• Modeling adversaries as non-uniform PPT Turing machines

• Negligible and noticeable functions

• Definitions of strong and weak OWFs

• Factoring assumption

• Candidate weak OWF $f_x$ based on factoring assumption
Today’s Class

- Proving $f_x$ 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 \( f : \{0, 1\}^* \rightarrow \{0, 1\}^* \) is a **weak one-way function** if it satisfies the following two conditions:

- **Easy to compute:** there is a PPT algorithm \( C \) s.t. \( \forall x \in \{0, 1\}^* \),
  \[
  \Pr[C(x) = f(x)] = 1.
  \]

- **Somewhat hard to invert:** there is a noticeable function \( \varepsilon : \mathbb{N} \rightarrow \mathbb{R} \) s.t. for every non-uniform PPT \( A \) and \( \forall n \in \mathbb{N} \),
  \[
  \Pr\left[ x \leftarrow \{0, 1\}^n, x' \leftarrow A(1^n, f(x)) : f(x') \neq f(x) \right] \geq \varepsilon(n).
  \]

**Noticeable (or non-negligible):** \( \exists c \) s.t. for infinitely many \( n \in \mathbb{N} \),
\[
\varepsilon(n) \geq \frac{1}{n^c}.
\]
Recall (contd.)

- Multiplication function \( f_{\times} : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N} \):

\[
f_{\times}(x, y) = \begin{cases} 
\bot & \text{if } x = 1 \lor y = 1 \\
x \cdot y & \text{otherwise}
\end{cases}
\]

Theorem

Assuming the factoring assumption, function \( f_{\times} \) is a weak OWF.
Proof Idea

• Let GOOD be the set of inputs $(x, y)$ to $f_x$ s.t. both $x$ and $y$ are prime numbers.

• When $(x, y) \in \text{GOOD}$, adversary cannot invert $f_x(x, y)$ (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 $q(n) = 8n^2$. Will show that no non-uniform PPT adversary can invert $f_x$ with probability greater than $1 - \frac{1}{q(n)}$. 
**Proof via Reduction**

**Goal:** Given an adversary $A$ that breaks weak one-wayness of $f_x$ 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. $x, y \leftarrow 0, 1^n$
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.

- **Think:** Show that when \( f \) is the \( f_x \) 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.,

$$\Pr_x [A \text{ fails on } f(x)] \geq \frac{1}{q(n)}$$
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.,

$$\Pr_x [A \text{ fails on } f(x)] \geq \frac{1}{q(n)}$$
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.,

$$\Pr_x [A \text{ fails on } f(x)] \geq \frac{1}{q(n)}$$
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.,

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

$\alpha := \alpha(n)$
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.,

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

\[\alpha := \alpha(n) \quad q := q(n)\]
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.,

$$\Pr_x [A \text{ fails on } f(x)] \geq \frac{1}{q(n)} \implies \alpha \geq \frac{1}{q} \quad (1)$$

- Main idea: 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:

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

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

$$\Pr_{(x_1,\ldots,x_N)}[B \text{ inverts } F(x_1,\ldots,x_N)] \geq \frac{1}{p(nN)}$$
Proof (continued)

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

$$\Pr_{(x_1,\ldots,x_N)} [B \text{ inverts } F(x_1,\ldots,x_N)] \geq \frac{1}{p(n,N)}$$
Proof (continued)

Now assume by contradiction that $F$ is not a strong OWF.

∃ efficient adversary $B$ and a polynomial $p(\cdot)$ s.t.

$$\Pr_{(x_1, \ldots, x_N)}[B \text{ inverts } F(x_1, \ldots, x_N)] \geq \frac{1}{p(n \cdot N)}$$
Proof (continued)

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

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

\[\beta := \beta(n \cdot N)\]
\[p := p(n \cdot N)\]
Proof (continued)

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

$$\Pr(x_1, \ldots, x_N) [B \text{ inverts } F(x_1, \ldots, x_N)] \geq \frac{1}{p(n \cdot N)} \implies \beta \geq \frac{1}{p} \quad (2)$$
Proof (continued)

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

$$\Pr_{(x_1,\ldots,x_N)}[B \text{ inverts } F(x_1, \ldots, x_N)] \geq \frac{1}{p(n \cdot N)} \implies \beta \geq \frac{1}{p} \quad (2)$$

\[
\begin{align*}
\beta := \beta(n \cdot N) \\
p := p(n \cdot N)
\end{align*}
\]

- 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.

$$\alpha < \frac{1}{q}.$$ 

- 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 $i \leftarrow [N]$ 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 $y$s, $B$ may have too low a chance of inverting.
- There can be many such $y$; we show they can’t be too many!
- Let $\text{BAD} =$ set of inputs $x$ s.t. $A_0$ has low chance of inverting $f(x)$.

$$\text{BAD} := \left\{ x \mid \Pr_{A_0}[A_0 \text{ inverts } f(x)] < \text{low} \right\}$$

- The lower the RHS the smaller will be the size of $\text{BAD}$.

Want to pick this chance so that $\Pr_x[x \in \text{BAD}] < 1/2q$. Let:

$$\text{BAD} := \left\{ x \mid \Pr_{A_0}[A_0 \text{ inverts } f(x)] < \frac{1}{4npq} \right\}$$
Proof (continued): Analysis of $A_0$

**Lemma**

$$\Pr_{x}[x \in \text{BAD}] < \frac{1}{2q}.$$ 

**Proof.** Suppose not, i.e., $\Pr_{x}[x \in \text{BAD}] > \alpha/2 \geq 1/2q$

$$\beta = \Pr_{x_1,\ldots,x_N} [B \text{ inverts } F(x_1,\ldots,x_N)]$$

$$= \Pr_{x_1,\ldots,x_N} [B \text{ inverts } F(x_1,\ldots,x_N) \land (\forall i : x_i \notin \text{BAD})] + \Pr_{x_1,\ldots,x_N} [B \text{ inverts } F(x_1,\ldots,x_N) \land (\exists i : x_i \in \text{BAD})]$$

$$\leq \Pr_{x_1,\ldots,x_N} [\forall i : x_i \notin \text{BAD}] + \sum_{i} \Pr_{x_1,\ldots,x_N} [B \text{ inverts } F(x_1,\ldots,x_N) \land (x_i \in \text{BAD})]$$

$$\leq \left(1 - \frac{1}{2q}\right)^{2nq} + N \cdot \Pr_{i,x_1,\ldots,x_N} [B \text{ inverts } F(x_1,\ldots,x_N) \land (x_i \in \text{BAD})]$$

$$\leq \left(1 - \frac{1}{2q}\right)^{2nq} + N \cdot \Pr_{x \leftarrow \{0,1\}^n,B} [A \text{ inverts } f(x) \land (x \in \text{BAD})]$$
Proof (continued): Analysis of $A_0$

$$\beta \leq \left(1 - \frac{1}{2q}\right)^{2nq} + N \cdot \Pr_{x \leftarrow \{0,1\}^n, B} [A \text{ inverts } f(x) \land (x \in \text{BAD})]$$

$$\leq e^{-n} + N \cdot \Pr[x \in \text{BAD}] \cdot \Pr_{x \leftarrow \{0,1\}^n, B} [A \text{ inverts } f(x) | x \in \text{BAD}]$$

$$\leq e^{-n} + 2nq \cdot 1 \cdot \frac{1}{4npq} = e^{-n} + \frac{1}{2p} < \frac{1}{2p} + \frac{1}{2p}$$

$$\implies \beta < \frac{1}{p}. \quad \text{(Contradicts (2)).} \quad \text{(QED)}$$
Proof (continued): Analysis of $A$

Failure probability of main adversary: $A$.

$$\alpha = \Pr_{x \leftarrow \{0,1\}^n} [A \text{ fails to invert } f(x)]$$

$$= \Pr_{x \leftarrow \text{BAD}} [x \in \text{BAD}] \cdot \Pr_{x \leftarrow \text{BAD}} [A \text{ fails to invert } f(x) | x \in \text{BAD}] + \Pr_{x \leftarrow \neg \text{BAD}} [x \notin \text{BAD}] \cdot \Pr_{x \leftarrow \neg \text{BAD}} [A \text{ fails to invert } f(x) | x \notin \text{BAD}]$$

$$\leq \frac{1}{2q} \cdot 1 + 1 \cdot \left( \Pr_{A_0} [A_0 \text{ fails to invert } f(x) | x \notin \text{BAD}] \right)^T$$

$$\leq \frac{1}{2q} + \left( 1 - \frac{1}{4npq} \right)^{4pqn^2}$$

$$\leq \frac{1}{2q} + e^{-n} < 1/q. \quad \text{(contradicts (1)) QED.}$$

$$\implies f \text{ is not a weak OWF if } F \text{ is not a strong OWF.}$$