# Pseudorandom functions and permutations

15-859I
Spring 2003

## Introduction

- Informally, a Pseudorandom function family (PRF) is a collection of functions which are indistinguishable from random functions.
- PRFs are enormously useful for cryptography. But they are also useful for several algorithmic applications.
- We'll show two equivalent characterizations of PRFs, and give a construction of a PRF from a secure PRG. (Thus we will have that if OWFs exist, then PRFs exist)

## Notations

- A collection of functions is a set $\{f_s\}_{s\in\{0,1\}^*}$ where each $f_s:\{0,1\}^{p(|s|)} \to \{0,1\}^{q(|s|)}$, for polynomials p(.), q(.). $f$ is efficiently computable if there exists a PPT F such that $F(s,x) = f_s(x)$ for all s, $x\in\{0,1\}^{p(|s|)}$.
- We denote by $F_{p,q}$ the set of all functions from $\{0,1\}^p \to \{0,1\}^q$.
- Given a PRG G: $\{0,1\}^k \to \{0,1\}^{2k}$ we define $G_0, G_1 : \{0,1\}^k \to \{0,1\}^k$ so that $G(x) = G_0(x), G_1(x)$.

## PRFs: Statistical Test formulation

An efficiently computable collection $f_s$ is statistical-test pseudorandom if for every probabilistic polynomial time oracle TM T,

$$Adv(T,k) = | Pr[T^f(1^k) = 1] - Pr[T^g(1^k) = 1] |$$

is negligible, where f denotes an oracle for $f_s$ for $s \leftarrow U_k$ and g denotes an oracle drawn uniformly from $F_{p(k),q(k)}$.

We call Adv(T,k) the *advantage* of T.

## PRFs: Polynomial inference formulation

Consider an experiment with t+1 stages:

First, choose $s \leftarrow U_k$.

Stage 1: A produces a query $x_1$.

Stage j, $1<j<t+1$: A is given $x_i, f_s(x_i)$, for $0 < i < j$, and produces query $x_j \neq x_i$.

Stage t+1: Choose $b \leftarrow U_1$. Choose $y_{1-b} \leftarrow U_{q(k)}$. Let $y_b = f_s(x_t)$. A is given $x_i$, $f_s(x_i)$, $x_t, y_0, y_1$.

We say that PPTM A *Q(k)-infers* $f$ if

$$Pr[A(f_s(x_1),\ldots,f_s(x_{t-1}),y_0,y_1) = b] > \tfrac{1}{2} + 1/Q(k)$$

## Equivalence

- Theorem: $f$ cannot be polynomially inferred if and only if it is statistical-test pseudorandom.
- Proof: both directions in the contrapositive:
  - (a) if $f$ can be Q(k) inferred, then there is a statistical test T which distinguishes between $f_S$ and $F_{p,q}$ with advantage 1/Q(k).
  - (b) if there is a statistical test T for $f$ that has advantage 1/Q(k) and makes P(k) queries then $f$ can be 2P(k)Q(k) inferred.

## Proof of (a)

- Let A be the algorithm that Q-infers $f$. Then the oracle PPTM $T_A^O$ works as follows:
  - Stage 1: run A to get query $x_1$, respond with $O(x_1)$.
  - Stage j: Let $x_j = A(O(x_1)\ldots O(x_{j-1}))$.
  - Stage t+1: Choose $b \leftarrow U_1$, let $y_b = O(x_t)$, draw $y_{1-b} \leftarrow U_{q(k)}$. If $A(O(x_1)\ldots O(x_{t-1}), y_0, y_1) = b$, output 1, else output 0.
- Notice that $\Pr[T_A^g(1^k) = 1] = ½$
- But since A Q(k)-infers $f$, $\Pr[T_A^f(1^k)=1] > ½ + 1/Q(k)$.
- So $T_A$ has advantage 1/Q(k) as claimed.

## Proof of (b)

- Suppose there is a statistical test T that has advantage 1/Q(k) against $f$ and makes P(k) (wlog, distinct) oracle queries. We will construct an algorithm $A_T$ that 2P(k)Q(k) infers $f$.

## Definition of $A_T$

- Choose $j \in_U \{0,1,\ldots P(k)-1\}$
- Stage i < j: Respond to T's query $q_{i-1}$ with $f_s(x_{i-1})$; run T until it makes query $q_i$ and return $x_i = q_i$.
- Stage j: Respond to T's query $q_{j-1}$ with $f_s(x_{i-1})$. Run T to get query $q_j$. Return $x_j \leftarrow U_{p(k)}$.
- Stage i, j<i<P(k): Return $x_i \leftarrow U_{p(|s|)}$.
- Stage P(k): Return $x_{P(k)} = q_j$.
- Stage P(k)+1: Run T with $y_0$ as $O(q_j)$ Respond to additional queries from T with uniform bits, until T outputs bit b'. Return 1-b'.

## Proof that this works:

- Consider doing an $(k,j,f_s)$ experiment: Run T, and on query $q_i$, respond with:
  - $f_s(q_i)$, if i < j.
  - $y_i \leftarrow U_{q(k)}$ otherwise.
- Let $p^j(k)$ denote the probability that T outputs 1 in an $(k,j,f_s)$ experiment. Then $p^0(k) = \Pr[T^g(1^k) = 1]$ and $p^{P(k)}(k) = \Pr[T^f(1^k) = 1]$.
- So we also have $\Sigma_j(p^j(k) - p^{j-1}(k)) = \mathrm{Adv}(T,k)$

…

## …

Then $\Pr[1\text{-}b' = b]$

$= \Sigma_i \Pr[j=i]\Pr[1\text{-}b' = b \mid j=i]$

$= 1/P(k)\ \Sigma_i \Pr[1\text{-}b' = b \mid j=i]$

$= 1/P(k)\ \Sigma_i (\Pr[b=1|j=i]\ \Pr[b'=0|y_0 \leftarrow U_{q(k)} \& j=i]$
$+ \Pr[b=0|j=i]\ \Pr[b'=1|y_0 = f_s(q_i) \& j=i])$

$= 1/P(k)\ \Sigma_i (½\ \Pr[T \text{ outputs } 0 \text{ in } (k,i,f_s) \text{ exp}]$
$+ ½\ \Pr[T \text{ outputs } 1 \text{ in } (k,i+1,f_s) \text{ exp}])$

$= 1/2P(k)\ \Sigma_i ((1-p^i(k)) + p^{i+1}(k))$

$= ½ + 1/2P(k)Q(k)$, QED.

## Recap

- So a function family is statistical test pseudorandom iff it is polynomially uninferrable. (Proof by Hybrid method!)
- From now on, we will just say that a function family is pseudorandom if it is statistical test pseudorandom.
- That is, a pseudorandom function family (PRF) is a function family which is statistical test pseudorandom.

## Constructing PRFs from PRGs

- Let $G:\{0,1\}^k \rightarrow \{0,1\}^{2k}$ be a PRG.
- Define $f_s$: $\{0,1\}^k \rightarrow \{0,1\}^k$ by
  $f_s(x_1x_2\ldots x_k) = G_{x_k}(G_{x_{k-1}}(\ldots(G_{x_2}(G_{x_1}(s)))\ldots))$
- Theorem: $f_s$ is a PRF.
- Proof: Suppose that there is a statistical test T that makes Q(k) queries and has Adv(T,k) = 1/P(k). We will construct a distinguisher for G with advantage 1/kQ(k)P(k).

## Proof of Theorem…

- Define a sequence of PPTs $A_0\ldots A_k$.
- $A_i$:
  - Let L be an associative array.
  - Run T, responding to queries $(y_1\ldots y_k)$:
    - If $(y_1\ldots y_i) \in L$, let $r = L(y_1\ldots y_i)$
    - Else choose $r \leftarrow U_k$ and set $L(y_1\ldots y_i) = r$
    - respond with $f_r(y_{i+1}\ldots y_k)$
  - Return output of T.
- Notice: $Pr[A_0 = 1] = Pr[T^f(1^k) = 1]$, $Pr[A_k = 1] = Pr[T^g(1^k) = 1]$; $E_i[A_i - A_{i-1}] = 1/kP(k)$

## Proof, continued

- Construct adversary $A_T(r_1,\ldots r_{Q(k)})$ $(r_i \in \{0,1\}^{2k})$:
- $A_T(r_1,\ldots,r_{Q(k)})$=
  - Choose $i \in_U \{0,\ldots k-1\}$. Let L be as in $A_j$. Set j = 0.
  - Run T, responding to queries $(y_1,\ldots,y_k)$:
    - If $(y_1\ldots y_{i+1}) \in L$, Let $r = L(y_1\ldots y_{i+1})$.
    - Else: increment j;
      - $L(y_1\ldots y_i0)$ = Left-Half($r_j$);$L(y_1\ldots y_i1)$ = Right-Half($r_j$)
      - $r = L(y_1\ldots y_{i+1})$
    - Respond with $f_r(y_{i+2}\ldots y_k)$
  - Return output of T.

## Properties of $A_T$

- Notice that
  - $Pr[A_T(U_{2kq(k)}) = 1] - Pr[A_T(G(U_{kq(k)}))] =$
  - $E_i[A_i\text{-}A_{i-1}] = 1/kP(k)$
- That is, if T distinguishes $f_s$ from $F_{p,q}$ with advantage 1/P(k), $A_T$ distinguishes Q(k) samples of $G(U_k)$ from q(k) samples of $U_{2k}$ with advantage at least 1/kP(k).
- But then $A_T$ can distinguish between a single sample of $G(U_k)$ and $U_{2k}$ with advantage 1/kQ(k)P(k), QED.

## Pseudorandom permutations

- A Pseudorandom Permutation family (PRP) is a PRF where every element $f_s$ is a bijection on $\{0,1\}^{p(|s|)}$. (PRPs have inverses)
- Typically for P: $\{0,1\}^k \times \{0,1\}^{L(k)} \rightarrow \{0,1\}^{L(k)}$ there is an efficient algorithm to compute $P_K^{-1}(x)$, given K.
- A Strong Pseudorandom Permutation family (SPRP) is a PRP which remains pseudorandom even when the adversary is given access to an oracle for $P_K$ and $P_K^{-1}$.
- Naor and Reingold show that given a PRF $f_s$: $\{0,1\}^k \rightarrow \{0,1\}^k$ we can construct a SPRP $P_f$ on $\{0,1\}^{2k}$.

## Notation, definitions

- Let $\Pi_L$ denote the uniform distribution on permutations on $\{0,1\}^L$.
- Define the Advantage of A against permutation family P on $\{0,1\}^{l(k)}$ by:
  $Adv(A,k) = |Pr[A^{P(K,.)}(1^k) = 1] - Pr[A^{\Pi_{l(k)}}(1^k)=1]|$
- For any function f: $\{0,1\}^k \rightarrow \{0,1\}^k$, define the permutation $D_f$: $\{0,1\}^{2k}$ by $D_f(l,r) = (r, l \oplus f(r))$
- Note that $D_f^{-1}$ is easy to compute.

## Intuition behind construction

- Let $f_1$, $f_2$ be chosen from a PRF family on $\{0,1\}^k$. Then $P = D_{f2} \circ D_{f1}$ is indistinguishable from a random permutation *on uniformly chosen inputs*.
- There *is* a distinguisher for P on chosen inputs: Left-half$(P(L_1,R) \oplus P(L_2,R)) = L_1 \oplus L_2$ whereas this holds with probability only $2^{-n}$ for a random permutation.
- But if we can insure that the inputs to P, $P^{-1}$ are unique (whp), then we get what we need.

## Construction: SPRP

- Let $h_1, h_2$ be chosen from a pairwise-independent family of permutations on $\{0,1\}^{2k}$
- Let $f_1, f_2 : \{0,1\}^k \rightarrow \{0,1\}^k$ be chosen from a PRF family.
- Define the permutation $P = h_2^{-1} \circ D_{f2} \circ D_{f1} \circ h_1$.
- SPRP Lemma: If $f_1$, $f_2$ are chosen according to $F_{k,k}$, then P is statistically close to a random permutation.
- SPRP Theorem: P is a SPRP.

## Proof tools

- Let M be a deterministic, unbounded oracle machine. M makes two types of queries:
  - (+,x): Querying G(x)
  - (-, y): Querying $G^{-1}(y)$
- Denote the $i^{th}$ query-answer pair of M by $(x_i,y_i)$.
- Wlog, assume M makes exactly m queries.
- Denote by $\{(x_1,y_1),\ldots,(x_m,y_m)\}$ the *transcript* of M with oracle G.
- Denote by $C_M[\{(x_1,y_1),\ldots,(x_{i-1},y_{j-1})\}]$ the deterministic function for the $i^{th}$ query of M, and let $C_M[\{(x_1,y_1)\ldots(x_m,y_m)\}]$ the output of M.

## Proof tools, cont'd.

- Call $\sigma = \{(x_1,y_1)\ldots(x_m,y_m)\}$ a *possible M transcript* if for every $1< i < m$, $C_M[\{(x_1,y_1),\ldots,(x_{i-1},y_{i-1})\}] \in \{(+,x_i),(-,y_i)\}$
- Let $T_P$ be a random variable denoting the transcript of M with an oracle for P.
- Let $T_\Pi$ be a random variable denoting the transcript of M with an oracle chosen from $\Pi_{2k}$.

## More Proof Tools.

- Let $T_R$ be a random variable denoting transcripts from a random process which responds to M's $i^{th}$ query as follows:
  - If the query is (+,x) and for some $j < i$ there is a query-answer pair $(x,y_j)$, return $y_j$.
  - If the query is (-,y) and for some $j < i$ there is a query-answer pair $(x_j,y)$ return $x_j$
  - Otherwise, return a uniformly chosen 2k-bit string.
- $T_R$ might contain responses inconsistent with a permutation: Call a possible M-transcript $\sigma = \{(x_1,y_1),\ldots(x_m,y_m)\}$ inconsistent if for some $1 < j < i < m$, $x_i=x_j$ and $y_i \neq y_j$ or $y_i=y_j$ and $x_i \neq x_j$.

## Lemma 1

- $Pr[C_M(T_R) = 1] - Pr[C_M(T_\Pi) = 1] \leq m^2/2^{2k+1}$.
- Proof:
  - $Pr[T_R = \sigma \mid T_R \text{ is consistent}] = Pr[T_\Pi = \sigma]$ (because a consistent transcript is a permutation, and $T_R$ chooses uniformly from transcripts)
  - So $Pr[C_M(T_R) = 1] - Pr[C_M(T_\Pi) = 1] \leq Pr[T_R \text{ is inconsistent}]$
  - $Pr[T_R \text{ is inconsistent}] \leq$ $\Sigma_{i,j} Pr[x_i=x_j \text{ and } y_i \neq y_j \text{ or } y_i=y_j \text{ and } x_i \neq x_j] =$ $(m(m-1)/2)2^{-2k} \leq m^2/2^{2k+1}$.

## BAD set

- For a fixed $h_1,h_2$, define the set BAD($h_1,h_2$) to be the set of possible and consistent M-transcripts $\{(x_1,y_1),\ldots,(x_m,y_m)\}$ such that:
  - There exist i<j with $h_1(x_i)|_R = h_1(x_j)|_R$. OR
  - There exist i<j with $h_2(y_i)|_L = h_2(y_j)|_L$ .
- Lemma 2. Fix a possible, consistent M-transcript $\sigma$. Then

  $\Pr_{h1,h2}[\sigma \in BAD(h_1,h_2)] < m^2/2^k$.

## Lemma 2 proof.

- $\sigma \in BAD(h_1,h_2)$ if there exist i < j with $h_1(x_i)|_R = h_2(x_j)|_R$ ($R_{i,j}$) or $h_2(y_i)|_L = h_2(y_j)|_L$ ($L_{i,j}$).
- So $\Pr[\sigma \in BAD(h_1,h_2)] \le \Sigma_{i<j}(\Pr[R_{i,j}] + \Pr[L_{i,j}])$
  $\le \Sigma_{i<j}(2^{-k} + 2^{-k})$
  $< m^2/2^k$, QED.

## Key Lemma

- Lemma 3: Let $\sigma = \{(x_1,y_1),\ldots,(x_m,y_m)\}$ be possible and consistent. Then
  $\Pr[T_P = \sigma \mid \sigma \notin BAD(h_1,h_2)] = \Pr[T_R = \sigma]$.
- Proof: Given $\sigma$ is consistent,
  $\Pr[T_R = \sigma] = 2^{-2km}$.
  But suppose $\sigma \notin BAD(h_1,h_2)$. Then for the $i^{th}$ query-answer pair, $y_i = P(x_i)$ exactly when:
  $h_1(x_i) = (L_i^0,R_i^0)$, $L_i^2 = L_i^0 \oplus f_1(R_i^0)$,
  $R_i^2 = R_i^0 \oplus f_2(L_i^2)$, and $L_i^2,R_i^2 = h_2(y_i)$.
  i.e., when $f_1(R_i^0) = L_i^0 \oplus L_i^2$, and $f_2(L_i^2) = R_i^0 \oplus R_i^2$

## Lemma Proof, Cont'd.

We get $y_i = P(x_i)$ when $f_1(R_i^0) = L_i^0 \oplus L_i^2$, and $f_2(L_i^2) = R_i^0 \oplus R_i^2$.

But since we never have $L_i^2 = L_j^2$ or $R_i^0 = R_j^0$ (otherwise $\sigma$ is bad), these probabilities are independent. And since $f_1,f_2$ are chosen randomly we get that $\Pr[T_P = \sigma \mid \sigma \notin BAD(h_1,h_2)] = 2^{-2km}$, QED.

## SPRP Lemma: Proof

- SPRP Lemma: if $f_1,f_2 \leftarrow F_{k,k}$ then for any oracle adversary M which makes at most m queries,
  $|\Pr[M^P(1^k) = 1] - \Pr[M^{\Pi 2k}(1^k) = 1]| \le m^2/2^{2k+1} + m^2/2^k$
- Proof: composition of previous three lemmas.

## SPRP Theorem: Proof

- SPRP Theorem: If $f_1,f_2 \leftarrow f_s$, where $f_s$ is a PRF, then P is a SPRP.
- Proof: Assume not. Let M be an efficient oracle machine distinguishing P from a randomly chosen permutation. Consider the hybrid distribution H where $f_1 \leftarrow F_{k,k}$, and $f_2 \leftarrow f_s$.
- Recall $\Pr[M^P=1] - \Pr[M^\Pi=1] > 1/\text{poly}(k)$.

## SPRP proof

- Then we must have either:
  (a) $|\Pr[M^{P}=1] - \Pr[M^{H}=1]| > 1/\text{poly}$; OR
  (b) $|\Pr[M^{H}=1] - \Pr[M^{\Pi}=1]| > 1/\text{poly}$.
- This gives a statistical test against $f_s$: run M, answering queries using P with either :
  - (a) a random function; or
  - (b) a pseudrandom function

  for $f_1$, and the function oracle for $f_2$.
- Then in either case, we distinguish $f_s$ with the same gap as M in distinguishing H from P or $\Pi$.