COM S 687 Introduction to Cryptography	September 5, 2006
Lecture 4: Hardness Amplification: From Weak to Strong OWFs	
Instructor: Rafael Pass	*Scribe: Jed Liu*

1 Review

In the previous lecture, we formalised the notion of a *strong one-way* function that is easy to compute, but hard to invert.

Definition 1 (Negligible function) *A function* $\epsilon : \mathbb{N} \to \mathbb{R}$ *is* negligible *if for any* $c \in \mathbb{N}$*, there is a* $k_0 \in N$ *such that we have* $\epsilon(k) < \frac{1}{k^c}$ *for all* $k > k_0$*.*

Definition 2 (Strong one-way function) *A function* $f : \{0,1\}^* \to \{0,1\}^*$ *is* strongly one-way *if it satisfies the following two conditions.*

1. **Easy to compute.** *There is a probabilistic polytime algorithm* $\mathcal{C} : \{0,1\}^* \to \{0,1\}^*$ *such that* $\mathcal{C}(x) = f(x)$ *on all inputs* $x \in \{0,1\}^*$*.*
2. **Hard to invert.** *Any efficient attempt to invert* f *on random input will succeed with only negligible probability. Formally, for any probabilistic polytime algorithm* $\mathcal{A} : \{0,1\}^* \to \{0,1\}^*$*, there exists a negligible function* ϵ *such that for any input length* $k \in \mathbb{N}$*,*

$$\mathsf{Pr}\left[x \leftarrow \{0,1\}^k;\ y = f(x);\ \mathcal{A}(1^k, y) = x' : f(x') = y\right] \leq \epsilon(k).$$

2 Weak One-Way Functions

Consider the function $f_{\mathsf{mult}} : \mathbb{N}^2 \to \mathbb{N}$ defined by $f_{\mathsf{mult}}(x, y) = xy$, with $|x| = |y|$. Is this a one-way function? Clearly, by the multiplication algorithm, f_{mult} is easy to compute. But f_{mult} is not always hard to invert! If at least one of x and y is even, then their product will be even as well. This happens with probability $\frac{3}{4}$ if the input (x, y) is picked uniformly at random from $\mathbb{N}^2$. So the following attack A will succeed with probability $\frac{3}{4}$:

$$A(z) = \begin{cases} (2, \frac{z}{2}) & \text{if } z \text{ even} \\ (0, 0) & \text{otherwise.} \end{cases}$$

Something is not quite right here, since f_{mult} is conjectured to be hard to invert on *some*, but not all, inputs. Our current definition of a one-way function is too restrictive to capture this notion, so we will define a weaker variant that relaxes the hardness condition on inverting the function. This weaker version only requires that all efficient attempts at inverting will fail with some non-negligible probability.

Definition 3 (Weak one-way function) *A function $f : \{0,1\}^* \to \{0,1\}^*$ is* weakly one-way *if it satisfies the following two conditions.*

1. **Easy to compute.** *(Same as that for a strong one-way function.) There is a probabilistic polytime algorithm $\mathcal{C} : \{0,1\}^* \to \{0,1\}^*$ such that $\mathcal{C}(x) = f(x)$ on all inputs $x \in \{0,1\}^*$.*

2. **Hard to invert.** *Any efficient algorithm will fail to invert f on random input with non-negligible probability. More formally, for any probabilistic polytime algorithm $\mathcal{A} : \{0,1\}^* \to \{0,1\}^*$, there exists a polynomial function $q : \mathbb{N} \to \mathbb{N}$ such that for any input length $k \in \mathbb{N}$,*

$$\mathsf{Pr}\left[x \leftarrow \{0,1\}^k;\ y = f(x);\ \mathcal{A}(1^k, y) = x' : f(x') = y\right] \leq 1 - \frac{1}{q(k)}$$

It is conjectured that f_{mult} is a weak one-way function.

3 Hardness Amplification

By falling back on the weak version of a one-way function, we actually haven't lost anything. As we will now show, a weak one-way function can be used to produce a strong one-way function by amplifying hardness. The main insight we will use is if we run a weak one-way function f with enough inputs, with luck, f will be hard to invert on least one of those inputs.

Theorem 1 *If there is a weak one-way function, then there is a strong one-way function. In particular, given a weak one-way function $f : \{0,1\}^* \to \{0,1\}^*$, there is a fixed $m \in \mathbb{N}$, polynomial in the input length $n \in \mathbb{N}$, such that the following function $f' : (\{0,1\}^n)^m \to (\{0,1\}^n)^m$ is strongly one-way:*

$$f'(x_1, x_2, \ldots, x_m) = (f(x_1), f(x_2), \ldots, f(x_m)).$$

We will prove this theorem by contradiction. We assume that f' is not strongly one-way so that there is an algorithm $\mathcal{A}'$ that inverts it with non-negligible probability. From this, we construct an algorithm $\mathcal{A}$ that inverts f with high probability.

Proof. Since f is weakly one-way, let $q : \mathbb{N} \to \mathbb{N}$ be a polynomial such that for any probabilistic polytime algorithm $\mathcal{A}$ and any input length $n \in \mathbb{N}$,

$$\mathsf{Pr}\left[x \leftarrow \{0,1\}^n;\ y = f(x);\ \mathcal{A}(1^n, y) = x' : f(x') = y\right] \leq 1 - \frac{1}{q(n)}.$$

Define $m = 2nq(n)$, dependent on the input length $n \in N$ to f.

Assume that f' as defined in the theorem is not strongly one-way. Then let $\mathcal{A}'$ be a probabilistic polytime algorithm and $p' : \mathbb{N} \to \mathbb{N}$ be a polynomial such that for infinitely many input lengths $n \in \mathbb{N}$ to f, $\mathcal{A}'$ inverts f' with probability $p'(n)$. i.e.,

$$\mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f(x_i) : f'(\mathcal{A}'(y_1, y_2, \ldots, y_m)) = (y_1, y_2, \ldots, y_m)\right] > \frac{1}{p'(m)}.$$

Since m is polynomial in n, then the function $p(n) = p'(m) = p'(2nq(n))$ is also a polynomial. Rewriting the above probability, we have

$$\mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f(x_i) : f'(\mathcal{A}'(y_1, y_2, \ldots, y_m)) = (y_1, y_2, \ldots, y_m)\right] > \frac{1}{p(n)}. \quad (1)$$

Define the algorithm $\mathcal{A}_0 : \{0,1\}^n \to \{0,1\}^n_\perp$, which will attempt to use $\mathcal{A}'$ to invert f, as follows.

(1) Input $y \in \{0,1\}^n$.

(2) Pick a random $i \leftarrow [1, m]$.

(3) For all $j \neq i$, pick a random $x_j \leftarrow \{0,1\}^n$, and let $y_j = f(x_j)$.

(4) Let $y_i = y$.

(5) Let $(z_1, z_2, \ldots, z_m) = \mathcal{A}(y_1, y_2, \ldots, y_m)$.

(6) If $f(z_i) = y$, then output z_i; otherwise, fail and output $\perp$.

To improve our chances of inverting f, we will run $\mathcal{A}_0$ multiple times. To capture this, define the algorithm $\mathcal{A} : \{0,1\}^n \to \{0,1\}^n_\perp$ to run $\mathcal{A}_0$ with its input $2nm^2p(n)$ times, outputting the first non-$\perp$ result it receives. If all runs of $\mathcal{A}_0$ result in $\perp$, then $\mathcal{A}$ outputs $\perp$ as well.

Given this, call an element $x \in \{0,1\}^n$ "good" if $\mathcal{A}_0$ will successfully invert $f(x)$ with non-negligible probability:

$$\mathsf{Pr}\left[\mathcal{A}_0(f(x)) \neq \perp\right] \geq \frac{1}{2m^2p(n)};$$

otherwise, call x "bad."

Note that the probability of $\mathcal{A}$ failing to invert $f(x)$ on a good x is small:

$$\mathsf{Pr}\left[\mathcal{A}(f(x)) \text{ fails} \mid x \text{ good}\right] \leq \left(1 - \frac{1}{2m^2 p(n)}\right)^{2m^2 np(n)} \approx e^{-n}.$$

We claim that there are a significant number of good elements—enough for $\mathcal{A}$ to invert f with sufficient probability to contradict the weakly one-way assumption on f. In particular, we claim there are at least $2^n\left(1 - \frac{1}{2q(n)}\right)$ good elements in $\{0,1\}^n$. If this holds, then

$$\begin{aligned}
&\mathsf{Pr}\left[\mathcal{A}(f(x)) \text{ fails}\right] \\
&\quad = \mathsf{Pr}\left[\mathcal{A}(f(x)) \text{ fails} \mid x \text{ good}\right] \cdot \mathsf{Pr}\left[x \text{ good}\right] + \mathsf{Pr}\left[\mathcal{A}(f(x)) \text{ fails} \mid x \text{ bad}\right] \cdot \mathsf{Pr}\left[x \text{ bad}\right] \\
&\quad \leq \mathsf{Pr}\left[\mathcal{A}(f(x)) \text{ fails} \mid x \text{ good}\right] + \mathsf{Pr}\left[x \text{ bad}\right] \\
&\quad \leq \left(1 - \tfrac{1}{2m^2 p(n)}\right)^{2m^2 np(n)} + \tfrac{1}{2q(n)} \\
&\quad \approx e^{-n} + \tfrac{1}{2q(n)} \\
&\quad < \tfrac{1}{q(n)}.
\end{aligned}$$

This contradicts the assumption that f is $q(n)$-weak.

It remains to be shown that there are at least $2^n\left(1 - \frac{1}{2q(n)}\right)$ good elements in $\{0,1\}^n$. Assume that there are more than $2^n\left(\frac{1}{2q(n)}\right)$ bad elements. We will contradict fact (1) that with probability $\frac{1}{p(n)}$, $\mathcal{A}'$ succeeds in inverting $f'(x)$ on a random input x. To do so, we establish an upper bound on the probability by splitting it into two quantities:

$$\begin{aligned}
&\mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds}\right] \\
&\quad = \mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds and some } x_i \text{ is bad}\right] \\
&\quad\quad + \mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds and all } x_i \text{ are good}\right]
\end{aligned}$$

For each $j \in [1, n]$, we have

$$\begin{aligned}
&\mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds and } x_j \text{ is bad}\right] \\
&\quad \leq \mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds} \mid x_j \text{ is bad}\right] \\
&\quad \leq m \cdot \mathsf{Pr}\left[\mathcal{A}_0(f(x_j)) \text{ succeeds} \mid x_j \text{ is bad}\right] \\
&\quad \leq \tfrac{m}{2m^2 p(n)} = \tfrac{1}{2mp(n)}.
\end{aligned}$$

So taking a union bound, we have

$$\begin{aligned}
&\mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds and some } x_i \text{ is bad}\right] \\
&\quad \leq \textstyle\sum_j \mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds and } x_j \text{ is bad}\right] \\
&\quad \leq \tfrac{m}{2mp(n)} = \tfrac{1}{2p(n)}.
\end{aligned}$$

Also,

$$\begin{aligned}
&\mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds and all } x_i \text{ are good}\right] \\
&\quad \leq \mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n : \text{all } x_i \text{ are good}\right] \\
&\quad < \left(1 - \tfrac{1}{2q(n)}\right)^m = \left(1 - \tfrac{1}{2q(n)}\right)^{2nq(n)} \approx e^{-n}.
\end{aligned}$$

Hence, $\mathsf{Pr}\left[x_i \leftarrow \{0,1\}^n; y_i = f'(x_i) : \mathcal{A}'(\vec{y}) \text{ succeeds}\right] < \frac{1}{2p(n)} + e^{-n} < \frac{1}{p(n)}$, thus contradicting (1). ■

This theorem indicates that the existence of weak one-way functions is equivalent to that of strong one-way functions. In the next lecture, we will identify a "universal" one-way function f_{Levin}. This function is universal in the sense that if one-way functions exist, then f_{Levin} is one-way.