COM S 6830 – Cryptography	Sep 24, 2009

Lecture 9: Pseudo-Random Generators

Instructor: Rafael Pass *Scribe: Matt Weinberg*

1 Recap

Here we will recap three properties of pseudo-randomness proved last class (the proofs are not repeated here).

1. Indistinguishability is preserved under **efficient operations**. More specifically, if two ensembles $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are indistinguishable, and M is any nuPPT, then $\{M(X_n)\}_{n\in\mathbb{N}}$ and $\{M(Y_n)\}_{n\in\mathbb{N}}$ are also indistinguishable.

2. **Hybrid Lemma**: If $\{X_n^1\}_{n\in\mathbb{N}}, \ldots, \{X_n^m\}_{n\in\mathbb{N}}$ is a finite sequence of ensembles, and there exists a distinguisher D that can distinguish $\{X_n^1\}_{n\in\mathbb{N}}$ and $\{X_n^m\}_{n\in\mathbb{N}}$ with advantage ϵ, then there exists some $i < m$ and D' that can distinguish $\{X_n^i\}_{n\in\mathbb{N}}$ and $\{X_n^{i+1}\}_{n\in\mathbb{N}}$ with advantage ϵ/m.

3. **Prediction Lemma**: If $\{X_n^0\}_{n\in\mathbb{N}}$ and $\{X_n^1\}_{n\in\mathbb{N}}$ are ensembles and there exists a distinguisher, D, that can distinguish $\{X_n^0\}_{n\in\mathbb{N}}$ and $\{X_n^1\}_{n\in\mathbb{N}}$ with advantage $\mu(n)$, then there exists a nuPPT A that can "predict" which distribution a sample came from. More specifically, $Pr[b \leftarrow \{0,1\} : t \leftarrow X_n^b : A(t) = b] \geq 1/2 + \mu(n)/2$.

2 Pseudo-Randomness and the Next-bit Test

2.1 Definitions

Definition 1 *An ensemble, $\{X_n\}_{n\in\mathbb{N}}$ is said to be **pseudorandom** if every element sampled from X_n has the same length, denoted $m(n)$, and $\{X_n\}_{n\in\mathbb{N}}$ is indistinguishable from $\{U_{m(n)}\}_{n\in\mathbb{N}}$ (the uniform distribution).*

Definition 2 *An ensemble, $\{X_n\}_{n\in\mathbb{N}}$ is said to pass the **next-bit test** if no nuPPT can guess the i^{th} bit given the first $(i-1)$. Specifically, $\forall$ nuPPT A, $\exists$ negligible ϵ, such that $\forall\ n \in \mathbb{N}$, $\forall\ i \in [0, m(n)-1]$, $Pr[t \leftarrow X_n : A(1^n, t_{0\to i}) = t_{i+1}] \leq 1/2 + \epsilon(n)$.*

2.2 The theorem

Theorem 1 *Let $\{X_n\}_{n\in\mathbb{N}}$ be an ensemble where every element sampled from X_n has length $m(n)$. Then $\{X_n\}_{n\in\mathbb{N}}$ passes the next-bit test IF and ONLY IF $\{X_n\}_{n\in\mathbb{N}}$ is pseudo-random.*

Proof. First, it was shown last class that if $\{X_n\}_{n\in\mathbb{N}}$ is pseudo-random, then it passes "all" randomness tests, so the IF direction is covered.

Next, say that $\{X_n\}_{n\in\mathbb{N}}$ pass the next-bit test, and assume for contradiction that $\{X_n\}_{n\in\mathbb{N}}$ is not pseudo-random. Then there exists some distinguisher D that can distinguish $\{X_n\}_{n\in\mathbb{N}}$ and $\{U_{m(n)}\}_{n\in\mathbb{N}}$ with advantage $1/p(n)$ for some polynomial p, for infinitely many n. The remainder of the proof will be showing that for exactly these n, $\{X_n\}_{n\in\mathbb{N}}$ fails the next-bit test. And because there are infinitely many such n, $\{X_n\}_{n\in\mathbb{N}}$ does in fact fail the next-bit test. So from now on we will focus on one of these such n.

We'll set up to try and use the Hybrid Lemma. Let the distribution $H_n^i = \{l \leftarrow X_n, r \leftarrow U_{m(n)} : l_{0\to i}||r_{i+1\to m(n)}\}$. In other words, H_n^i samples from X_n and $U_{m(n)}$. It uses the first i bits from its sample of X_n, and the remaining bits from its sample of $U_{m(n)}$. then $H_n^0 = U_{m(n)}$ and $H_n^{m(n)} = X_n$. So we know that there exists a D that distinguishes H_n^0 from $H_n^{m(n)}$, so by the Hybrid Lemma, there exists a D' that distinguishes H_n^i from H_n^{i+1} with advantage $1/p(n)m(n)$, which is still polynomial.

Now we're going to define $G_n^{i+1} = \{l \leftarrow X_n, r \leftarrow U_{m(n)} : l_{0\to i}||1 - l_{i+1}||r_{i+2\to m(n)}\}$. Notice that G_n^{i+1} and H_n^{i+1} pull every bit from exactly the same distribution, except for the $i+1$ bit. Intuitively, if we could tell apart G_n^{i+1} and H_n^{i+1}, then we could tell whether the $i+1$ bit was "right" or "wrong," and we could use such a distinguisher to guess that $i+1$ bit given only the first i. First we have to show that we can in fact distinguish G_n^{i+1} from H_n^{i+1}.

We can write $H_n^i = \frac{1}{2}G_n^{i+1} + \frac{1}{2}H_n^{i+1}$. This is because the first i bits and the last $m(n)-i-1$ bits are sampled exactly the same in all three distributions. In addition, the $i+1$ bit of H_n^i is uniformly at random. The $i+1$ bit of H_n^{i+1} and G_n^{i+1} will always be opposites, so choosing each with probability $1/2$ is exactly a uniform distribution. Now we can use the fact that we can tell apart H_n^i from H_n^{i+1} to show that we can, in fact, tell apart G_n^{i+1} and H_n^{i+1}.

$$1/p(n)m(n) \leq |Pr[t \leftarrow H_n^{i+1} : D(t) = 1] - Pr[t \leftarrow H_n^i : D(t) = 1]|$$

$$= |Pr[t \leftarrow H_n^{i+1} : D(t) = 1] - (\frac{1}{2}Pr[t \leftarrow H_n^{i+1} : D(t) = 1] + \frac{1}{2}Pr[t \leftarrow G_n^{i+1} : D(t) = 1])|$$

$$= \frac{1}{2}|Pr[t \leftarrow H^{i+1} : D(t) = 1] - Pr[t \leftarrow G^{i+1} : D(t) = 1]|$$

The last line exactly says that the exactly same distinguisher, D, can distinguish H_n^{i+1} and G_n^{i+1} with advantage $2/p(n)m(n)$, which is still inverse polynomial. Now we want to use the prediction lemma, which says that there exists a machine A, such that if we randomly sample from H_n^{i+1} and G_n^{i+1} (each with probability $1/2$), then A can guess which sample we chose from with probability at least $1/2 + 1/p(n)m(n)$.

Now that we have this A, we will build and A' that can guess that $i+1$ bit of X_n given only the first i bits with probability $1/2 + 1/p(n)m(n)$, contradicting the fact that X_n passed the next-bit test. $A'(1^n, y)$ will do the following:

Let $b \leftarrow \{0,1\}$, and $r \leftarrow \{0,1\}^{m(n)-i-1}\}$. b is our "guess" at what the next-bit of y is. Our guess can either be right or wrong. If our guess is right, then $y||b||r$ is EXACTLY a sample from H_n^{i+1} (because the first $i+1$ bits are exactly a sample from X_n, and the remaining bits are uniform at random). If our guess is wrong, then $y||b||r$ is EXACTLY a sample from G_n^{i+1} (because the first i bits are exactly a sample from X_n, the $i+1$ bit is flipped, and the remaining bits are uniform at random). So now we run $A(y||b||r)$. If A outputs H_n^{i+1}, then A' outputs b. If A outputs G_n^{i+1}, then A' outputs $1-b$. Notice that A' is correct EXACTLY when A is correct by the argument above. If $y||b||r$ came as a sample from H_n^{i+1}, then b is the next bit of y. If $y||b||r$ came as a sample from G_n^{i+1}, then $1-b$ is the next bit of y. So A' succeeds with probability $1/2 + 1/p(n)m(n)$ in guessing the $i+1$ bit of a sample from X_n after seeing only the first i bits.
This proof was for a specific n, and holds for the infinitely many n such that the distinguisher D succeeded with probability $1/p(n)$. So for infinitely many n, A' can guess the $i+1$ bit given the first i bits (note that the i is non-uniform and depends on n) with advantage $1/p(n)m(n)$, so this is a contradiction and $\{X_n\}_{n\in\mathbb{N}}$ does not pass the next-bit test.
So now we have shown that $\{X_n\}_{n\in\mathbb{N}}$ is pseudo-random IF and ONLY IF it passes the next-bit test.

3 Constructing PRGs

3.1 Definition

A function $g : \{0,1\}^* \to \{0,1\}^*$ is a **Pseudo-Random Generator** if it satisfies the following three conditions:

1. **Efficiency**: g is (PPT)-computable. (and deterministic)
2. **Expansion**: $|g(x)| = l(|x|)$, and $l(k) > k$, $\forall\ k$.
3. **Pseudo-Random**: $\{x \leftarrow \{0,1\}^n g(n)\}_{n\in\mathbb{N}}$ is indistinguishable from $\{U_{l(n)}\}_{n\in\mathbb{N}}$.

3.2 First attempt

Shamir proposed the following PRG. Let f be a OWP, then let $g(s) = f^n(s)||f^{n-1}(s)||\ldots||s$. This is appealing because it is, by definition, hard to predict future "blocks" of $g(s)$ given only the first "blocks" of $g(s)$. However, it is easy to distinguish even $f(s)||s$ from a uniform distribution over $2|s|$. This is because from a random distribution, the chance that U_{2n} is of the form $f(s)||s$ is $1/2^n$. So a distinguisher just needs to check if the first half of the bits are equal to f applied on the second half. If so, then the probability that this happened by chance from a uniform distribution is negligible, so it is safe to say that it came from the distribution of $g(s)$. So g fails to be a pseudo-random generator.

3.3 Second attempt

Instead, let f be a OWP, with a hardcore predicate b. Then let $g(s) = f(s)||b(s)$. This time, g is provably a PRG. It is clear that g is **efficient** and **expanding**. Assume for contradiction that g is not a pseudo-random. Then g must fail the next-bit test by the work in the previous section. What possibilities are there for the i such that g can guess the $i+1$ bit given the first i bits? It clearly cannot be $i < |s|$. This is because f is a permutation, so $f(s)$, where $s \leftarrow U_n$ is EXACTLY a uniform distribution, which means that it is statistically impossible to guess the $i+1$ bit with probability better than $1/2$. However, we also cant have $i = |s|$, or else this means we could guess $b(s)$ given $f(s)$, contradicting the fact that $b(s)$ is a hardcore predicate for f. So no matter that i is, there is a contradiction, so no such i can exist, and g must pass the next-bit test, and therefore g is pseudo-random.