Lecture 4: One Way Functions - II

Instructor: Omkant Pandey

Spring 2017 (CSE 594)

Last Class

- Modeling adversaries as non-uniform PPT Turing machines
- Negligible and noticeable functions
- Definitions of strong and weak OWFs
- Factoring assumption
- Candidate weak OWF $f_{\times}$ based on factoring assumption

Today's Class

- Proving $f_\times$ is a weak OWF
- Yao's hardness amplification: from weak to strong OWFs
- Volunteer for today's scribes?

Recall

Definition (Weak One Way Function)

A function $f : \{0,1\}^* \to \{0,1\}^*$ is a *weak one-way function* if it satisfies the following two conditions:

- **Easy to compute:** there is a PPT algorithm $\mathcal{C}$ s.t. $\forall x \in \{0,1\}^*$,

$$\Pr\left[\mathcal{C}(x) = f(x)\right] = 1.$$

- **Somewhat hard to invert:** there is a noticeable function $\varepsilon : \mathbb{N} \to \mathbb{R}$ s.t. for every non-uniform PPT $\mathcal{A}$ and $\forall n \in \mathbb{N}$:

$$\Pr\left[x \leftarrow \{0,1\}^n, x' \leftarrow \mathcal{A}(1^n, f(x)) : f(x') \neq f(x)\right] \geqslant \varepsilon(n).$$

Noticeable (or non-negligible): $\exists c$ s.t. for infinitely many $n \in \mathbb{N}$, $\varepsilon(n) \geqslant \frac{1}{n^c}$.

Recall (contd.)

- Multiplication function $f_\times : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$:

$$f_\times(x, y) = \begin{cases} \perp & \text{if } x = 1 \vee y = 1 \\ x \cdot y & \text{otherwise} \end{cases}$$

Theorem

Assuming the factoring assumption, function $f_\times$ is a weak OWF.

Proof Idea

- Let Good be the set of inputs (x, y) to $f_\times$ s.t. both x and y are prime numbers
- When $(x, y) \in \text{Good}$, adversary cannot invert $f_\times(x, y)$ (due to hardness of factoring)
- Suppose adversary inverts with probability 1 when $(x, y) \notin \text{Good}$
- But if $\Pr[(x, y) \in \text{Good}]$ is noticeable, then overall, adversary can only invert with a bounded noticeable probability
- Formally: let Let $q(n) = 8n^2$. Will show that no non-uniform PPT adversary can invert $f_\times$ with probability greater than $1 - \frac{1}{q(n)}$

Proof via Reduction

Goal: Given an adversary A that breaks weak one-wayness of $f_\times$ with probability *at least* $1 - \frac{1}{q(n)}$, we will construct an adversary B that breaks the factoring assumption with non-negligible probability

Adversary $B(z)$:

1. $x, y \xleftarrow{\$} 0, 1^n$
2. If x and y are primes, then $z' = z$
3. Else, $z' = x \cdot y$
4. $w \leftarrow A(1^n, z')$
5. Output w if x and y are primes

Analysis of B:

- Since A is non-uniform PPT, so is B (using polynomial-time primality testing)
- A fails to invert with probability at most $\frac{1}{q(n)} = \frac{1}{8n^2}$
- B fails to pass z to A with probability at most $1 - \frac{1}{4n^2}$ (by Chebyshev's Thm.)
- Union bound: B fails with probability at most $1 - \frac{1}{8n^2}$
- B succeeds with probability at least $\frac{1}{8n^2}$: Contradiction to factoring assumption!

Weak to Strong OWFs

Theorem (Yao)

Strong OWFs exist if and only weak OWFs exist

- This is called hardness amplification: convert a somewhat hard problem into a really hard problem
- Intuition: Use the weak OWF *many* times
- Think: Is $f(f(...f(x)))$ a good idea?
  - Hint 1: what happens when f is not injective?
  - Hint 2: f could behave strangely on special inputs.

Weak to Strong OWFs

- GOOD inputs: hard to invert, BAD inputs: easy to invert
- A OWF is weak when the fraction of BAD inputs is **noticeable**
- In a strong OWF, the fraction of BAD inputs is **negligible**
- To convert weak OWF to strong, use the weak OWF on **many** (say N) inputs independently
- In order to successfully invert the new OWF, adversary must invert ALL the N outputs of the weak OWF
- If N is sufficiently large and the inputs are chosen independently at random, you'll hit one of the "hard to invert" inputs with high probability.
- $\Rightarrow$ the probability of inverting all of them will be very small

Weak to Strong OWFs

Theorem

For any weak one-way function $f : \{0,1\}^n \rightarrow \{0,1\}^n$*, there exists a polynomial* $N(\cdot)$ *s.t. the function* $F : \{0,1\}^{n \cdot N(n)} \rightarrow \{0,1\}^{n \cdot N(n)}$ *defined as*

$$F(x_1, \ldots, x_N(n)) = (f(x_1), \ldots, f(x_N(n)))$$

is strongly one-way.

- <u>Think:</u> Show that when f is the $f_\times$ function, then F is a strong one-way function

Proof

- Since f is weakly one-way, there exists a polynomial $q(\cdot)$ s.t. every efficient A fails to invert f with at least $1/q(n)$ probability. I.e.,

  $$\Pr_x \left[A \text{ fails on } f(x) \right] \geqslant \frac{1}{q(n)}$$

Proof

- Since f is weakly one-way, there exists a polynomial $q(\cdot)$ s.t. every efficient A fails to invert f with at least $1/q(n)$ probability. I.e.,

  $$\Pr_x\left[A \text{ fails on } f(x)\right] \geqslant \frac{1}{q(n)}$$

Proof

- Since f is weakly one-way, there exists a polynomial $q(\cdot)$ s.t. every efficient A fails to invert f with at least $1/q(n)$ probability. I.e.,

  $\Pr_x\left[A \text{ fails on } f(x)\right] \geqslant \frac{1}{q(n)}$

Proof

- Since f is weakly one-way, there exists a polynomial $q(\cdot)$ s.t. every efficient A fails to invert f with at least $1/q(n)$ probability. I.e.,

$$\underbrace{\Pr_x\left[A \text{ fails on } f(x)\right]}_{\alpha:=\alpha(n)} \geqslant \frac{1}{q(n)}$$

Proof

- Since f is weakly one-way, there exists a polynomial $q(\cdot)$ s.t. every efficient A fails to invert f with at least $1/q(n)$ probability. I.e.,

$$\underbrace{\Pr_x\left[A \text{ fails on } f(x)\right]}_{\alpha:=\alpha(n)} \geqslant \underbrace{\frac{1}{q(n)}}_{q:=q(n)}$$

Proof

- Since f is weakly one-way, there exists a polynomial $q(\cdot)$ s.t. every efficient A fails to invert f with at least $1/q(n)$ probability. I.e.,

$$\underbrace{\Pr_x\left[A \text{ fails on } f(x)\right]}_{\alpha:=\alpha(n)} \geqslant \underbrace{\tfrac{1}{q(n)}}_{q:=q(n)} \qquad \Longrightarrow \alpha \geqslant \tfrac{1}{q} \qquad (1)$$

- <u>Main idea</u>: large N should almost always hit a "hard to invert x"
- Choose N s.t. $\alpha^N = \left(1 - \frac{1}{q}\right)^N$ is small. Let:

$$N = 2nq \quad \Longrightarrow \left(1 - \frac{1}{q}\right)^{2nq} \quad \approx e^{-2n}$$

Proof (continued)

- Now assume by contradiction that F is not a strong OWF.
- $\exists$ efficient adversary B and a polynomial $p(\cdot)$ s.t.

$$\Pr_{(x_1,\ldots,x_N)}\left[B \text{ inverts } F(x_1,\ldots,x_N)\right] \geqslant \frac{1}{p(n \cdot N)}$$

Proof (continued)

- Now assume by contradiction that F is not a strong OWF.
- $\exists$ efficient adversary B and a polynomial $p(\cdot)$ s.t.

$$\Pr_{(x_1,\ldots,x_N)}\left[B \text{ inverts } F(x_1,\ldots,x_N)\right] \geqslant \frac{1}{p(n\cdot N)}$$

Proof (continued)

- Now assume by contradiction that F is not a strong OWF.
- $\exists$ efficient adversary B and a polynomial $p(\cdot)$ s.t.

$$\Pr_{(x_1,\ldots,x_N)}\left[B \text{ inverts } F(x_1,\ldots,x_N)\right] \geqslant \frac{1}{p(n\cdot N)}$$

Proof (continued)

- Now assume by contradiction that F is not a strong OWF.
- $\exists$ efficient adversary B and a polynomial $p(\cdot)$ s.t.

$$\underbrace{\Pr_{(x_1,\ldots,x_N)}\left[B \text{ inverts } F(x_1,\ldots,x_N)\right]}_{\beta:=\beta(n\cdot N)} \geqslant \underbrace{\frac{1}{p(n\cdot N)}}_{p:=p(n\cdot N)}$$

Proof (continued)

- Now assume by contradiction that F is not a strong OWF.
- $\exists$ efficient adversary B and a polynomial $p(\cdot)$ s.t.

$$\underbrace{\Pr_{(x_1,\ldots,x_N)}\left[B \text{ inverts } F(x_1,\ldots,x_N)\right]}_{\beta:=\beta(n\cdot N)} \geqslant \underbrace{\frac{1}{p(n\cdot N)}}_{p:=p(n\cdot N)} \implies \beta \geqslant \frac{1}{p} \qquad (2)$$

Proof (continued)

- Now assume by contradiction that F is not a strong OWF.
- $\exists$ efficient adversary B and a polynomial $p(\cdot)$ s.t.

$$\underbrace{\Pr_{(x_1,\ldots,x_N)}\left[B \text{ inverts } F(x_1,\ldots,x_N)\right]}_{\beta:=\beta(n\cdot N)} \geqslant \underbrace{\frac{1}{p(n\cdot N)}}_{p:=p(n\cdot N)} \implies \beta \geqslant \frac{1}{p} \qquad (2)$$

- Think: How to use B to contradict that f is weak one-way?
  - Build an adversary A that uses B to break f with prob.

$$\alpha < 1/q.$$

  - Feed $(y,\ldots,y)$ to B?
  - Feed $(y, y_2, \ldots, y_N)$ to B where each $y_i = f(x_i)$ for a random x_i?
  - Feed y in a random location i to balance out probabilities.

Adversary A for inverting f

Adversary $A_0(y)$**:**

1. Choose $i \xleftarrow{\$} [N]$ and set $y_i = y$
2. $\forall j \neq i$, sample $x_j \in \{0,1\}^N$ and set $y_j = f(x_j)$
3. Let $(z_1, \ldots, z_N) \leftarrow B(y_1, \ldots, y_n)$
4. If $f(z_i) = y$, output z_i, else output $\perp$.

How good is A_0?

- Good but not good enough to give $\alpha < 1/q$.
- Good: even for each fixed y, B still gets somewhat random inputs.
- Idea: run A_0 many times to improve chances for inverting y!

Main Adversary $A(y)$**:**

1. Run $A_0(y)$ for $T := T(n)$ times $= 4n^2 \times q(n) \times p(n \cdot N)$.
2. Output the first non-$\perp$ answer

Proof (continued, Analysis of A_0)

Analysis of A_0

- Why A_0 is not good enough?

  ...because for some ys, B may have too low a chance of inverting.
- There can be many such y; we show they can't be too many!
- Let $\textsc{Bad}$ = set of inputs x s.t. A_0 has low chance of inverting $f(x)$.

$$\textsc{Bad} := \left\{ x \,\middle|\, \Pr_{A_0}[A_0 \text{ inverts } f(x)] < \text{low} \right\}$$

- The lower the RHS the smaller will be the size of $\textsc{Bad}$.
- Want to pick this chance so that $\Pr_x[x \in \textsc{Bad}] < 1/2q$. Let:

$$\textsc{Bad} := \left\{ x \,\middle|\, \Pr_{A_0}[A_0 \text{ inverts } f(x)] < \frac{1}{4npq} \right\}$$

Proof (continued): Analysis of A_0

Lemma

$$\Pr_x[x \in \text{Bad}] < \frac{1}{2q}.$$

Proof. Suppose not, i.e., $\Pr_x[x \in \text{Bad}] > \alpha/2 \geqslant 1/2q$

$$\begin{aligned}
\beta &= \Pr_{x_1,\ldots,x_N}[B \text{ inverts } F(x_1,\ldots,x_N)] \\
&= \Pr_{x_1,\ldots,x_N}[B \text{ inverts } F(x_1,\ldots,x_N) \wedge (\forall i : x_i \notin \text{Bad})] \quad + \\
&\quad \Pr_{x_1,\ldots,x_N}[B \text{ inverts } F(x_1,\ldots,x_N) \wedge (\exists i : x_i \in \text{Bad})] \\
&\leqslant \Pr_{x_1,\ldots,x_N}[\forall i : x_i \notin \text{Bad}] + \sum_i \Pr_{x_1,\ldots,x_N}[B \text{ inverts } F(x_1,\ldots,x_N) \wedge (x_i \in \text{Bad}) \\
&\leqslant \left(1 - \frac{1}{2q}\right)^N + N \cdot \Pr_{i,x_1,\ldots,x_N}[B \text{ inverts } F(x_1,\ldots,x_N) \wedge (x_i \in \text{Bad})] \\
&\leqslant \left(1 - \frac{1}{2q}\right)^{2nq} + N \cdot \Pr_{x \xleftarrow{\$} \{0,1\}^n, B}[A \text{ inverts } f(x) \wedge (x \in \text{Bad})]
\end{aligned}$$

Proof (continued): Analysis of A_0

$$
\begin{aligned}
\beta &\leqslant \left(1-\frac{1}{2q}\right)^{2nq} + N \cdot \Pr_{x \xleftarrow{\$} \{0,1\}^n, B}\left[A \text{ inverts } f(x) \wedge (x \in \textsc{Bad})\right] \\
&\leqslant e^{-n} + N \cdot \Pr[x \in \textsc{Bad}] \cdot \Pr_{x \xleftarrow{\$} \{0,1\}^n, B}\left[A \text{ inverts } f(x) | x \in \textsc{Bad}\right] \\
&\leqslant e^{-n} + 2nq \cdot 1 \cdot \frac{1}{4npq} = e^{-n} + \frac{1}{2p} < \frac{1}{2p} + \frac{1}{2p} \\
\Longrightarrow \beta &< \frac{1}{p}. \quad \text{(Contradicts (2)).} \qquad \text{(QED)}
\end{aligned}
$$

Proof (continued): Analysis of A

Failure probability of main adversary: A.

$$
\begin{aligned}
\alpha &= \Pr_{x \xleftarrow{\$} \{0,1\}^n} [A \text{ fails to invert } f(x)] \\
&= \Pr_x[x \in \textsc{Bad}] \cdot \Pr_x[A \text{ fails to invert } f(x) | x \in \textsc{Bad}] + \\
&\quad \Pr_x[x \notin \textsc{Bad}] \cdot \Pr_x[A \text{ fails to invert } f(x) | x \notin \textsc{Bad}] \\
&\leqslant \frac{1}{2q} \cdot 1 + 1 \cdot \left(\Pr_{A_0}[A_0 \text{ fails to invert } f(x) | x \notin \textsc{Bad}] \right)^T \\
&\leqslant \frac{1}{2q} + \left(1 - \frac{1}{4npq} \right)^{4pqn^2} \\
&\leqslant \frac{1}{2q} + e^{-n} < 1/q. \quad \text{(contradicts (1))} \quad \text{QED.}
\end{aligned}
$$

$\Longrightarrow$ f is not a weak OWF if F is not a strong OWF.