6.875J/18.425J Cryptography and Cryptanalysis	October 31, 2001
Lecture 15	
Instructor: Anna Lysyanskaya	*Scribed by: Josh Marron*

1 Recap

1.1 PRF definition

A random function is one whose output is fixed at random. A pseudorandom function, intuitively, is one that is not really random, but you cannot tell. This notion was introduced by Goldreich, Goldwasser, and Micali, who also gave a construction.

Let $\mathcal{F} : D \mapsto R$ be a function; let $A(\cdot)$ be and algorithm. By $A^{\mathcal{F}}(\cdot)$ we denote an algorithm that makes oracle calls to the function $\mathcal{F}$. More precisely, we may imagine that A has an additional tape Q (for "query"), on which it can write queries to $\mathcal{F}$, and as soon as it is done writing such a query q, an answer to it, namely $\mathcal{F}(q)$ magically appears on this tape.

Definition 1 *A family of functions $F.(\cdot)$ is pseudorandom if F is efficiently computable and for a randomly chosen S, F_S cannot be distinguished from a random function. More precisely, let $\rho(k, l(k), m(k))$ denote a family of random functions from $\{0,1\}^{l(k)}$ to $\{0,1\}^{m(k)}$. $F.(\cdot)$ determines a family of pseudorandom functions from $l(k)$ bits to $m(k)$ bits if $\forall$ PPTF $\{A_k\}$, there exists a negligible function $\nu(k)$ such that*

$$\begin{aligned} \Pr[S \leftarrow \{0,1\}^k; O_0 = F_S &\quad ; \\ R \leftarrow \rho(k, l(k), m(k)); O_1 = R &\quad ; \\ b \leftarrow \{0,1\}; b' \leftarrow A_k^{O_b}(1^k) &\quad : \quad b = b'] \leq 1/2 + \nu(k) \end{aligned}$$

Intuitively, the Turing machine for F and the random k bit seed implicitly defines a string of length $2^{l(k)} \cdot m(k)$.

1.2 The GGM construction of PRFs from PRGs

Suppose G is a PRG that expands k random bits into $2k$ pseudorandom bits. Then here is a tree-like construction of a function, F_S, from $\{0,1\}^k$ to $\{0,1\}^k$ with a k-bit seed:

Since $G(x)$ is length $2k$, we can divide it in half and create G_0 and G_1. $G_0(x)$ is the first k bits of $G(x)$ and $G_1(x)$ is the last k bits of $G(x)$

Let $S \leftarrow \{0,1\}^k$ be the random seed and $x = x_1 \circ x_2 \circ \ldots \circ x_k$. Then $F_s(x) \triangleq l_x$ where

$$\begin{aligned} l_x &= l_{x_1 \circ x_2 \circ \ldots \circ x_k} \\ &= G_{x_k}(l_{x_1 \circ x_2 \circ \ldots \circ x_{k-1}}) \\ &= G_{x_k} G_{x_{k-1}}(l_{x_1 \circ x_2 \circ \ldots \circ x_{k-2}}) \end{aligned}$$

$$\begin{aligned} l_0 &= G_0(S) \\ l_1 &= G_1(S) \end{aligned}$$

1.3 The Decisional Diffie-Hellman assumption

Let $P(1^k)$ be an algorithm that generates a prime p of length k and an element $g \in \mathbb{Z}_p^*$ such that g has large multiplicative order q. (Let's say that $q = \Omega(\sqrt{p})$.) Let $G =< g >$. Let's write $(p, q, g) \leftarrow P(1^k)$.

The decisional Diffie-Hellman assumption is that, for all adversaries $\{A_k\}$, there exists a negligible function $\nu(k)$ such that

$$\Pr[(p,q,g) \leftarrow P(1^k); x \leftarrow \mathbb{Z}_q; y \leftarrow Z_q; z_0 = xy \bmod q; z_1 \leftarrow \mathbb{Z}_q; b \leftarrow \{0,1\}; b' \leftarrow A_k(p,q,g,g^x,g^y,g^{z_b}) : b = b'] \leq 1/2 + \nu(k)$$

In contrast, the Diffie-Hellman assumption is that, if $P'(1^k)$ is an algorithm that generates a prime p and an element $g \in \mathbb{Z}_p^*$, then for all adversaries $\{A_k\}$, there exists a negligible function $\nu(k)$ such that,

$$\Pr[(p,g) \leftarrow P'(1^k); x \leftarrow \mathbb{Z}_p; y \leftarrow Z_p; Z \leftarrow A_k(p,g,g^x,g^y) : Z = g^{xy}] \leq \nu(k)$$

2 Three notions of indistinguishability

2.1 Perfect indistinguishability

Intuitively, two distributions X_1 and X_2 over a finite set D are perfectly indistinguishable if no adversary, even computationally unbounded, can tell whether he is sampling from X_1 or from X_2, no matter how many samples he obtains.

The formal definition is like this: two distributions X_1 and X_2 are perfectly indistinguishable if they are identical.

Definition 2 *X and Y are perfectly indistinguishable if* $\forall a \in D \Pr[X = a] = \Pr[Y = a]$.

Theorem 3 *Let Z_0 and Z_1 be perfectly indistinguishable. Then,* $\forall A$
$\Pr[b \leftarrow \{0,1\}; a \leftarrow Z_b; b' = A(a) : b' = b] = 1/2$

Proof First, let us show that if $Pr[b' = b|b = 0] = \alpha$ then $Pr[b' = b|b = 1] = 1 - \alpha$. More formally the first probability is $\Pr[a \leftarrow Z_0; b' = A(a) : b' = 0]$. Since Z_0 and Z_1 are perfectly indistinguishable this probability is the same as $\Pr[a \leftarrow Z_1; b' = A(a) : b' = 0]$. Since this probability is just α we now have $Pr[b' = b|b = 1] = 1 - \alpha$. Now we decompose $\Pr[b' = b]$ into $\Pr[b' = b|b = 0]\Pr[b = 0] + \Pr[b' = b|b = 1]\Pr[b = 1] = \alpha \times 1/2 + (1 - \alpha) \times 1/2 = 1/2$ ■

2.2 Statistical indistinguishability

The intuition here is that even an unbounded adversary, given only polynomially many samples from either X_1 or X_2, cannot tell which it is with an advantage better than negligible.

Definition 4 *X and Y are ϵ-close if* $\sum_{a \in D} |\Pr[X = a] - \Pr[Y = a]| \leq \epsilon$.

Definition 5 *$\{X_k\}$ and $\{Y_k\}$ are statistically indistinguishable families of probability distributions over sets $\{D_k\}$ if for all k, X_k and Y_k are $\nu(k)$-close for some negligible function $\nu(k)$.*

Theorem 6 *Let $\{X_k\}$ and $\{Y_k\}$ be statistically indistinguishable. Then, for any polynomial p, given $p(k)$ independently chosen no adversary can tell X_k from Y_k with advantage better than negligible.*

Proof First, let us show that if X and Y are ϵ-close, then, given one sample from either X or Y, given just one sample, the advantage in distinguishing which distribution it came from is ϵ. Why? Because the best strategy that any A can use to determine which distribution a sample came from is to choose the one with the maximum likelihood. Using this strategy $\Pr[A\, correct|a] = max(\Pr[X|a], \Pr[Y|a]) = 1/2 + 1/2|\Pr[X|a] - \Pr[Y|a]|$ and so $\forall A \Pr[A\, correct|a] \leq 1/2 + 1/2|\Pr[X|a] - \Pr[Y|a]|$.
$\Pr[X|a] = \Pr[X, a]/\Pr[a] = \Pr[a|X]\Pr[X]/\Pr[a] = \Pr[a|X]1/2/\Pr[a]$ and so
$Pr[A\, correct] = \sum_{a\in D} \Pr[A\, correct|a]\Pr[a]$
$\leq \sum_{a\in D}(1/2 + 1/2|\Pr[X|a] - \Pr[Y|a]|)\Pr[a]$
$\leq 1/2 + 1/2\sum_{a\in D}|\Pr[a|X]1/2/\Pr[a] - \Pr[a|Y]1/2/\Pr[a]|\Pr[a]$
$\leq 1/2 + 1/4\sum_{a\in D}|\Pr[a|X] - \Pr[a|Y]|$
$\leq 1/2 + 1/4\epsilon$

And now we can use the hybrid argument to finish the proof. Let our $p(k)$ hybrids be as follows:
$H_0 = X^{p(k)}$
$H_1 = Y \circ X^{p(k)-1}$
$H_i = Y \circ X^{p(k)-i}$
$H_{p(k)} = Y^{p(k)}$

Suppose that the sets of samples from X and Y are δ-close for $\delta > \nu(k)p(k)$. The hybrid argument says that at least on pair of adjacent hybrids is not $\nu(k)$-close and since these hybrids only differ by one sample X and Y cannot be $\nu(k)$-close. ■

2.3 Computational indistinguishability

This is the notion we have seen quite a bit already.

2.4 Examples

Identical distributions: Shannon's one-time pad.

Statistical closeness: generate k bits uniformly at random or by picking a $4k/3$-bit prime and trunkating the first $k/3$ bits.

Computational closeness: all the stuff we've been staring at.

Note: Perfect indistinguishability $\Rightarrow$ Statistical indistinguishability $\Rightarrow$ Computational indistinguishability