: } m_i \leftarrow M] \geq 1 - \varepsilon$$

Define: $M' = \{(m_1, \ldots, m_d): \text{Comp succeeds}\}$

• Note that $|M'| \geq (1 - \varepsilon) |M|^d$
The Protocol

Find $k, r_1, ..., r_d$ such that
$c_i^* = Enc(k, m_i; r_i) \ \forall i$

For each $i$,
Let $m_i \leftarrow Dec(k, c_i^*)$
Output $(m_1, ..., m_d)$

By previous analysis,
• Alice always successfully compresses
• Bob always successfully decompresses
Final Touches

Can compress messages in $\mathcal{M}'$ into keys in $\mathcal{K}$

Therefore, it must be that $|\mathcal{M}'| \leq |\mathcal{K}|

 Meaning $t = \log |\mathcal{K}|$

$\geq \log |\mathcal{M}'|$

$\geq \log \left[ (1-\varepsilon) |\mathcal{M}|^d \right]$

$= d \log |\mathcal{M}| + \log [1-\varepsilon]$

$\geq dn - 2\varepsilon$

$\geq dn$ (as long as $\varepsilon < \frac{1}{2}$)
Takeaway

If you don’t want to physically exchange keys frequently, you cannot obtain statistical security

So, now what?
Computational Security

We are ok if adversary takes a really long time

Usually measure in machine operations
• Though depends on architecture, so rough approx
• $2^{80}$, $2^{128}$, or maybe $2^{256}$ are probably ok

For comparison:
• Lifetime of universe in nanoseconds: $2^{58}$
• Number of atoms in known universe: $2^{265}$
Brute Force Attacks

Simply try every key until find right one

Relevant as long as key length is smaller than total length of messages encrypted

If keys have length $\lambda$, $2^\lambda$ is upper bound on attack
Crypto and P vs NP

What if P = NP?

From this point forward, almost all crypto we will see depends on computational assumptions
[TRANSLTR]’s three million processors would all work in parallel ... trying every new permutation as they went.
“What’s the longest you’ve ever seen TRANSLTR take to break a code?”

“About an hour, but it had a ridiculously long key—ten thousand bits”
Defining Security

Consider an attacker as a probabilistic efficient algorithm

Attacker gets to choose the messages

All attacker has to do is distinguish them
Security Experiment/Game
(One-time setting)

Challenger

\[ m_0, m_1 \in M \]

\[ b \]

IND-Exp_b(\(\square\))
Security Definition  (One-time setting)

Definition: \((\text{Enc}, \text{Dec})\) has \((t, \varepsilon)\)-ciphertext indistinguishability if, for all \(\beta\) running in time at most \(t\)

\[
\left| \Pr[1 \leftarrow \text{IND-Exp}_0(\beta)] - \Pr[1 \leftarrow \text{IND-Exp}_1(\beta)] \right| \leq \varepsilon
\]
Construction with $|k| \ll |m|$

Idea: use OTP, but have key generated by some expanding function $G$
What Do We Want Out of $G$?

Definition: $G: \{0,1\}^\lambda \rightarrow \{0,1\}^n$ is a $(t,\varepsilon)$-secure pseudorandom generator (PRG) if:

- $n > \lambda$
- $G$ is deterministic
- For all $\mathcal{A}$ running in time at most $t$,

$$\Pr[\mathcal{A}(G(s)) = 1 : s \leftarrow \{0,1\}^\lambda] - \Pr[\mathcal{A}(x) = 1 : x \leftarrow \{0,1\}^n] \leq \varepsilon$$
Secure PRG $\rightarrow$ Ciphertext Indistinguishability

$$K = \{0,1\}^\lambda$$
$$M = \{0,1\}^n$$
$$C = \{0,1\}^n$$

$$\text{Enc}(k,m) = \text{PRG}(k) \oplus m$$
$$\text{Dec}(k,c) = \text{PRG}(k) \oplus c$$
Security?

Intuitively, security is obvious:
- \( \text{PRG}(k) \) ”looks” random, so should completely hide \( m \)
- However, formalizing this argument is non-trivial.

Solution: reductions
- Assume toward contradiction an adversary for the encryption scheme, derive an adversary for the PRG
Security

Assume towards contradiction that there is an $m_0, m_1 \in M$ such that

$$|\Pr[W_0] - \Pr[W_1]| \geq \varepsilon,$$

non-negligible

$W_b$: $b' = 1$ in IND-Exp$_b$
Security

Use $\mathbf{m}_0, \mathbf{m}_1 \in \mathcal{M}$ to build $\mathbf{c}$. $\mathbf{b}$ will run $\mathbf{b}'$ as a subroutine, and pretend to be $\mathcal{M}$.

$m_0, m_1 \in \mathcal{M}$

$\mathbf{b} \leftarrow \{0, 1\}$

$\mathbf{c} \leftarrow x \oplus m_b$

$1 \oplus b \oplus b'$

(either $G(s)$ or truly random)
Security

Case 1: \( x = \text{PRG}(s) \) for a random seed \( s \)

- \( \text{Nexx} \) “sees” \( \text{IND-Exp}_b \) for a random bit \( b \)

\[
\begin{align*}
\text{m}_0, \text{m}_1 & \in \mathcal{M} \\
b & \leftarrow \{0,1\} \\
s & \leftarrow K \\
c & \leftarrow \text{PRG}(s) \oplus \text{m}_b
\end{align*}
\]
Security

Case 1: \( x = \text{PRG}(s) \) for a random seed \( s \)

- \( \text{sees} \) IND-Exp \( b \) for a random bit \( b \)
- \( \Pr[1 \oplus b \oplus b' = 1] = \Pr[b = b'] \)
  
  \[
  = \frac{1}{2} \Pr[b' = 1 \mid b = 1] 
  + \frac{1}{2} (1 - \Pr[b' = 1 \mid b = 0])
  \]
  
  \[
  = \frac{1}{2}(1 + \Pr[W_0] - \Pr[W_1])
  = \frac{1}{2}(1 \pm \varepsilon)
  \]
Security

Case 2: $\times$ is truly random

- “sees” OTP encryption

\[
m_0, m_1 \in M_\lambda \quad b \leftarrow \{0,1\} \quad x \leftarrow \{0,1\}^n \\
\quad c \leftarrow x \oplus m_b \\
b' \quad c \leftarrow b \oplus m_b\]
Security

Case 2: $\times$ is truly random

- 🕵️‍♂️ “sees” OTP encryption
- Therefore $\Pr[b'=1 \mid b=0] = \Pr[b'=1 \mid b=1]$
- $\Pr[1 \oplus b \oplus b'=1] = \Pr[b=b']$
  
  \[
  = \frac{1}{2} \Pr[b'=1 \mid b=1] \\
  \quad + \frac{1}{2} (1 - \Pr[b'=1 \mid b=0])
  \]

  \[
  = \frac{1}{2}
  \]
Security

Putting it together:

• $\Pr[ (G(s)) = 1 : s \leftarrow \{0, 1\}^\lambda ] = \frac{1}{2} (1 \pm \varepsilon(\lambda))$

• $\Pr[ (x) = 1 : x \leftarrow \{0, 1\}^n ] = \frac{1}{2}$

• Absolute Difference: $\frac{1}{2}\varepsilon$, $\Rightarrow$ Contradiction!
Security

**Thm:** If $G$ is a $(t + t', \epsilon/2)$-secure PRG, then $(Enc, Dec)$ is has $(t, \epsilon)$-ciphertext indistinguishability, where $t'$ is the time to:

- Flip a random bit $b$
- XOR two $n$-bit strings
Thm: If $G$ is a $(t+\text{poly},\varepsilon/2)$-secure PRG, then $(\text{Enc},\text{Dec})$ is has $(t,\varepsilon)$-ciphertext indistinguishability.
An Alternate Proof: Hybrids

Idea: define sequence of “hybrid” experiments “between” IND-Exp\(_0\) and IND-Exp\(_1\)

In each hybrid, make small change from previous hybrid

Hopefully, each small change is undetectable

Using triangle inequality, overall change from IND-Exp\(_0\) and IND-Exp\(_1\) is undetectable
An Alternate Proof: Hybrids

Hybrid 0: IND-Exp₀

m₀, m₁ ∈ M → c → k ← K

k ← G(k) ⊕ m₀
An Alternate Proof: Hybrids

Hybrid 1:

\[ m_0, m_1 \in M \]

\[ x \leftarrow \{0,1\}^n \]

\[ c \leftarrow x \oplus m_0 \]
An Alternate Proof: Hybrids

Hybrid 2:

\[ m_0, m_1 \in M \]

\[ x \leftarrow \{0,1\}^n \]

\[ c \leftarrow x \oplus m_1 \]

\[ b' \]
An Alternate Proof: Hybrids

Hybrid 3: $\text{IND-Exp}_1$

$m_0, m_1 \in M$

$k \leftarrow K$

$c \leftarrow G(k) \oplus m_1$

$b'$
An Alternate Proof: Hybrids

\[
\left| \Pr[b'=1 : \text{IND-Exp}_0] - \Pr[b'=1 : \text{IND-Exp}_1] \right|
\]
\[
= \left| \Pr[b'=1 : \text{Hyb } 0] - \Pr[b'=1 : \text{Hyb } 3] \right|
\]
\[
\leq \left| \Pr[b'=1 : \text{Hyb } 0] - \Pr[b'=1 : \text{Hyb } 1] \right|
+ \left| \Pr[b'=1 : \text{Hyb } 1] - \Pr[b'=1 : \text{Hyb } 2] \right|
+ \left| \Pr[b'=1 : \text{Hyb } 2] - \Pr[b'=1 : \text{Hyb } 3] \right|
\]

If \(\left| \Pr[b'=1 : \text{IND-Exp}_0] - \Pr[b'=1 : \text{IND-Exp}_1] \right| \geq \varepsilon\),
Then for some \(i=0,1,2,\)
\[
\left| \Pr[b'=1 : \text{Hyb } i] - \Pr[b'=1 : \text{Hyb } i+1] \right| \geq \frac{\varepsilon}{3}
\]
An Alternate Proof: Hybrids

Suppose \( b \) distinguishes Hybrid 0 from Hybrid 1 with advantage \( \varepsilon / 3 \)

\[
k \leftarrow K \\
m_0, m_1 \in M \\
c \leftarrow G(k) \oplus m_0 \\
\downarrow b' \\
x \leftarrow \{0,1\}^n \\
m_0, m_1 \in M \\
c \leftarrow x \oplus m_0 \\
\downarrow b' \\
\]
An Alternate Proof: Hybrids

Suppose distinguishes Hybrid 0 from Hybrid 1 with advantage $\varepsilon/3 \implies$ Construct

$m_0, m_1 \in M$

c

(x or truly random)

$G(s)$

$c \leftarrow x \oplus m_0$

b'}
An Alternate Proof: Hybrids

Suppose \( \text{嫣} \) distinguishes \textbf{Hybrid 0} from \textbf{Hybrid 1} with advantage \( \varepsilon/3 \) \( \Rightarrow \) Construct

If \( \text{嫣} \) is given \( G(s) \) for a random \( s \), \( \text{嫣} \) sees \textbf{Hybrid 0}
If \( \text{嫣} \) is given \( x \) for a random \( x \), \( \text{嫣} \) sees \textbf{Hybrid 1}

Therefore, advantage of \( \text{嫣} \) is equal to advantage of \( \text{嫣} \) which is at least \( \varepsilon/3 \) \( \Rightarrow \) Contradiction!
An Alternate Proof: Hybrids

Suppose \( \mathcal{A} \) distinguishes Hybrid 1 from Hybrid 2 with advantage \( \varepsilon/3 \)

\[
x \leftarrow \{0,1\}^n
\]

\[
m_0, m_1 \in M
\]

\[
c \leftarrow x \oplus m_0
\]

\[
b'
\]
An Alternate Proof: Hybrids

Suppose distinguishes Hybrid 1 from Hybrid 2 with advantage $\varepsilon \cdot \frac{1}{3}$

Impossible by OTP security
An Alternate Proof: Hybrids

Suppose distinguishes Hybrid 2 from Hybrid 3 with advantage $\varepsilon/3$

$x \leftarrow \{0,1\}^n$

$k \leftarrow K$

$m_0, m_1 \in M$

$c \leftarrow x \oplus m_1$

Proof essentially identical to Hybrid 0/Hybrid 1 case
Reminders

PR1 Part 1 Due Tuesday, Feb 20th