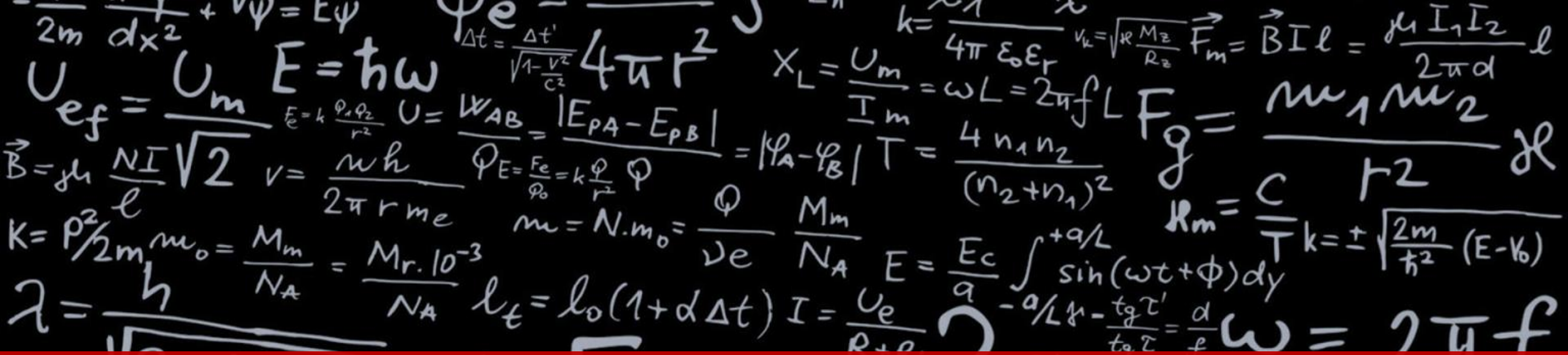


2-Definitions of security & stream ciphers

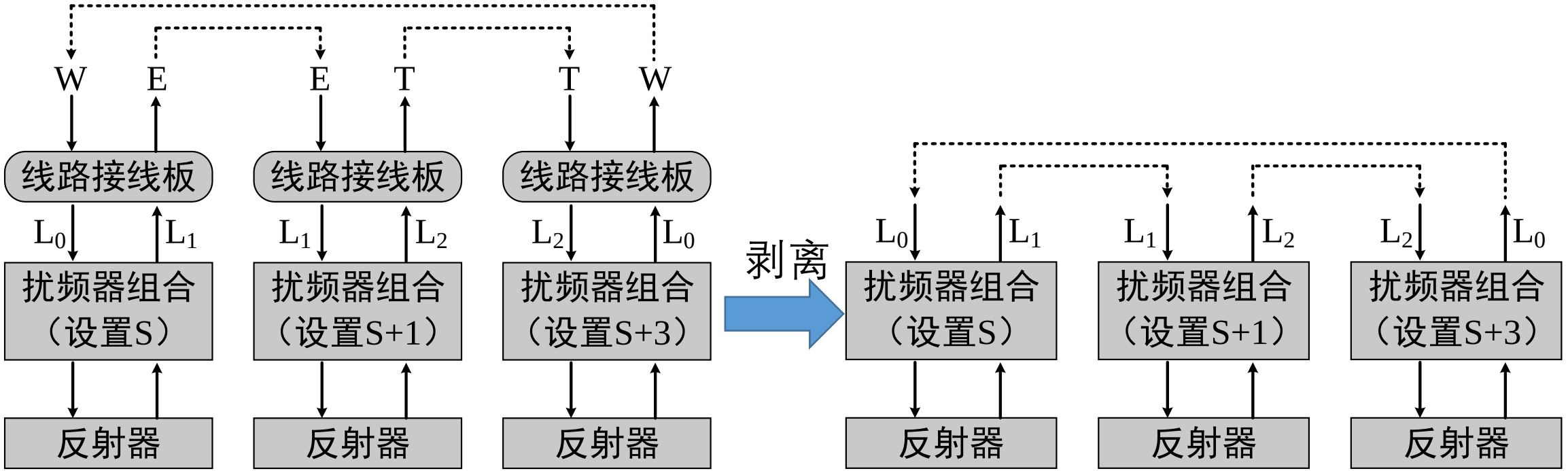
宋凌

邮箱：songlingcs@163.com

办公室：南海楼114

回顾上次课——破解Enigma

- 图灵发现特殊的明密文对 —— 形成环路
- 形成环时，可以先只考虑扰频器和反射器部分，各个击破



1.59×10^{20}种设置

1.05×10^{14}种设置

回顾上次课——关于机密性

- 足够大的密钥空间
- 密文需要看起来随机，避免出现可被利用的统计特征
  - 但随机难以定义也难以达到。
- 密码部件之间的关系要足够复杂，避免“分治攻击”

Outline

Def. of security	Ciphers	Examples & attacks
• Perfect secrecy • Unpredictability • Indistinguishability • Semantic security	• One-time pad (OTP) • PRG & stream ciphers	• Venona project • MS PPTP • WEP • CSS • RC4

One-time pad

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

Encryption: Plaintext ⊕ Key = Ciphertext

	h	e	i	l	h	i	t	l	e	r
Plaintext:	001	000	010	100	001	010	111	100	000	101
Key:	111	101	110	101	111	100	000	101	110	000
Ciphertext:	110	101	100	001	110	110	111	001	110	101
	s	r	l	h	s	s	t	h	s	r

One-time pad

e=000 h=001 i=010 k=011 l=100 r=101 s=110 t=111

Decryption: Ciphertext ⊕ Key = Plaintext

	s	r	l	h	s	s	t	h	s	r
Ciphertext:	110	101	100	001	110	110	111	001	110	101
Key:	111	101	110	101	111	100	000	101	110	000
Plaintext:	001	000	010	100	001	010	111	100	000	101
	h	e	i	l	h	i	t	l	e	r

A quiz

1. You are given a ciphertext (c) from one-time pad encryption Can you learn something about m?

   A. No

   B. Yes

2. You are given a message (m) and its one-time pad encryption (c). Can you compute the key from m and c ?

   A. No, I cannot compute the key.

   B. Yes, the key is $k = m \oplus c$.

   C. I can only compute half the bits of the key.

One-Time Pad (OTP) (Vernam 1917)

Very fast enc/dec !!

… but long keys (as long as plaintext)

Why not distribute message the same way as the pad?

It is not practical. However, it has been used.

Is the OTP secure? What is a secure cipher?

What is a secure cipher?

Attacker's abilities: **CT only attack** (for now)

Possible security requirements:

attempt #1: **attacker cannot recover secret key**

$E(\mathrm{k}, \mathrm{m}) = \mathrm{m}$ would be secure

attempt #2: **attacker cannot recover all of plaintext**

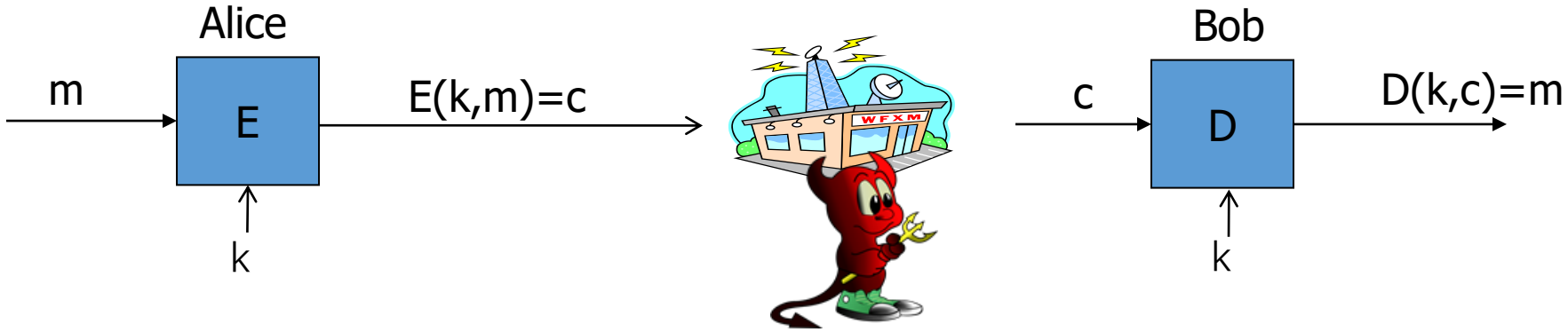
$E(\mathrm{k}, m_0||m_1) = m_0||\mathrm{k} \oplus m_1$ would be secure

Shannon's idea:

CT should reveal no "info" about PT

Formalize the security—perfect secrecy

Notation



- A cipher is a pair of functions $\mathcal{E} = (E, D)$ defined over $(\mathcal{K}, \mathcal{M}, \mathcal{C})$ where $\mathcal{K}, \mathcal{M}, \mathcal{C}$ are the key space, message space and ciphertext space respectively,

$$E\colon \mathcal{K} \times \mathcal{M} \to \mathcal{C}$$

$$D\colon \mathcal{K} \times \mathcal{C} \to \mathcal{M}$$

such that

$$\forall m \in \mathcal{M}, k \in \mathcal{K}, D\big(k, E(k, m)\big) = m.$$

Information theoretic security (Shannon 1949)

A cipher $\mathcal{E} = (E, D)$ over $(\mathcal{K}, \mathcal{M}, \mathcal{C})$ has **perfect secrecy** if

$\forall m_0, m_1 \in \mathcal{M}$ ($|m_0| = |m_1|$) and $\forall c \in \mathcal{C}$

$$Pr[\, E(\mathbf{k}, m_0) = c \,] = Pr[\, E(\mathbf{k}, m_1) = c \,] \quad \text{where } \mathbf{k} \xleftarrow{R} \mathcal{K}$$

The random variable **k** is uniformly distributed over $\mathcal{K}$

Equivalently,

$$\frac{\#\{k \in \mathcal{K} : E(k, m_0) = c\}}{|\mathcal{K}|} = \frac{\#\{k \in \mathcal{K} : E(k, m_1) = c\}}{|\mathcal{K}|}$$

- From c one can't tell if the message is m_0 or m_1 (for all m_0 or m_1)
- Most powerful attacker learns nothing about PT from CT
- No CT only attack

The one-time pad has perfect secrecy

A question first

Let $m \in \mathcal{M}$ and $c \in \mathcal{C}$.

How many keys map m to c for one-time pad ?

None

1

2

Depends on m

The one-time pad has perfect secrecy

- Proof

Suppose $\mathcal{K} = \mathcal{M} = \mathcal{C} = \{0,1\}^L$, for any fixed message m and ciphertext c, there is a unique k satisfying $k \oplus m = c$.

That is, for all m_0, m_1, and all c

$$\frac{\#\{k \in \mathcal{K}: E(k, m_0) = c\}}{|\mathcal{K}|} = \frac{\#\{k \in \mathcal{K}: E(k, m_1) = c\}}{|\mathcal{K}|} = \frac{1}{|\mathcal{K}|}$$

Equivalently, $Pr[\, E(\mathbf{k}, m_0) = c\,] = Pr[\, E(\mathbf{k}, m_1) = c\,] = 2^{-L}$

(so, $\forall$m, c, $\#\{k \in \mathcal{K}: E(k, m) = c\} = \text{const} \Longrightarrow$ perfect secrecy)

Bad news

- Thm: perfect secrecy $\Rightarrow$ $|\mathcal{K}| \geq |\mathcal{M}|$

  Idea: Assume $|\mathcal{K}| < |\mathcal{M}|$, then $\#\{k \in \mathcal{K}: E(k,m) = c\} = \text{const}$ does not hold.

- $|\mathcal{K}| \geq |\mathcal{M}|$ makes it hard to use for ciphers with perfect secrecy

Examples & attacks on OTP

Two-time pad is insecure !!

Never use the key more than once !!

$$C1 \leftarrow m1 \oplus k$$

$$C2 \leftarrow m2 \oplus k$$

Eavesdropper does:

$$C_1 \oplus C_2 \rightarrow$$ 

Enough redundancy in English and ASCII encoding that:

$$m_1 \oplus m_2 \rightarrow m_1 , m_2$$

Real-world one-time pad

- Project VENONA
  - Soviet spy messages from U.S. between 1941-1946
  - Nuclear espionage, etc.
  - Thousands of messages
- Spy carried one-time pad into U.S.

> *During WWII the Soviet Union could not produce enough one-time pads . . . to keep up with the enormous demand . . . . So, they used a number of one-time pads twice, thinking it would not compromise their system.*

Stream ciphers

Review

- Cipher over $(\mathcal{K},\mathcal{M},\mathcal{C})$: a pair of functions $\mathcal{E}=(E, D)$

$$\forall m\in\mathcal{M},\ k\in\mathcal{K},\ D(k,E(k,m))=m.$$

- **Weak ciphers**: substitution ciphers, e.g., Vigener
- **A good cipher**: one-time pad c = k $\oplus$ m

Lemma: the one-time pad has perfect secrecy (i.e. no CT only attacks)

Bad news: perfect-secrecy ⇒ key-len ≥ msg-len

Stream ciphers: making OTP practical

Idea: replace the "random" key of OTP by "pseudorandom" key

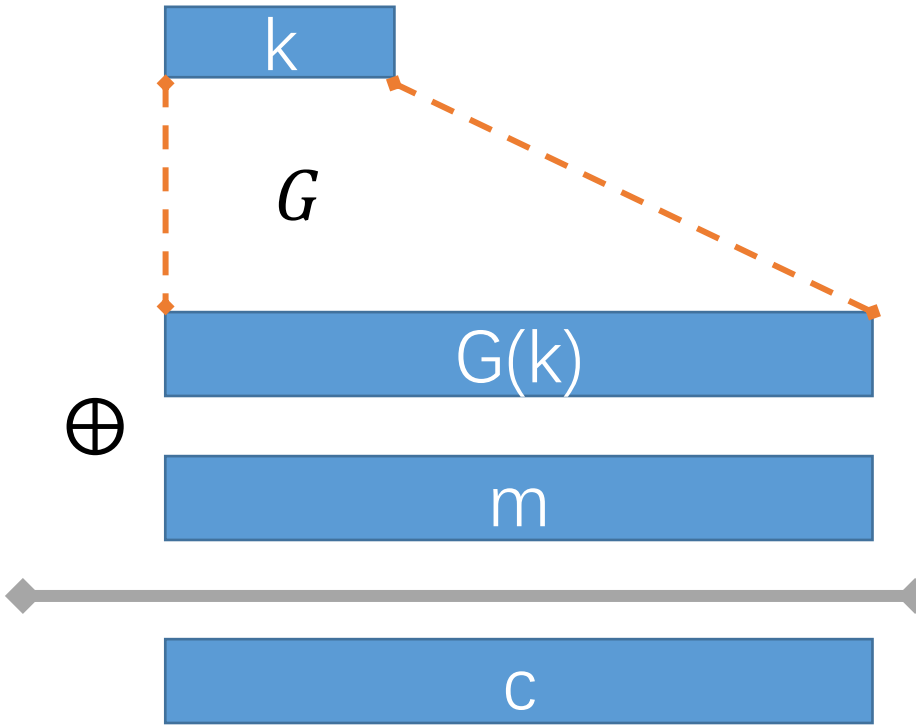
- A pseudorandom generator (PRG) is an algorithm

$G: \{0,1\}^s \rightarrow \{0,1\}^n$, where $n \gg s$

- Stream cipher:

$$c = E(k, m) = m \oplus G(k)$$

$$m = D(k, c) = c \oplus G(k)$$



Can a stream cipher have perfect secrecy?

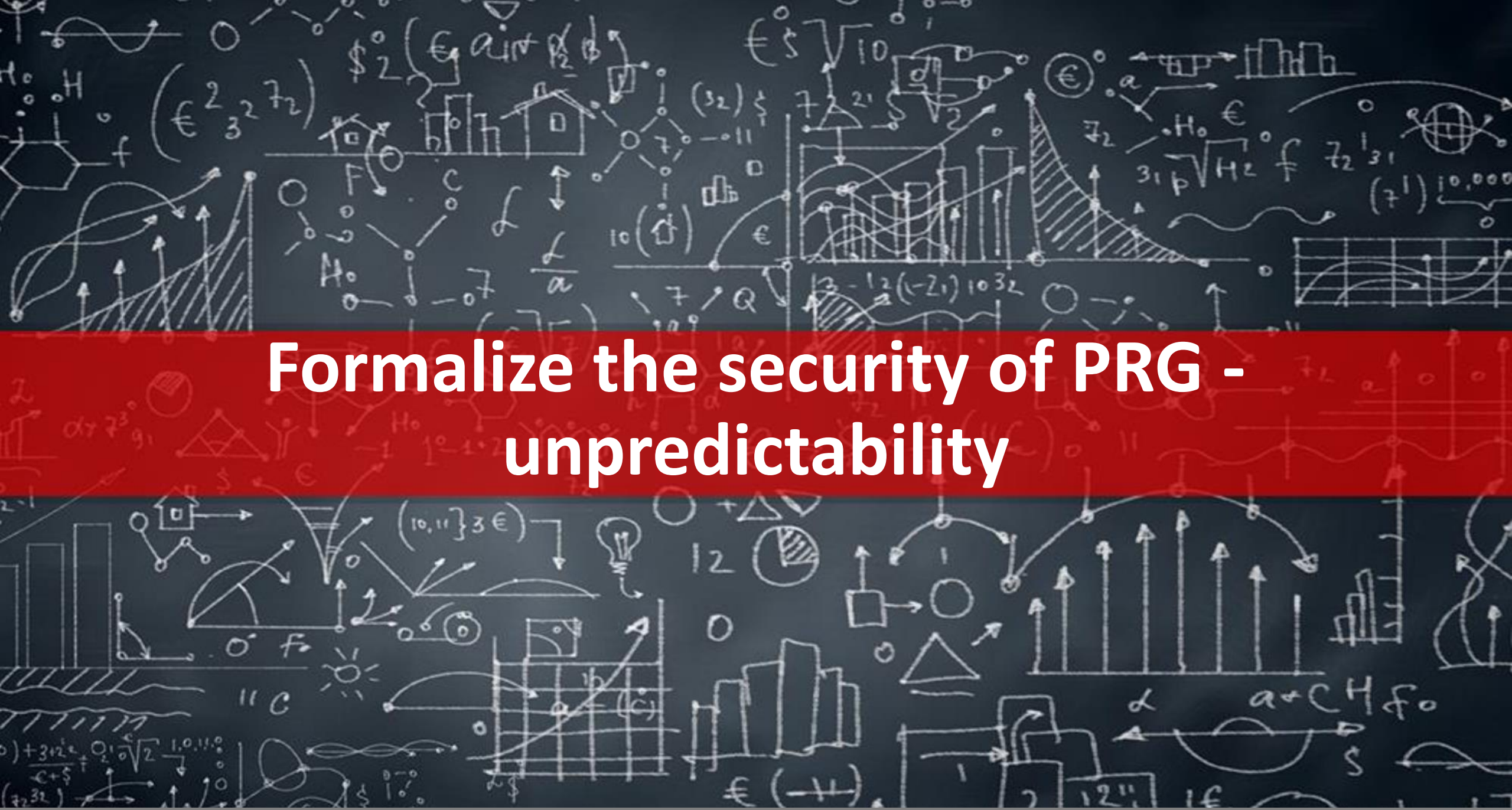
A. Yes, if the PRG is really “secure”
B. No, there are no ciphers with perfect secrecy
C. Yes, every cipher has perfect secrecy
D. No, since the key is shorter than the message

Stream ciphers: almost OTP

Stream ciphers cannot have perfect secrecy !!

- Need a different definition of security

- Security will depend on specific PRG



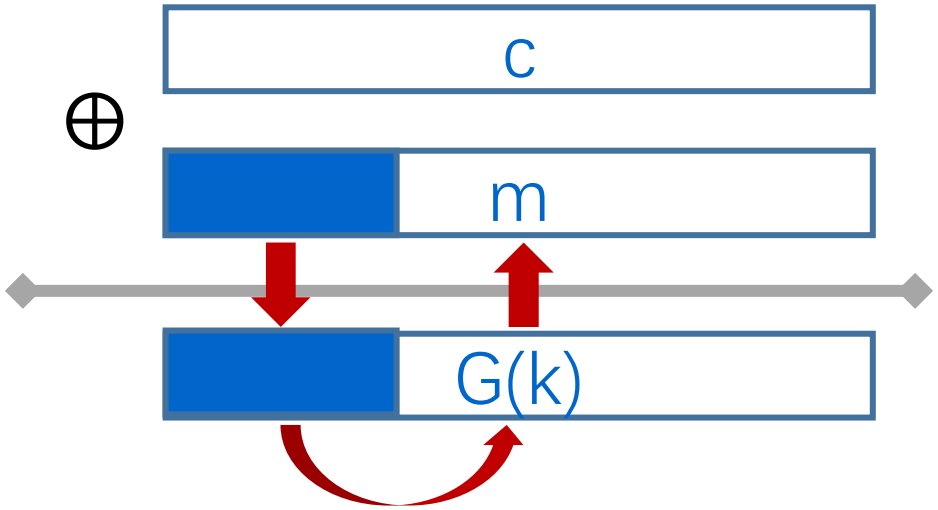
Formalize the security of PRG - unpredictability

PRG must be unpredictable

- Suppose PRG is predictable. That is, there is an algorithm A

$$\exists i, G(k)|_{1,\ldots,i} \overset{A}{\rightarrow} G(k)|_{i+1,\ldots,n}$$

Then



Even the next bit prediction
$G(k)|_{1,\ldots,i} \overset{A}{\rightarrow} G(k)|_{i+1}$ is a problem!

PRG must be unpredictable

G: K $\longrightarrow \{0,1\}^n$ is **predictable** if: $\exists$ efficient alg. A and $\exists 1 \leq i < n$ s.t.

$$\Pr\left[A\left(G(k)|_{1,\ldots,i}\right) = G(k)|_{i+1}\right] > \frac{1}{2} + \varepsilon$$

Where ε is non-negligible (say $\varepsilon = \frac{1}{2^{32}}$)

PRG is **unpredictable** if it is not predictable
$\Rightarrow$ $\forall$i: no "eff" adv. can predict bit (i+1) for "non-neg" ε

Negligible and non-negligible

- In practice: ε is a scalar and
  - ε non-neg: $\varepsilon \geq 1/2^{30}$ (likely to happen over 1GB of data)
  - ε negligible: $\varepsilon \leq 1/2^{80}$ (won't happen over life of key)

- In theory: ε is a function $\varepsilon: Z^{\geq 0} \longrightarrow R^{\geq 0}$ and
  - ε non-neg: $\exists d: \varepsilon(\lambda) \geq 1/\lambda^d$ inf. often ($\varepsilon \geq$ 1/poly, for many λ)
  - ε negligible: $\forall d, \lambda \geq \lambda_d: \varepsilon(\lambda) \leq 1/\lambda^d$ ($\varepsilon \leq$ 1/poly, for large λ)

A quiz

Suppose G: $K \longrightarrow \{0,1\}^n$ is such that for all k: **XOR(G(k)) = 1**

Is G predictable ?

A. Yes, given the first bit I can predict the second

B. No, G is unpredictable

C. Yes, given the first (n-1) bits I can predict the n'th bit

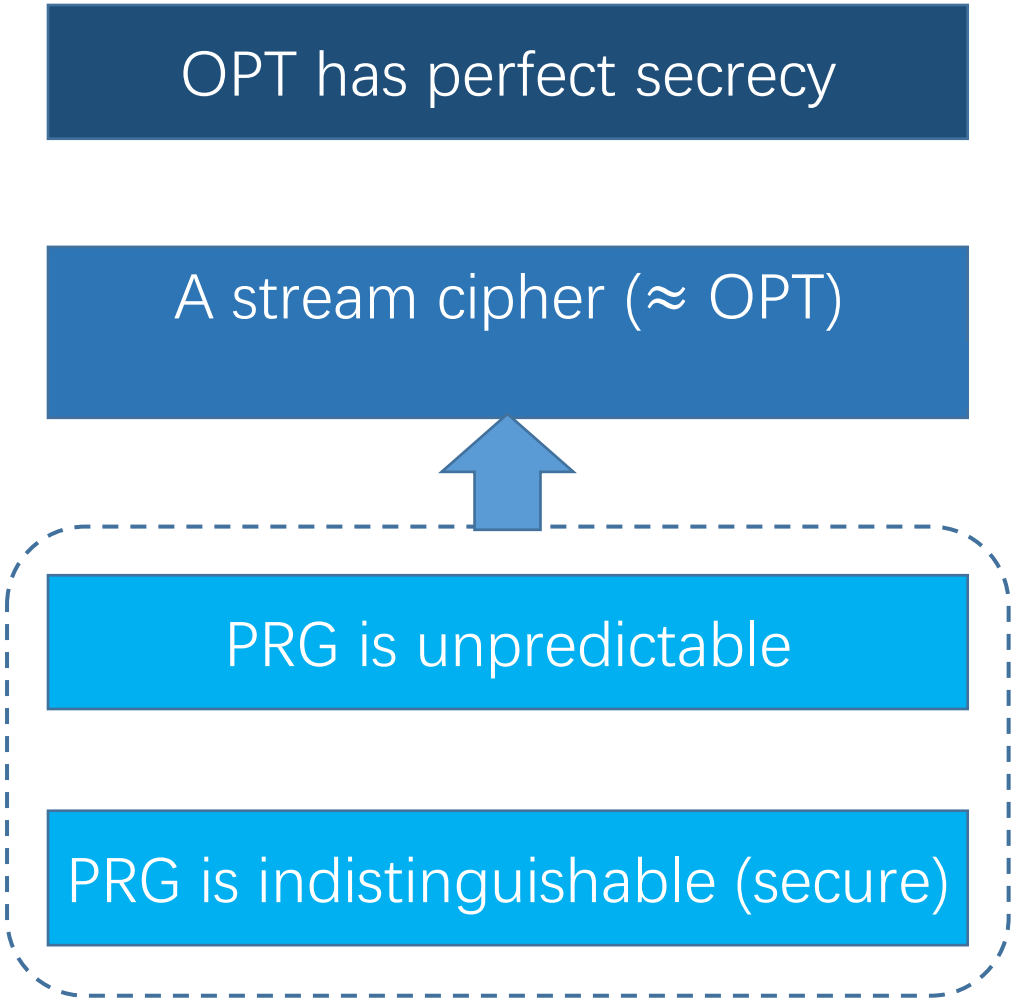
D. It depends

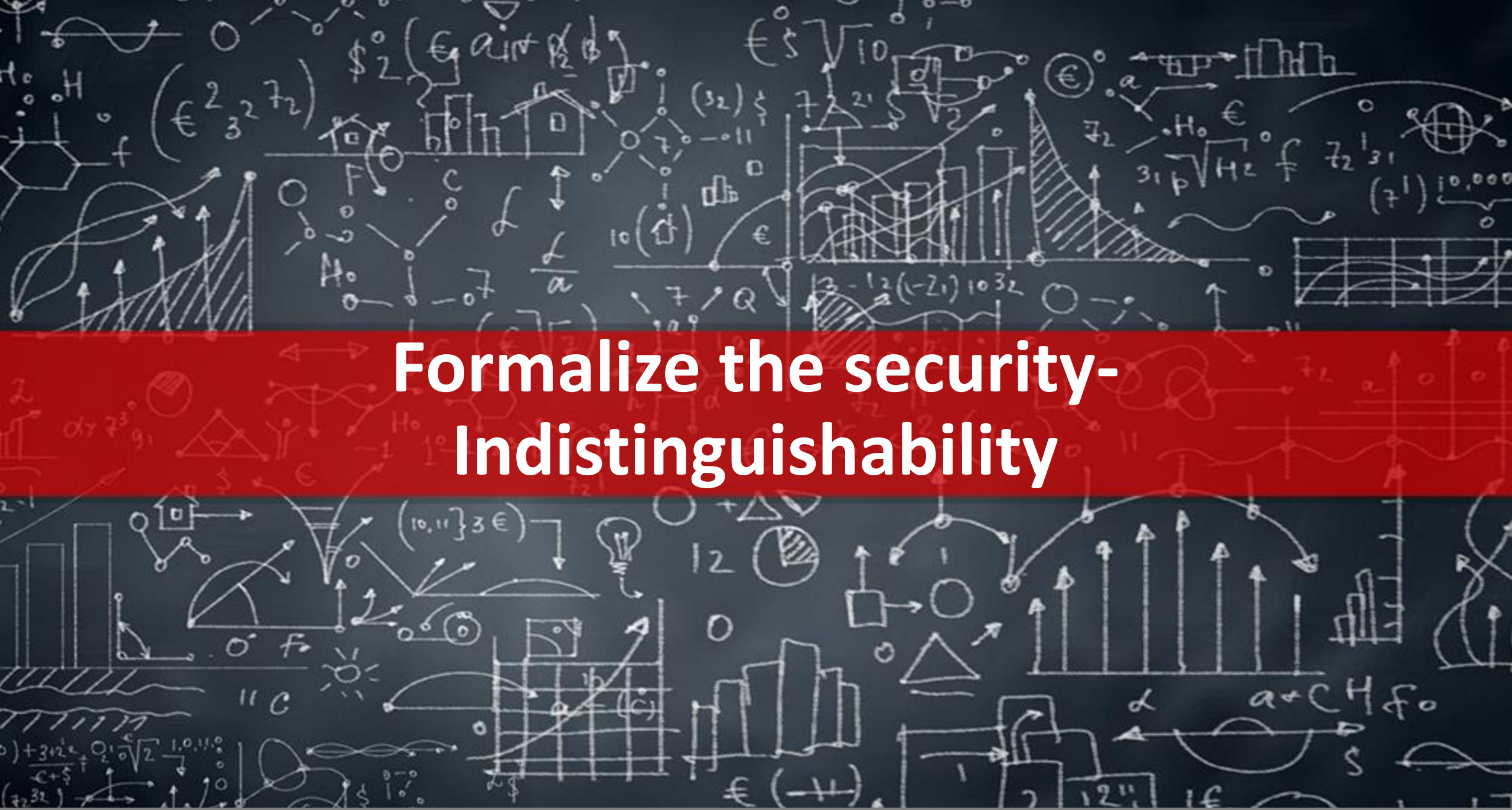
Review

Let (E,D) be a cipher over (K,M,C)

(E,D) has **perfect secrecy** if $\forall$ m0, m1 $\in$ M (|m0| = |m1|)

$\{ E(k,m_0) \} = \{ E(k,m_1) \}$ where $k \xleftarrow{R} K$





Formalize the security-Indistinguishability

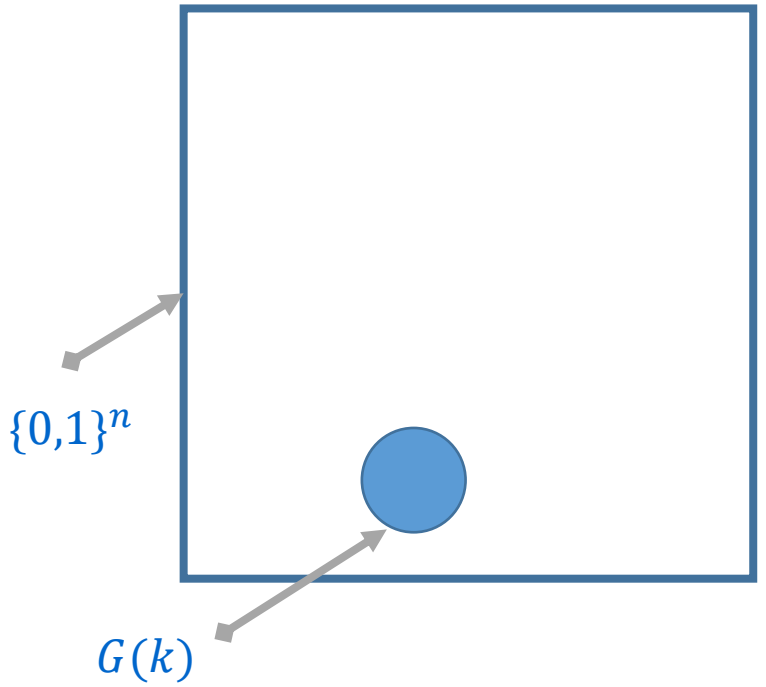
Definition of PRG

Let $G:K \longrightarrow \{0,1\}^n$ be a pseudorandom generator

Goal: pseudorandom is "**indistinguishable**" from random

1. $r \xleftarrow{R} \{0,1\}^n$, output r
2. $k \xleftarrow{R} \mathcal{K} = \{0,1\}^s$, output $G(k)$

$s \ll n$



Statistical tests

Statistical test on $\{0,1\}^n$:

$$\text{an alg. A s.t. } A(x) = \begin{cases} 0 & \text{If x is not random} \\ 1 & \text{If x is random} \end{cases}$$

Examples:

1. $A(x) = 1$ iff $|\#0(x) - \#1(x)| \leq \lambda(n)$, say $\lambda(n) = 10 \times \sqrt{n}$
2. $A(x) = 1$ iff $\left|\#00(x) - \frac{n}{4}\right| \leq \lambda(n)$, say $\lambda(n) = 10 \times \sqrt{n}$
3. $A(x) = 1$ iff max-run-of-0(x)$\leq \lambda(n)$, say $\lambda(n) = 10 \times \log_2 n$
4. …

Advantage

Let G:K $\longrightarrow\{0,1\}^n$ be a PRG and A a stat. test on $\{0,1\}^n$

Define:

$$\mathbf{PRG_{adv}}[\boldsymbol{A},\boldsymbol{G}] \coloneqq \left|\Pr_{k\xleftarrow{R}\mathcal{K}}\left[A\big(G(k)\big)=1\right]-\Pr_{r\xleftarrow{R}\{0,1\}^n}[A(r)=1]\right| \in [0,1]$$

Adv close to 1 ⇒ A can distinguish G from random

Adv close to 0 ⇒ A cannot

A quiz

Suppose $G:K \longrightarrow \{0,1\}^n$ satisfies msb(G(k)) = 1 for 2/3 of keys in K

Define statistical test A(x) as:

if [msb(x)=1] output “1” else output “0”

Then

PRGadv [A,G] = | Pr[A(G(k))=1] - Pr[A(r)=1] | =

Secure PRG (indistinguishability)

Def: $G:K \longrightarrow \{0,1\}^n$ is a **secure PRG** if for **any** efficient statistical test A, $\mathbf{PRG_{adv}}[A, G]$ is negligible.

Are there provably secure PRGs?

but we have heuristic candidates.

Easy fact: a secure PRG is unpredictable

PRG predictable ⇒ PRG is insecure (Proof by contradiction!)

Suppose A is an efficient algorithm s.t.

$$\Pr_{k \xleftarrow{R} \mathcal{K}} \left[A\left(G(k)|_{1,\ldots,i} \right) = G(k)|_{i+1} \right] > \frac{1}{2} + \varepsilon$$

for non-negligible ε (e.g. ε = 1/1000)

Easy fact: a secure PRG is unpredictable

Define statistical test B as:

$$\mathrm{B}(\mathrm{x}) = \begin{cases} 1 & \text{If } A(x_{1,\cdots,i}) = x_{i+1} \text{ (happens with probability } > \frac{1}{2} + \varepsilon \text{)} \\ 0 & \text{ese} \end{cases}$$

Then

$$\mathrm{PRG}_{\mathrm{adv}}[B, G] := \left| \Pr_{k \xleftarrow{R} \mathcal{K}}[B(G(k)) = 1] - \Pr_{r \xleftarrow{R} \{0,1\}^n}[B(r) = 1] \right|$$

$$> \frac{1}{2} + \varepsilon - \frac{1}{2} = \varepsilon$$

Thm (Yao'82): an unpredictable PRG is secure

Let $G:K \longrightarrow \{0,1\}^n$ be PRG

"Thm": if $\forall i \in \{0, \ldots, n-1\}$ PRG G is unpredictable at pos. i

then G is a secure PRG.

If next-bit predictors cannot distinguish G from random

then no statistical test can !!

A quiz

Let $G: K \longrightarrow \{0,1\}^n$ be a PRG such that

from the last n/2 bits of G(k)

it is easy to compute the first n/2 bits.

Is G predictable for some $i \in \{0, \cdots, n-1\}$?

Yes

No

Summary

Let (E,D) be a cipher over (K,M,C)

(E,D) has **perfect secrecy** if $\forall$ m0, m1 $\in$ M (|m0| = |m1|)

$\{ E(k,m_0) \} = \{ E(k,m_1) \}$ where $k \xleftarrow{R} K$