Meta Complexity

Lecture 2

Ronald de Haan
me@ronalddehaan.eu

University of Amsterdam

January 8, 2023

What will we cover in this lecture?

- One-way functions
- Average-case complexity

Most forms of cryptography depend on $\text{P} \neq \text{NP}$

- Whenever there is a private key with the property that an encoded message can be decoded efficiently with the private key, this is an NP problem
- So if $\text{P} = \text{NP}$, breaking the cryptographic scheme can be done in polynomial time

One-way functions (OWFs)

Definition (one-way functions)

A polynomial-time computable function $f : \{0,1\}^* \to \{0,1\}^*$ is a *one-way function* if for every polynomial-time probabilistic TM $\mathbb{M}$ there is a negligible function $\epsilon : \mathbb{N} \to [0,1]$ such that for every $n \in \mathbb{N}$:

$$\mathop{\mathbb{P}}_{\substack{x \in_{\mathsf{R}} \{0,1\}^n \\ y = f(x)}} \left[\ \mathbb{M}(y) = x' \text{ such that } f(x') = y\ \right] < \epsilon(n)$$

where a function $\epsilon : \mathbb{N} \to [0,1]$ is *negligible* if $\epsilon(n) = \frac{1}{n^{\omega(1)}}$, that is, for every c and sufficiently large n, $\epsilon(n) < \frac{1}{n^c}$.

- Conjecture: there exist one-way functions (implying $\mathsf{P} \neq \mathsf{NP}$)
- OWFs can be used to create private-key cryptography

Levin's universal OWF

- Consider the function f_U that is defined as follows.
  If there exists any OWF f, then f_U is also an OWF.
  - Treat the input x as a list $x_1, \ldots, x_{\log n}$ of $n/\log n$ bit long strings.
  - Output $\mathbb{M}_1^{n^2}(x_1) \ldots \mathbb{M}_{\log n}^{n^2}(x_{\log n})$.
  - Here $\mathbb{M}_i^t(y)$ denotes the output that the ith TM $\mathbb{M}_i$ gives on input y, or $0^{|y|}$ if $\mathbb{M}_i$ takes more than t steps on input y.
- Main idea:
  - If there is an OWF, then there is one that runs in time n^2—using padding.
  - The function that concatenates the output of several (polynomial-time computable) functions $f_1, \ldots, f_k$ is an OWF if and only if at least one of $f_1, \ldots, f_k$ is an OWF.
  - Whenever n gets large enough, there is some $\mathbb{M}_i$ that is an OWF that runs in time at most n^2, and so therefore is f_U.

OWFs as building blocks for cryptography

Definition

An *encryption scheme* is a pair (E, D) of algorithms, each taking a key k and a message x, such that $D_k(E_k(x)) = x$.

The scheme is *perfectly secret*, for messages of length m and keys of length n, if for every pair $x, x' \in \{0,1\}^m$ of messages, the distributions $E_{U_n}(x)$ and $E_{U_n}(x')$ are identical.

The scheme is *computationally secure* if for every probabilistic polynomial-time algorithm A, there is a negligible function $\epsilon : \mathbb{N} \to [0,1]$ such that

$$\underset{\substack{k \in_R \{0,1\}^n \\ x \in_R \{0,1\}^m}}{\mathbb{P}} [\ A(E_k(x)) = (i, b) \text{ s.t. } x_i = b\ ] < 1/2 + \epsilon(n).$$

- Suppose that OWFs exist. Then for every $c \in \mathbb{N}$ there exists a computationally secure encryption scheme (E, D) using n-length keys for n^c-length messages.

Worst-case complexity

- A problem $L \subseteq \{0,1\}^*$ can be solved in *worst-case running time* $T(n)$ if there exists an algorithm A that solves L and that halts within time $T(|x|)$ for each $x \in \{0,1\}^*$.
- In other words, the worst-case running time $T(n)$ is the maximum of the running times for all inputs of size n.

Distributional problems

Definition (distributional problems)

A *distributional problem* $\langle L, \mathcal{D} \rangle$ consists of a language $L \subseteq \{0,1\}^*$ and a sequence $\mathcal{D} = \{\mathcal{D}_n\}_{n \in \mathbb{N}}$ of probability distributions, where each $\mathcal{D}_n$ is a probability distribution over $\{0,1\}^n$.

The class distP / avgP

Definition (distP)

$\langle L, \mathcal{D} \rangle$ is in the class distP (also called: avgP) if there exists a deterministic TM $\mathbb{M}$ that decides L and a constant $\epsilon > 0$ such that for all $n \in \mathbb{N}$:

$$\mathop{\mathbb{E}}_{x \in_{\mathsf{R}} \mathcal{D}_n} [\ \mathrm{time}_{\mathbb{M}}(x)^{\epsilon}\ ] \text{ is } O(n).$$

- The ϵ is there for technical reasons—to invert a polynomial to $O(n)$.

Polynomial-time computable distributions

Definition (P-computable distributions)

A sequence $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ of distributions is P-*computable* if there exists a polynomial-time TM that, given $x \in \{0,1\}^n$, computes:

$$\mu_{\mathcal{D}_n}(x) = \sum_{\substack{y\in\{0,1\}^n \\ y\leq x}} \mathbb{P}_{\mathcal{D}_n}[y],$$

where $y \leq x$ if the number represented by the binary string y is at most the number represented by the binary string x.

Polynomial-time samplable distributions

Definition (P-samplable distributions)

A sequence $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ of distributions is P-*samplable* if there exists a polynomial-time probabilistic TM $\mathbb{M}$ such that for each $n \in \mathbb{N}$, the random variables $\mathbb{M}(1^n)$ and $\mathcal{D}_n$ are equally distributed.

The class distNP and sampNP

Definition (distNP)

A problem $\langle L, \mathcal{D} \rangle$ is in distNP if $L \in \mathsf{NP}$ and $\mathcal{D}$ is P-computable.

Definition (sampNP)

A problem $\langle L, \mathcal{D} \rangle$ is in sampNP if $L \in \mathsf{NP}$ and $\mathcal{D}$ is P-samplable.

- The questions "distNP $\stackrel{?}{=}$ distP" and "sampNP $\stackrel{?}{=}$ distP" are average-case analogues of the question "NP $\stackrel{?}{=}$ P"

Zero-error average-case hardness

Definition (zero-error heuristics)

A *zero-error heuristic H for f* is a probabilistic polynomial-time algorithm that for each $x \in \{0,1\}^*$, when given x as input, it outputs either $f(x)$ or "?".

Definition (zero-error average-case hardness)

Let $\alpha : \mathbb{N} \to [0,1]$ be a function. A function $f : \{0,1\}^n \to \{0,1\}$ is *zero-error* α*-hard-on-average* if for all zero-error heuristics H for f and all sufficiently large $n \in \mathbb{N}$, it holds that:

$$\mathbb{P}_{x \in_R \{0,1\}^n} [\ H(x) = \text{“?”}\ ] \geq \alpha(n).$$

Summary

- One-way functions
- Average-case complexity