CIS 5560

Cryptography
Lecture 8

Course website:

pratyushmishra.com/classes/cis-5560-s24/

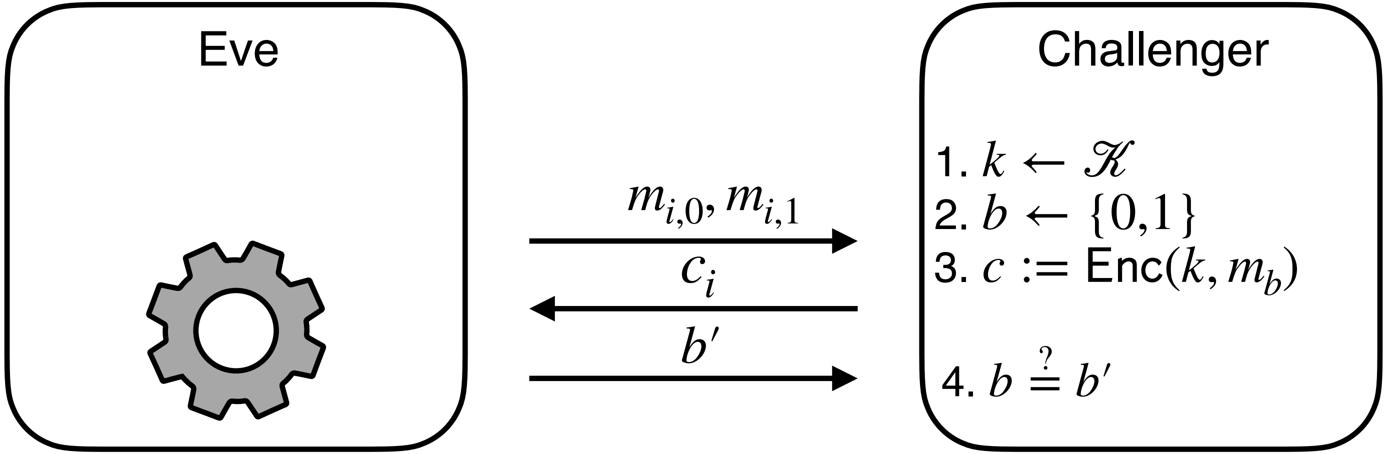
Slides adapted from Dan Boneh and Vinod Vaikuntanathan

Announcements

- **HW 4 out after lecture**
  - Due **Tuesday**, Feb 20 at 1PM on Gradescope
  - Covers PRFs, IND-CPA

Recap of last lecture

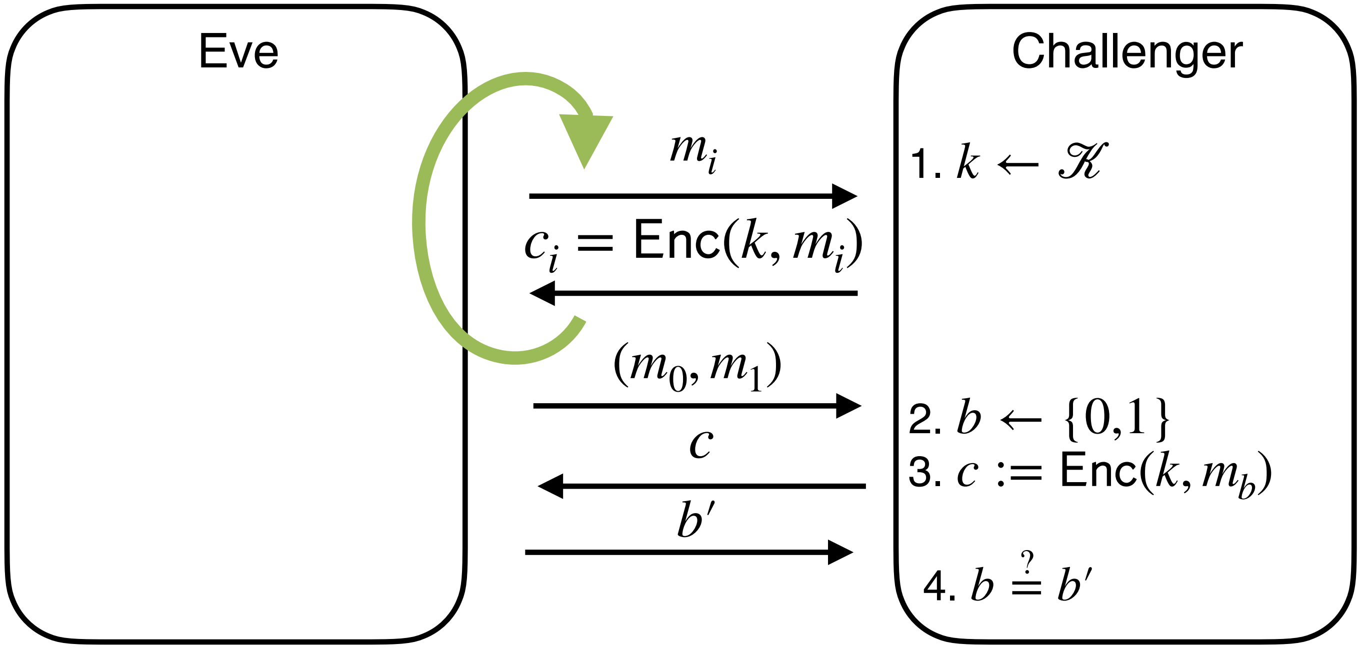
Semantic Security for Many Msgs



For every **PPT** Eve, there exists a negligible fn ε,

$$\Pr\left[\text{Eve}(c_q) = b \;\middle|\; \begin{array}{r} k \leftarrow \mathscr{K} \\ b \leftarrow \{0,1\} \\ \color{blue}{\text{For } i \text{ in } 1,\ldots,q:} \\ (m_{i,0}, m_{i,1}) \leftarrow \text{Eve}(c_{i-1}) \\ c_i = \text{Enc}(k, m_{i,b}) \end{array}\right] < \frac{1}{2} + \color{red}{\varepsilon(n)}$$

Alternate (Stronger?) definition



Also called "IND-CPA": Indistinguishability under Chosen-Plaintext Attacks

Equivalent to previous definition: just set $m_{i,0} = m_{i,1} = m_i$

Pseudorandom Functions

Collection of functions $\mathscr{F}_\ell = \{F_k : \{0,1\}^\ell \to \{0,1\}^m\}_{k \in \{0,1\}^n}$

- indexed by a key k
- n: key length, ℓ: input length, m: output length.
- Independent parameters, all poly(sec-param) = poly(n)
- #functions in $\mathscr{F}_\ell \leq 2^n$ (singly exponential in n)

$\mathbf{Gen}(1^n)$: Generate a random n-bit key k.

$\mathbf{Eval}(k, x)$ is a poly-time algorithm that outputs $F_k(x)$

Security: Cannot distinguish from random function

$$\left| \Pr\left[A^{f_k}(1^n) = 1 \mid k \leftarrow \{0,1\}^{\ell}\right] - \Pr\left[A^{F}(1^n) = 1 \mid F \leftarrow \mathsf{Fns}\right] \right| \leq \mathsf{negl}(n)\,.$$

Randomized encryption w/ PRFs

Gen(1^n): Generate a random n-bit key k that defines

$$F_k : \{0,1\}^\ell \to \{0,1\}^m$$

Enc(k, m): Pick a random x and
let the ciphertext c be the pair $(x, y = F_k(x) \oplus m)$

Dec($k, c = (x, y)$):

Output $F_k(x) \oplus c$

Indistinguishable distributions

Definition: Two distributions X and Y are *computationally indistinguishable* if for every efficient distinguisher

$$\left| \Pr[D(x) = 1 \mid x \leftarrow X] - \Pr[D(y) = 1 \mid y \leftarrow Y] \right| = \mathsf{negl}(n)$$

Denoted by $X \approx Y$

Eg: PRG security says that $X := \{G(x) | x \leftarrow \{0,1\}^n\} \approx Y := \{y | y \leftarrow \{0,1\}^m\}$

Eg: Single msg security says that

$$\{c \leftarrow \mathsf{Enc}(k, m_0) \mid k \leftarrow \mathscr{K}\} \approx \{c \leftarrow \mathsf{Enc}(k, m_1) \mid k \leftarrow \mathscr{K}\}$$

Proof by hybrid argument

$\mathsf{Enc}(k, m)$: Pick a random x and output $(x, y = F_k(x) \oplus m)$

$\mathsf{Dec}(k, c = (x, y))$: Output $F_k(x) \oplus c$

Single msg security says that the following dists are indistinguishable.

$$\{c \leftarrow \mathsf{Enc}(k, m_0) \mid k \leftarrow \mathscr{K}\} \text{ and } \{c \leftarrow \mathsf{Enc}(k, m_1) \mid k \leftarrow \mathscr{K}\}$$

How to do this? Let's create more (supposedly) indistinguishable distributions:

$$H_0 = \{c := (r, m_0 \oplus F_k(r) \mid r \leftarrow \{0,1\}^n; k \leftarrow \mathscr{K}\}$$

$\approx$ by PRF security

$$H_1 = \{c := (r, m_0 \oplus R(r) \mid r \leftarrow \{0,1\}^n; R \leftarrow \mathsf{Fns}\}$$

$\approx$ defn of random fn

$$H_2 = \{c := (r, m_0 \oplus r' \mid r \leftarrow \{0,1\}^n; r' \leftarrow \{0,1\}^n\}$$

$\approx$ one time pad

$$H_3 = \{c := (r, m_1 \oplus r' \mid r \leftarrow \{0,1\}^n; r' \leftarrow \{0,1\}^n\}$$

$\approx$ defn of random fn

$$H_4 = \{c := (r, m_1 \oplus R(r) \mid r \leftarrow \{0,1\}^n; R \leftarrow \mathsf{Fns}\}$$

$\approx$ by PRF security

$$H_5 = \{c := (r, m_1 \oplus F_k(r) \mid r \leftarrow \{0,1\}^n; k \leftarrow \mathscr{K}\}$$

Hybrid argument

The key steps in a hybrid argument are:

1. Construct a sequence of poly many distributions b/w the two target distributions.
2. Argue that each pair of neighboring distributions are indistinguishable.
3. Conclude that the target distributions are indistinguishable via contradiction:
    A. Assume the target distributions are distinguishable
    B. Must be the case that an intermediate pair of distributions is distinguishable
    C. This contradicts 2 above.

Hybrid argument

B. Must be the case that an intermediate pair of distributions is distinguishable

Lemma: Let $\cancel{p_0, p_1, p_2, \ldots, p_m}$ be advantage of distinguishing $(H_0, H_1), (H_1, H_2), \ldots, (H_{n-1}, H_n)$

If $p_0 - p_m \geq \epsilon$ there is an index i such that $p_i - p_{i+1} \geq \epsilon/m$.

Proof:

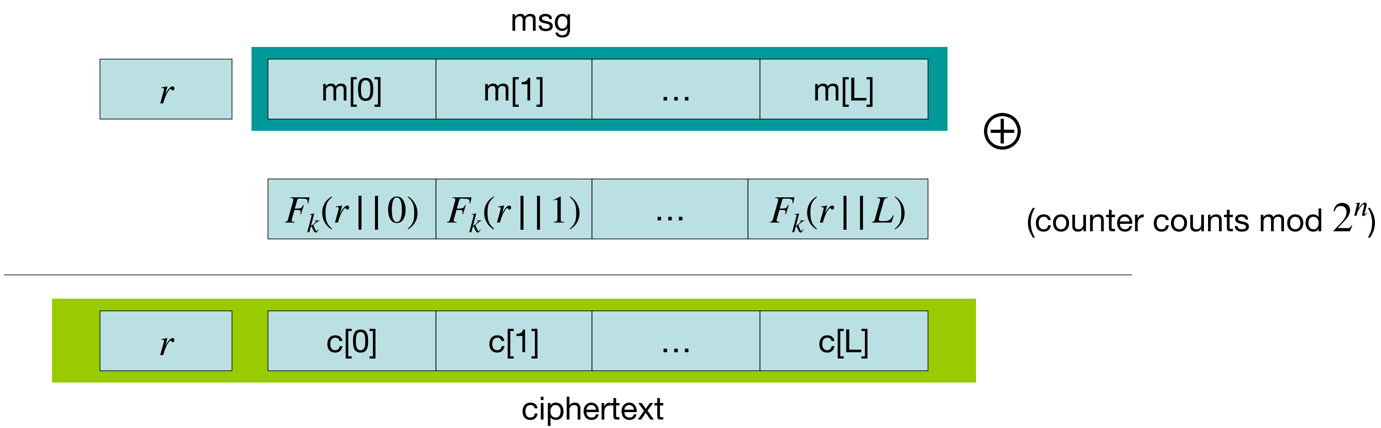
$$p_m - p_0 = (p_m - p_{m-1}) + (p_{m-1} - p_{m-2}) + \cdots + (p_1 - p_0) \geq \epsilon$$

At least one of the m terms has to be at least ε/m (averaging).

Construction 2: rand ctr-mode

F: PRF defined over (K, X, Y) where $X = \{0,1\}^{2n}$ and $Y = \{0,1\}^n$

(e.g., n=128)



r - chosen at random for every message

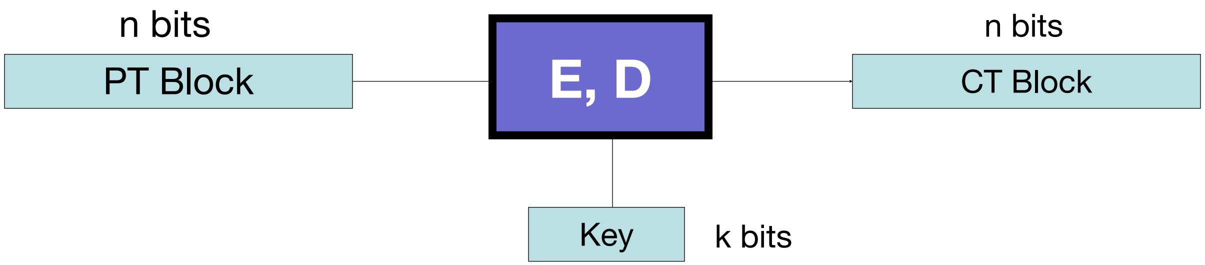
note: parallelizable

Today's Lecture

- PRPs and block cipher modes of operation
- PRGs → PRFs
- Message Integrity

Also called a Block Cipher

A **block cipher** is a pair of efficient algs. (E, D):



Canonical examples:

1. **AES**: n=128 bits, k = 128, 192, 256 bits
2. **3DES**: n= 64 bits, k = 168 bits (historical)

Running example

- Example PRPs: 3DES, AES, …

  AES128: $K \times X \to X$ where $K = X = \{0,1\}^{128}$

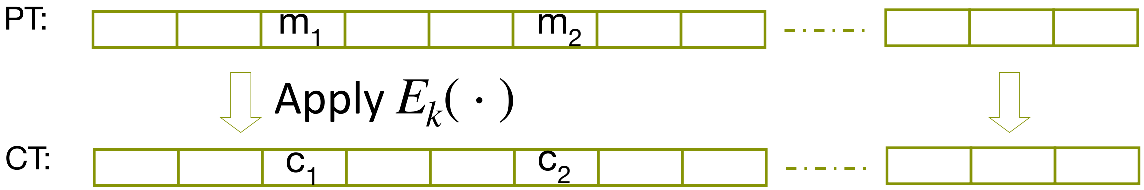
  DES: $K \times X \to X$ where $X = \{0,1\}^{64}$, $K = \{0,1\}^{56}$

  3DES: $K \times X \to X$ where $X = \{0,1\}^{64}$, $K = \{0,1\}^{168}$

- Functionally, any PRP where K and X are large is also a PRF.
  - A PRP is a PRF where X=Y and is efficiently invertible

Incorrect use of a PRP

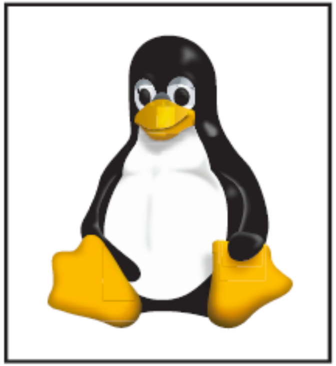
Electronic Code Book (ECB):



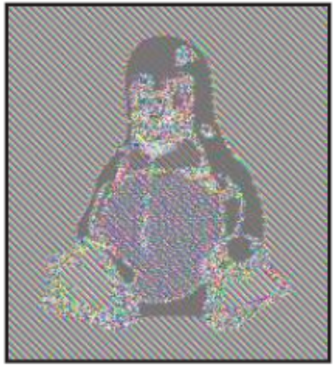
<u>Problem</u>:

– if $m_1 = m_2$ then $c_1 = c_2$

In pictures



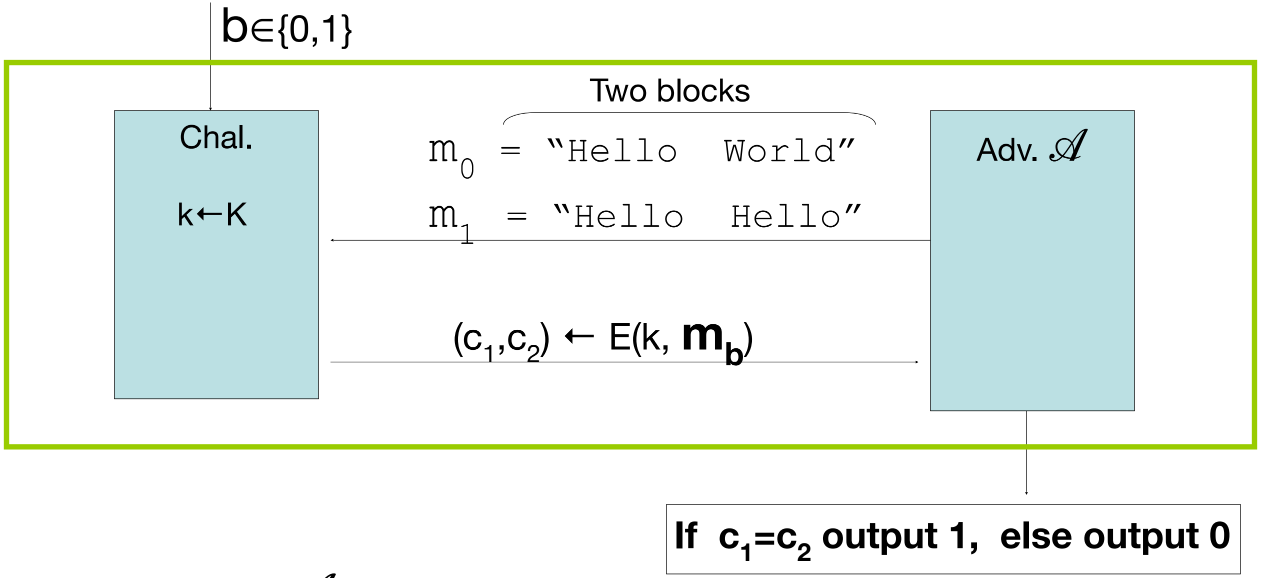
Original penguin



ECB encrypted penguin

ECB is not Semantically Secure even for 1 msg

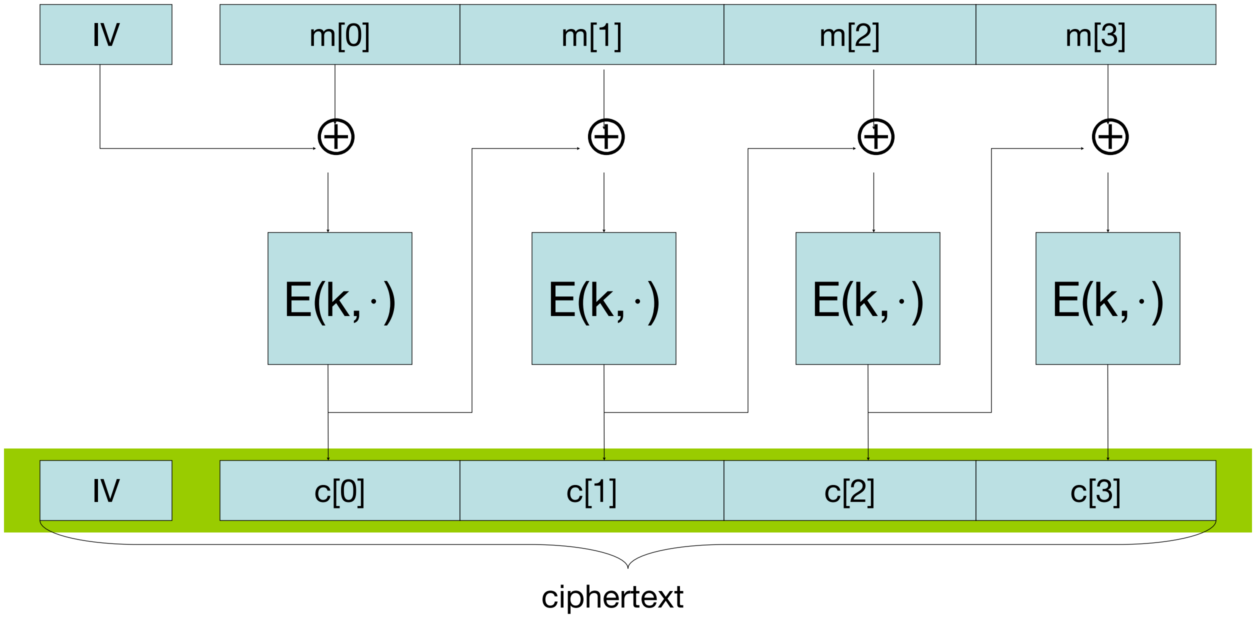
ECB is not semantically secure for messages that contain two or more blocks.



Then $\text{Adv}_{SS}[\mathscr{A}, \text{ECB}] = 1$

Secure Construction 1: CBC with random nonce

Cipher block chaining with a random IV (IV = nonce)



note: CBC where attacker can predict the IV is not CPA-secure. HW.

CBC: CPA Analysis

<u>CBC Theorem</u>: For any L>0,

If E is a secure PRP over (K,X) then

E_{CBC} is a sem. sec. under CPA over (K, X^L, X^{L+1}).

In particular, for a q-query adversary A attacking E_{CBC}

there exists a PRP adversary B s.t.:

$$\mathrm{Adv}_{CPA}[A, E_{CBC}] \leq 2 \cdot \mathrm{Adv}_{PRP}[B, E] + 2\, q^2 L^2 / |X|$$

Note: CBC is only secure as long as $q^2 \cdot L^2 \ll |X|$

messages enc. with key

max msg length

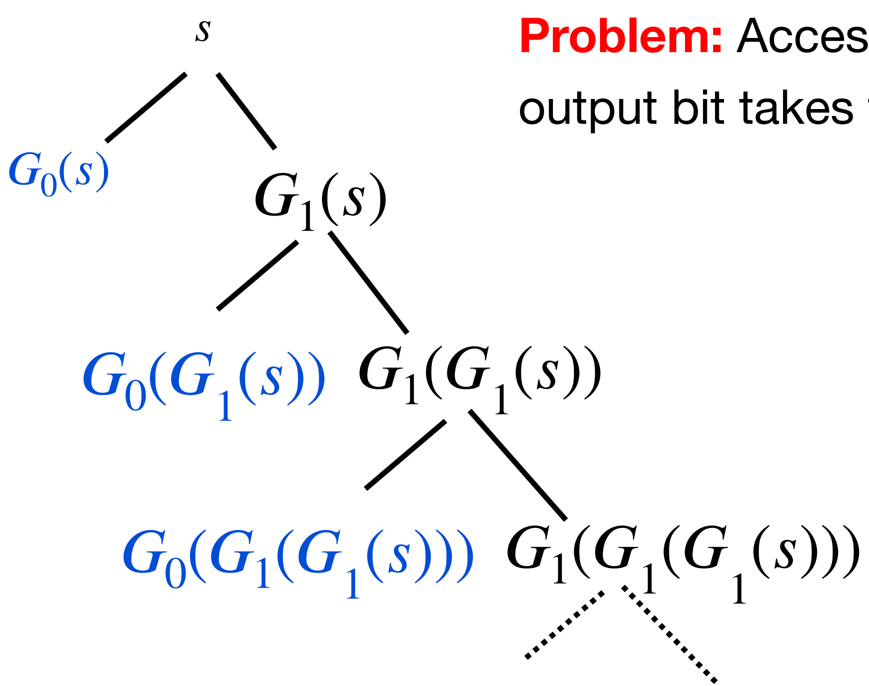
- PRPs and block cipher modes of operation
- PRGs → PRFs
- MACs, if we have time

Let's Look Back at Length Extension…

Theorem: Let G: $\{0,1\}^n \rightarrow \{0,1\}^{n+1}$ be a PRG. Then, for every polynomial m(n), there is a PRG G': $\{0,1\}^n \rightarrow \{0,1\}^{m(n)}$.

Let's Look Back at Length Extension…

Construction: Let G(s) = $G_0(s) \,||\, G_1(s)$ where $G_0(s)$ is 1 bit and $G_1(s)$ is n bits .



Problem: Accessing the i^{th} output bit takes time $\approx i$.

Goldreich-Goldwasser-Micali PRF

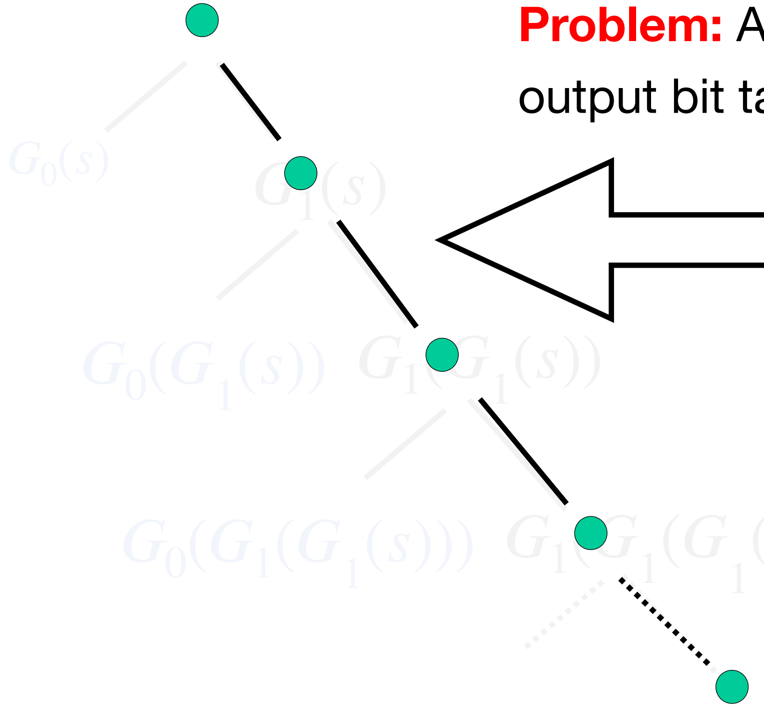
Theorem: Let G be a PRG. Then, for every polynomials $\ell = \ell(n)$, $m = m(n)$, there exists a PRF family $\mathcal{F}_\ell = \{f_s : \{0,1\}^\ell \rightarrow \{0,1\}^m\}_{s \in \{0,1\}^n}$.

Note: We will focus on $m = \ell$.
The output length could be made smaller (by truncation) or larger (by expansion with a PRG).

What is the standard way to improve

Let's Look Back at Length Extension…

Construction: Let G(s) = $G_0(s) \,||\, G_1(s)$ where $G_0(s)$ is 1 bit and $G_1(s)$ is n bits .



Problem: Accessing the i^{th} output bit takes time $\approx i$.

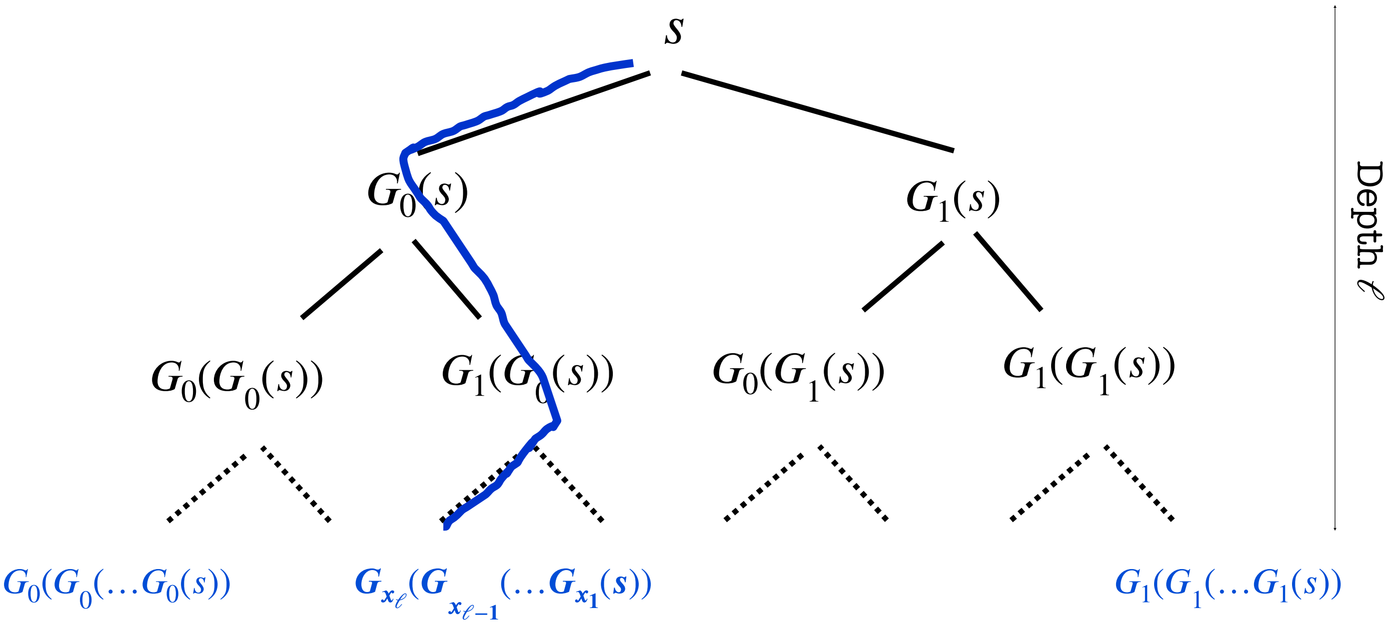
What data structure does this remind you of?

Ans: a list!
No wonder it's linear time!

What is the standard technique to do better?

Goldreich-Goldwasser-Micali PRF

Construction: Let G(s) = $G_0(s)||G_1(s)$ where $G_0(s)$ and $G_1(s)$ are both n bits each.



Each path/leaf labeled by $x \in \{0,1\}^\ell$ corresponds to $f_s(x)$.

Goldreich-Goldwasser-Micali PRF

Construction: Let G(s) = $G_0(s)||G_1(s)$ where $G_0(s)$ and $G_1(s)$ are both n bits each.

The pseudorandom function family $\mathscr{F}_\ell$ is defined by a collection of functions f_s where:

$$f_s\left(x_1x_2\ldots x_\ell\right) = \boldsymbol{G}_{\underbrace{\boldsymbol{x}_\ell}(\boldsymbol{G}_{\boldsymbol{x}_{\ell-1}}(\ldots\boldsymbol{G}_{\boldsymbol{x}_1}}_{\ell\text{-bit input}}(\boldsymbol{s}))$$

- f_s defines 2^ℓ pseudorandom bits.
- The x^{th} bit can be computed using ℓ evaluations of the PRG G (as opposed to $x \approx 2^\ell$ evaluations as before.)

PRG Repetition Lemma

Lemma: Let G be a PRG. Then, for every polynomial L=L(n), the following two distributions are computationally indistinguishable:

$$(G(s_1), G(s_2), \ldots, G(s_L)) \approx (u_1, u_2, \ldots, u_L)$$

Proof: **By Hybrid Argument.**

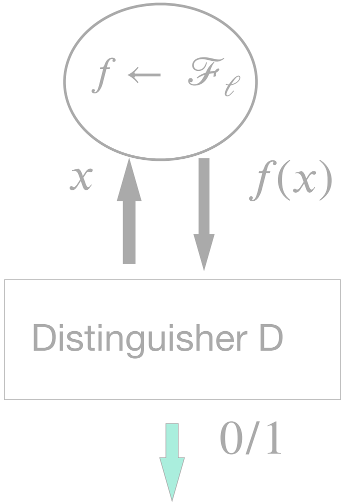
If there is a ppt distinguisher between the two distributions with distinguishing advantage ε, then there is a ppt distinguisher for G with advantage $\geq \varepsilon/L$.

GGM PRF: Proof of Security

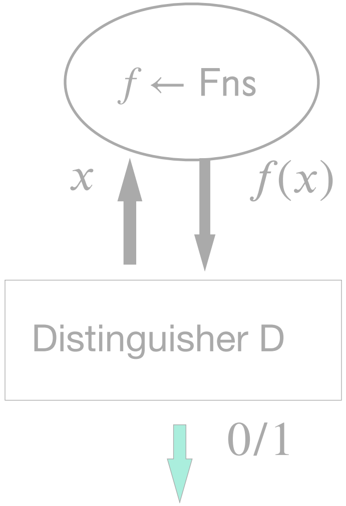
By contradiction. Assume there is a ppt D and a poly function p s.t.

$$\left| \Pr\left[A^{f_k}(1^n) = 1 \mid k \leftarrow \{0,1\}^{\ell}\right] - \Pr\left[A^{F}(1^n) = 1 \mid F \leftarrow \mathsf{Fns}\right] \right| \geq 1/p(n)\,.$$

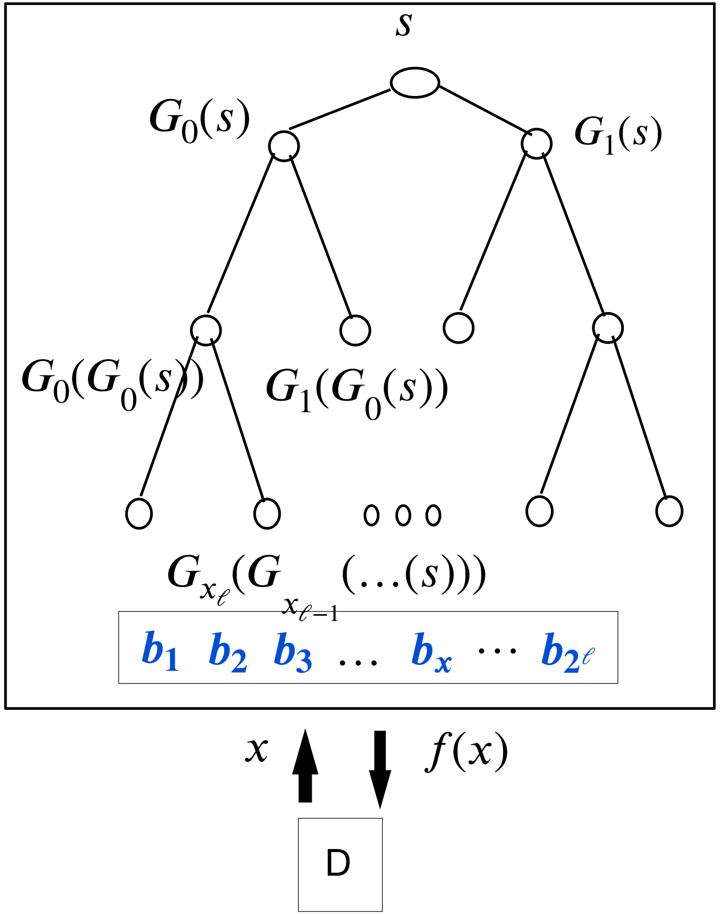
The pseudorandom world



The random world

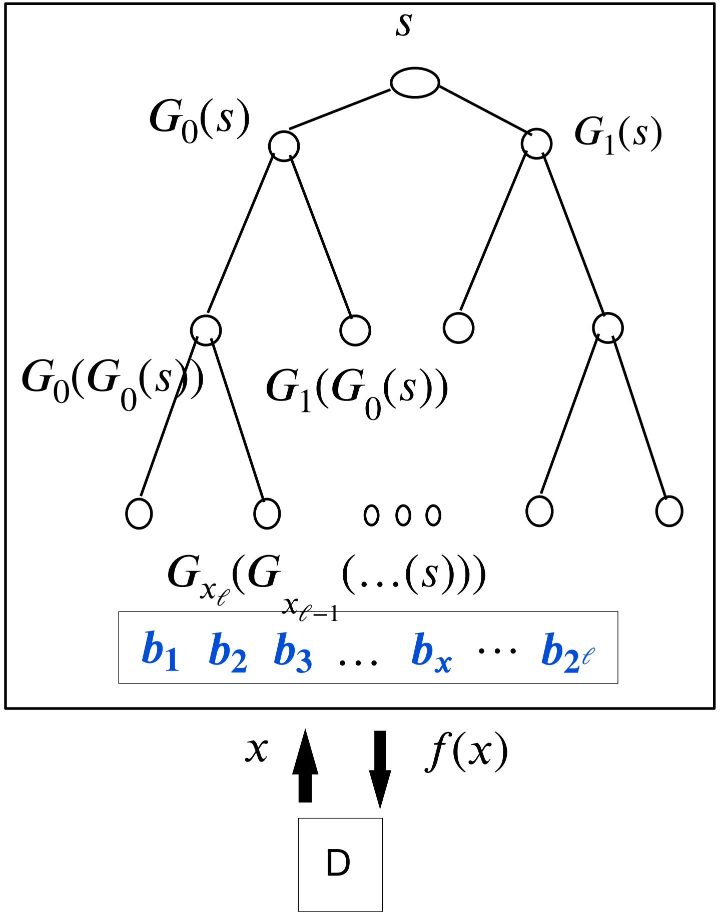


The pseudorandom world: Hybrid 0



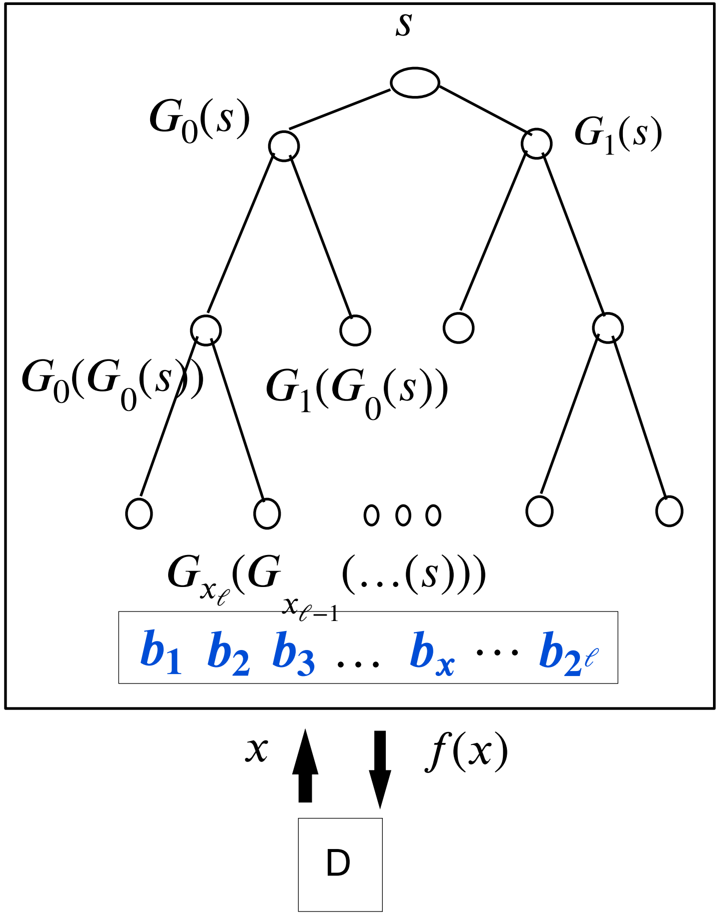
Problem: Hybrid argument on leaves doesn't work. Why?

The pseudorandom world:
Hybrid 0

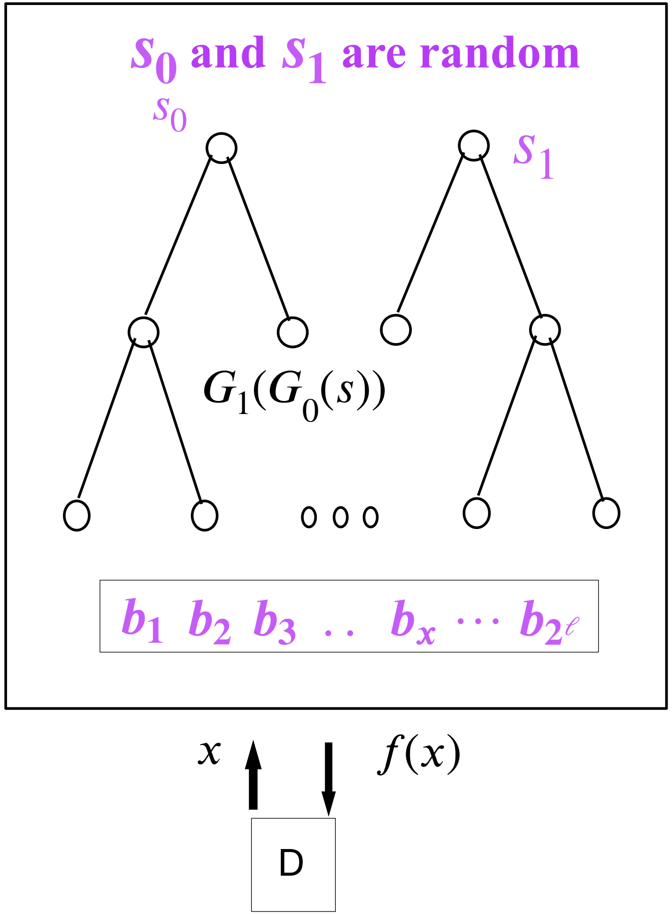


Key Idea:
Hybrid argument by levels of the tree

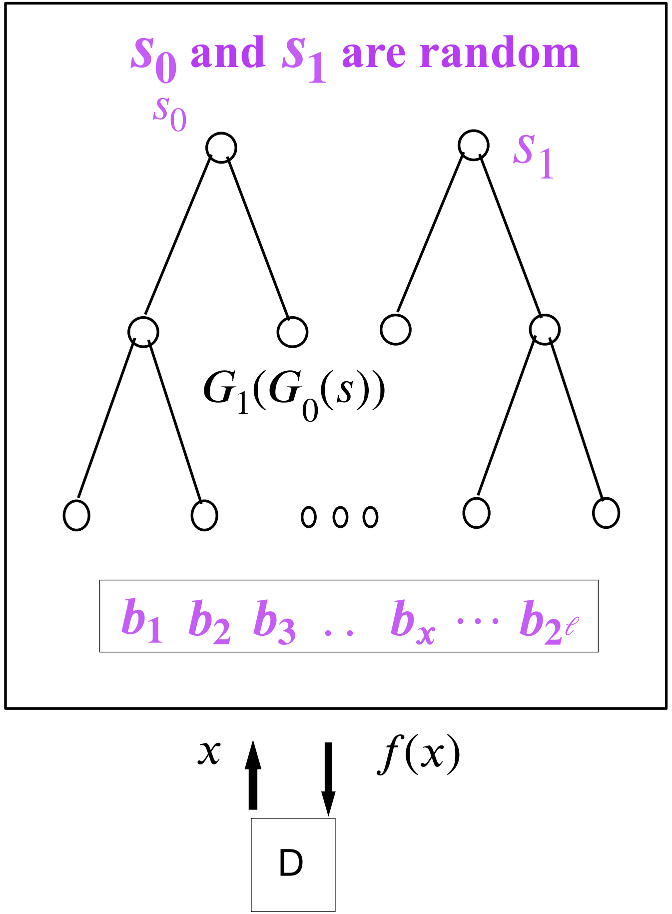
The pseudorandom world:
Hybrid 0



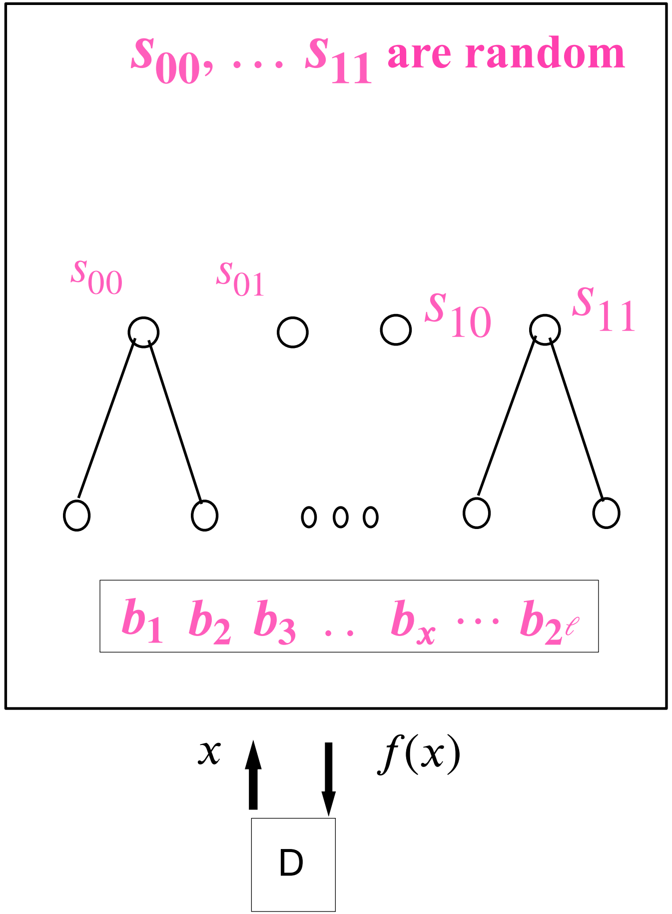
Hybrid 1



Hybrid 1

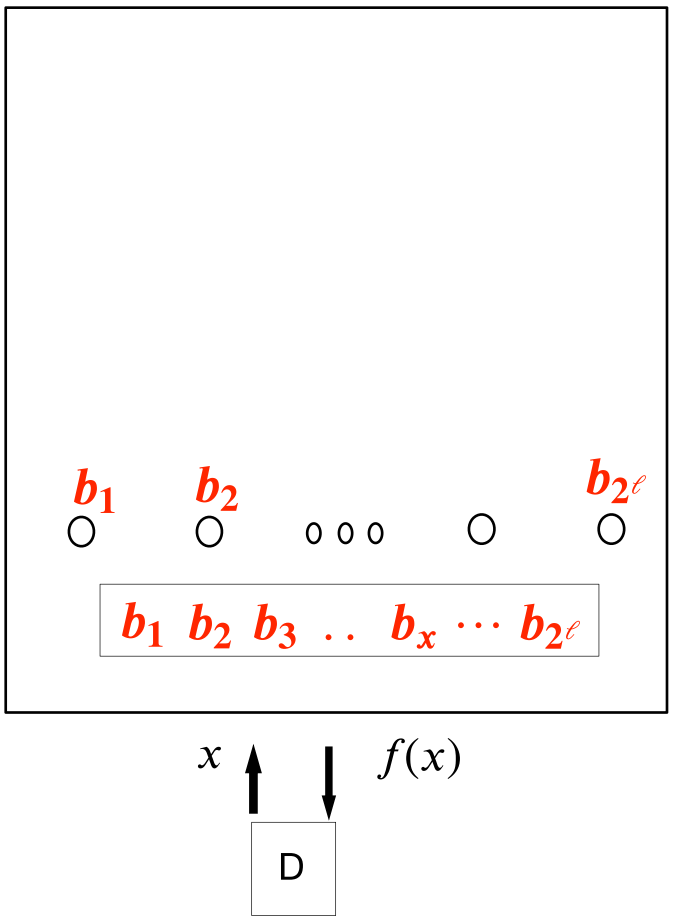


Hybrid 2

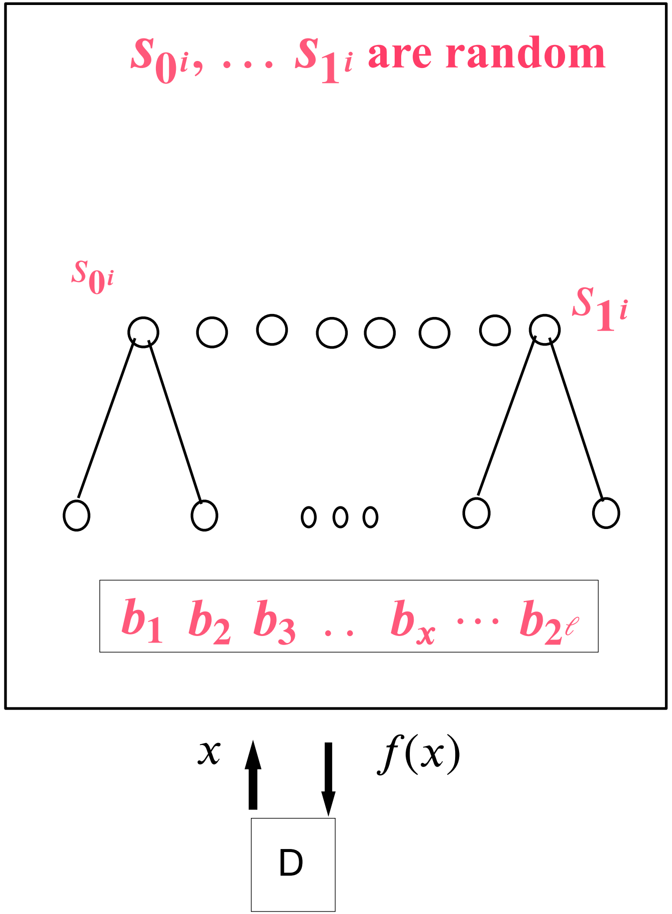


■ ■ ■

The random world:
Hybrid ℓ



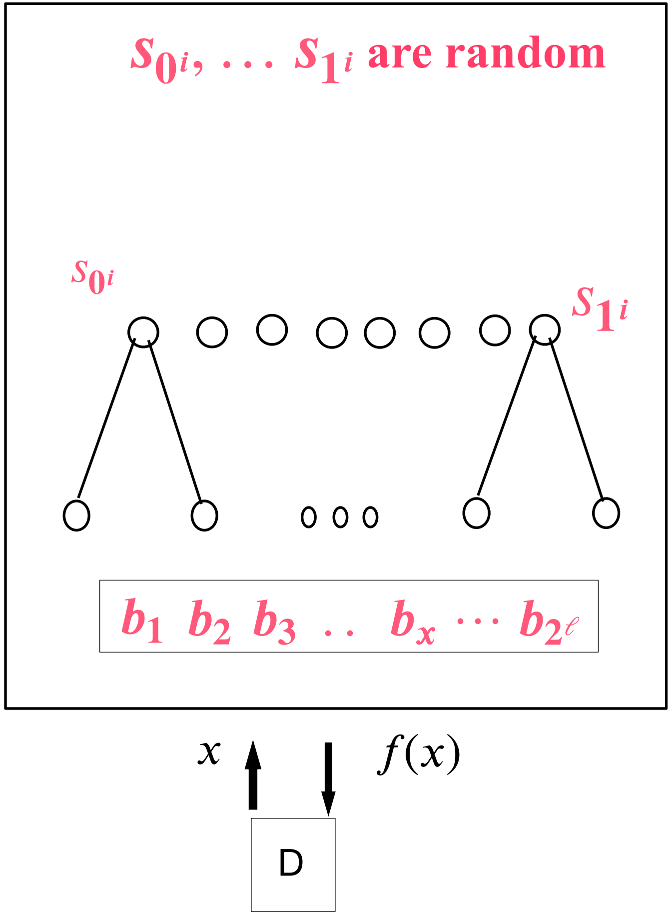
Hybrid i



Q: **Are the hybrids efficiently computable?**

A: **Yes! Lazy Evaluation.**

Hybrid i



Let $p_i = \Pr[f \leftarrow H_i : D^f(1^n) = 1]$

We know: $p_0 - p_\ell \geq \varepsilon$

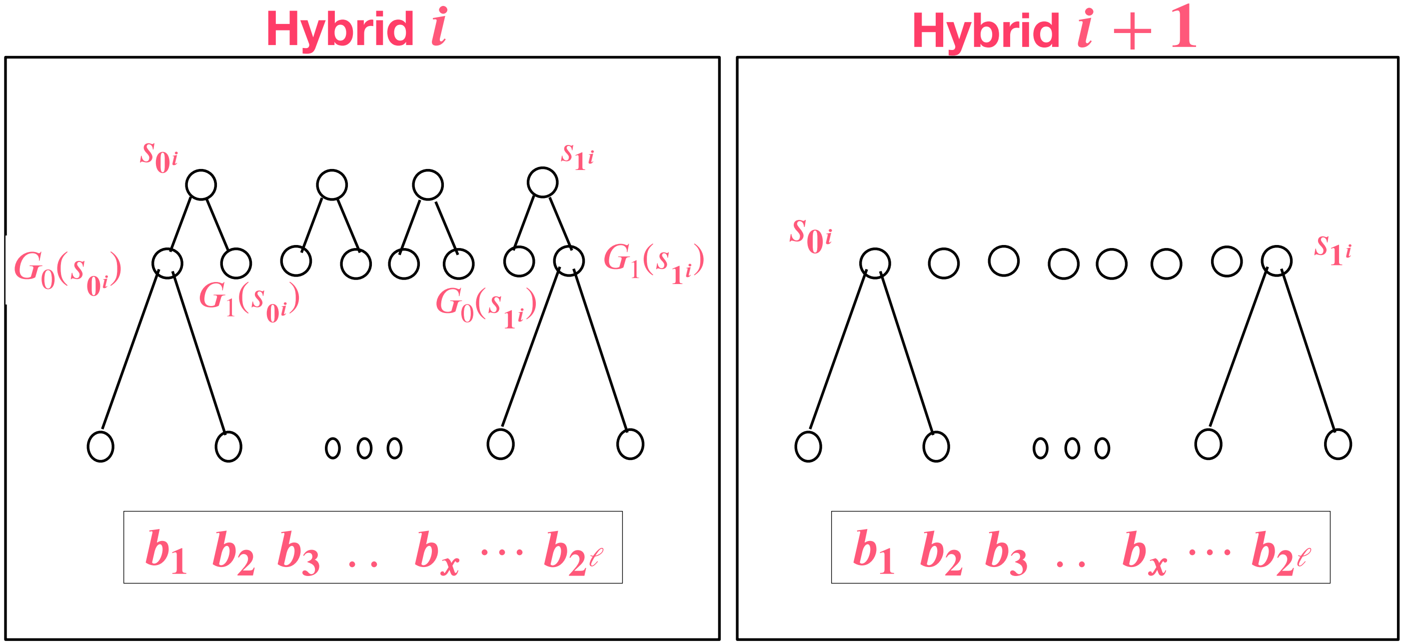
By a hybrid argument:

For some i: $p_i - p_{i+1} \geq \varepsilon / \ell$

(use the PRG repetition lemma)

A distinguisher with advantage ε/ℓ between the hybrids implies a distinguisher with advantage $\geq \varepsilon/q\ell$ for the PRG.

(where q is the number of queries that D makes)



GGM PRF

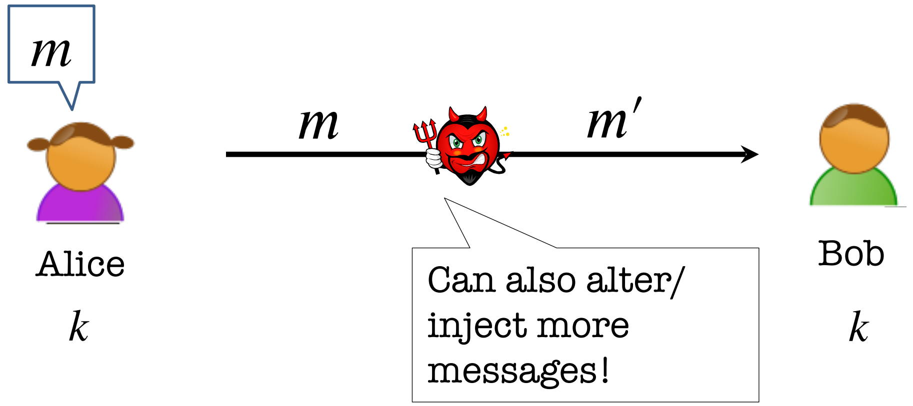
Theorem: Let G be a PRG. Then, for every polynomials ℓ, m, there exists a PRF family $\mathscr{F}_\ell = \{f_s : \{0,1\}^\ell \to \{0,1\}^m\}_{s \in \{0,1\}^n}$.

Some nits:

- *Expensive*: ℓ invocations of a PRG.
- *Sequential*: bit-by-bit, ℓ sequential invocations of a PRG.
- *Loss in security reduction*: break PRF with advantage $\varepsilon \implies$ break PRG with advantage $\varepsilon / q\ell$, where q is an arbitrary polynomial = #queries of the PRF distinguisher.

Tighter reduction? Avoid the loss?

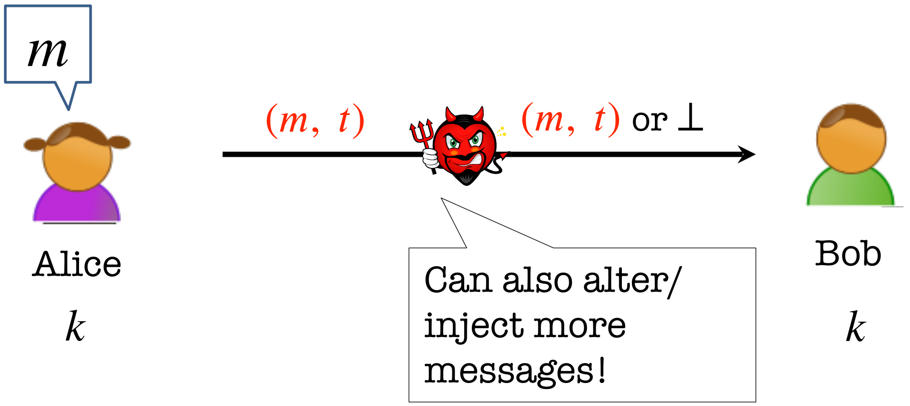
The authentication problem



This is known as a **man-in-the-middle attack.**

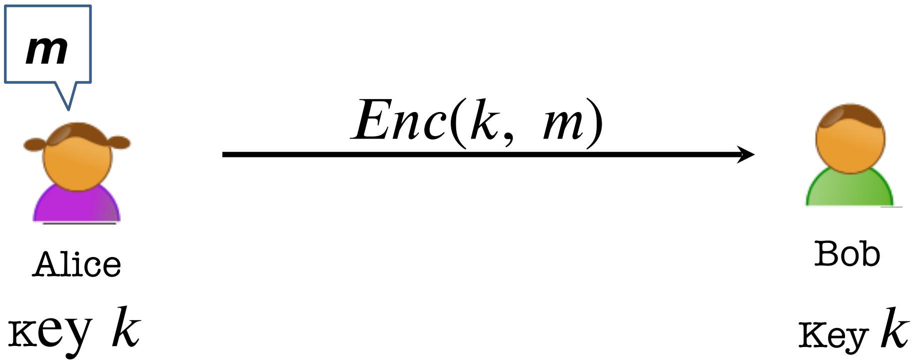
How can Bob check if the message is indeed from Alice?

The authentication problem

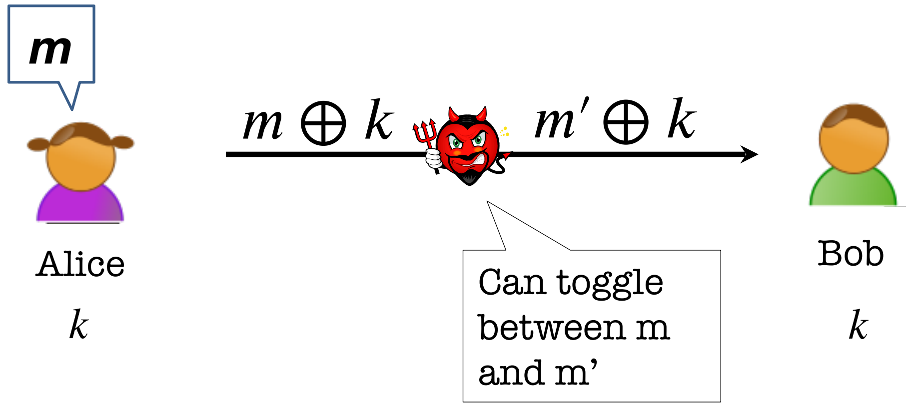


We want Alice to generate a tag for the message *m* which is **hard to generate** without the secret key *k*.

Wait… Does encryption not solve this?

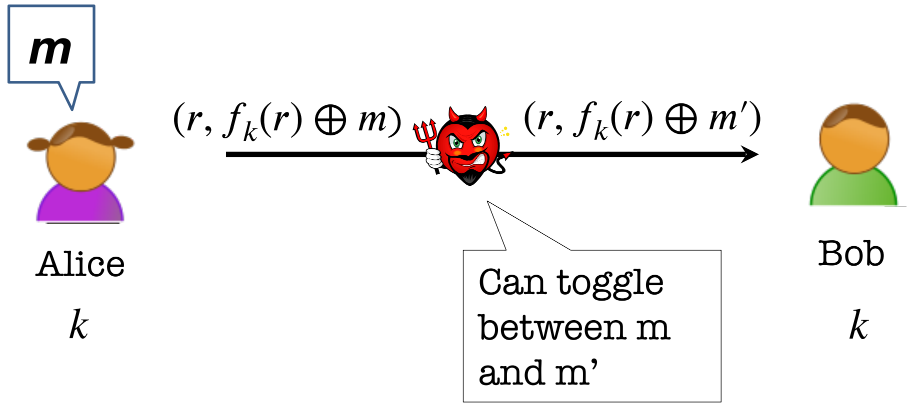


Wait… Does encryption not solve this?



One-time pad (and encryption schemes in general) are ***malleable***.

Wait... Does encryption not solve this?



One-time pad (and encryption schemes in general) are ***malleable***.

Privacy and Integrity are very **different goals!**

Message Authentication Codes (MACs)

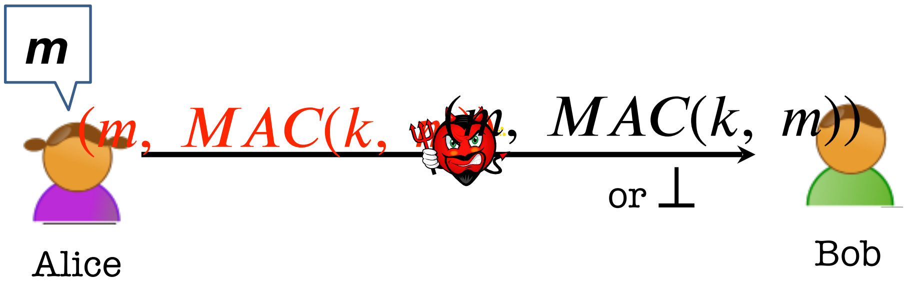
A triple of algorithms (Gen, MAC, Ver):

- Gen(1^n): Produces a key $k \leftarrow \mathcal{K}$.
- MAC$(k,\ m)$: Outputs a tag t (may be deterministic).
- Ver$(k,\ m,\ t)$: Outputs Accept or Reject.

Correctness: $\Pr[\text{Ver}(k, m, \text{MAC}(k, m)) = 1] = 1$

Security: *Hard to forge.* Intuitively, it should be hard to come up with a new pair *(m', t')* such that Ver accepts.

What is the power of the adversary?



- Can see many pairs $(m, \ MAC(k, \ m))$.
- Can access a MAC oracle $MAC(k, \ \bullet \ )$
  - Obtain tags for message of choice.

This is called a *chosen message attack (CMA).*

Defining MAC Security

- **Total break:** The adversary should not be able to recover the key *k*.
- **Universal break:** The adversary can generate a valid tag for every message.
- **Existential break:** The adversary can generate a new valid tag *t* for some message *m.*

We will require MACs to be secure against the existential break!!

EUF-CMA Security

Existentially Unforgeable against Chosen Message Attacks

$k \leftarrow K$

m_1 →

← $t_1 = \text{MAC}(k, m_1)$

m_2 →

← $t_2 = \text{MAC}(k, m_2)$

$\vdots$

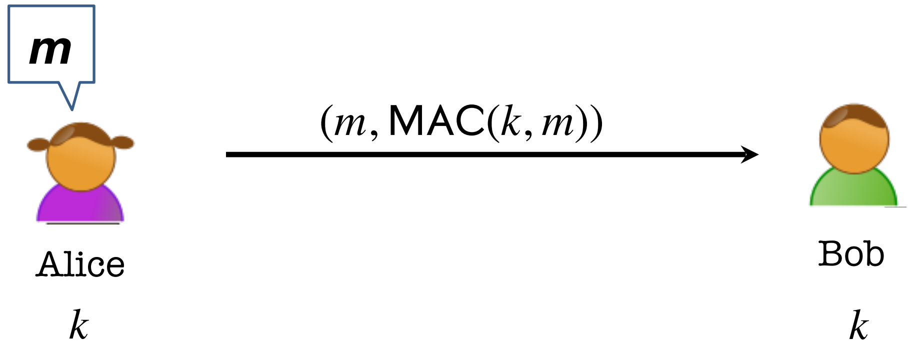
$(m,\ t)$ →

Accept if $(m,\ t) \neq (m_i,\ t_i)$ for all i, and $\text{Ver}(k, m, t) = 1$

Want: $\Pr((m,\ t) \leftarrow A^{MAC(k,\ \bullet)}(1^n),\ Ver(k,\ m,\ t) = 1,\ (m,\ t) \notin Q)) = negl(n)$.

where Q is the set of queries $\left\{ (m_i,\ t_i) \right\}_i$ that A makes.

Constructing a MAC



Gen(1^n): Produces a PRF key $k \leftarrow K$.
MAC(k, m): Output $f_k(m)$.
Ver(k, m, t): Accept if $f_k(m) = t$, reject otherwise.

Security: Our earlier unpredictability lemma about PRFs essentially proves that this is secure!

Dealing with Replay Attacks

- The adversary could send an old valid *(m, tag)* at a later time.
  – In fact, our definition of security does not rule this out.

- **In practice:**
  – Append a time-stamp to the message. Eg. (m, T, MAC(m, T)) where T = 21 Sep 2022, 1:47pm.
  – Sequence numbers appended to the message (this requires the MAC algorithm to be *stateful*).