1 Review

Last lecture, we defined strong and weak one-way functions (OWFs) as follows:

**Definition 1** A function $\epsilon : \mathbb{N} \rightarrow \mathbb{R}$ is **negligible** if, for every $c \in \mathbb{N}$ there exists $n_0$ such that for all $n > n_0$, $\epsilon(n) < \frac{1}{n^c}$.

**Definition 2** A function $f : \{0,1\}^* \rightarrow \{0,1\}^*$ is **strongly one-way** if:

- $f$ can be computed in probabilistic polynomial time (PPT).
- for every non-uniform probabilistic polynomial time (N.U. PPT) algorithm $A$, there exists a negligible function $\epsilon$ such that for every $n \in \mathbb{N}$,
  $$\Pr[x \leftarrow \{0,1\}^n; A(1^n, f(x)) \in f^{-1}(f(x))] < \epsilon(n)$$

So $f$ can be inverted only with negligible probability on a random input – it is hard to invert on all but a negligible fraction of inputs.

**Definition 3** A function $f : \{0,1\}^* \rightarrow \{0,1\}^*$ is **weakly one-way** if:

- $f$ can be computed in probabilistic polynomial time (PPT).
- for every non-uniform probabilistic polynomial time (N.U. PPT) algorithm $A$, there exists a polynomial $q : \mathbb{N} \rightarrow \mathbb{R}$ such that for every input length $n \in \mathbb{N}$,
  $$\Pr[x \leftarrow \{0,1\}^n; A(1^n, f(x)) \in f^{-1}(f(x))] < 1 - \frac{1}{q(n)}$$

So $f$ can be inverted with non-negligible probability on a random input – this suggests that $f$ is easy to invert on some non-negligible fraction of inputs.

We considered the function $f_{\text{mult}} : \mathbb{N}^2 \rightarrow \mathbb{N}$ defined by $f_{\text{mult}}(x, y) = xy$ for $|x| = |y|$, and showed that this is at best weakly one-way, being easy to invert when one of the inputs is even.
We will now show that we don’t lose anything by relaxing our requirements from strong to weak one-wayness, since a weak OWF can in fact be used to construct a strong OWF; this is called hardness amplification. The intuition behind this is that, if we evaluate a weak one-way function on a sufficiently large number of inputs, it is likely to be hard to invert on at least one of those inputs.

**Lemma 1** Let $f : \{0,1\}^* \rightarrow \{0,1\}^*$ be a weak OWF. Then there exists a polynomial $m : \mathbb{N} \rightarrow \mathbb{N}$ such that for input length $n \in \mathbb{N}$, the following function $g : \{0,1\}^{mn} \rightarrow \{0,1\}^{mn}$ (where $m = m(n)$) is a strong OWF:

$$g(x_1, x_2, ..., x_m) = f(x_1)f(x_2)...f(x_m)$$

**Proof.** By contradiction. We will assume that $g$ is not strongly one-way and construct an algorithm that inverts $f$ with high probability, contradicting its weak one-wayness.

Let $q : \mathbb{N} \rightarrow \mathbb{N}$ be the polynomial in the definition of a weak OWF, such that for any PPT algorithm $A$ and any input length $n \in \mathbb{N}$,

$$\Pr[x \leftarrow \{0,1\}^n; A(1^n, f(x)) \in f^{-1}(f(x))] < 1 - \frac{1}{q(n)}$$

We want to define $m$ such that $(1 - \frac{1}{q(n)})^m$ tends to 0 for large $n$. $(1 - \frac{1}{q(n)})^{nq(n)} \approx (\frac{1}{e})^n$, so we choose $m = 2nq(n)$.

Assume for a contradiction that $g$ is not a strong OWF. Then there exists a N.U. PPT algorithm $A$ and a polynomial $p' : \mathbb{N} \rightarrow \mathbb{R}$ such that for infinitely many input lengths $n' \in \mathbb{N}$, $A$ can invert $g$ with probability at least $\frac{1}{p'(n')}$. Formally,

$$\Pr[x \leftarrow \{0,1\}^n; A(1^n, g(x_1, x_2, ..., x_m)) \in g^{-1}(g(x_1, x_2, ..., x_m))] \geq \frac{1}{p'(mn)}$$

Since $m$ is polynomial in $n$, the function $p : \mathbb{N} \rightarrow \mathbb{R}$ defined as $p(n) = p'(mn) = p'(2n^2q(n))$ is also a polynomial in $n$. So we can rewrite the above probability as

$$\Pr[x \leftarrow \{0,1\}^n; A(1^{mn}, g(x_1, x_2, ..., x_m)) \in g^{-1}(g(x_1, x_2, ..., x_m))] \geq \frac{1}{p(n)}$$

For convenience, we will rewrite the above inequality as

$$\Pr[x_i \leftarrow \{0,1\}^n; A \text{ succeeds}] \geq \frac{1}{p(n)}$$

**Goal:** Given $A$ that takes $y_1y_2...y_m$ as input and outputs $z_1, ..., z_m$ such that $f(z_i) = y_i$ for all $i$ with probability $\geq \frac{1}{p(n)}$, we want to construct an adversary $A'$ that uses $A$ to
invert $f$ with probability $\geq 1 - \frac{1}{q(n)}$. In other words, given $y = f(x)$ for random $x$, we want algorithm $A'$ to return $z$ such that $f(z) = y$ with probability $\geq 1 - \frac{1}{q(n)}$.

**Approach 1:** Given $y$, give $A$ as input $yy...y$ (i.e. $y_i = y$ for all $i$). However, it is possible that the algorithm $A$ always fails when the input has the format above, i.e. consists of a string repeated $m$ times (these strings form a very small fraction of all strings of length $mn$). So this won’t work.

**Approach 2:** Give $A$ as input, set $y_1 = y$ and for $j \neq 1$, select a random $x_j \in \{0, 1\}^n$ and set $y_j = f(x_j)$. Again, it is possible that the algorithm $A$ always returns garbage for the first “spot”.

**Correct Approach:** Pick a random $i \in \{1, ..., m\}$ and set $y_i = y$. For $j \neq i$, select a random $x_j \in \{0, 1\}^n$ and set $y_j = f(x_j)$. Formally, $A''(y)$:

- Pick $i \leftarrow [1, m]$ and let $y_i = y$.
- For all $j \neq i$, pick $x_j \leftarrow \{0, 1\}^n$ and let $y_j = f(x_j)$.
- Let $z_1, ..., z_m = A'(1^{mn}, y_1, ..., y_m)$.
- If $f(z_1) = y$, output $z$. Otherwise output $\bot$ (fail).

The algorithm $A'$ does what we want, but the probability of success is not high enough. To improve the probability of inverting $f$, we need to run $A'$ multiple times. To do so, define the algorithm $A'$ as follows:

$A'(y)$:

- Run $A''(y) 2nm^2p(n)$ times.
- Output the first answer that is not $\bot$.
- If no such answer exists, output $\bot$.

However, we are using the same input $y$ in all $2nm^2p(n)$ runs, so the runs are not independent. Therefore it may be the case that $y$ is simply a bad input for $A''$, and therefore fails on every run.

To overcome this, we define the notion of “good” and “bad” inputs. We say that element $x \in \{0, 1\}^n$ is “good” if $\Pr[A''(f(x))$ succeeds] $\geq \frac{1}{2nm^2p(n)}$. Otherwise $x$ is “bad”.

Thus,

$$\Pr[A'(f(x))$ fails$|x$ is good$] \leq \left(1 - \frac{1}{2m^2p(n)}\right)^{2nm^2p(n)} \approx \left(\frac{1}{e}\right)^n$$

3-3
Claim: There are at least $2^n \left( 1 - \frac{1}{2q(n)} \right)$ good inputs $x \in \{0, 1\}^n$.

Why does proving this claim give us what we want?

If the above claim is true, then $Pr[x \text{ is good}] \geq \left( 1 - \frac{1}{2q(n)} \right)$, so $Pr[x \text{ is bad}] \leq \frac{1}{2q(n)}$. So

$$Pr[A'(f(x)) \text{ fails}] = \Pr[A'(f(x)) \text{ fails}|x \text{ is good}] \cdot Pr[x \text{ is good}] + Pr[x \text{ is bad}] \leq Pr[x \text{ is good}] + \Pr[A'(f(x)) \text{ fails}|x \text{ is bad}] \cdot Pr[x \text{ is bad}]$$

$$\approx \left( \frac{1}{e} \right)^n + \frac{1}{2q(n)} \leq \frac{1}{q(n)}$$

So $A'$ succeeds in inverting $f(x)$ for random $x$ with probability $\geq 1 - \frac{1}{q(n)}$, contradicting the weak one-wayness of $f$. So if we can prove that there are a significant number of good elements, $A'$ can invert $f$ with high probability, contradicting the assumption that $f$ is weakly one-way.

**Proof.** Assume for a contradiction that the number of bad inputs $> \frac{2^n}{2q(n)}$. We wish to contradict the fact that $A$ inverts $g$ with probability $\geq \frac{1}{p(n)}$.

$$Pr[A(1^{mn}, g(x_1, ..., x_m)) \text{ succeeds}] = Pr[A(1^{mn}, g(x_1, ..., x_m)) \text{ succeeds} \land \exists i \text{ is bad x}_i]$$

$$+ Pr[A(1^{mn}, g(x_1, ..., x_m)) \text{ succeeds} \land \forall i, x_i \text{ is good}] \tag{2}$$

(1) $Pr[A \text{ succeeds} \land \exists i \text{ bad } x_i] \leq \sum_i Pr[A \text{ succeeds} \land x_i \text{ is bad}]$ (via the union bound)

For each $i \in [1, m]$,

$$Pr[A \text{ succeeds} \land x_i \text{ is bad}] \leq Pr[A \text{ succeeds}|x_i \text{ is bad}]$$

$$< m \cdot Pr[A'(f(x_i)) \text{ succeeds}|x_i \text{ is bad}] \quad (A'' \text{ can put } x_i \text{ at one of } m \text{ positions})$$

$$\leq \frac{m}{2m^2p(n)} = \frac{1}{2mp(n)}$$

So $Pr[A \text{ succeeds} \land \exists i \text{ bad } x_i] \leq \sum_i \frac{1}{2mp(n)} = \frac{1}{2p(n)}$
(2) $\Pr[A \text{ succeeds } \land \forall x_i, x_i \text{ is good}]$

\[
\begin{align*}
&< \Pr[\forall x_i, x_i \text{ is good}] \\
&\leq \left(1 - \frac{1}{2q(n)}\right)^m \\
&= \left(1 - \frac{1}{2q(n)}\right)^{2nq(n)} \\
&\approx \left(\frac{1}{e}\right)^n
\end{align*}
\]

Hence, $\Pr[A(1^{mn}, g(x_1, ..., x_m)) \text{ succeeds}] < \frac{1}{2p(n)} + \frac{1}{e} < \frac{1}{p(n)} \Rightarrow \Leftarrow$

Therefore the number of bad inputs $\leq \frac{2n}{2q(n)}$, and the number of good inputs $\geq 1 - \frac{2n}{2q(n)}$. ■

This proves Lemma 1, showing that the existence of weak one-way functions implies that of strong one-way functions.