iO - powered proofs / impossibility of iO

obfuscator → PPT compiler that takes as input a program or circuit P and outputs a new program O(P) with same functionality but "unintelligible" in some sense (e.g., "virtual black box")

Turing Machine → tape + head + states | sequential

Boolean Circuit → logic gates AND, OR, NOT. input: (fixed) string of bits | output: 1 bit.

- If C is a circuit with [n inputs, m outputs] and $x \in \{0,1\}^n$, $C(x) \in \{0,1\}^m$. C computes a function $f: \{0,1\}^n \to \{0,1\}^m$ if $\forall x \in \{0,1\}^n \to$ [C(x) "f(x)"]
- $\forall$ algorithms A, M and string x, $A^M(x)$ is the output of A when executed on input x and oracle access to M.
  - if M circuit → $A^M(x) = A(M(x))$
  - if M, TM → $A^M(x) = \begin{cases} A(M(x)) & \text{if } M \text{ halts within } t \text{ steps on } x \\ A(\perp) & \text{otherwise} \end{cases}$

  | - if A is TM, $A(x;r) \to A(x)$ with random tape r.
  | $A(x) \to$ distribution by choosing $r \xleftarrow{\$} U$ and running $A(x;r)$
- if D distribution, $Supp(D) = \{x : P(x) \neq 0\}$
- if M is TM, $\langle M \rangle : 1^* \times \{0,1\}^* \to \{0,1\}^*$

  $(1^t, x) \to \begin{cases} y & \text{if } M(x) \text{ halts with output } y \text{ after at most } t \text{ steps} \\ \perp & \text{otherwise} \end{cases}$

  | • if C circuit, $[C] \equiv$ function it computes.
  | • if M is TM, $[M] \equiv$ (possibly partial) function it computes

---

Virtual Black Box - based obfuscation: (O(P) should reveal only black box access to P) — Barak et al

- (computational indistinguishability) [strongest require] $\begin{bmatrix} \forall A, \exists S: \{A(O(P))\} \approx \{S^P(1^{|P|})\} \\ \forall P \end{bmatrix}$
- (computing a predicate) [weakest require] $\begin{bmatrix} \forall \alpha & \forall \pi: \mathcal{F} \to \{0,1\} \\ \forall P & \forall P \text{ computing a function } f \in \mathcal{F} \end{bmatrix} \left| P(A(O(C)) = \pi(f)) - P(S^C(1^{|C|}) = \pi(C)) \right| \leq negl(|C|)$

Def: Inherently "unobfuscatable" function ensemble

$\{H_k\}_{k \in \mathbb{N}}$ distributions of finite functions $f: \{0,1\}^{in(k)} \to \{0,1\}^{out(k)}$ s.t

| • $f \xleftarrow{\$} H_k$ efficiently computable
| • $\exists \pi: \bigcup_{k \in \mathbb{N}} Supp(H_k) \to \{0,1\}$ s.t
| 1. $\forall$ PPT S, $P[S^f(1^k) = \pi(f)] \leq \frac{1}{2} + negl(k)$ $f \xleftarrow{\$} H_k$
| 2. $\exists$ PPT A: $[\forall f \in \bigcup_{k=1}^{\infty} Supp(H_k)] \Rightarrow A(C) = \pi([C])$ $\forall C: [C] = f$

Theorem 3.9: [$\exists$ OWF $\Rightarrow \exists$ inherently unobfuscatable function ensemble]

Def:
- 2-TM/C obfuscator: $\forall$ PPT A, $\exists$ [PPT S: $\forall$ TM/C M, N] [negl α] → $|P[A(O(M), O(N)) = 1] - P[S^{M,N}(1^{|M|+|N|}) = 1]| \leq \alpha(\min\{|M|, |N|\})$
  - ↳ $\nexists$ (counterexample: $\alpha, \beta \in \{0,1\}^k$, $\begin{bmatrix} C_{\alpha,\beta}(x) = \begin{cases} \beta & x = \alpha \\ 0^k & \text{—} \end{cases} \\ D_{\alpha,\beta}(C_{\cdot,\cdot}) = \begin{cases} 1 & C(\alpha) = \beta \\ 0 & \text{—} \end{cases} \end{bmatrix}$

  | Let $A(C, D) = D(C)$
  | → $P[A(O(C_{\alpha,\beta}), O(D_{\alpha,\beta})) = 1] = 1$ $\forall \alpha, \beta \in \{0,1\}^k$
  | However, $\forall$ PPT S cannot query exponential of inputs
  | ↳ $|P[S^{C_{\alpha,\beta}, D_{\alpha,\beta}}(1^k) = 1] - P[S^{\tilde{z}_k, D_{\alpha,\beta}}(1^k) = 1]| \leq 2^{-\Omega(k)}$ (TM outputs 0^k)
- $\nexists$ TM obfuscators