Introduction to Cryptography Feb 1, 2018

Lecture 6: Hard-core predicates

Instructor: Vipul Goyal *Scribe: Apoorva Bhagwat*

1 Motivation

We've seen that if $f : \{0,1\}^n \to \{0,1\}^m$ is a one-way function, then the function $F : \{0,1\}^{2n} \to \{0,1\}^{m+n}$ given by

$$F(x_1||x_2) = f(x_1)||x_2 \text{ (where } |x_1| = |x_2|)$$

is also a one-way function, yet it leaks half of its input bits by outputting them in clear.

One useful question we can ask is : is there some function of the input bits that a one-way function hides? In other words, for a OWF $f : \{0,1\}^n \to \{0,1\}^m$, we would like a function $h : \{0,1\}^n \to \{0,1\}$ such that given $f(x)$, it's hard to predict $h(x)$. Such a function h is called a *hard-core predicate* for f. We formalize this idea in the following definition.

Definition 1 *(Hard-core predicates) Given a function $f : \{0,1\}^n \to \{0,1\}^m$, we say that $h : \{0,1\}^n \to \{0,1\}$ is a hard-core predicate (or a hard-core bit) for f if —*

1. *h is computable in polynomial time.*
2. *There is a negligible function $\nu(\cdot)$ such that for every PPT (probabilistic polynomial time) adversary $\mathcal{A}$ and for every n, we have :*

$$\mathbb{P}[x \xleftarrow{\$} \{0,1\}^n : \mathcal{A}(f(x)) = h(x)] \leq \frac{1}{2} + \nu(n)$$

where " $x \xleftarrow{\$} S$ " means 'choose x uniformly at random from the set S'.

In other words, given $f(x)$, a computationally bounded adversary who wants to figure out $h(x)$ cannot do much better than just randomly guessing 0 or 1 with equal probability.

Remark 1 *Note that the second condition in the above definition is based on an experiment, i.e. we demand that for a randomly chosen string x, the adversary should not be able to guess the h-value of that particular x better than chance.*

We can imagine a different definition of a hard-core predicate that has the following requirement instead :

$$\mathbb{P}[x \xleftarrow{\$} \{0,1\}^n : \mathcal{A}(f(x)) = h(x') \text{ such that } f(x) = f(x')] \leq \frac{1}{2} + \nu(n)$$

It is easier for an adversary to break this requirement if the OWF f has collisions (i.e. if it's not injective).

From here on, we will abbreviate hard-core predicates as HCPs.

2 Trivial and Non-trivial HCPs

Let $f : \{0,1\}^n \to \{0,1\}^m$ be a OWF, and let $F : \{0,1\}^{n+1} \to \{0,1\}^m$ be the OWF given by $F(b||x) = f(x)$, where b is a single bit and x is an n-bit string.

Now consider the HCP for F given by $h(b||x) = b$. This is certainly a HCP for F. It is *trivial* in the sense that for any string s, $F(s)$ contains no information about $h(s)$ — even a computationally unbounded adversary will not be able to confidently recover $h(s)$ given $F(s)$; not only is it computationally infeasible to recover $h(s)$, it is information-theoretically impossible to do so. Such HCPs are not interesting from the point of view of cryptography.

This motivates the following definition.

Definition 2 *(Non-trivial HCPs) An HCP h for a OWF f is said to be non-trivial if for every string s, $f(s)$ contains full-information about $h(s)$, i.e. given $f(s)$, a computationally unbounded adversary can compute $h(s)$ exactly. This is equivalent to saying that if $f(a) = f(b)$, then $h(a) = h(b)$.*

3 HCPs based on Inner Products

Throughout the course, we will use a HCP based on the inner product in the vector space of n-bit strings.

The inner product of two strings $x, y \in \{0,1\}^n$ is given by :

$$\langle x, y \rangle = \left(\sum_{i=1}^{n} x_i y_i \right) \mod 2$$

4 Goldreich-Levin Theorem

We still haven't seen whether it is actually possible to construct HCPs. It turns out that we can, assuming the existence of one-way functions. This was shown by Oded Goldreich and Leonid Levin in 1989.

Theorem 1 *(Goldreich-Levin) Given a OWF f, we can construct another OWF g and a HCP for g. Specifically, suppose $f : \{0,1\}^n \to \{0,1\}^n$ is a OWF. Define $g : \{0,1\}^{2n} \to \{0,1\}^{2n}$ as*

$$g(x||r) = f(x)||r \text{ (where } |x| = |r| = n)$$

and $h : \{0,1\}^{2n} \to \{0,1\}$ as

$$h(x||r) = \langle x, r \rangle \text{ (again } |x| = |r| = n)$$

Then, g is a OWF and h is a HCP for g.

We have already seen in the lecture on OWFs how to show that g is a OWF. It remains to show that h is a HCP for g. We won't prove this theorem in its full generality, but rather two weaker versions of it. If we were to prove the theorem in full generality, we would start by assuming the existence of a PPT adversary $\mathcal{A}$ with the following property :

$$\mathbb{P}[s \xleftarrow{\$} \{0,1\}^{2n} : \mathcal{A}(g(s)) = h(s)] \geq \frac{1}{2} + \epsilon(n)$$

where ϵ is a noticeable function of n. Then, we would contradict the assumption that f is a OWF.

However, to make our life easier, we start by assuming something stronger about the adversary $\mathcal{A}$, namely that $\mathcal{A}$ is perfect (i.e. $\epsilon(n) = \frac{1}{2}$). Note that this amounts to proving a weaker version of the theorem.

Theorem 2 *(Goldreich-Levin, Warmup 1) There is no perfect adversary against the HCP h given by Goldreich-Levin.*

Proof. Suppose there was such a perfect adversary[1] $\mathcal{A}$. This means that :

$$\text{For every } s \in \{0,1\}^{2n},\ \mathbb{P}[\mathcal{A}(g(s)) = h(s)] = 1$$

We use $\mathcal{A}$ to construct $\mathcal{B}$, an adversary that inverts f with some noticeable probability.

Now we define $\mathcal{B}$. On input $y \in \{0,1\}^m$, $\mathcal{B}$ does the following. For every $i \in \{1, \cdots, n\}$, $\mathcal{B}$ computes $a_i = \mathcal{A}(y||e_i)$[2]. Finally, it outputs $a_1 a_2 \cdots a_n$.

Now we investigate why $\mathcal{B}$ successfuly inverts f. First, observe that $g(x||e_i) = f(x)||e_i$ by definition. Then notice that since $\mathcal{A}$ is perfect, for any $x \in \{0,1\}^n$, $\mathcal{A}(f(x)||e_i)$ returns $h(x||e_i)$. This is equal to $\langle x, e_i \rangle = x_i$, i.e. the i^{th} bit of x. This means that if $\mathcal{B}$ receives $f(x)$ as input for some x, then it outputs $x_1 x_2 \cdots x_n$, i.e. x, thus inverting f with probability 1. ■

Now, we weaken our assumption that the adversary is perfect, and prove a stronger version of Theorem 2.

Theorem 3 *(Goldreich-Levin, Warmup 2) There is no adversary $\mathcal{A}$ with the following property (against the HCP h given by Goldreich-Levin) :*

$$\textit{For every } x \in \{0,1\}^n,\ \mathbb{P}[r \xleftarrow{\$} \{0,1\}^n : \mathcal{A}(g(x||r)) = h(x||r)] \geq \frac{3}{4} + \epsilon(n)$$

where ϵ is a noticeable function of n.

Proof. Suppose there was such an adversary $\mathcal{A}$. We want to use $\mathcal{A}$ to construct $\mathcal{B}$, an adversary that inverts f with some noticeable probability.

Note that unlike in Theorem 2, we cannot simply plug in e_i as the last n bits of our input to $\mathcal{A}$. This is because $\mathcal{A}$ is only guaranteed to succeed with a high probability if the last n bits of its inputs are chosen uniformly at random; for all we know, $\mathcal{A}$ might have the property that it always fails when the last n bits of its input are of the form e_i. However, we can get around this problem by cleverly choosing these n bits.

Before we define $\mathcal{B}$, let's first define a random experiment $E(y, i)$ ($y \in \{0,1\}^n$ and $i \in \{1, \cdots, n\}$ are parameters to the experiment). The experiment involves choosing a string r uniformly at random from $\{0,1\}^n$, and then computing the two bits $b = \mathcal{A}(y||e_i)$ and $b' = \mathcal{A}(y||(r \oplus e_i))$. The bit $b \oplus b'$ is considered to be the outcome of the experiment.

Then, we define $\mathcal{B}$. On input $y \in \{0,1\}^m$, $\mathcal{B}$ does the following. For each $i \in \{1, \cdots, n\}$, $\mathcal{B}$ runs the experiment $E(y, i)$ $\frac{2n}{\epsilon(n)}$ times (each time with fresh, independent random bits), and computes the more frequent outcome. This bit is called a_i. Finally, it outputs $a_1 a_2 \cdots a_n$.

[1] Note that this is only possible if $g(s)$ contains full information about $h(s)$

[2] e_i is the n-bit string with a 1 at the i^{th} bit and 0s everywhere else

Now we investigate why $\mathcal{B}$ inverts f with noticeable probability. Suppose $\mathcal{B}$ is given $f(x)$ for some x. First, observe that in any single run of the experiment $E(f(x), i)$, $b = \mathcal{A}(f(x)||r) = \mathcal{A}(g(x||r)) = h(x||r) = \langle x, r\rangle$ with probability at least $\frac{3}{4} + \epsilon(n)$, because r is chosen uniformly at random. Similarly, $b' = \mathcal{A}(f(x)||(r \oplus e_i)) = \mathcal{A}(g(x||(r \oplus e_i))) = h(x||(r \oplus e_i)) = \langle x, r \oplus e_i\rangle$ with probability at least $\frac{3}{4} + \epsilon(n)$, because $r \oplus e_i$ also has a uniformly random distribution (even though it is certainly not independent from r). Then, note that if indeed $b = \langle x, r\rangle$ and $b' = \langle x, r \oplus e_i\rangle$, then the outcome of the experiment is

$$b \oplus b' = \langle x, r\rangle \oplus \langle x, r \oplus e_i\rangle = \langle x, r \oplus (r \oplus e_i)\rangle = \langle x, e_i\rangle = x_i$$

Thus, we have :

$$\begin{aligned}
\mathbb{P}[E(f(x), i) \text{ doesn't output } x_i] &\leq \mathbb{P}[b \neq \langle x, r\rangle \text{ or } b \neq \langle x, r \oplus e_i\rangle] \\
&\leq \mathbb{P}[b \neq \langle x, r\rangle] + \mathbb{P}[b \neq \langle x, r \oplus e_i\rangle] \qquad \text{(by a union bound[3])} \\
&\leq \left(\frac{1}{4} - \epsilon(n)\right) + \left(\frac{1}{4} - \epsilon(n)\right) \\
&= \boxed{\frac{1}{2} - \epsilon(n)}
\end{aligned}$$

Thus, the outcome of $E(f(x), i)$ is a Bernoulli random variable whose value equals x_i with probability at least $\frac{1}{2} + \epsilon(n)$. So, if we repeat the experiment $\frac{2n}{\epsilon(n)^2}$ times (which is poly(n)), and take the majority outcome, we get the value of x_i with probability at least $1 - 2^{-n}$. This can be shown by using Chernoff bounds. Please see the appendix for more details on how the bound is applied.

This means that for each i, the bit a_i computed by $\mathcal{B}$ equals x_i with overwhelming probability $(1 - 2^{-n})$. Thus, we have :

$$\begin{aligned}
\mathbb{P}[\mathcal{B} \text{ doesn't output } x] &= \mathbb{P}[\text{There is some } i \text{ such that } a_i \neq x_i] \\
&\leq \sum_{i=1}^{n} \mathbb{P}[a_i \neq x_i] \qquad \text{(By a union bound)} \\
&\leq \sum_{i=1}^{n} 2^{-n} \\
&= \boxed{n/2^n}
\end{aligned}$$

So, $\mathcal{B}$ in fact inverts f with very high probability.

5 An application of HCPs

We'll apply HCPs to construct an encryption scheme for 1-bit messages. This scheme is not really interesting as an encryption scheme (since it only encodes 1-bit messages and doesn't do anything that a one-time pad can't do). However, it has a funny property that will be of interest to us.

[3]'Union bound' refers to the following fact : if two events A and B happen with probability p and q respectively, then their union, $A \cup B$, happens with probability at most $p + q$.

We will design a scheme[4] (E, D) such that for every PPT adversary $\mathcal{A}$, we have a negligible function $\nu(\cdot)$ such that —

$$\mathbb{P}[k \xleftarrow{\$} \{0,1\}^n, b \xleftarrow{\$} \{0,1\}, c = E(b,k) : \mathcal{A}(c) = b] \leq \frac{1}{2} + \nu(n)$$

i.e. when given a randomly chosen bit encrypted with a randomly chosen key, a computationally bounded adversary has little advantage over chance of guessing the unencrypted bit. Moreover, this advantage goes down rapidly as the key size increases.

Let $f : \{0,1\}^n \to \{0,1\}^n$ be a one-way permutation and let $h : \{0,1\}^n \to \{0,1\}$ be a HCP for f. We claim that the following scheme works :

$$E(b,k) \to f(k)||(h(k) \oplus b)$$

$$D(c,k) \to h(k) \oplus c_{n+1} \text{ (note that } c_{n+1} \text{ is the last bit of } c)$$

It is an easy exercise to show that if an adversary could break this scheme with non-negligible property, then h wouldn't be a HCP for f.

Remark 2 *The remarkable property that this system has is the following. The ciphertext from this scheme (which is an $(n+1)$-bit string) contains full information about both the key and the unencrypted message. In other words, a computationally unbounded adversary could exactly recover both k and b from the ciphertext, yet PPT adversaries cannot do this with significant advantage over chance (assuming the existence of one-way functions). This theme comes up in computational indistinguishability and pseudorandom generators, which we will study later in the course.*

6 Appendinx : applying Chernoff bound

Chernoff bounds have many general forms that can be applied in a wide variety of situations. Here, we'll state a special case of Chernoff bounds and show how we applied it in our attempt to prove Goldreich-Levin. This application can be thought of as an algorithm to boost confidence in the outcome of a noisy experiment.

Theorem 4 *Suppose there is some bit b (unknown to us), and we have a noisy random experiment E whose outcome is weakly correlated with the bit b. More formally, suppose that there is some ϵ $(0 < \epsilon < 1/2)$ such that —*

$$\mathbb{P}[\textit{The outcome of } E \textit{ is } b] \geq \frac{1}{2} + \epsilon$$

In other words, the experiment E lets us estimate the value of b slightly better than chance. Then, if we repeat the experiment N times (with fresh random bits each time) and take the majority answer, we get the true value of b with very high probability, namely $1 - 2^{-\Omega(\epsilon^2 N)}$. The upshot of this is that the probability of error goes down exponentially in the number of trials.

[4] E is the encryption algorithm. It takes in the key and a single bit to encrypt. D is the decryption algorithm. It takes in the key, an encrypted bit, and decrypts the bit using the key

Proof. One form of Chernoff bounds states the following — if X is a random variable that is the sum of N independent Bernoulli(p) random variables, and $\mu = Np$ denotes the expected value of X, then for all $0 < \delta < 1$ —

$$\mathbb{P}[X > (1+\delta)\mu] \leq \exp(-\delta^2\mu/3)$$

Now consider the case in our application when $b = 0$. Then, $p = \frac{1}{2} - \epsilon$. Choosing δ to be $\frac{2\epsilon}{1-2\epsilon}$, which is a reverse-engineered value, we have :

$$\mathbb{P}[X > (1+\delta)\mu] \leq \exp(-\delta^2\mu/3)$$

$$\therefore \mathbb{P}\left[X > \left(1 + \frac{2\epsilon}{1-2\epsilon}\right)\left(\frac{1}{2} - \epsilon\right)N\right] \leq \exp\left(\frac{-\left(\frac{2\epsilon}{1-2\epsilon}\right)^2\left(\frac{1}{2} - \epsilon\right)N}{3}\right)$$

$$\therefore \mathbb{P}\left[X > \left(\frac{1}{1-2\epsilon}\right)\left(\frac{1-2\epsilon}{2}\right)N\right] \leq \exp\left(\frac{-\left(\frac{2\epsilon}{1-2\epsilon}\right)^2\left(\frac{1}{2} - \epsilon\right)N}{3}\right)$$

$$\therefore \mathbb{P}\left[X > \frac{N}{2}\right] \leq \exp\left(\frac{-2\epsilon^2 N}{3(1-2\epsilon)}\right)$$

Since $0 < \epsilon < 1/2$, we have $1 - 2\epsilon \leq 1$. Thus,

$$\leq \exp(-2\epsilon^2 N/3)$$
$$= \boxed{\exp(-\Omega(\epsilon^2 N))}$$