BASIC INFORMATION THEORETIC TOOLS-II

Information-theoretic MACs, Randomness Extractors

ACM Summer School 2024

Message Authentication

Can Bob find out if the message is indeed from Alice or not? (Accept if from Alice, else not)

k

Hi Bob

Mac(k, "Hi Bob") = t

Eve can add, delete, modify, etc

k

If Verify(k, m, t) = 1, accept
Else, reject

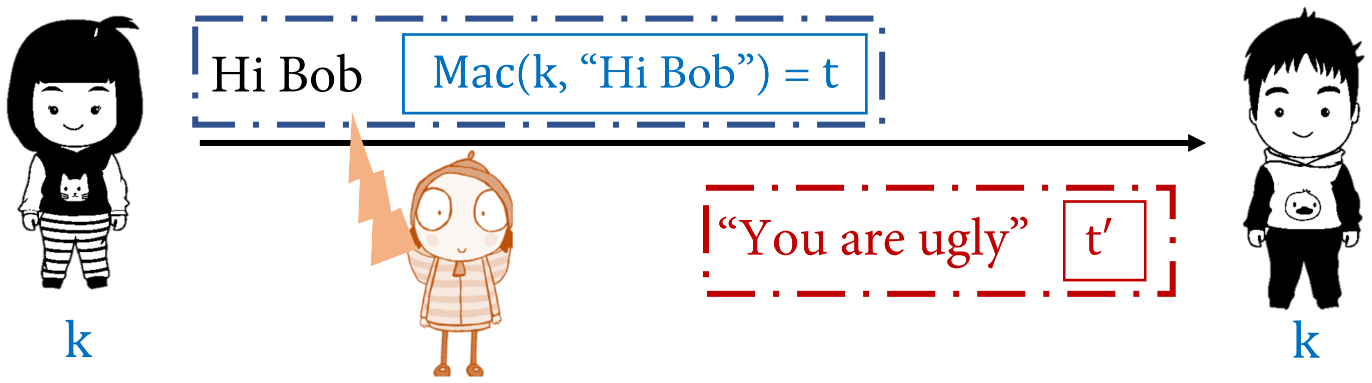
- Gen: generates a secret key k
- Mac(k,m): Takes key k and message m and outputs a tag t
- Verify(k,m,t): Take key k along with the received m,t and output 0/1

Correctness $\forall k \leftarrow \text{Gen}\ \forall\, m, \forall\, t \leftarrow \text{Mac}(k, m), \text{Verify}(k,m,t) = 1$

Security ? (Eve is all powerful or computationally unbounded)

Message Authentication Codes (MAC)

Can Bob find out if the message is indeed from Alice or not? (Accept if from Alice, else not)



Eve wins if Bob accepts her message and tag

One-time Security of Information-theoretic MAC

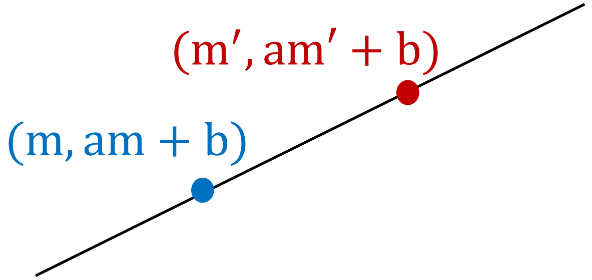
Given $(m, t = \mathrm{Mac}(k, m))$ Eve wins if she produces a (m', t') such that:
$m' \neq m$ and $\mathrm{Verify}(k, m', t') = 1$

(ε-secure) $\forall$ unbounded Eve, $\Pr[\text{Eve wins}] \leq \varepsilon$

A Simple Information-theoretic MAC

- Gen: $k = (a, b) \leftarrow \mathbb{Z}_p^2$
- Mac(k,m): $(am + b) \bmod p$
- Verify(k,m,t): If $t = (am + b) \bmod p$, output 1, else output 0

$(\mathbb{Z}_p = \{0,1, \ldots, p-1\}, + \bmod p)$



Theorem

$(\text{Gen}, \text{Mac}, \text{Verify})$ is a $1/p$-secure one-time MAC.

Given one point on a random line, can you find another point on it?

Proof Sketch

Given (m, t) such that $t = (am + b) \bmod p$, and for any $m' \neq m$, what is the probability that Eve can find $t' = (am' + b) \bmod p$?

$$1/p$$

Universal Hash Functions

Definition

$H: \mathcal{K} \times \mathcal{M} \to \mathcal{T}$ is a universal hash function if for all $m' \neq m \in \mathcal{M}$ and all $t', t \in \mathcal{T}$

$$\Pr_{k \leftarrow \mathcal{K}}[H_k(m) = t \wedge H_k(m') = t'] = 1/|\mathcal{T}|^2$$

Uniformly and independently distributed in $\mathcal{T}$ when k is a uniform key.

Example

For $k = (a, b) \leftarrow (\mathbb{Z}_p \times \mathbb{Z}_p)$, $m \in \mathbb{Z}_p$

$$H_k(m) := (am + b) \bmod p$$

($\mathbb{Z}_p = \{0,1, \ldots, p-1\}, +\bmod p$)

Exercise 1

Prove that H is a universal hash function.

MAC from Universal Hash Function

Given: $H: \mathcal{K} \times \mathcal{M} \to \mathcal{T}$ is a universal hash function

- Gen: $k \leftarrow \mathcal{K}$
- Mac(k, m): For $m \in \mathcal{M}$, $t \coloneqq H_k(m)$
- Verify(k, m, t): If $t = H_k(m)$, output 1, else output 0

$H_k(m) := (am + b) \bmod p$

Gen: $k = (a, b) \leftarrow \mathbb{Z}_p^2$
Mac(k,m): $(am + b) \bmod p$
Verify(k,m,t): If $t = (am + b) \bmod p$, output 1, else output 0

Recall

Theorem

If $H: \mathcal{K} \times \mathcal{M} \to \mathcal{T}$ is a universal hash function, then (Gen, Mac, Verify) is a $1/|\mathcal{T}|$-secure MAC.

Exercise 2: Prove it!

Hint: Since for $m' \neq m$, $H_k(m)$ and $H_k(m')$ are independently and uniformly distributed in $\mathcal{T}$, use a similar argument as before!

Limitations of Information-theoretic MACs

Gen: $k = (a, b) \leftarrow \mathbb{Z}_p^2$
Mac(k,m): $(am + b) \bmod p$
Verify(k,m,t): If $t = (am + b) \bmod p$, output 1, else output 0

Recall

Security: $\varepsilon = 1/p$
Key Length: $2p$

Theorem

Let (Gen, Mac, Verify) be $1/2^n$-secure MAC where all keys output by Gen are of same length. Then, the keys output by Gen must have a length of at least $2n$.

Intuition

Exercise 3

- Fix two distinct messages $m \neq m'$. There must be at least 2^n possibilities for the tag of m (or else Eve could guess it with probability better than 2^{-n})
- Further conditioned on the value of tag for m, there must be 2^n possibilities for the tag of m' (or else Eve could forge a tag on m' with probability better than 2^{-n})
- Since each key defines a tag on m and m', there must be at least $2^n \times 2^n$ keys!

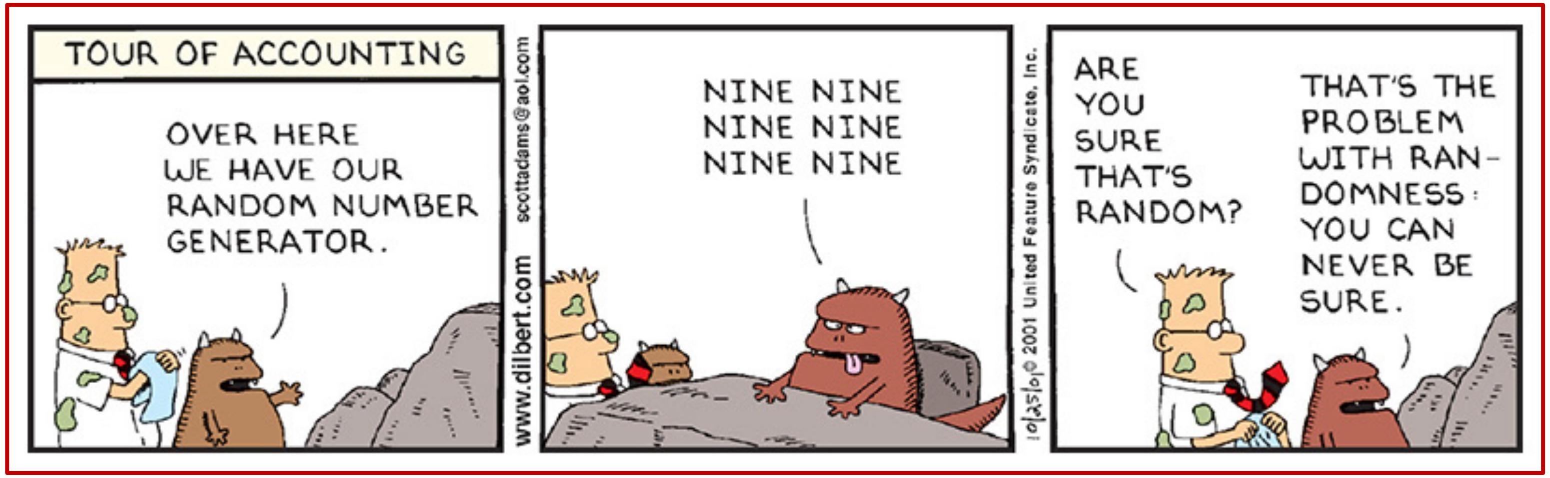
RANDOMNESS EXTRACTORS

Quest for Perfect Randomness

- Uniform randomness is crucial in many applications
  - Truly uniform bits are used to generate secret keys in Cryptography (One time pad)
  - Randomized algorithms assume access to truly uniform bits.

- In reality, random sources are not perfect
  - Correlated and biased bits (partial secrecy)
  - Physical sources, system RNGs, biometric data, etc.

Can we convert imperfect sources into (almost) uniform bits?



Credits
Dilbert: Scott Adams

Imperfect source: Examples

IID-Bit Source

$X = X_1, X_2, \dots, X_n \in \{0,1\}$: identical and independent, but biased

$\forall i, \Pr[X_i = 1] = \delta$ for some unknown δ

How to convert into a source of independent unbiased bits?

consider X in pairs, $X_i X_{i+1} = \begin{cases} 01 \Longrightarrow \text{output } 0 \\ 10 \Longrightarrow \text{output } 1 \\ 00/11 \Longrightarrow \text{discard} \end{cases}$

Independent-Bit Source

$X = X_1, X_2, \dots, X_n \in \{0,1\}$: identical and independent, but different biased

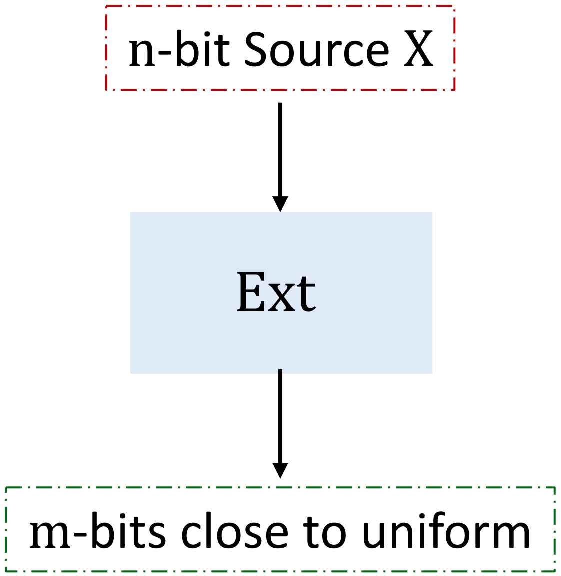
$\forall i, \Pr[X_i = 1] = \delta_i$ for different δ_i s.t. $0 < \delta < \delta_i \leq 1 - \delta$ for some constant δ

How to convert into a source of independent unbiased bits?

Output parity of each t bits: $\left|\Pr\left[\bigoplus_{i=1}^{t} X_i = 1\right] - \frac{1}{2}\right| \leq 2^{-\Omega(t)}$

Randomness Extraction

- Source: Random variable X over $\{0,1\}^{\mathrm{n}}$ in certain class $\mathcal{C}$
  - $\mathrm{IndBits}_{\mathrm{n},\delta}$: $\mathrm{X} = \mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_{\mathrm{n}} \in \{0,1\}$ independent bits, $\forall i, \Pr[\mathrm{X}_{\mathrm{i}} = 1] = \delta_i$ for $0 < \delta < \delta_i \leq 1 - \delta$
  - $\mathrm{IIDBits}_{\mathrm{n},\delta}$: assume all δ_i are same



Deterministic Extractor

A function $\mathrm{Ext}: \{0,1\}^{\mathrm{n}} \longrightarrow \{0,1\}^{m}$ such that
$\forall$ source $\mathrm{X} \in \mathcal{C}$, Ext(X) is "$\varepsilon$-close" to uniform.

How do you define closeness?

Statistical Distance

Definition I

X, Y be random variables over a range U. The statistical distance between X and Y is

$$\Delta(X, Y) := \frac{1}{2} \sum_{u \in U} |\Pr[X = u] - \Pr[Y = u]|$$

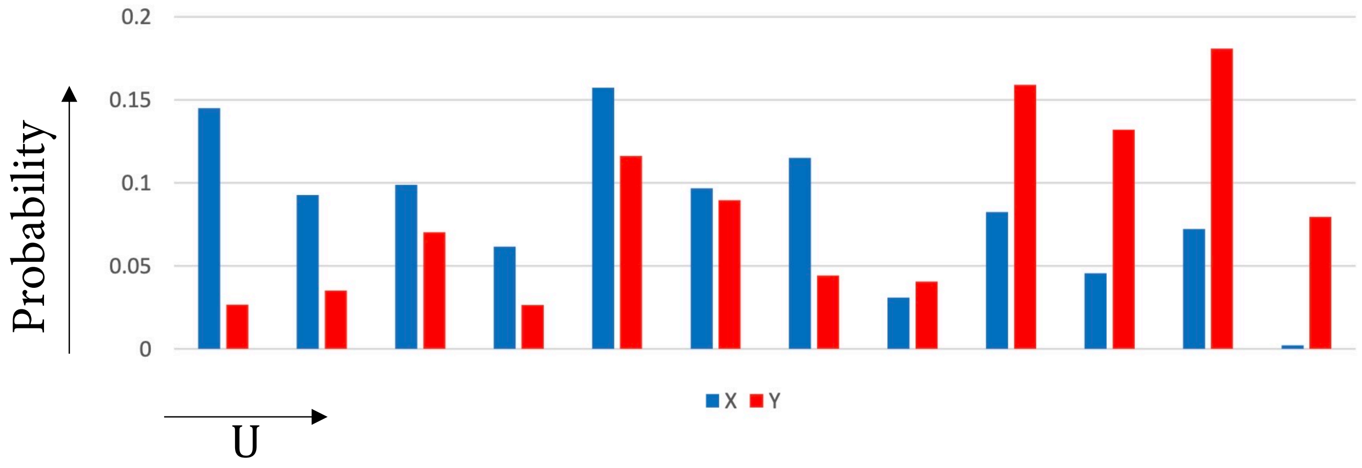
We say that X is ε-close to Y if $\Delta(X, Y) \leq \varepsilon$.

X is ε-close to Y iff we can transform X into Y by "shifting" at most ε fraction of the probability mass.

Example

$X = (.15, .09, .10, .06, .16, .09, .11, .03, .08, .04, .078, .002)$

$Y = (.03, .04, .07, .03, .11, .09, .04, .04, .16, .13, .18, .08)$



Statistical Distance: Properties

Operational Definition II: Max advantage to distinguish X and Y

$$\Delta(X,Y) := \max_{T \subset U} |\Pr[X \in T] - \Pr[Y \in T]|$$

If X is ε-close to Y, then for every event T

$$\Pr[X \in T] \leq \Pr[Y \in T] + \varepsilon$$

Exercise 4

Show equivalence of Definitions I and II

Data processing inequality: For any function f, $\Delta\big(f(X), f(Y)\big) \leq \Delta(X,Y)$

- i.e., post-processing only decreases the statistical distance!
- When f is bijective, equality holds. Why?

Exercise 5

Prove this inequality!
Hint: use the Def II

Extractor for $\mathrm{IndBits}_{n,\delta}$

$\mathrm{IndBits}_{n,\delta}$: $X = X_1, X_2, \ldots, X_n \in \{0,1\}$ independent bits,
$\forall i, \Pr[X_i = 1] = \delta_i$ for $0 < \delta < \delta_i \leq 1 - \delta$

$X = X_1, X_2, \ldots, X_n$

$\downarrow$

Ext(X): Breaks X into m blocks of length $\lfloor n/m \rfloor$ and outputs parity of each block

$\downarrow$

m-bit $Y \approx_{\varepsilon}$ Uniform $\quad \varepsilon = m2^{-\Omega(n/m)}$

Extractor for General Sources?

Can we extract truly uniform bits from any source X?

No, not if the source is not random, e.g. $X = 0^n$ w. p. 1

Hope is Ext works whenever X has sufficient “entropy”

What entropy?

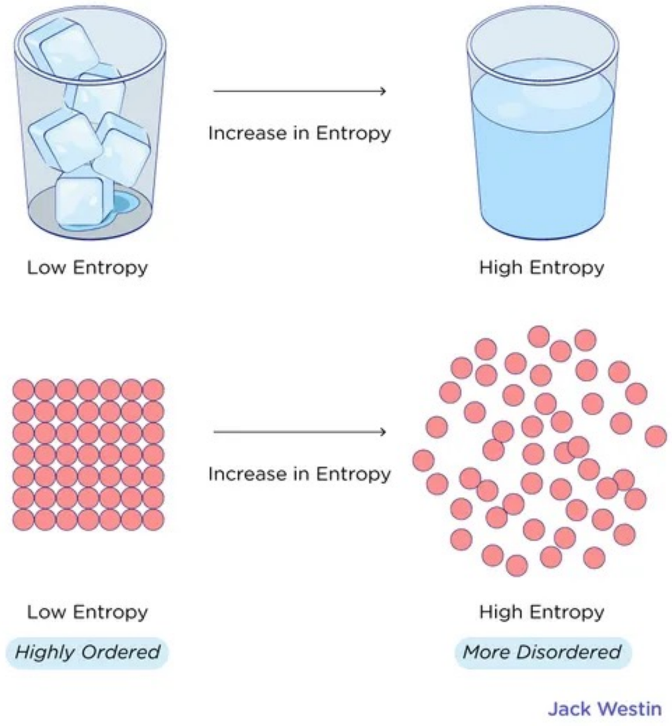
n-bit Source X

↓

Ext

↓

m-bits close to uniform



Attempt I: Shannon Entropy

Definition

Shannon entropy

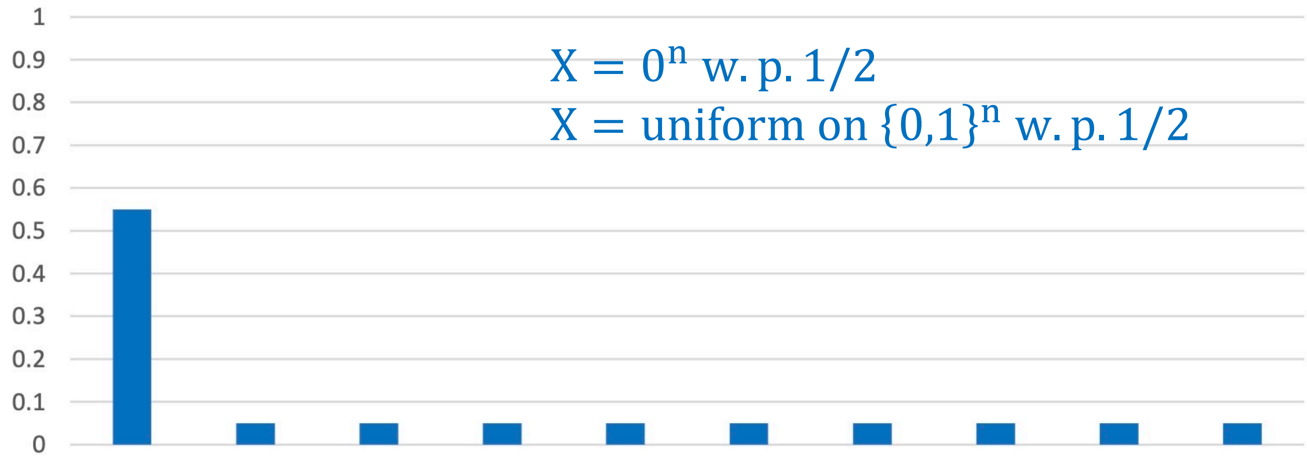
the average number of bits required to represent a string drawn from X

$$H_{sh}(X) := \sum_{x} \Pr[X = u] \log \frac{1}{\Pr[X = x]} = E_{x \leftarrow X}\left[\log \frac{1}{\Pr[X = x]}\right]$$

Is this the right notion of entropy?

$H_{sh}(X) \geq n/2$ but $\Pr[X = 0^n] > 1/2$

Can't extract from X



Attempt II: Min-Entropy

Definition

Min-entropy

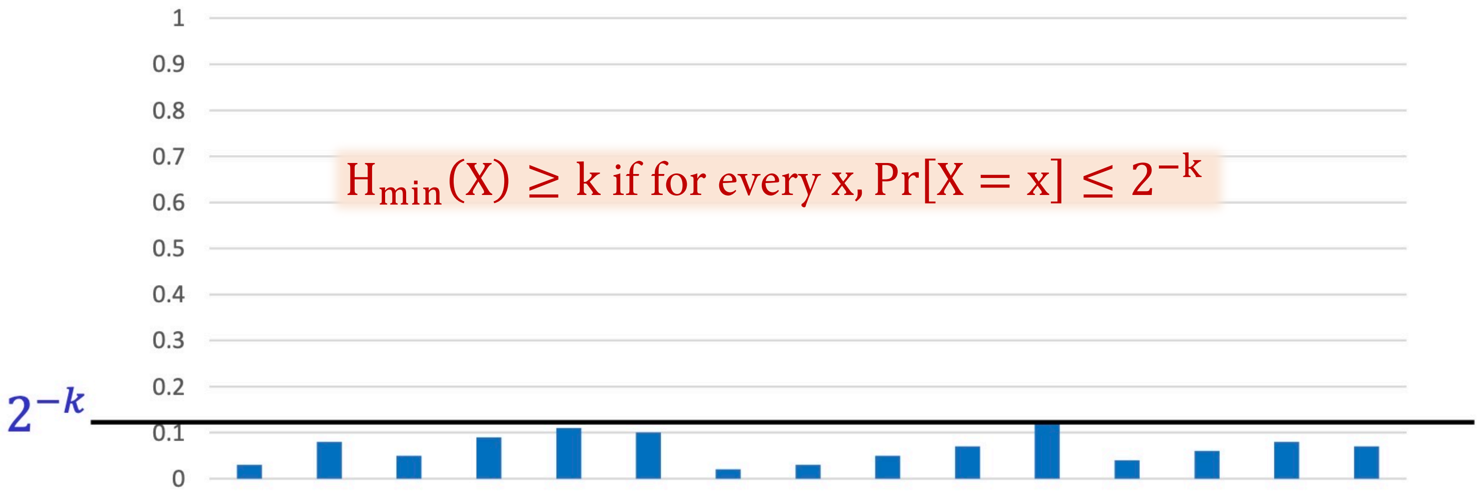
Worst-case measure of unpredictability of X

$$H_{min}(X) := \max_x \left[\log \frac{1}{\Pr[X = x]}\right]$$

X is a k-source if $H_{min}(X) \geq k$

Extractor for the class of k-sources?

$H_{min}(X) \geq k$ if for every x, $\Pr[X = x] \leq 2^{-k}$



Impossibility of Deterministic Extraction

Theorem

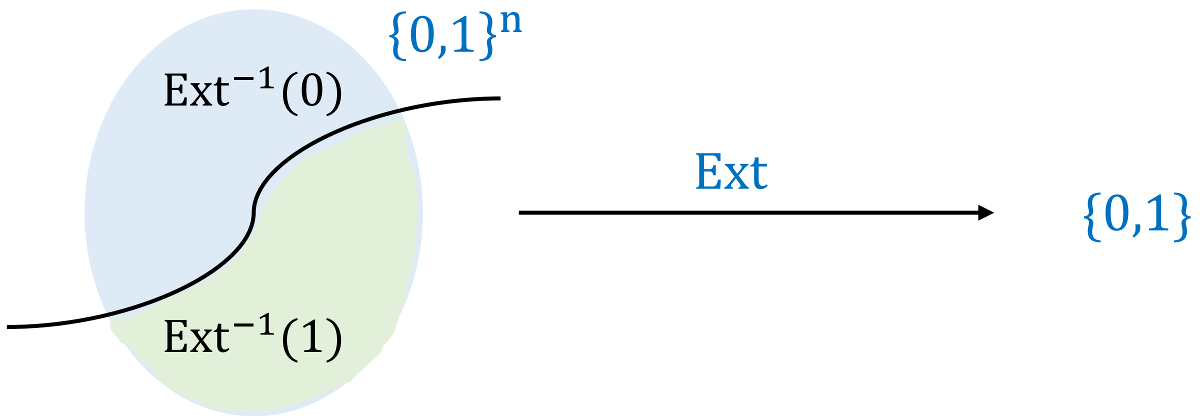
For any $\mathrm{Ext}: \{0,1\}^n \longrightarrow \{0,1\}$ there exists an $(n-1)$-source X such that $\mathrm{Ext}(X) = \text{constant}$

Proof

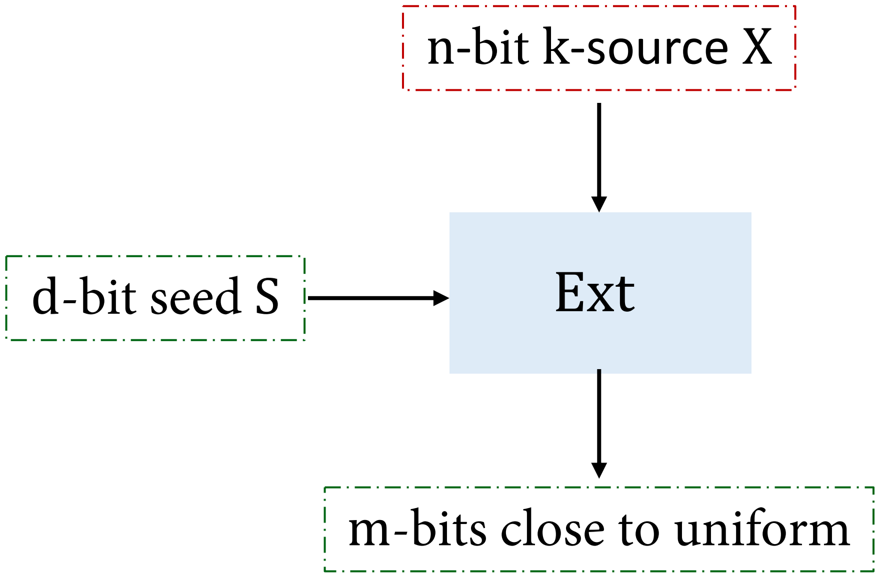
Consider $X_b = \text{uniform on } \mathrm{Ext}^{-1}(b)$

- $\mathrm{Ext}(X_b) = \text{constant}$
- Either $H_{\min}(X_0)$ or $H_{\min}(X_1) \geq n - 1$

Deterministic extractor for k-source is impossible even for extracting 1 bit and even for $k = n - 1$



Seeded Extractors

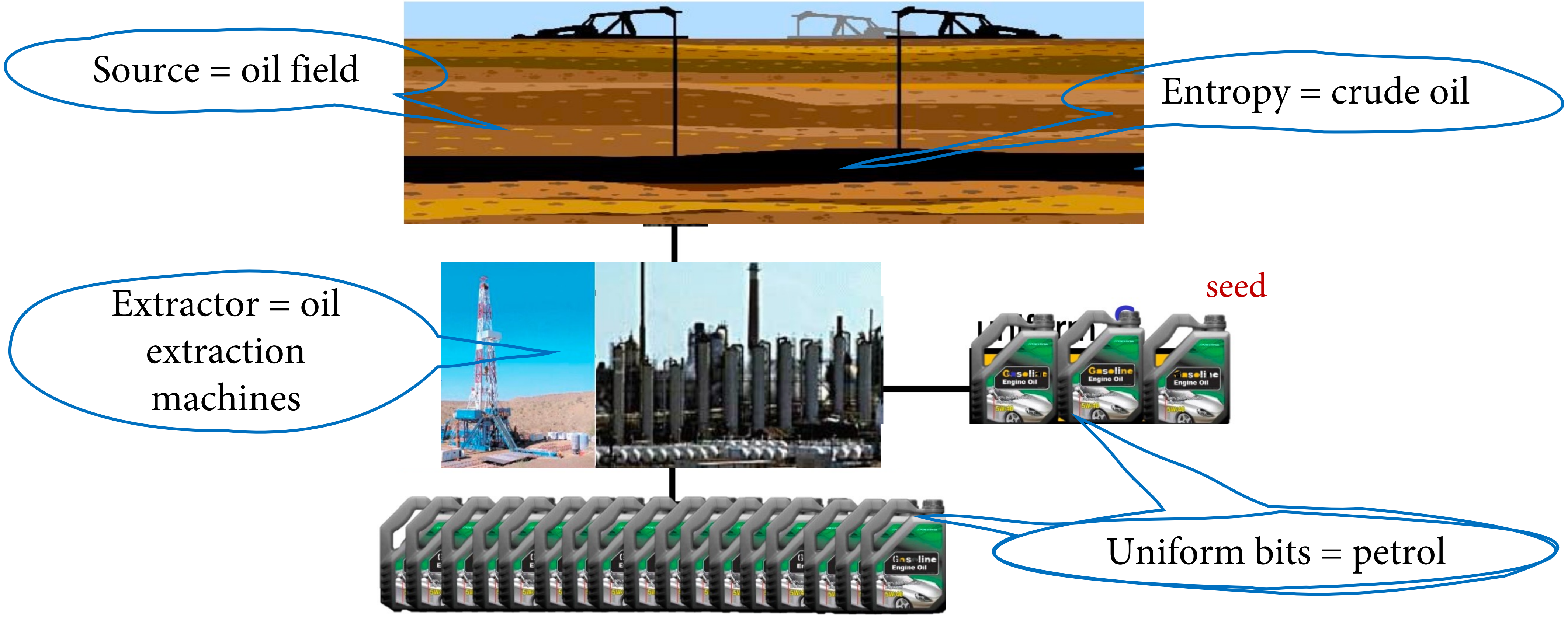


A short uniform seed as a catalyst for extraction

Seeded Extractor

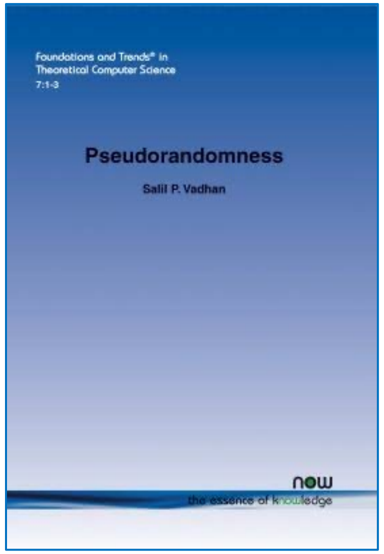
$\mathrm{Ext}: \{0,1\}^n \times \{0,1\}^d \longrightarrow \{0,1\}^m$ is a (k, ε)-seeded extractor if
$\forall$ k-source X, Ext(X; S) is ε-close to uniform.

Seeded Extractors: An Analogy



Pervasive Applications

- Diverse topics in Theoretical Computer Science
  - Cryptography, Derandomization & pseudorandomness, Distributed Algorithms Data Structures, Hardness of Approximation,…
- Many applications in Cryptography
  - Privacy Amplification, Bounded-storage model, PRG, Biometrics, Leakage-resilient crypto



Pseudorandomness
Salil P. Vadhan