CS 7880 Graduate Cryptography	October 6, 2015

Lecture 7: Goldreich-Levin Theorem

Lecturer: Daniel Wichs *Scribe: Hridam Basu*

1 Topic Covered

- Hard core Predicate
- Goldreich-Levin Theorem

2 Hard Core Predicate

We are going to provide two definitions of hard core predicate and show that the two definitions are equivalent:-

DEFINITION 1 [Indistinguishability] A polynomial time function $hc : \{0,1\}^* \to \{0,1\}$ is a hard core predicate of f if $(f(x), hc(x)) \approx (f(x), b)$ where $x \leftarrow \{0,1\}^n, b \leftarrow \{0,1\}$ $\diamondsuit$

Now one might ask whether there exists a hard core predicate for every One-Way Function(OWF)? There is a good news and a bad news to this question. At first, let us reveal the bad news. There is no single function hc which is a hard core predicate for every OWF. Because if f is a OWF then $f'(x) = (f(x), hc(x))$ is also a OWF but hc is not a hard core predicate for f'. But the good news is that given any one-way function f we can construct a new one-way function g and a hard-core predicate for g.

Now we present an alternative deifinition of hard core predicate which is easier to work with.

DEFINITION 2 [Unpredictability] A polynomial time function $hc : \{0,1\}^* \to \{0,1\}$ is a hard core predicate of f if $\forall$ PPT "predictor" P

$$Pr[P(f(x)) = hc(x) : x \leftarrow \{0,1\}^n] \leq 1/2 + \mathsf{negl}(n)$$

$\diamondsuit$

This definition means that an adversary can't do much better in predicting $hc(x)$ than simply guessing a random bit.

Lemma 1 *Indistinguishability implies Unpredictability.*

Proof: We prove that if hc does not satisfy unpredictability than it does not satisfy indistinguisahability.

Assume $\exists$ PPT "predictor" P such that $\Pr[P(f(x)) = hc(x)] \geq 1/2 + \varepsilon(n)$. Define a distinguisher D via

$$D(y, b) = \{\text{If } P(y) = b, \text{ output 1, else output 0}\}$$

Then

$$\Pr[D(f(x), hc(x)) = 1] - \Pr[D(f(x), b) = 1] \geq \frac{1}{2} + \varepsilon(n) - \frac{1}{2} = \varepsilon(n)$$

where all probabilities are over $x \leftarrow \{0,1\}^n, b \leftarrow \{0,1\}$.

So if we can predict with non-negligible advantage ε , then we can distinguish by non-negligible advantage ε.

$\square$

Lemma 2 *Unpredictability implies indistinguishability.*

Proof: We prove that if hc does not satisfy indistinguisahability than it does not satisfy unpredictability. Suppose $\exists$ PPT "distinguisher" D and $\varepsilon(n) \neq negl(n)$ such that

$$|\Pr[D(f(x), hc(x)) = 1] - \Pr[D(f(x), b) = 1]| \geq \varepsilon(n)$$

where $x \leftarrow \{0,1\}^n, b \leftarrow \{0,1\}$.

Without loss of generality, we can remove the absolute value of the above equation by potentially flipping the output bit of D to ensure that the difference is positive. In slightly more detail, we know that $|\Pr[D(f(x), hc(x)) = 1] - \Pr[D(f(x), b) = 1]| > 1/n^c$ for some constant c and infintiely many n. Therefore either $\Pr[D(f(x), hc(x)) = 1] - \Pr[D(f(x), b) = 1] > 1/n^c$ for infinitely many n or $\Pr[D(f(x), hc(x)) = 0] - \Pr[D(f(x), b) = 0] > 1/n^c$ for infinitely many n. In the latter case, we can flip the output bit of D.

Define

$$P(y) = \{ \text{ Choose } b \leftarrow \{0,1\} \quad : \quad \text{If } D(y,b) = 1, \text{ output } b, \text{ else } \bar{b} \ \}$$

First note that:

$$\begin{aligned}
\Pr[D(f(x), b) = 1] &= \Pr[D(f(x), b) = 1, b = hc(x)] + \Pr[D(f(x), b) = 1, b = \overline{hc}(x)] \\
&= \frac{1}{2}(\Pr[D(f(x), hc(x)) = 1] + \Pr[D(f(x), \overline{hc}(x) = 1]) \\
&\Rightarrow \Pr[D(f(x), \overline{hc}(x) = 1] = 2\Pr[D(f(x), b) = 1] - \Pr[D(f(x), hc(x)) = 1]
\end{aligned}$$

This implies

$$\begin{aligned}
\Pr[P(f(x)) = hc(x)] &= \Pr[D(f(x), hc(x)) = 1, b = hc(x)] + \Pr[D(f(x), \overline{hc}(x)) = 0, b = \overline{hc}(x)] \\
&= \frac{1}{2}(\Pr[D(f(x), hc(x) = 1] + \Pr[D(f(x), \overline{hc}(x) = 0]) \\
&= \frac{1}{2}(\Pr[D(f(x), hc(x) = 1] + 1 - \Pr[D(f(x), \overline{hc}(x) = 1]) \\
&= \frac{1}{2} + \frac{1}{2}(\Pr[D(f(x), hc(x) = 1] - \Pr[D(f(x), \overline{hc}(x) = 1]) \\
&= \frac{1}{2} + \frac{1}{2}(2\Pr[D(f(x), b) = 1] - \Pr[D(f(x), hc(x)) = 1]) \\
&= \frac{1}{2} + \varepsilon(n)
\end{aligned}$$

where the second to last line follows by substituting for $\Pr[D(f(x), \overline{hc}(x) = 1]$ using the previous derivation.

$\square$

3 Goldreich Levin Theorem

Theorem 1 *If f is a one way function, then $g(x,r) = (f(x), r)$ is also a one way function and $hc(x,r) = \langle x, r\rangle = \sum(x_i \cdot r_i)(\mod 2)$ is a hard core predicate of g.*

As an alternate interpretation of the Goldreich-Levin theorem, we can think of $hc(x,r) = \langle x, r\rangle$ as a randomized hard core predicate for any one way function f, meaning that

$$(f(x), r, hc(x,r)) \approx (f(x), r, b)$$

where $x, r \leftarrow \{0,1\}^n, b \leftarrow \{0,1\}$.

We will finish the poof of the Goldreich-Levin theorem in the next lecture, but let's start to build some intuition for the proof and see what the main components are.

We do a proof by contradiction. Suppose hc is not a hard core predicate of g, then we wish to show that f is not a one-way function. By the unpredictability definition of hard-core predictates we know that $\exists$ PPT P, $\varepsilon(n) \neq$ negl(n) such that
$\Pr[P(f(x), r) = \langle x, r\rangle] \geq \frac{1}{2} + \varepsilon(n)$
We want to show that we can invert f. We first explore some simple cases that make the proof much easier.

Simple Case 1 : Suppose $\Pr[P(f(x), r) = \langle x, r\rangle] = 1$

The Algorithm to invert OWF f is:-
$A(y)$:
for $i = 1, ..., n$
$\tilde{x}_i = P(y, e_i)$
Output $x = (\tilde{x}_1, ...\tilde{x}_n)$

Here e_i denotes the i'th standard basis vector (all 0 except for 1 in i'th position). The algorithm is correct since we are guaranteed that $\tilde{x}_i = P(y, e_i) = \langle x, e_i\rangle = x_i$.

Simple Case 2 : Suppose $\forall x$ (ie, not only for any random x), $\Pr[P(f(x), r) = \langle x, r\rangle] \geq \frac{3}{4} + \frac{1}{p(n)}$ where the probability is over $r \leftarrow \{0,1\}^n$. In this case, we have no guarantees on $P(y, e_i)$ giving us correct answers since e_i is not random. Here is a smarter strategy.

Call $b_1 = P(y, r)$, $b_2 = P(y, r + e_i)$ where $r \leftarrow \{0,1\}^n$.
Output $x_i = b_2 - b_1$
Note: r and $r + e_i$ are individually random but not independent.
If $P(y, r)$, $P(y, r + e_i)$ are both "correct" then: $x_i = b_2 - b_1 = \langle x, r + e_i\rangle - \langle x, r\rangle = \langle x, e_i\rangle$ is also correct. Moreover:

Pr[Both b_1 and b_2 are correct]
= 1 - Pr[At least one of them is wrong]
$= 1 - (\frac{1}{4} - \frac{1}{p(n)} + \frac{1}{4} - \frac{1}{p(n)}) = \frac{1}{2} + \frac{2}{p(n)}$
We have to run the above procedure many times for the i-th bit and take the majority vote. If there are enough votes, majority is correct with high probability (Chernoff bound).

There are two main differences between Simple Case 2 and what we need to prove. Most importantly, our predictor is only correct with probability $1/2 + \varepsilon(n)$ rather than $3/4 + 1/p(n)$. Secondly, in our case the probability is over a random x, r whereas in simple case 2 it's only over random r for worst-case x. We show how to handle the second problem. Essentially, this is an "averaging argument" which shows that if some probability is high over random x, r then for many x the probability is high over a random r.

Claim 1 *$\forall n \in \mathbb{N}$, $\exists G_n \subseteq \{0,1\}^n$ of size $|G_n| \geq \frac{\varepsilon(n)}{2}.2^n$ ($\frac{\varepsilon(n)}{2}$ is the density) such that $\forall x \in G_n$:*

$$\Pr_{r \leftarrow \{0,1\}^n}[P(f(x), r) = \langle x, r \rangle] \geq \frac{1}{2} + \frac{\varepsilon(n)}{2} \tag{1}$$

Proof: Define $G_n = \{x : \text{equation 1 holds }\}$. Then

$$\begin{aligned}
\frac{1}{2} + \frac{\varepsilon(n)}{2} &\leq \Pr_{x,r}[P(f(x), r) = \langle x, r \rangle] \\
&= \Pr_{x,r}[P(f(x), r) = \langle x, r \rangle, x \in G_n] + \Pr_{x,r}[P(f(x), r) = \langle x, r \rangle, x \notin G_n] \\
&\leq \Pr_x[x \in G_n] + \frac{1}{2} + \frac{\varepsilon(n)}{2} \\
&\Rightarrow \Pr_x[x \in G_n] \geq \frac{\varepsilon(n)}{2} \\
&\Rightarrow |G_n| \geq \frac{\varepsilon(n)}{2} \cdot 2^n
\end{aligned}$$

□

So there are many good values x for which $P(f(x), r)$ answers correctly on most r. In the next lecture we will show that this is sufficient to invert $f(x)$. This is essentially a *decoding* problem which we abstract in the next claim (to be proved next time):

Claim 2 *For any $\delta(n) = \frac{1}{\mathsf{poly}(n)}$ there exists a PPT algorithm Dec^O and a polynomial $p(n) = \mathsf{poly}(n)$ such that for all $n \in \mathbb{N}, \forall x \in \{0,1\}^n$:*
If $\Pr[O(r) = \langle x, r \rangle] \geq \frac{1}{2} + \delta(n)$
then $\Pr[Dec^O(1^n) = x] \geq \frac{1}{p(n)}$.
(The notation Dec^O denotes that Dec has oracle access to O meaning that it can call O on arbitrary values r.)

We will prove this claim and discuss a connection to coding theory next time.