COS433/Math 473: Cryptography

Mark Zhandry

Princeton University

Spring 2017

Announcements

Homework 1 up

- You should be able to complete it after today's lecture

Office Hours

Mark: Mondays 3pm-4pm in COS 314

Fermi: Fridays 2-3pm in Theory Lounge

(3rd floor of COS)

Last Time

Many examples of classical cryptosystems

- Substitution ciphers
- Disk-based ciphers
- Transposition ciphers
- Anagrams
- Enigma

Essentially all completely insecure

Today

Some basic principles

Defining encryption: perfect security

The one-time pad

Kerckhoffs's Principle

Security should only depend on the secrecy of the **key**

- Should still be secure if attacker knows encryption procedure

Why?

- Bad things happen
- Hard to update entire system, easy to update key
- System can be analyzed by crypto community
- Easier to formalize security notions

Designing Crypto Is Hard

Cannot discern security through program analysis

- Just because it compiles doesn't mean it's secure
- Just because you can't see how to break it doesn't mean someone else can't

Even experts get it wrong

Unexpected attack vectors

- Known/chosen plaintext attack
- Chosen *ciphertext* attack
- Timing attack
- Power analysis
- Acoustic cryptanalysis

Solution

When designing new crypto, should have a formal argument why is should resist ALL attacks

- Not always possible; if not, use crypto standards vetted by crypto community

Better yet: only use well-known crypto libraries

- Don’t implement crypto yourself
- You’ll probably get it wrong and introduce side-channels

Defining Crypto

Formal security notion a necessary step before proving anything

Syntax: The algorithms in the cryptosystem and their inputs

- E.g. **Enc**, **Dec**, keys, messages, etc

Correctness/Completeness: Functional relationships between algorithms

- E.g. **Dec(k, Enc(k, m)) = m**

Security: What an attacker should not be able to do

Defining Encryption

Encryption Basics (for now)

Syntax:

- Key space $\mathbf{K}$ (usually $\{0,1\}^\lambda$)
- Message space $\mathbf{M}$ (usually $\{0,1\}^n$)
- Ciphertext space $\mathbf{C}$ (usually $\{0,1\}^m$)
- $\mathbf{Enc}: \mathbf{K} \times \mathbf{M} \rightarrow \mathbf{C}$
- $\mathbf{Dec}: \mathbf{K} \times \mathbf{C} \rightarrow \mathbf{M}$

Correctness:

- For all $\mathbf{k} \in \mathbf{K}$, $\mathbf{m} \in \mathbf{M}$, $\mathbf{Dec}(\mathbf{k}, \mathbf{Enc}(\mathbf{k},\mathbf{m})) = \mathbf{m}$

Encryption Security?

Questions to think about:

<u>What kind of messages?</u>
<u>What does the adversary already know?</u>
<u>What information are we trying to protect?</u>

Examples:

- Messages are always either “attack at dawn” or “attack at dusk”, trying to hide which is the case
- Messages are status updates (“<person> reports <event> at <location>”). Which data is sensitive?

Encryption Security?

Questions to think about:

<u>What kind of messages?</u>

<u>What does the adversary already know?</u>

<u>What information are we trying to protect?</u>

Goal:

Rather than design a separate system for each use case, design a system that works in all possible settings

Semantic Security

Idea:

- Plaintext comes from an arbitrary distribution
- Adversary initially has some information about the plaintext
- Seeing the ciphertext should not reveal any more information

(Perfect) Semantic Security

Definition: A scheme **(Enc,Dec)** is **(perfectly) semantically secure** if, for all:

- Distributions **D** on **M** ← Plaintext distribution
- Functions $\mathbf{I}:\mathbf{M}\rightarrow\{0,1\}^*$ ← Info adv gets
- Functions $\mathbf{f}:\mathbf{M}\rightarrow\{0,1\}^*$ ← Info adv tries to learn
- Functions $\mathbf{A}:\mathbf{C}\times\{0,1\}^*\rightarrow\{0,1\}^*$ ← Adversary

"Simulator"

There exists a function $\mathbf{S}:\{0,1\}^*\rightarrow\{0,1\}^*$ such that

$$\Pr[\ A(\ \mathrm{Enc}(k,m)\ ,\ I(m)\ )\ =\ f(m)\ ]$$
$$=\Pr[\ S(\ I(m)\ )\ =\ f(m)\ ]$$

where probabilities are taken over $\mathbf{k}\leftarrow\mathbf{K},\ \mathbf{m}\leftarrow\mathbf{D}$

Semantic Security

Captures what we want out of an encryption scheme

But, complicated, with many moving parts

Want: something simpler…

Perfect Secrecy [Shannon'49]

Definition: A scheme **(Enc,Dec)** has **perfect secrecy** if, for all:

- Two messages $\mathbf{m_0}, \mathbf{m_1} \in \mathbf{M}$
- Ciphertext $\mathbf{c} \in \mathbf{C}$

$$\Pr[\ \mathrm{Enc}(k,\ m_0) = c] = \Pr[\ \mathrm{Enc}(k,\ m_1) = c]$$

where probabilities are taken over $\mathbf{k} \leftarrow \mathbf{K}$

Notation

Two random variables $\mathbf{X,Y}$ over a finite set $\mathbf{S}$ have identical distributions if, for all $\mathbf{s} \in \mathbf{S}$,

$$\Pr[\, X = s \,] \;=\; \Pr[\, Y = s \,]$$

In this case, we write

$$X \overset{d}{=} Y$$

Perfect Secrecy [Shannon'49]

Definition: A scheme **(Enc,Dec)** has **perfect secrecy** if, for any two messages $\mathbf{m_0}, \mathbf{m_1} \in \mathbf{M}$

$$\mathbf{Enc(K,\ m_0)} \overset{d}{=} \underbrace{\mathbf{Enc(K,\ m_1)}}$$

Random variable corresponding to uniform distribution over **K**

Random variable corresponding to encrypting $\mathbf{m_1}$ using a uniformly random key

Obtaining Perfect Secrecy: The One-Time Pad

Key space $\mathcal{K} = \{0,1\}^n$
Message space $\mathcal{M} = \{0,1\}^n$
Ciphertext space $\mathcal{C} = \{0,1\}^n$

$\mathrm{Enc}(k, m) = k \oplus m$
$\mathrm{Dec}(k, c) = k \oplus c$

Correctness:

$$
\begin{aligned}
\mathrm{Dec}(k, \mathrm{Enc}(k, m)) &= k \oplus (k \oplus m) \\
&= (k \oplus k) \oplus m \\
&= 0 \oplus m \\
&= m
\end{aligned}
$$

Example:

$k = 0011010110$
$m = 1001010101$
$c = 1010000011$

Obtaining Perfect Secrecy: The One-Time Pad

Security?

Theorem: For any message $\mathbf{m} \in \{0,1\}^n$ and ciphertext $\mathbf{c} \in \{0,1\}^n$,

$$\Pr[\ \mathrm{Enc}(k, m) = c] = 2^{-n}$$

Proof:

$$\begin{aligned}\Pr[\ \mathrm{Enc}(k, m) = c] &= \Pr[\ k \oplus m = c\ ] \\ &= \Pr[\ k = c \oplus m\ ] \\ &= 2^{-n}\end{aligned}$$

Obtaining Perfect Secrecy: The One-Time Pad

Security?

Theorem: For any message $\mathbf{m} \in \{0,1\}^n$ and ciphertext $\mathbf{c} \in \{0,1\}^n$,

$$\Pr[\ \mathrm{Enc}(k,\ m) = c] = 2^{-n}$$

In other words, for any **m**, $\mathrm{Enc}(K,m) \stackrel{d}{=} C$

Perfect secrecy easily follows:

$$\mathrm{Enc}(K,\ m_0) \stackrel{d}{=} C \stackrel{d}{=} \mathrm{Enc}(K,\ m_1)$$

Meaning of Perfect Secrecy

Perfect secrecy is a great definition

- Simple
- Easy to prove

However, it doesn’t obviously capture what we need

- What does adversary learn about plaintext?

Semantic Security = Perfect Secrecy

Theorem: A scheme **(Enc,Dec)** is semantically secure if and only if it has perfect secrecy

Perfect Secrecy $\Rightarrow$ Semantic Security

Given arbitrary:

- Distribution $\mathbf{D}$ on $\mathbf{M}$
- Function $\mathbf{I:M \rightarrow \{0,1\}^*}$
- Function $\mathbf{f:M \rightarrow \{0,1\}^*}$
- Function $\mathbf{A:C \times \{0,1\}^* \rightarrow \{0,1\}^*}$

Know: $\mathbf{E(K,\ m_0) \stackrel{d}{=} E(K,\ m_1)}$

Goal: Construct $\mathbf{S:\{0,1\}^* \rightarrow \{0,1\}^*}$ such that

$$\mathbf{Pr[\ A(\ Enc(k,m)\ ,\ I(m)\ ) = f(m)\ ]}$$
$$\mathbf{= Pr[\ S(\ I(m)\ ) = f(m)\ ]}$$

Perfect Secrecy $\Rightarrow$ Semantic Security

S(i):

- Choose random **k** $\leftarrow$ **K**
- Set **c** $\leftarrow$ **Enc(k,0)**
- Run and output **A(c,i)**

$$
\begin{aligned}
\Pr[\ S(\ I(m)\ ) = f(m)\ ] \\
&= \Pr[\ A(\ Enc(k,0)\ ,\ I(m)\ ) = f(m) : m \leftarrow D\ ] \\
&= \Sigma_{m,c} \Pr[D=m] \Pr[Enc(K,0)=c] \Pr[\ A(c,I(m)) = f(m)\ ] \\
&= \Sigma_{m,c} \Pr[D=m] \Pr[Enc(K,m)=c] \Pr[\ A(c,I(m)) = f(m)\ ] \\
&= \Pr[\ A(\ Enc(k,m)\ ,\ I(m)\ ) = f(m)\ ]
\end{aligned}
$$

Semantic Security $\Rightarrow$ Perfect Secrecy

Proof by contrapositive:

- Assume $\exists$ m_0, m_1 s.t. $Enc(K, m_0) \stackrel{d}{\neq} Enc(K, m_1)$
- Devise **D,I,f,A** such that no **S** exists

D: pick $b \leftarrow \{0,1\}$ at random, output m_b

I: empty

$f(m_b) = b$

$A(c) = 1$ iff $Pr[Enc(K,m_1) = c] > Enc(K,m_0) = c]$

Semantic Security $\Rightarrow$ Perfect Secrecy

Let $\mathbf{T} = \{c\text{: } \Pr[\mathrm{Enc}(K,m_1) = c] > \mathrm{Enc}(K,m_0) = c]\}$

$$\Pr[\ A(\ \mathrm{Enc}(K,m)\ ) = f(m) : m \leftarrow D]$$

$$= \tfrac{1}{2} \Pr[A(\ \mathrm{Enc}(K,m_0)\ ) = 0\ ] + \tfrac{1}{2} \Pr[A(\ \mathrm{Enc}(K,m_1)\ ) = 1\ ]$$

$$= \tfrac{1}{2} \Pr[\ \mathrm{Enc}(K,m_0) \notin T] + \tfrac{1}{2} \Pr[\ \mathrm{Enc}(K,m_1) \in T]$$

$$= \tfrac{1}{2} + \tfrac{1}{2} (\Pr[\ \mathrm{Enc}(K,m_1) \in T] - \Pr[\ \mathrm{Enc}(K,m_0) \in T])$$

Semantic Security $\Rightarrow$ Perfect Secrecy

$$
\begin{aligned}
\Pr[\,\mathrm{Enc}(K,m_b) \in T\,] \\
&= \textstyle\sum_{c\in T} \Pr[\mathrm{Enc}(K,m_b) = c] \\
&= 1 - \textstyle\sum_{c\notin T} \Pr[\mathrm{Enc}(K,m_b) = c]
\end{aligned}
$$

$$
\begin{aligned}
\Pr[\,\mathrm{Enc}(K,m_1) \in T] - \Pr[\,\mathrm{Enc}(K,m_0) \in T] \\
&= \textstyle\sum_{c\in T} \Pr[\mathrm{Enc}(K,m_1) = c] - \Pr[\mathrm{Enc}(K,m_0) = c] \\
&= \textstyle\sum_{c\notin T} \Pr[\mathrm{Enc}(K,m_0) = c] - \Pr[\mathrm{Enc}(K,m_1) = c] \\
&= \tfrac{1}{2} \textstyle\sum_{c} \mid \Pr[\Pr[\mathrm{Enc}(K,m_1)=c] - \Pr[\mathrm{Enc}(K,m_0)=c] \mid
\end{aligned}
$$

Notation: Statistical Distance

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

$$\Delta(D_1,D_2) = \tfrac{1}{2}\Sigma_x \mid Pr[D_1=x] - Pr[D_2=x] \mid$$

Observations:

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

$$\Delta(D_1,D_2) = 0 \quad \Leftrightarrow \quad D_1 = D_2$$

Semantic Security $\Rightarrow$ Perfect Secrecy

$$\Pr[\ \mathrm{Enc}(K,m_1) \in T] - \Pr[\ \mathrm{Enc}(K,m_0) \in T]$$
$$= \tfrac{1}{2} \Sigma_c \mid \Pr[\Pr[\mathrm{Enc}(K,m_1)=c] - \Pr[\mathrm{Enc}(K,m_0)=c] \mid$$
$$= \Delta(\ \mathrm{Enc}(K,m_0)\ ,\ \mathrm{Enc}(K,m_1)\ )$$

Therefore,

$$\Pr[\ A(\ \mathrm{Enc}(K,m)\ ) = f(m)\ ]$$
$$= \tfrac{1}{2} + \tfrac{1}{2}\,\Delta(\ \mathrm{Enc}(K,m_0)\ ,\ \mathrm{Enc}(K,m_1)\ )$$

Since $E(K,\ m_0) \overset{d}{\neq} E(K,\ m_1)$,

$$\Rightarrow\ \Delta(\ \mathrm{Enc}(K,m_0)\ ,\ \mathrm{Enc}(K,m_1)\ ) > 0$$
$$\Rightarrow\ \Pr[\ A(\ \mathrm{Enc}(K,m)\ ) = f(m)\ ] > \tfrac{1}{2}$$

Semantic Security $\Rightarrow$ Perfect Secrecy

Pr[A(Enc(K,m)) = f(m)] > ½

However, for any **S**,

$$\Pr[\, S() = f(m) \,] = \Pr[\, S() = b : b \leftarrow \{0,1\} \,]$$
$$= ½$$

Therefore, contradicts semantic security

Another View of One-Time Pad

Can be thought of as instance of Vigenère cipher

- Alphabet = **{0,1}**
- Shift by **0** means identity
- Shift by **1** means negation
- **|Message| = |Key|**

What Happens if **|Message| > |Key|**?

Use Vigenère's convention: repeat key bits as necessary

Example:

```
k = 00110001100011000110
m = 10010101011001010010
c = 10100100111010010100
```

Does this satisfy perfect secrecy/semantic security?

What Happens if $|\mathbf{Message}| > |\mathbf{Key}|$?

No perfect secrecy/semantic security

Example:

- $\mathbf{m_0 = 0^{|k|}0}$
- $\mathbf{m_1 = 0^{|k|}1}$
- $\mathbf{Enc(k,m_0)}$ will always have first and last bit identical
- $\mathbf{Enc(k,m_1)}$ will always have first and last bit different
- Therefore, distributions are not the same

Variations

$K = M = C = \mathbb{Z}_N := \{0, 1, \ldots , N-1\}$

$\mathrm{Enc}(k, m) = (m + k) \bmod N$
$\mathrm{Dec}(k, c) = (c - k) \bmod N$

Correctness:

$$\begin{aligned}\mathrm{Dec}(k, \mathrm{Enc}(k, m)) &= (m + k)-k \bmod N \\ &= m \bmod N\end{aligned}$$

Security:

$$\mathrm{Enc}(K,m_0) \stackrel{d}{=} \mathbb{Z}_N \stackrel{d}{=} \mathrm{Enc}(K,m_1)$$

Variations

$K = M = C = \mathbb{Z}_N^* := \{x \in \mathbb{Z}_N : GCD(x,N)=1\}$

$Enc(k, m) = (m \times k) \bmod N$
$Dec(k, c) = (c/k) \bmod N$

Correctness:

$$\begin{aligned} Dec(k, Enc(k, m)) &= (m \times k)/k \bmod N \\ &= m \bmod N \end{aligned}$$

Security:

$$Enc(K, m_0) \stackrel{d}{=} \mathbb{Z}_N^* \stackrel{d}{=} Enc(K, m_1)$$

Other Examples

$\mathbf{K} = \mathbf{M} = \{0,1\}^n$, $\mathbf{C} = \{0,1\}^{3n}$

Enc(k,m) =

For $i=1,\ldots,n$:

Let $t_i = m_i \oplus k_i$

Choose random bits c_{i1}, c_{i2}, c_{i3}

such that $c_{i1} \oplus c_{i2} \oplus c_{i3} = t_i$

Output $c_{11}c_{12}c_{13}c_{21}c_{22}c_{23}\ldots$

Dec(k,c) =

For $i=1,\ldots,n$:

Let $m_i = c_{i1} \oplus c_{i2} \oplus c_{i3} \oplus k_i$

Other Examples

$K = M = \{0,1\}^n$, $C = \{0,1\}^{3n}$

Enc(k,m) =

For i

c_{i2}, c_{i3}

${}_{i3} = t_i$

Outp

Example:

$k =$ 0011010110

$m =$ 1001010101

$c =$ 1100011100
1010101101
1100110010

Dec(k,c) =

For $i=1,\dots,n$:

Let $m_i = c_{i1} \oplus c_{i2} \oplus c_{i3} \oplus k_i$

Probabilistic Functions

Intuition: a function that flips random coins

Definition: A probabilistic function $\mathbf{f}$ with domain $\mathbf{X}$ and co-domain $\mathbf{Y}$ is a function $\mathbf{F:X \rightarrow Dist(Y)}$.

We will write $\mathbf{f(x)}$ to denote running $\mathbf{D \leftarrow F(x)}$, and then choosing a random sample $\mathbf{y}$ according to $\mathbf{D}$

Example:

- $\mathbf{f:\{0,1\} \rightarrow \{0,1\}^3}$, on input $\mathbf{b}$, choose random $\mathbf{c_1, c_2, c_3}$ such that $\mathbf{b = c_{i1} \oplus c_{i2} \oplus c_{i3}}$

Probabilistic Functions

Intuition: a function that flips random coins

Definition: A probabilistic function **f** with domain **X** and co-domain **Y** is a function **F:X → Dist(Y)**.

We will write **f(x)** to denote running **D←F(x)**, and then choosing a random sample **y** according to **D**

In cryptography, we generally allow all functions/algorithms to be probabilistic, including cryptosystem procedures and adversaries

Randomized Encryption

Syntax:

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

Correctness:

- ~~For all $\mathbf{k \in K,\ m \in M,\ Dec(k,\ Enc(k,m)\ ) = m}$~~

Randomized Encryption

Syntax:

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

Correctness:

- For all $\mathbf{k \in K, \; m \in M,}$

$$\mathbf{Pr[\; Dec(k, \; Enc(k,m) \;) = m \;] \quad = \quad 1}$$

Back To Our Example

$K = M = \{0,1\}^n$, $C = \{0,1\}^{3n}$

Enc(k,m) =

For $i=1,\ldots,n$:

Let $t_i = m_i \oplus k_i$

Choose random bits c_{i1}, c_{i2}, c_{i3}

such that $c_{i1} \oplus c_{i2} \oplus c_{i3} = t_i$

Output $c_{11}c_{12}c_{13}c_{21}c_{22}c_{23}\ldots$

Dec(k,c) =

For $i=1,\ldots,n$:

Let $m_i = c_{i1} \oplus c_{i2} \oplus c_{i3} \oplus k_i$

Security Proof

Distribution of **Enc(K,m)**?

- Given any ciphertext **c**, exactly one **k** that gives **Enc(k,m)=c** $(k_i = m_i \oplus c_{i1} \oplus c_{i2} \oplus c_{i3})$
- If encrypting with this **k**, prob of seeing **c** is 4^{-n}
  (For position **i**, **4** possibilities for c_{i1}, c_{i2}, c_{i3})
- $\Pr[\mathrm{Enc}(K,m) = c] = 2^{-n} \times 4^{-n} = 8^{-n}$
- Meaning **Enc(K,m)** $\stackrel{d}{=} C$

Alternate Security Proof

Let $\mathbf{f:\{0,1\} \rightarrow \{0,1\}^3}$, on input $\mathbf{b}$, choose random $\mathbf{c_1, c_2, c_3}$ such that $\mathbf{b = c_{i1} \oplus c_{i2} \oplus c_{i3}}$

$$\mathbf{Enc(k,m) = f(OTP(k,m))}$$

Therefore,

$$\begin{aligned} \mathbf{Enc(K,m_0)} &= \mathbf{f(OTP(K,m_0))} && \text{(Definition)} \\ &= \mathbf{f(OTP(K,m_1))} && \text{(OTP Security)} \\ &= \mathbf{Enc(K,m_1)} && \text{(Definition)} \end{aligned}$$

Example: Homophonic Substitution

$\mathbf{M} = \{0,1\}^n$
$\mathbf{C} = \Gamma^n$ where $\Gamma = \{A,B,C,\ldots,Z\}$
$\mathbf{K} =$ Partition of Γ into two sets S_0,S_1 of size 13

$\mathbf{Enc(k,m)} =$

For $i=1,\ldots, n$, let c_i be random element of S_{m_i}

$\mathbf{Dec(k,c)} =$

For $i=1,\ldots, n$: find c_i in S_b, let $m_i = b$

Example: Homophonic Substitution

$\mathbf{M} = \{0,1\}^n$

Example:

$\mathbf{k} = \{S_0 = \{$B,D,H,I,J,L,O,P,R,T,V,W,Z$\}$,

$S_1 = \{$A,C,E,F,G,K,M,N,Q,S,U,X,Y$\}\ \}$

$\mathbf{m} =$ 1001010101

$\mathbf{c} =$ MJWEZKPXOA

For $\mathbf{i=1,\ldots,\ n}$: find $\mathbf{c_i}$ in $\mathbf{S_b}$, let $\mathbf{m_i = b}$

Example: Homophonic Substitution

Q: Does hom. substitution have perfect secrecy?

A: NO

Proof:

$m_0 = 00$, $m_1 = 01$

$\Pr[\mathrm{Enc}(K,m_0)$ repeats$] = 1/13$

$\Pr[\mathrm{Enc}(K,m_1)$ repeats$] = 0$

Variable-Length Messages

So far, assumed all messages are same length

Not reasonable in practice

Likely want to allow variable-length messages

$$\mathcal{M} = \{0,1\}^* \text{ or } \{0,1\}^{\leq n}$$

Variable-Length OTP

Key space $\mathbf{K} = \{0,1\}^n$
Message space $\mathbf{M} = \{0,1\}^{\leq n}$
Ciphertext space $\mathbf{C} = \{0,1\}^{\leq n}$

$$\mathbf{Enc}(k, m) = k_{[1, |m|]} \oplus m$$
$$\mathbf{Dec}(k, c) = k_{[1, |m|]} \oplus c$$

Example:

$k = 0011010110$
$m = 10010$
$c = 10100$

Variable-Length OTP

Q: Is it secure?

A: NO, but for an unavoidable reason

Theorem: If $\mathbf{M} = \{0,1\}^*$, perfect secrecy is impossible for any encryption scheme*

Proof

Let $\mathbf{m_0} = \mathbf{0}$, let $\mathbf{t_0} = \mathbb{E}[\ |\mathbf{Enc}(\mathbf{K},\mathbf{m_0})|\ ]$, $\mathbf{u}=\mathbf{t_0}+\mathbf{3}$

Claim: $\mathbb{E}[\ |\mathbf{Enc}(\mathbf{K},\mathbf{M})\ |\ ] \geq \mathbf{u}-\mathbf{2}$, where $\mathbf{M}=\{\mathbf{0},\mathbf{1}\}^{\mathbf{u}}$

Fix **k**. Each possible **m** must map to different ciphertexts (by correctness).

At most $\mathbf{2^i}$ ciphertexts of length **i**

$\mathbb{E}[\ |\mathbf{c}|\ ] \geq \Sigma_{0\leq i<u}\ 2^{i-u}\times i \geq u-2$

Therefore, $\exists\, \mathbf{m_1} \in \mathbf{M}$ where

$\mathbf{t_1} := \mathbb{E}[\ |\mathbf{Enc}(\mathbf{K},\mathbf{m_1})|\ ] \geq \mathbf{u}-\mathbf{2} = \mathbf{t_0}+\mathbf{1}$

Thus, $\mathbf{Enc}(\mathbf{K},\mathbf{m_0}) \overset{d}{\neq} \mathbf{Enc}(\mathbf{K},\mathbf{m_1})$

Variable-Length OTP

Q: Is it secure?
A: NO, but for an unavoidable reason

Theorem: If $\mathbf{M} = \{0,1\}^*$, perfect secrecy is impossible for any encryption scheme*

Message length always leaked to some extent

Therefore, we will explicitly leak message length in security definition

* Assuming finite expected message length

(Perfect) Semantic Security for Variable Length Messages

Definition: A scheme **(Enc,Dec)** is **(perfectly) semantically secure** if, for all:

- Distributions **D** on **M**
- (Probabilistic) Functions $\mathbf{I:M \rightarrow \{0,1\}^*}$
- (Probabilistic) Functions $\mathbf{f:M \rightarrow \{0,1\}^*}$
- (Probabilistic) Functions $\mathbf{A:C \times \{0,1\}^* \rightarrow \{0,1\}^*}$

There exists (probabilistic) func $\mathbf{S:\{0,1\}^* \rightarrow \{0,1\}^*}$ s.t.

$$\Pr[\, A(\, Enc(k,m)\, ,\, I(m)\,) = f(m)\,]$$
$$= \Pr[\, S(\, I(m),\, |m|\,) = f(m)\,]$$

where probabilities are taken over $\mathbf{k \leftarrow K,\ m \leftarrow D}$

Perfect Secrecy For Variable Length Messages

Definition: A scheme **(Enc,Dec)** has **perfect secrecy** if, for any two messages $\mathbf{m_0}$, $\mathbf{m_1}$ where $|\mathbf{m_0}| = |\mathbf{m_1}|$,

$$\mathbf{Enc(K,\ m_0)} \stackrel{d}{=} \mathbf{Enc(K,\ m_1)}$$

Easy to adapt earlier proof to show:

Theorem: A scheme **(Enc,Dec)** is semantically secure if and only if it has perfect secrecy

Encrypting Variable Length Messages

Leakage of message length unavoidable

However, this can lead to exploits:

- CRIME/BREACH attacks:
  - Leverage compression in HTTP protocol
  - Compression before encrypting
  - Higher compression means shorter ciphertext
  - Able to gain some info about plaintext by amount of compression seen

Other Limitations of OTP

<u>It is only one-time</u>

- Try to encrypt two messages, security will fail

$$\begin{aligned} \mathbf{Enc(k,m_0)} \oplus \mathbf{Enc(k,m_1)} &= \mathbf{(k \oplus m_0) \oplus (k \oplus m_1)} \\ &= \mathbf{m_0 \oplus m_1} \end{aligned}$$

<u>Key length ≥ message length</u>

- Limited use in practice: if I can securely transmit **n**-bit key, why don't I just use that to transmit **n**-bit message?

Next Week

Multiple message security

Limitations of perfect secrecy/semantic security

- $|k| \geq |m|$ is inherent

How do we fix this?

For next time: brush up on your number theory