CS195-7: Introduction to Cryptography	DUE: November 12, 2002
Solution Set 7	
Instructor: Anna Lysyanskaya	

PRF Candidates

Suppose that $\{F_S : \{0,1\}^k \mapsto \{0,1\}^k \mid s \in \{0,1\}^k\}$ is a pseudo-random family of functions from k-bit input to k-bit output, indexed by a k-bit key ("seed"). We would like to construct a new PRF family. Consider the following constructions, and for each show whether it is good or bad (namely whether the specified family is pseudo-random or not, for the appropriate domain and range). If your answer is "yes," give an outline of a reduction. If your answer is "no," give an example F, some convincing evidence that F is a PRF, and an attack on the construction of F^i when it is based on your F.

Below, "$\circ$" denotes concatenation, and $\bar{x}$ denotes the bitwise negation of x.

(a) $F_S^1(x) = F_S(x) \circ F_S(\bar{x})$.

This is not a PRF. $F_S^2(\bar{x}) = F_S(\bar{x}) \circ F_S(x)$ can be obtained from $F_S^2(x)$ by flipping the first and the last k bits, i.e. $F_S^2(\bar{x})$ is determined from $F_S^2(x)$. For a random function, for any x, the probability that this happens is $1/2^{2k}$. Thus, advantage is $1 - \frac{1}{2^{2k}}$.

(b) $F_S^2(x) = F_{0^k}(x) \circ F_S(x)$.

This is not a PRF. $F_S^2(x)$ has the same first k bits for every $x \in \{0,1\}^k$. A distinguishing test that checks if $f(0^k)$ and $f(1^k)$ have the same first k bits, distinguishes with probability $1 - \frac{1}{2^k}$. Alternatively, $F_S^1(0^k) = F_S(0^k) \circ F_S(0^k)$ – a concatenation of 2 identical strings, which happens with probability $1/2^k$ for a random f.

(c) $F_S^3(x) = F_{S_1}(x) \circ F_{S_2}(x)$, where $S_1 = F_S(0^k)$ and $S_2 = F_S(1^k)$.

Yes, $F_S^3(x)$ is a PRF. Let us denote by $R_1, R_2, R = R_1 \circ R_2$ truly random functions from k to k, k and $2k$ bits, respectively; by $g_1, g_2, g = g_1 \circ g_2$ – functions F_{S_1}, F_{S_2}, $F_{S_1} \circ F_{S_2}$, where S_1 and S_2 are chosen independently and uniformly at random from $\{0,1\}^k$; by $h_1, h_2, h = h_1 \circ h_2$ – functions F_{S_1}, F_{S_2}, F_S^3, where $S_1 = F_S(0^k)$, $S_2 = F_S(1^k)$ and S is chosen at random from $\{0,1\}^k$. Given an algorithm C with a black-box access to a function f chosen according to some distribution, denote by $C(f)$ the random variable denoting the value of the output of C when f is chosen according to this distribution. Say that $f_1 \approx f_2$ if they are distributed according to the same distribution.

Now, we are ready to prove that $\{F_S^3\}$ is a PRF family. Assume not, i.e. there is a distinguisher $A(\cdot)$ for $\{F_S^3\}$ s.t. for infinitely many k,

$$\Big| \Pr(A^h = 1) - \Pr(A^R = 1) \Big| > \frac{1}{poly(k)} \qquad (1)$$

Let us obtain a contradiction by constructing a distinguisher B for $\{F_S\}$. Given a black-box access to a function $f : \{0,1\}^k \mapsto \{0,1\}^k$, define $f_1 = F_{f(0^k)} \circ F_{f(1^k)}$,

$f_2 = g_1 \circ f$ (where g_1 is simulated by B by choosing random S_1) and $f_3 = f \circ R_2$ (where R_2 is a random function simulated by B). B picks a random $i \in \{1,2,3\}$ and outputs $A(f_i)$.

We note that when $f = F_S$ for a random S, then: $f_1 \approx h$, $f_2 \approx g$ and $f_3 \approx g_1 \circ R_2$; if f is a random function, then $f_1 \approx g$, $f_2 \approx g_1 \circ R_2$ and $f_3 \approx R$. Hence, the advantage of B can be written as:

$$\begin{aligned} p &= \frac{1}{3} \cdot \mid (\Pr(A(h)=1) - \Pr(A(g)=1)) + (\Pr(A(g)=1) - \Pr(A(g_1 \circ R_2)=1)) + \\ &\quad (\Pr(A(g_1 \circ R_2)=1) - \Pr(A(R)=1)) \mid = \frac{1}{3} \cdot \mid \Pr(A(h)=1) - \Pr(A(R)=1) \mid \\ &\overset{(1)}{\geq} \frac{1}{3 \cdot poly(k)} \end{aligned}$$

(d) $F_S^4(x) = F_S(x) \oplus x$.

$F_S^4(x)$ is a PRF. Suppose that we have an adversary who, with oracle access to a function f, can distinguish if f is F_S^4 for a randomly chosen seed S, or a truly random function R.

Let us use this adversary to distinguish F_S from R. Suppose we are given oracle access to some function f' and our task is to determine, in polynomial time, whether f' is computed according to F_S, or is a truly random function. Here is how we proceed: we invoke the adversary. If the adversary's query to his oracle f is the value x, we query our oracle f' on the value x. Suppose our oracle's answer is y. We reply to the adversary's query with the value $y \oplus x$. In the end, if the adversary outputs "Random," we also output "Random," otherwise, we output "Not random."

Claim 1: if our oracle f' is the function F_S for a randomly chosen seed S then the oracle that the reduction implements for the adversary is for the function F_S^4 for the same seed S.

Proof of claim: this is immediate, since on input x, we give the adversary the value $F_S(x) \oplus x = F_S^4(x)$.

Claim 2: if our oracle f' is a random function R, then the oracle that the reduction implements for the adversary is also a random function R'.

Proof of claim: If $R(x)$ is a random function, then $R'(x) = R(x) \oplus x$ is a random function as well, and this is what we give to the adversary.

From these two claims, it is clear that the reduction produces the correct answer whenever the adversary does. Therefore, if the adversary's success is non-negligible, F_S is not a PRF, which is a contradiction.

(e) $F_S^5(x) = F_S(x) \oplus S$.

It does not follow that F_S^5 is a PRF. For example, suppose we are given a pseudorandom function family $F'_S(x) : \{0,1\}^k \mapsto \{0,1\}^k$ which is indexed by a $k-1$-bit seed. Suppose $F_S(x)$ is formed as follows. For a k-bit string u, let $u^{(k-1)}$ denote the first $k-1$ bits of u, and u^k denote the kth bit of u. $F_S(x) = F'_{S^{(k-1)}}(x)$ if $x \neq 0^k$, and

$F_S(0^k) = (F'_S(0^k))^{(k-1)} \circ S^k$. $F_S(x)$ is a pseudorandom function because an adversary that distinguishes it from a truly random function can be plugged in directly to distinguish $F'_S(x)$ from a truly random function. On the other hand:

$$
\begin{aligned}
F_S^5(0^k) &= ((F'_S(0^k))^{(k-1)} \circ S^k) \oplus S \\
&= ((F'_S(0^k))^{(k-1)} \oplus S^{(k-1)}) \circ (S^k \oplus S^k) \\
&= ((F'_S(0^k))^{(k-1)} \oplus S^{(k-1)}) \circ 0
\end{aligned}
$$

thus the last bit of $F_S^5(0^k)$ is always 0, while for a random function, this happens with probability 1/2. This allows us to distinguish.