An Introduction to Extractors

David Fuchs
fuchsd@student
May 14, 2005

Randomness

Why use randomized algorithms and protocols?

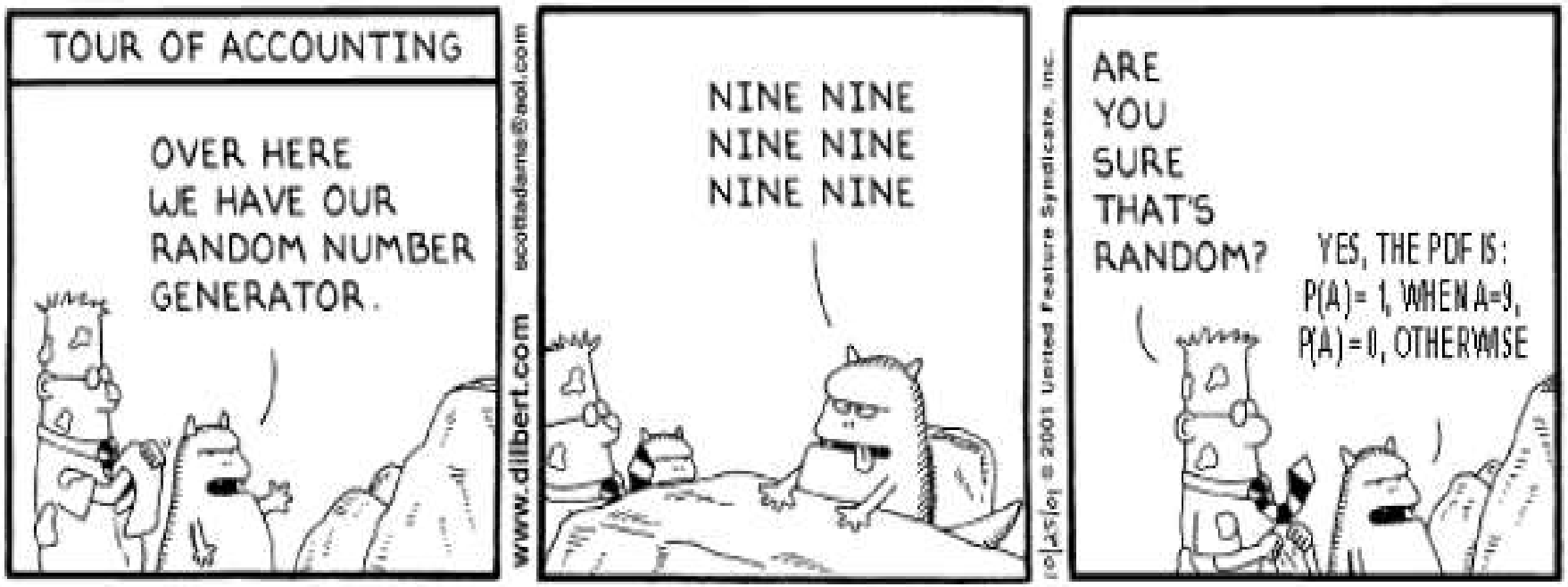
- Faster algorithms:
  e.g., PCP theorem (a slow computer can efficiently check computations done by a supercomputer)

- Better algorithms:
  Algorithms can have, in expectation, a better solution than the best deterministic algorithms

- Cryptography:
  Many protocols require random keys or challenges

Sources of Randomness

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



Sources of Randomness

- Coin tosses?
- Pseudo-Random Generator?
- Environment variables?
  Temperature, time, CPU state, ...
- Physical processes?
  Zener diode, Geiger counter, ...

Many sources are somewhat random, but not completely random.

An Extractor is a deterministic procedure that extracts truly random bits from such a weak source.

Measuring Randomness

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

- A source is described by the probability distribution S of its output
- "Truly random bits": A sequence of uniformly distributed and independent bits
- Randomness of a source S is measured by Min-Entropy:

$$H_\infty(S) = \min_x \left\{ \log \left(\frac{1}{\Pr_S[x]} \right) \right\}$$

- A k-source is a source S with $H_\infty(S) \geq k$

Comparing Distributions

Extractor's output distribution O should be "close to uniform"

Output of extractor is compared to uniform distribution with statistical difference Δ:

$$\Delta(P,Q) := \frac{1}{2}\sum_x \left| \Pr_P(x) - \Pr_Q(x) \right|$$

O is ϵ-close to uniform distribution U iff

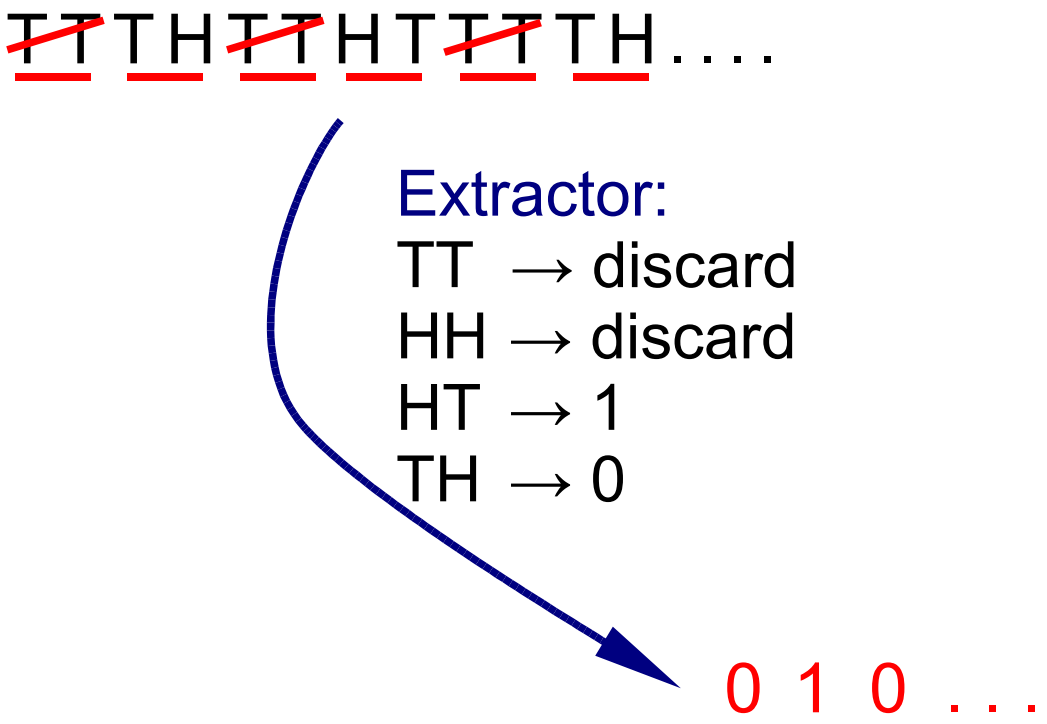
$$\Delta(O,U) \leq \epsilon$$

Extractor: Simple Example

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

We have: Sequence of {H,T} by biased coin:
$\Pr[H] = h$, $\Pr[T] = t$, $t \neq h$

We want: Uniformly distributed {0,1}-Sequence.



Output is Uniform, since Pr[HT] = Pr[TH] = ht

Extractors: Formal Definition

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

$$Ext : \{0,1\}^n \times \{0,1\}^d \rightarrow \{0,1\}^m$$

so that for an arbitrary source S on $\{0,1\}^n$ with $H_\infty(S) \geq k$, $Ext(S, U_d)$ is ϵ-close to U_m.

Weak randomness
(n Bits)

0001010001...

Extractor

Random output
(m Bits)

0010...

Truly random seed
(d Bits)

Why do we need a Seed?

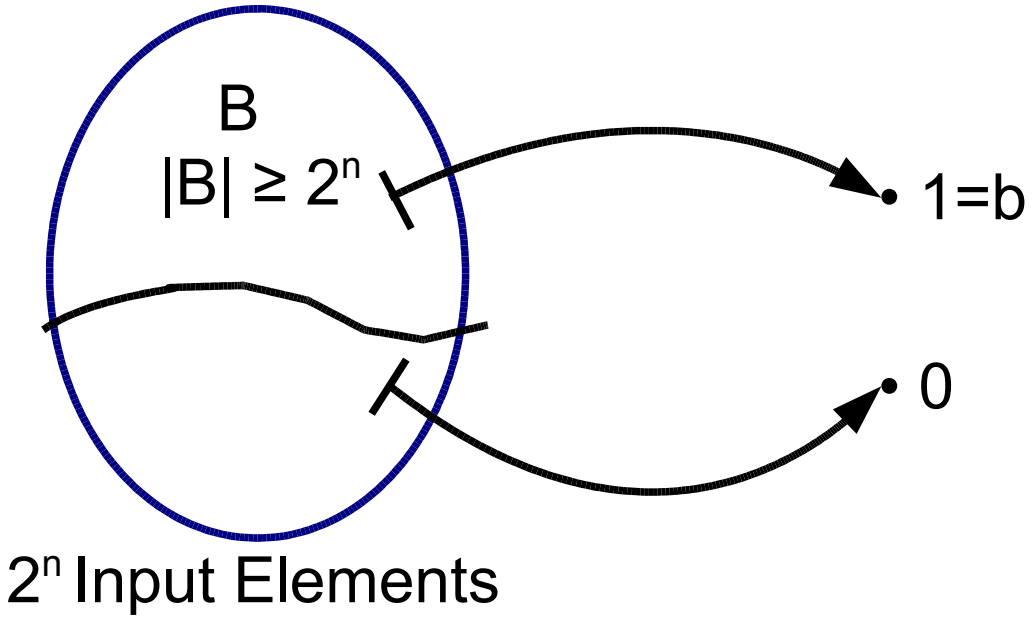
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Claim: for any $Ext : \{0,1\}^n \rightarrow \{0,1\}$, $\exists$ (n-1)-source S s.t. $Ext(S) = const$.

B, b as defined in figure. Let S be the uniform distribution on B

$\Rightarrow H_\infty(S) \geq \log \frac{1}{1/2^{n-1}} = n - 1$

$\Rightarrow Ext(S) = b = const$



Optimal Extractors

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

The following lower bounds apply to extractors for arbitrary sources. (i.e., any n, $k \leq n$, $\epsilon > 0$.)

- Seed length:
  $d \geq \log(n-k) + 2\log(\frac{1}{\epsilon}) - O(1)$
  [Nisan, Zuckermann, 96]

- Entropy loss:
  no non-trivial extractor can extract all randomness present in a source.
  $k + d - m \geq 2\log(\frac{1}{\epsilon}) - O(1)$
  [Radhakrishnan, Ta-Shma, 00]

Asymptotically optimal constructions are known, although not for arbitrary k.

Simple Extractor Construction

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Construction is based on a universal family of hash functions:

$$H = \left\{h : \{0,1\}^n \to \{0,1\}^l\right\} \text{ s.t. } \Pr_{h \in H, x \neq y}[h(x) = h(y)] \leq \frac{1}{2^l}$$

Example: $H = \{H_{ab}\} = last\ l\ bits\ of\ ax + b$ in GF(2^n).

$$Ext(x,h) := h||h(x)$$

is Extractor with Error ϵ and min-entropy-threshold k if $l = k - 2\log(\frac{1}{\epsilon}) - O(1)$.

Proof: show that $H||H(X)$ is ϵ-close to $U_d \times U_l$.

Proof(1): Collision Prob. is small

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Claim: $cp\,(H||H(X)) \simeq cp\,(U_d \times U_l)$.

Facts:

$|H| = 2^d$, because d bits of seed define 2^d hash functions.

$H_\infty(X) \geq k \Rightarrow \Pr[X = X'] \leq \frac{1}{2^k}$.

$\Pr\,[H(X) = H(X')|X \neq X'] = \frac{1}{2^l}$ by definition of universal hash functions.

Proof: $cp\,(H||H(X)) = \Pr\,[(H||H\,(X)) = (H'||H'\,(X'))]$

$$= \frac{1}{|H|}\left(\Pr\,[X = X'] + \Pr\,[H(X) = H(X')|X \neq X']\,\Pr\,[X \neq X']\right)$$

$$= \frac{1}{2^d}\left(\frac{1}{2^k} + \frac{1}{2^l}\left(1 - \frac{2}{2^k}\right)\right) \leq \frac{1}{2^d}\left(\frac{1}{2^k} + \frac{1}{2^l}\right) = \frac{1}{2^d 2^l}\left(1 + \epsilon^2\right)$$

(with $\epsilon = \frac{2^l}{2^k}$)

Proof(2): Error is small

Let $U := U_d \times U_l$, $V := H||H(X)$. Let Θ be the output domain.

Claim: $cp(V) \simeq cp(U) \Rightarrow V$ is ϵ-close to U

Proof: $\Delta(V,U) = \frac{1}{2}\sum_{x\in\Theta} |\Pr_U[x] - \Pr_V[x]|$

$$\leq \frac{\sqrt{|\Theta|}}{2}\sqrt{\sum_{x\in\Theta} |\Pr_V[x] - \Pr_U[x]|^2}$$

$$= \frac{\sqrt{|\Theta|}}{2}\sqrt{\sum_{x\in\Theta} \left(\Pr_V[x]^2 - 2\frac{\Pr_V[x]}{|\Theta|} + \frac{1}{|\Theta|^2}\right)}$$

$$= \frac{\sqrt{|\Theta|}}{2}\sqrt{cp(V) - \frac{1}{|\Theta|}}$$

$$\leq \frac{\sqrt{2^d 2^l}}{2}\sqrt{\frac{\epsilon^2}{2^d 2^l}} = \frac{\epsilon}{2}$$

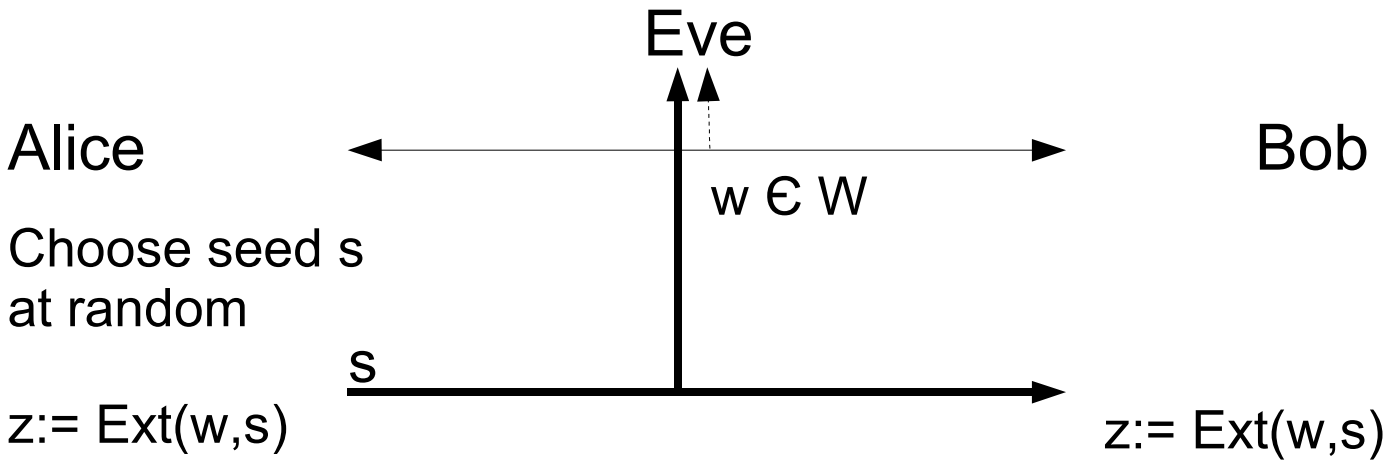
Privacy Amplification

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Setting: Alice and Bob know variable w, about which Eve has partial knowledge, i.e. Eve knows v according to a distribution P_{VW}. v contains at most t bits of information about w. Assume $W = U_{2^n}$.

Alice & Bob want a function f s.t. Eve has almost no information about $z = f(w)$.

Solution: Extractor function!



Privacy Amplification: Proof

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Claim: $\Delta((Z||V),(U \times V)) \leq \epsilon$

Proof: $\Delta((Z||V),(U \times V)) = \ldots$
$= \sum_v \Pr_V(v) * \Delta((Z|V=v),U)$

Because Eve knows at most t bits about W, $H_\infty(W|V=v) \geq n-t$.

$\Rightarrow$we can use an Extractor with min-entropy-threshold $n-t$ and error ϵ to extract a random Z from W
$\Rightarrow \forall v: \Delta((Z|V=v),U) \leq \epsilon$
$\Rightarrow \sum \Pr_V(v) * \Delta((Z|V=v),U) \leq \epsilon$

Whatever algorithm Eve uses to discover z, her advantage over blindly guessing is not more than ϵ!

Thanks for your attention!

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Further reading:

- Salil Vadhan, Lectures on Pseudorandomness (12-14).
  `http://www.courses.fas.harvard.edu/~cs225/` .
- Ronen Shaltiel. Recent developments in extractors.
  `http://www.cs.haifa.ac.il/~ronen/online_papers/survey.ps` .
- Valentine Kabanets, Lectures on Pseudorandomness (12-15).
  `www.cs.sfu.ca/~kabanets/cmpt881/lec/` .
- Slides: `http://n.ethz.ch/student/fuchsd/extractors.pdf` .