Notes for Lecture 18

1 Overview

So far in this course we have seen that $\mathsf{iO} + \mathsf{OWF}$ gives you lots of cryptopgraphy. The question today is: to what extent are one way functions necessary?

Intuitively, obfuscation is very powerful while one way functions are a weak object. Most of the cryptographic tools we've been using imply one way functions. So it seems like iO should be able to imply OWF.

To start off today, we'll see that iO doesn't necessarily imply OWF. This means that we can imagine a world consistent with the current state of complexity theory where iO exists and OWFs do not.

2 $\mathsf{P} = \mathsf{NP}$

Imagine if $\mathsf{P} = \mathsf{NP}$. In this world, OWFs don't exist. To see this, suppose we have a one way function $y = F(x)$. x is a witness for the fact that y is the image. If $\mathsf{P} = \mathsf{NP}$, we can find this witness in polynomial time.

However, you can still show that iO exists if $\mathsf{P} = \mathsf{NP}$. We can do this with canonical circuits, defined as follows. If I have $C_0 \equiv C_1$, then $Canon(C_0) = Canon(C_1)$. This easily implies a very perfect version of iO where we get complete indistinguishability, not just computational indistinguishability. One way to build canonical circuits is to find a minimal equivalent circuit.

This motivates the problem, CIRCUIT-MIN: Given C and s, does there exist a $C' \equiv C$ such that $|C'| \leq s$? Note that circuit equivalence is in coNP, since we can give a poly-size witness for non-equivalence. So we can solve this problem in $\mathsf{NP}^{\mathsf{coNP}}$ by non-deterministically picking C' and checking for equivalence. If $\mathsf{P} = \mathsf{NP}$, then this class is in P.

We can show that $\mathsf{P} \neq \mathsf{NP}$ is basically all we need to get OWF from iO. We will show that $\mathsf{iO} + \mathsf{BPP} \neq \mathsf{NP} \rightarrow \mathsf{OWF}$.

Consider, circuit satisfiability, which is in NP but not BPP. We will use a trivial circuit Z that always outputs 0. Let the one way function be $F(r) = \mathsf{iO}(Z; r)$, the obfuscation of Z under random coins r.

To see that this is one way, we'll show that being able to invert this function gives a contradiction. Assume A inverts F with non-negligible probability (so A is a BPP adversary),

$$Pr[A(F(r)) = r' \text{s.t. } F(r') = F(r)] \geq \epsilon$$

We use this to build a PPT algorithm $B(C)$ that checks if circuit C is satisfiable. It runs $r' \leftarrow A(\mathsf{iO}(C))$. B will check that $\mathsf{iO}(Z, r') = \hat{C}$. If so, it outputs *unsatisfiable*. Else, it outputs *satisfiable*.

To see why this is correct, suppose C is satisfiable. Then $C \neq Z$, so it is impossible for $\mathsf{iO}(Z, r') = \hat{C}$, so the algorithm will output *satisfiable*.

Suppose C is unsatisfiable. Consider another adversary B' that is given $\hat{C}$ instead of C, and checks that $\mathsf{iO}(Z, A(\hat{C})) = \hat{C}$. **View 1.** If $\hat{C} = \mathsf{iO}(C; r)$, then we note that $B'(\hat{C}) = B(C)$. **View 2.** If $\hat{C} = \mathsf{iO}(Z; r)$, then $\hat{C} = F(r)$, so $B'(\hat{C})$ is checking $\mathsf{iO}(Z, A(F(r))) = \mathsf{iO}(Z, r)$, and so with non-negligible probability $A(F(r)) = r$, so $B'(\hat{C})$ outputs *unsatisfiable* with probability at least ϵ.

By iO, these two views are indistinguishable. Since view 1 is equivalent to $B(C)$, we have

$$Pr[B(C) = \textit{unsatisfiable}] \geq \epsilon - negl$$

So we output *unsatisfiable* with non-negligible probability.

3 Statistically Secure iO

For statistically secure iO, we want $\forall C_0 \equiv C_1$ that $\mathsf{iO}(C_0) \approx_S \mathsf{iO}(C_1)$. In other words, we want the distributions of $\mathsf{iO}(C_0)$ and $\mathsf{iO}(C_1)$ to be exactly the same. It turns out that statistically secure iO probably doesn't exist, since it implies the collapse of the polynomial hierarchy.

We'll show that the existence of canonical circuits collapses the polynomial hierarchy. Note that this is a much weaker result, since canonical circuits imply perfectly secure iO (statistical iO in the case where the distribution is just one point), and perfectly secure iO clearly implies statistical iO.

To see this, observe that we can test circuit equivalence by just checking if the two circuits have the same canonical circuit. This implies $\mathsf{P} = \mathsf{NP}$.

If we want to prove this result for statistical iO, we note that if $C_0 \equiv C_1$, then the support of $\mathsf{iO}(C_0)$ and $\mathsf{iO}(C_1)$ is the same. So that means there exists randomness r_0, r_1 such that $\mathsf{iO}(C_0; r_0) = \mathsf{iO}(C_1; r_1)$. On the flipside, if $C_0 \not\equiv C_1$, then the supports are disjoint. So the outputs can never be the same no matter what random coins you toss.

So we can say that $C_0 \equiv C_1$ if and only if there exists r_0, r_1 such that $\mathsf{iO}(C_0; r_0) =$

$\mathsf{iO}(C_1; r_1)$. So this shows that circuit equivalence is in NP, since the random string is a witness. But circuit equivalence is in coNP, since a differing input is a witness for non-equivalence. Since circuit equivalence is complete for coNP, this actually shows that $\mathsf{NP} = \mathsf{NP} \cap \mathsf{coNP}$, which collapses the polynomial hierarchy to its first level.

These proofs rely crucially on perfect correctness. If we had approximately correct iO, these proofs won't work. Approximately correct iO is defined as $\forall x, C$, $Pr[\mathsf{iO}(C)(x) = C(x)] \geq 1 - negl$. For any particular point, with high probability the obfuscated circuit outputs the correct result. Since this is a pointwise definition, note that it leaves open the possibility that the obfuscated circuit is wrong on many points.

It turns out that approximately correct iO is enough to get most applications. With a lot of hard work, you can recover essentially all these results from today (some are a bit weaker, such as $\mathsf{iO} + \mathsf{BPP} \neq \mathsf{NP} \rightarrow \mathsf{OWF}$.