COM S 6830 – Cryptography September 22, 2011

Lecture 8: Computational Indistinguishability

Instructor: Rafael Pass *Scribe: Nick Alessi*

1 Motivation

Recall the one-time pad with message m and key k the code is $m \oplus k$. The main problem is that $|k| = |m|$, but what if we could expand a short key into a long one then this could make a good encryption scheme.

Lets say we expand a short random string into a long random string, what properties should that string have?

- Roughly as many 0's as 1's
- Any subset of the bits has roughly equal probability of being any bit string
- Any subset of the bits is "unbiased"
- Knowing some prefix we shouldn't be able to learn the next bit

These are all statistical tests of randomness. So if a string can pass these tests then it is pretty random. This is good enough for simulations, but for cryptography all possible tests must be considered.

2 Indistinguishability

The first thought would be to try to define indistinguishable by passing any statistical test. This does not work because $\nexists g : \{0,1\}^n \to \{0,1\}^{n+1}$ such that $U_{n+1} = \{x \leftarrow \{0,1\}^n; g(x)\} = g(U_n)$. Where $U_n = \{x \leftarrow \{0,1\}^n; x\}$.

Proof. Assume that such a g existed. Then take $k \leftarrow \{0,1\}^n$. Then $g(k)$ is going to be sampled with the distribution U_{n+1}. Then $g(k)$ can be used to encrypt a $n+1$ bit message as the key to a OTP. Since g samples uniformly we know that this is perfectly secure. However this contradicts Shannon's theorem that the OTP requires a key the length of the message. Thus no such g exists. ■

2.1 Computational Indistinguishability

We now try the same idea but instead of passing any statistical test (because that would be impossible) just pass the statistical tests in $nuPPT$.

First define an **Ensemble of Distributions** $\{X_n\}_{n\in\mathbb{N}}$ (when being lazy it may be written $\{X_n\}$), as a sequence $X_1, X_2, \ldots$ of distributions.

Definition: Let $\{X_n\}$ and $\{Y_n\}$ be ensembles of distributions over $\{0,1\}^{l(n)}$ where l is a polynomial. We say that $\{X_n\}$ and $\{Y_n\}$ are **computationally indistinguishable** ($\{X_n\} \approx \{Y_n\}$) if: $\forall\, \mathcal{D} \in nuPPT\ \exists\, \epsilon \in neg$ such that $\forall\, n \in \mathbb{N}$

$$|\Pr[t \leftarrow X_n; \mathcal{D}(1^n, t) = 1] - \Pr[t \leftarrow Y_n; \mathcal{D}(1^n, t) = 1]| \leq \epsilon(n)$$

Also say that $\mathcal{D}$ **distinguishes** X_n and Y_n with probability ϵ if:

$$|\Pr[t \leftarrow X_n; \mathcal{D}(1^n, t) = 1] - \Pr[t \leftarrow Y_n; \mathcal{D}(1^n, t) = 1]| > \epsilon(n)$$

$\mathcal{D}$ distinguishes $\{X_n\}$ and $\{Y_n\}$ with probability $\mu(\cdot)$ if $\forall\, n \in \mathbb{N}$:

$$|\Pr[t \leftarrow X_n; \mathcal{D}(1^n, t) = 1] - \Pr[t \leftarrow Y_n; \mathcal{D}(1^n, t) = 1]| > \mu(n)$$

First observe that if $\{X_n\} = \{Y_n\}$ then the probabilities in the above are equal, so $\{X_n\} \approx \{Y_n\}$. Also, if $\{X_n\}$ is statistically close to $\{Y_n\}$ then $\{X_n\} \approx \{Y_n\}$.

In fact two distributions can be disjoint and still computationally indistinguishable:

$$X_n = \left\{ p \leftarrow prime_n; g \leftarrow gen(Z_p^*); x \leftarrow \left[0, \frac{p-1}{2}\right] : g^x \right\}$$
$$Y_n = \left\{ p \leftarrow prime_n; g \leftarrow gen(Z_p^*); x \leftarrow \left[\frac{p-1}{2} + 1, p-1\right] : g^x \right\}$$

Since knowing this tells us the first bit being able to distinguish these would bread the discrete log assumption. Thus by contradiction $\{X_n\} \approx \{Y_n\}$.

2.2 Properties of Computational Indistinguishability

2.2.1 Sunglasses Lemma

Computational Indistinguishability is preserved under efficient operations. If $\{X_n\} \approx \{Y_n\}$ and $M \in nuPPT$ then $\{M(X_n)\} \approx \{M(Y_n)\}$

Proof. Assume $\mathcal{D} \in nuPPT$ and p a polynomial such that for infinitely many n $\mathcal{D}$ distinguishes $M(X_n)$ and $M(Y_n)$ with probability $\frac{1}{p(n)}$. Then the machine $\mathcal{D}' = \mathcal{D} \circ M$ distinguishes X_n and Y_n with probability $\frac{1}{p(n)}$. This contradicts that $\{X_n\} \approx \{Y_n\}$, so $\{M(X_n)\} \approx \{M(Y_n)\}$. ■

2.2.2 Transitivity

The **hybrid lemma**: Let $X_1, X_2, \cdots, X_m$ be a sequence of probability distributions. Assume that $\mathcal{D}$ distinguished X_1 and X_m with probability ϵ. Then $\exists\, i \in [m-1]$ such that $\mathcal{D}$ distinguishes X_i and X_{i+1} with probability $\frac{\epsilon}{m}$.

Proof. Let $g_i = \Pr[t \leftarrow X_i : \mathcal{D}(t) = 1]$. So using the triangle inequality:

$$\begin{aligned} \epsilon < |g_1 - g_m| &= |g_1 - g_2 + g_2 - g_3 \cdots + g_{m-1} - g_m| \\ &\leq |g_1 - g_2| + \cdots + |g_{m-1} - g_m| \end{aligned}$$

Thus if all of the terms $|g_i - g_{i+1}| \leq \frac{\epsilon}{m}$ then we get $\epsilon < (m-1) \cdot \frac{\epsilon}{m}$. This is a contradiction so there is an i such that $|g_i - g_{i+1}| > \frac{\epsilon}{m}$. ■

2.2.3 Application of Above

Let $\{X_n\} \approx \{Y_n\} \approx \{Z_n\}$ assume that all of these are PPT computable, then $\{X_n Y_n\} \approx \{Z_n Z_n\}$.

Proof. Assume that $\mathcal{D}$ distinguishes $\{X_n Y_n\}$ and $\{X_n Z_n\}$. Define M as the machine that samples from the correct X_n and concatenates that to the beginning of its input. Then by the sunglasses lemma $\{X_n Y_n\} = \{M(Y_n)\} \approx \{M(Z_n)\} = \{X_n Z_n\}$. Similarly redefine M as the machine that samples from the appropriate Z_n and concatenates that to the end of its input. Again the sunglasses lemma gives $\{X_n Z_n\} \approx \{Z_n Z_n\}$.

Define $H_1 = X_n Y_n$, $H_2 = X_n Z_n$, and $H_3 = Z_n Z_n$. Assume that $\mathcal{D}$ distinguishes H_1 and H_3 with non-negligible probability for infinitely many n. Then either $\mathcal{D}$ distinguishes H_1 and H_2 or H_2 and H_3 with non-negligible probability by the hybrid lemma. However either of these options contradicts the above, so no such $\mathcal{D}$ exists. Thus $\{X_n Y_n\} \approx \{Z_n Z_n\}$. ■