## COS433/Math 473: Cryptography

Mark Zhandry
Princeton University
Spring 2017

## Previously on COS 433...

## Takeaway: Crypto is Hard

Designing crypto is hard, even experts get it wrong

 Just because I don't know how to break it doesn't mean someone else can't

#### Unexpected attack vectors

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

## Takeaway: Crypto is Hard

Don't design your own crypto

- You'll probably get it wrong
- Use peer-reviewed schemes instead

Actually, don't even implement your own crypt

 Instead, use well studied crypto library built and tested by many experts

### Takeaway: Need for Formalism

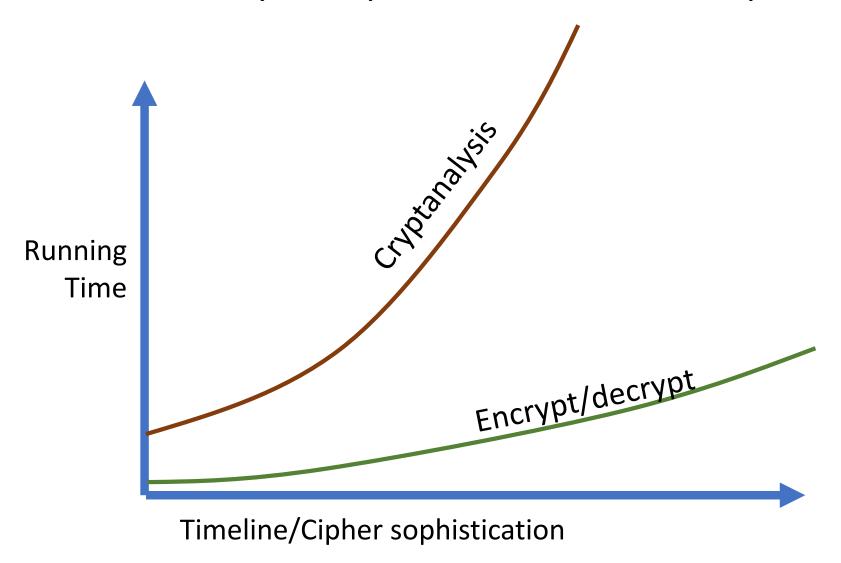
For most of history, cipher design and usage based largely on intuition

Intuition in many cases false

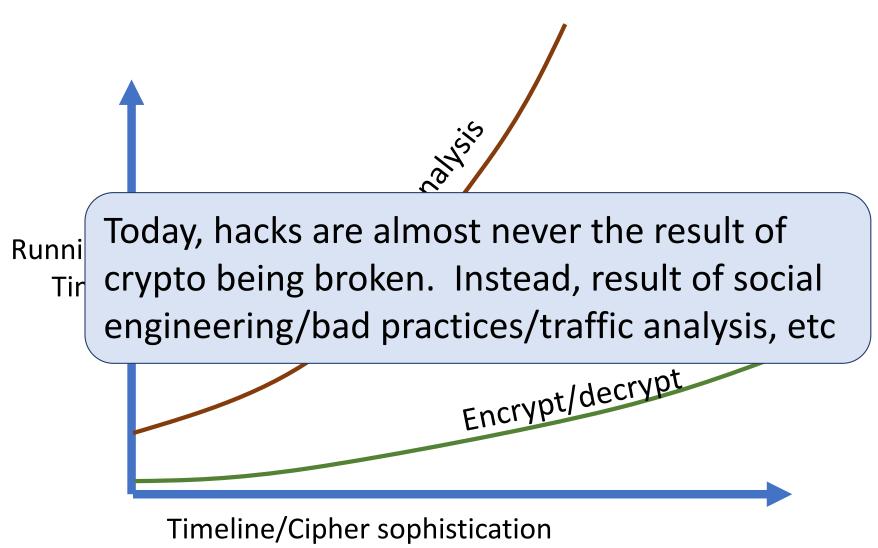
Instead, need to formally define the usage scenario

- Prove that scheme is secure in scenario
- Only use scheme in that scenario

#### Takeaway: Importance of Computers



#### Takeaway: Importance of Computers



## Modern Cryptography

## Encryption Basics (for now)

#### **Syntax:**

- Key space K (usually {0,1}<sup>λ</sup>)
- Message space M (usually {0,1}<sup>n</sup>)
- Ciphertext space C (usually {0,1}<sup>m</sup>)
- Enc:  $K \times M \rightarrow C$
- Dec: K×C → M

#### **Correctness (aka Completeness):**

• For all  $k \in K$ ,  $m \in M$ , Dec(k, Enc(k,m)) = m

#### The One-Time Pad

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

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

#### Correctness:

Dec(k, Enc(k, m)) = 
$$k\oplus(k\oplus m)$$
  
=  $(k\oplus k)\oplus m$   
=  $0\oplus m$   
=  $m$ 

### **Encryption Security?**

#### Questions to think about:

What kind of messages?

What does the adversary already know?

What information are we trying to protect?

#### **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:

What kind of messages?

What does the adversary already know?

What information are we trying to protect?

#### 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
- Model unknown key by assuming it is chosen uniformly at random

## (Perfect) Semantic Security

```
Definition: A scheme (Enc, Dec) is (perfectly)
semantically secure if, for all:
                                  Plaintext distribution
  Distributions D on M
                                    Info adv gets
 Functions I:M \rightarrow \{0,1\}^*
                                   Info adv tries to learn
 Functions f:M \rightarrow \{0,1\}^*
  Functions A:C×{0,1}*→{0,1}*
There exists a function S:\{0,1\}^* \rightarrow \{0,1\}^* such that
     Pr[A(Enc(k,m),I(m))=f(m)]
           = Pr[S(I(m)) = f(m)]
```

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

### Semantic Security

Captures what we want out of an encryption scheme

But, complicated, with many moving parts

Want: something simpler... like perfect secrecy

## 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 perfectly semantically secure if and only if it has perfect secrecy

**Corollary:** the One-Time Pad is perfectly semantically secure

#### Perfect Secrecy ⇒ Semantic Security

#### Given arbitrary:

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

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

Goal: Construct 
$$S:\{0,1\}^* \rightarrow \{0,1\}^*$$
 such that  $Pr[A(Enc(k,m), I(m)) = f(m)]$  =  $Pr[S(I(m)) = f(m)]$ 

#### Perfect Secrecy ⇒ Semantic Security

#### **S(i)**:

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

#### Semantic Security ⇒ Perfect Secrecy

Proof by contrapositive:

- Assume  $\exists m_0, m_1$  s.t.  $Enc(K, m_0) \neq enc(K, m_1)$
- 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 ⇒ Perfect Secrecy

```
Let T = \{c: Pr[Enc(K,m_1) = c] > Enc(K,m_0) = c]\}
Pr[A(Enc(K,m)) = f(m) : m \leftarrow D]
        = \frac{1}{2} Pr[A(Enc(K,m_0)) = 0]
           + \frac{1}{2} Pr[A(Enc(K,m_1)) = 1]
        = \frac{1}{2} Pr[ Enc(K,m<sub>0</sub>) \notin T]
           + \frac{1}{2} Pr[ Enc(K,m<sub>1</sub>) \in T]
        = \frac{1}{2} + \frac{1}{2} (Pr[ Enc(K,m<sub>1</sub>) \in T]
                         - Pr[ Enc(K,m<sub>0</sub>) \in T])
```

#### Semantic Security → Perfect Secrecy

```
Pr[ Enc(K,m<sub>b</sub>) \in T ]
= \Sigma_{c \in T} Pr[Enc(K,m<sub>b</sub>) = c]
= 1 - \Sigma_{c \notin T} Pr[Enc(K,m<sub>b</sub>) = c]
```

```
Pr[ Enc(K,m<sub>1</sub>) \in T] - Pr[ Enc(K,m<sub>0</sub>) \in T]

= \sum_{c \in T} Pr[Enc(K,m<sub>1</sub>) = c] - Pr[Enc(K,m<sub>0</sub>) = c]

= \sum_{c \notin T} Pr[Enc(K,m<sub>0</sub>) = c] - Pr[Enc(K,m<sub>1</sub>) = c]

= \frac{1}{2} \sum_{c} | Pr[Pr[Enc(K,m<sub>1</sub>)=c] - Pr[Enc(K,m<sub>0</sub>)=c] |
```

#### Perfect Secrecy vs Semantic Security

Perfect secrecy is much easier to reason about, so we will usually analyze schemes for perfect secrecy

However, semantic security is really the definition we care about, so always keep in mind

#### Proper Use Case for Perfect Security

- Message can come from any distribution
- Adversary can know anything about message
- Encryption hides anything
- But, definition only says something about an adversary that sees a single message
   ⇒ If two messages, no security guarantee
- Assumes no side-channels
- Assumes key is uniformly random

#### One-time Pad

We know OTP is perfectly semantically secure

But, we know it is insecure if:

- Used to encrypt multiple messages
- Key length shorter than message

## Variable-Length Messages

OTP has message-length {0,1}<sup>n</sup> where **n** is the key length

In practice, fixing the message size is unreasonable

So instead, will allow for smaller messages to be encrypted

## Variable-Length OTP

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

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

#### Correctness:

Dec(k, Enc(k, m)) = 
$$k\oplus(k\oplus m)$$
  
=  $(k\oplus k)\oplus m$   
=  $0\oplus m$   
=  $m$ 

# Does the variable length OTP have perfect secrecy according to our definition?

## Ciphertext Size

Theorem: For scheme with perfect secrecy, the expected ciphertext size for any message, **E**[ | Enc(K,m)| ], is at least (log<sub>2</sub> |M|) - 3

Fix a key **k**.

Let  $C_{k,m}$  be set of ciphertexts c s.t. Pr[Enc(k,m)=c]>0

By correctness, each  $C_{k,m}$  as  $\mathbf{m}$  varies are disjoint and non-empty

• If  $c \in C_{k,m}$  and  $c \in C_{k,m'}$ , then m' = Dec(k,c) = m

Therefore, therefore  $| \bigcup_{m} C_{k,m} | \ge |M|$ 

$$| \cup_{m} C_{k,m} | \geq |M|$$

Therefore, if we encrypt a random message, the expects size of a ciphertext is at least

$$\Sigma_{m}$$
 min(  $|c|:c \in C_{k,m}$ ) /  $|M|$ 

min( $|c|: c \in C_{k,m}$ ) = † for at most 2† different m

```
Let r = floor(log_2|M|)
```

$$\sum_{m} \min(|c| : c \in C_{k,m}) / |M|$$

$$= (1 \times 0 + 2 \times 1 + 4 \times 2 + ... + 2^{r-1} \times (r-1)$$

$$+ (|M| - (2^{r} - 1)) \times r) / |M|$$

$$= (2^{r}(r-2) + 2 + (|M| - (2^{r} - 1)) \times r) / |M|$$

$$= (r-2(2^{r} - 1) + |M| \times r) / |M|$$

$$\ge (0 - 2|M| + |M| \times r) / |M| = r-2$$

Therefore, for a random message, the expected ciphertext length for any key is at least log<sub>2</sub> |M|-3

Must also be true for a random key **k** 

```
By perfect secrecy, for any messages \mathbf{m}_0, \mathbf{m}_1

\mathbb{E}_{\mathsf{K}}[|\mathsf{Enc}(\mathsf{K},\mathsf{m}_0)|] = \mathbb{E}_{\mathsf{K}}[|\mathsf{Enc}(\mathsf{K},\mathsf{m}_1)|]
```

Therefore,  $\mathbb{E}_{K}[|\text{Enc}(K,m_{0})|]$   $= \mathbb{E}_{K,M}[|\text{Enc}(K,M)|] \ge \log_{2}|M|-3$ 

## Variable-Length Messages

For perfect secrecy of variable length messages, must have expected ciphertext length for short messages almost as long as longest messages

In practice, very undesirable

 What if I want to either send a 100mb attachment, or just a message "How are you?"

Therefore, we usually allow message length to be revealed

# (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  $I:M \rightarrow \{0,1\}^*$
- (Probabilistic) Functions **f:M→{0,1}**\*
- (Probabilistic) Functions A:C×{0,1}\*→{0,1}\*

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

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

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

# 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  $|\mathbf{m}_0| = |\mathbf{m}_1|$ ,

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

Easy to adapt earlier proof to show:

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

## Variable-Length OTP

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

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

**Theorem:** Variable-Length OTP has perfect secrecy

What if we have a **100mb** long key **k**, but encrypt only **1mb**?

Can't use first **1mb** of **k** again, but remaining **99mb** is still usable

However, basic OTP definition does not allow us to re-use the key ever

## Syntax for Stateful Encryption

#### **Syntax:**

- Key space K, Message space M, Ciphertext space C
- State Space **S**
- Init:  $\{\} \rightarrow S$
- Enc: K×M×S → C×S
- Dec: K×C×S → M×S

```
State<sub>0</sub> \leftarrow Init()
(c<sub>0</sub>, state<sub>1</sub>) \leftarrow Enc(k,m<sub>0</sub>,state<sub>0</sub>)
(c<sub>1</sub>, state<sub>2</sub>) \leftarrow Enc(k,m<sub>1</sub>,state<sub>1</sub>)
```

•••

<

m



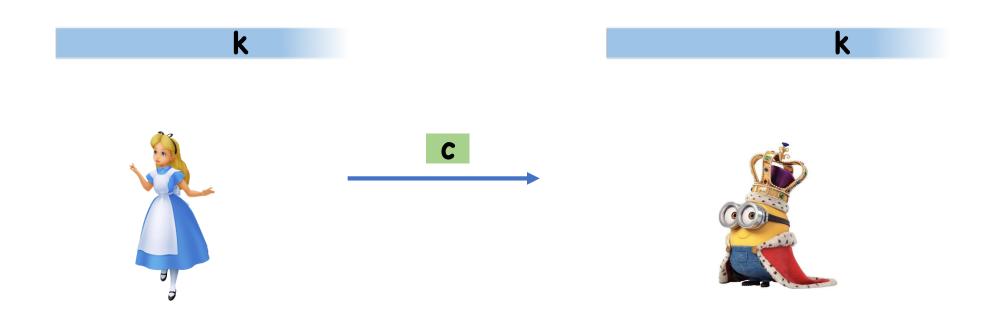




C

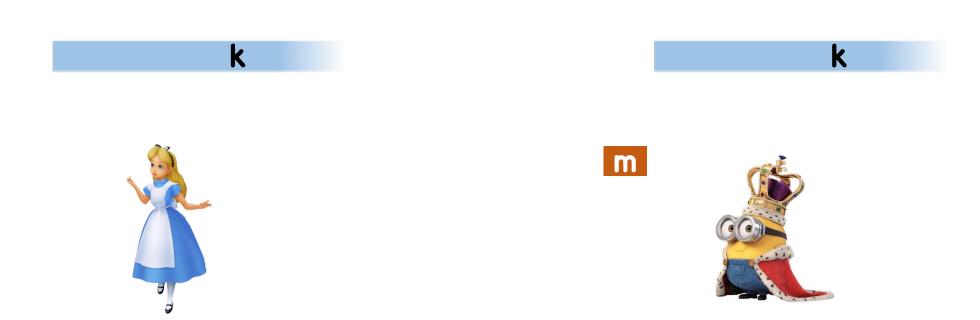






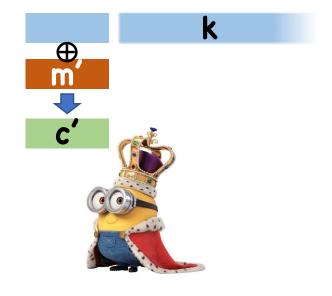


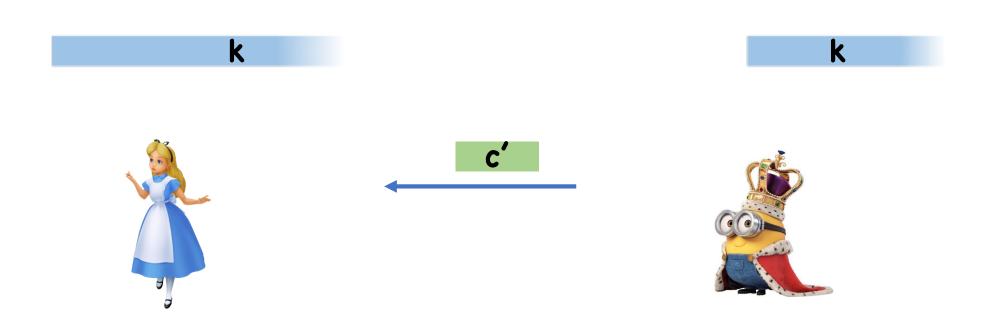




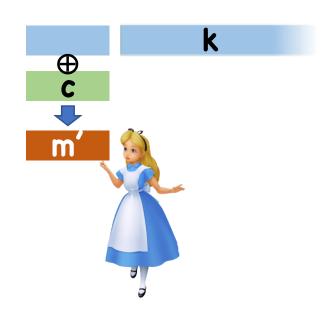














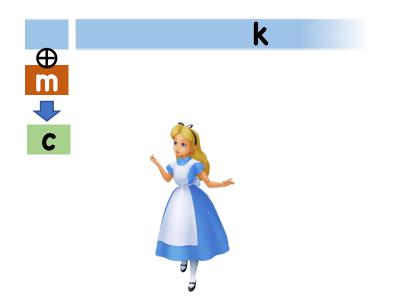


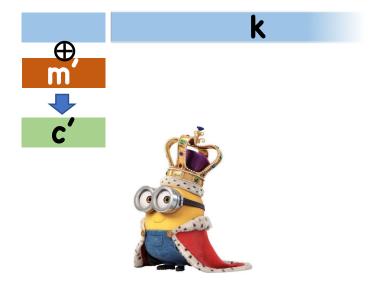
In real world, messages aren't always synchronous

What happens if Alice and Bob try to send message at the same time?

They will both use the same part of the key!

k k k m'



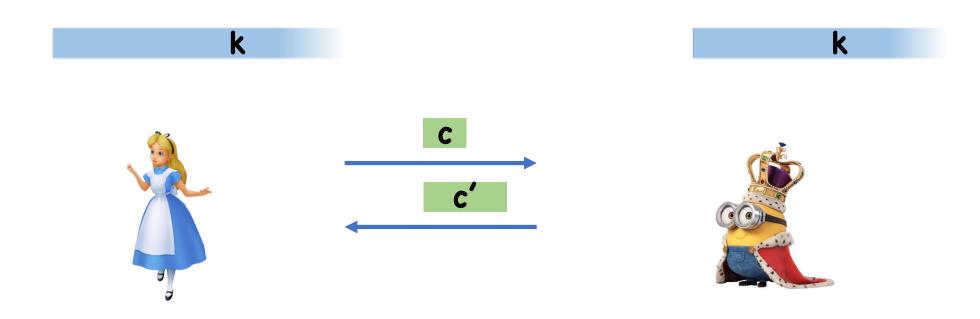


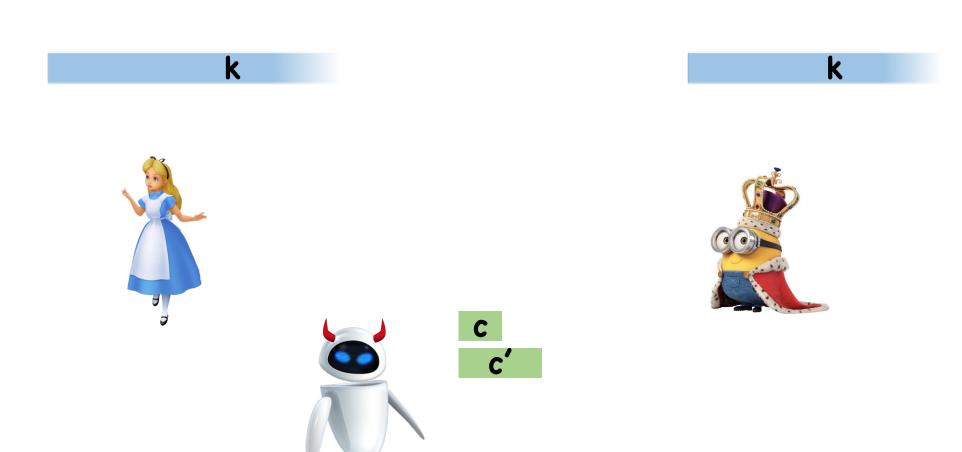
, k

C









k







#### Solution

Alice and Bob have two keys

- One for communication from Alice to Bob
- One for communication from Bob to Alice

Can obtain two logical keys from one by splitting key in half

• Ex: odd bits form  $k_{A\rightarrow B}$ , even bits form  $k_{B\rightarrow A}$ 

 $k_{A\rightarrow B}$ 

 $k_{B\rightarrow A}$ 

 $k_{A\rightarrow B}$ 

 $k_{B\rightarrow A}$ 





#### Still A Problem

In real world, messages aren't always synchronous

Also, sometimes messages arrive out of order or get dropped

Need to be very careful to make sure decryption succeeds

These difficulties exist in any stateful encryption

 For this course, we will generally consider only stateless encryption

### Perfect Security for Multiple Messages

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

$$(Enc(K, m_0^{(i)}))_{i \in [n]} \stackrel{d}{=} (Enc(K, m_1^{(i)}))_{i \in [n]}$$

Notation:  $(f(i))_{i \in [n]} = (f(1), f(2), ..., f(n))$ 

#### Stateless Encryption with Multiple Messages

Ex:

$$M = C = \mathbb{Z}_p$$
 (p a prime)  
 $K = \mathbb{Z}_p^* \times \mathbb{Z}_p$   
 $Enc((a,b), m) = (am + b) \mod p$   
 $Dec((a,b), c) = (c-b)/a \mod p$ 

Q: Is this perfectly secure for two messages?

**Theorem:** No stateless deterministic encryption scheme can have perfect security for multiple messages

## Randomized Encryption

#### **Syntax:**

- Key space K (usually {0,1}<sup>λ</sup>)
- Message space M (usually {0,1}<sup>n</sup>)
- Ciphertext space C (usually {0,1}<sup>m</sup>)
- Enc: K×M → C (potentially probabilistic)
- Dec: K×C → M (usually deterministic)

#### **Correctness:**

· For all k⊂K, m⊂M, Dec(k, Enc(k,m) ) = m

## Randomized Encryption

#### **Syntax:**

- Key space **K** (usually  $\{0,1\}^{\lambda}$ )
- Message space M (usually {0,1}<sup>n</sup>)
- Ciphertext space C (usually {0,1}<sup>m</sup>)
- Enc: K×M → C (potentially probabilistic)
- Dec: K×C → M (usually deterministic)

#### **Correctness:**

• For all  $k \in K$ ,  $m \in M$ , Pr[ Dec(k, Enc(k,m)) = m] = 1

#### Stateless Encryption with Multiple Messages

Ex:

$$M = \mathbb{Z}_p$$
 (p a prime)  
 $C = \mathbb{Z}_p^2$   
 $K = \mathbb{Z}_p^2$   
 $Enc( (a,b), m) = (r, (ar+b) + m )$   
 $Dec( (a,b), (r,c) ) = c - (ar+b)$ 

Q: Is this perfectly secure for two messages?

## Proof of Easy Case

Let (Enc, Dec) be stateless, deterministic

Let 
$$\mathbf{m}_0^{(0)} = \mathbf{m}_0^{(1)}$$
  
Let  $\mathbf{m}_1^{(0)} \neq \mathbf{m}_1^{(1)}$ 

Consider distributions of encryptions:

• ( 
$$c^{(0)}$$
 ,  $c^{(1)}$  ) = (  $Enc(K, m_0^{(0)})$ ,  $Enc(K, m_0^{(1)})$ )

 $\Rightarrow c^{(0)} = c^{(1)}$  (by determinism)

• (  $c^{(0)}$  ,  $c^{(1)}$  ) = (  $Enc(K, m_1^{(0)})$ ,  $Enc(K, m_1^{(1)})$ )

 $\Rightarrow c^{(0)} \neq c^{(1)}$  (by correctness)

### Generalize to Randomized Encryption

Let (Enc, Dec) be stateless, deterministic

Let 
$$\mathbf{m}_0^{(0)} = \mathbf{m}_0^{(1)}$$
  
Let  $\mathbf{m}_1^{(0)} \neq \mathbf{m}_1^{(1)}$ 

Consider distributions of encryptions:

• ( 
$$c^{(0)}$$
 ,  $c^{(1)}$  ) = (  $Enc(K, m_0^{(0)})$ ,  $Enc(K, m_0^{(1)})$ )  
• (  $c^{(0)}$  ,  $c^{(1)}$  ) = (  $Enc(K, m_1^{(0)})$ ,  $Enc(K, m_1^{(1)})$ )  
•  $c^{(0)} \neq c^{(1)}$  (by correctness)

### Generalize to Randomized Encryption

$$(c^{(0)}, c^{(1)}) = (Enc(K, m), Enc(K, m))$$

$$Pr[c^{(0)} = c^{(1)}]$$
?

- Fix **k**
- Conditioned on k,  $c^{(0)}$ ,  $c^{(1)}$  are two independent samples from same distribution Enc(k, m)

Lemma: Given any distribution D over a finite set X,  $Pr[Y=Y': Y\leftarrow D, Y'\leftarrow D] \ge 1/|X|$ 

• Therefore,  $Pr[c^{(0)} = c^{(1)}]$  is non-zero

### Generalize to Randomized Encryption

Let (Enc, Dec) be stateless, deterministic

Let 
$$\mathbf{m}_0^{(0)} = \mathbf{m}_0^{(1)}$$
  
Let  $\mathbf{m}_1^{(0)} \neq \mathbf{m}_1^{(1)}$ 

Consider distributions of encryptions:

• ( 
$$c^{(0)}$$
 ,  $c^{(1)}$  ) = (  $Enc(K, m_0^{(0)})$ ,  $Enc(K, m_0^{(1)})$ )
$$\Rightarrow Pr[c^{(0)} = c^{(1)}] > 0$$
• (  $c^{(0)}$  ,  $c^{(1)}$  ) = (  $Enc(K, m_1^{(0)})$ ,  $Enc(K, m_1^{(1)})$ )
$$\Rightarrow Pr[c^{(0)} = c^{(1)}] = 0$$

### 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<sup>-80</sup>, 2<sup>-128</sup>, or maybe 2<sup>-258</sup>

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

#### Reminders

HW1 Due Tomorrow (Feb 13<sup>th</sup>)

Start working on PR1

Part 1 is due next week (Feb 20<sup>th</sup>)