CSE 5852: Problem Set 5

Due: October 31, 2016

This homework combines ideas from pseudorandom generators, pseudorandom functions and message authentication codes. Please be clear what definitions are used in your proofs. Also since there are many definitions feel free to copy and paste from the problem set latex source.

1 Pseudorandom Number Generators

Recall the second definition we gave of a pseudorandom generator.

All efficient tests That is consider two experiments: $\texttt{exp} - \texttt{pr}$ and $\texttt{exp} - \texttt{r}$. Let T be some PPT test that outputs either 1 or 0.

Experiment $\texttt{exp} - \texttt{pr}^{G,T}$:	**Experiment** $\texttt{exp} - \texttt{r}^{T}$:
Select random s of length n.	Select random y of length m
Compute $y = G(s)$	Run $T(y)$ and output whatever it does.
Run $T(y)$ and output whatever it does.	

Definition 1. *G passes all statistical tests if for all PPT T, there exists negligible function $\epsilon(n)$ such that for all n,*

$$\left|\Pr[\texttt{exp} - \texttt{pr}^{G,T} = 1] - \Pr[\texttt{exp} - \texttt{r}^{T} = 1]\right| \leq \epsilon(n).$$

Let $G_1 : \{0,1\}^n \to \{0,1\}^{2n}, G_2 : \{0,1\}^{2n} \to \{0,1\}^{4n}$ be pseudorandom generators.

Answer the following questions using a proof by reduction. The following proofs require the use of a hybrid argument.

Hybrid Argument Introduce an intermediate experiment and show that if the two end experiments are distinguishable by an inverse polynomial then the distance between some end experiment and the intermediate experiment is distinguishable by an inverse polynomial. (See Lecture 13 notes, page 3). Then use show a distinguisher for some other problem based on each of these intermediate problems.

a) **35 pts** If G_1, G_2 are pseudorandom generators then the function $G_3 : \{0,1\}^n \to \{0,1\}^{4n}$ defined as $G_3(s) = G_2(G_1(s))$ is a pseudorandom generator.

Hint: Your intermediate experiment should consider when the input to G_2 is replaced by a truly random string.

(a) **5 pts** State, using mathematics, what it means for G_3 to not be a pseudorandom generator.

(b) **5 pts** State in your own words the two parts of any hybrid argument. Feel free to assign names to experiments. What is the largest number of experiments that can be introduced safely?

(c) **5 pts** Describe a hybrid experiment between $\texttt{exp} - \texttt{pr}$ and $\texttt{exp} - \texttt{r}$.

(d) **5 pts** Show what must be true about one of the hybrids if the distance between $\texttt{exp} - \texttt{pr}$ and $\texttt{exp} - \texttt{r}$ is an inverse polynomial.

(e) **5 pts** Build a distinguisher for G_1 in one setting of the hybrid.

(f) **5 pts** Build a distinguisher for G_2 in the other setting of the hybrid.

(g) **5 pts** Describe how you have reached a contradiction overall.

b) **20 pts** If G_1 is a pseudorandom generator then the function $G_3 : \{0,1\}^{2n} \to \{0,1\}^{4n}$ defined as $G_3(s_1||s_2) = G_1(s_1)||G_1(s_2)$ is a pseudorandom generator.

Hint: Your intermediate experiment should consider when one of the outputs of G_1 is replaced by a truly random string.

(a) **5 pts** State, using mathematics, what it means for G_3 to not be a pseudorandom generator.

(b) **5 pts** Describe a hybrid experiment between $\texttt{exp} - \texttt{pr}$ and $\texttt{exp} - \texttt{r}$ and show what must be true about the distinguisher in one of the two settings.

(c) **5 pts** Build a distinguisher for G_1 in either setting of the hybrid.

(d) **5 pts** Describe how you have reached a contradiction overall.

2 Pseudorandom Functions are good MACs

In this question we will show that pseudorandom functions work as a good MAC function (under the appropriate definition). **Read the entire question before beginning!**

Pseudorandom Functions First, the definition of a pseudorandom function considers two experiments: $\texttt{exp} - \texttt{prf}^f$ and $\texttt{exp} - \texttt{r}$. Let $\mathcal{A}$ be some PPT algorithm that outputs either 1 or 0.

Experiment $\texttt{exp} - \texttt{prf}^{f,\mathcal{A}}$:
Select random s of length κ.
Repeat an arbitrary number of times:
 Receive x_i from $\mathcal{A}$.
 Give $y_i = f_s(x_i) = f(s, x_i)$ to $\mathcal{A}$.
When $\mathcal{A}$ outputs "finished" and
a bit b, output b.

Experiment $\texttt{exp} - \texttt{r}^{\mathcal{A}}$:
Initialize an empty table of values.
Repeat an arbitrary number of times:
 Receive x_i from $\mathcal{A}$.
 Lookup x_i in the table of values
  if it exists return y_i the stored value.
  else randomly select y_i and
  store (x_i, y_i) in the table.
When $\mathcal{A}$ outputs "finished" and
a bit b, output b.

Definition 2. *A family of functions* $f : \{0,1\}^\kappa \times \{0,1\}^n \to \{0,1\}^n$ *is called a pseudorandom function if for all PPT* $\mathcal{A}$ *there exists a negligible function* $\epsilon(n)$ *such that*

$$|\Pr[\mathtt{exp} - \mathtt{prf}^{f,\mathcal{A}} = 1] - \Pr[\mathtt{exp} - \mathtt{r}^{\mathcal{A}} = 1]| < \epsilon(n).$$

Information-theoretic MAC Furthermore, recall the definition of a good information-theoretic MAC function. Consider the following experiment: $\mathtt{Mac} - \mathtt{forge}^{\mathcal{A},\mathsf{Mac}}$:

1. Key $k \leftarrow \mathsf{Gen}()$.
2. The adversary, $\mathcal{A}$ outputs m and is given $t \leftarrow \mathsf{Mac}(k, m)$.
3. The adversary outputs (m', t').
4. The output of the experiment is 1 if $m' = m$ and $\mathsf{Vfy}(k, m', t') = 1$. Otherwise the output is 0.

We say a Mac scheme is secure if no adversary can win this game:

Definition 3. *A scheme* $(\mathsf{Mac}, \mathsf{Vfy})$ *is* ϵ*-unforgeable under chosen message attack if*

$$\forall \mathcal{A}, \Pr_K[\mathtt{Mac} - \mathtt{forge}^{\mathcal{A},\mathsf{Mac}} = 1] < \epsilon.$$

a) **10 pts** Adapt the above definition of unforgeability to consider only computationally bounded adversaries. If necessary, provide a new experiment and definition.

b) **10 pts** Adapt your definition from part a to additionally consider multiple messages. If necessary, provide a new experiment and definition. State what you are assuming about how the adversary receives messages in your definition. That is, provide an informal definition of what you are protecting and a mathematical experiment.

Hint: You may want to refer to the solution to Problem Set 2.

c) **25 pts** Consider the following $(\mathsf{Gen}, \mathsf{Mac}, \mathsf{Vfy})$ scheme for messages:

- Gen : sample random k of length κ.
- $\mathsf{Mac}(k, m) = f(k, m) = t$.
- $\mathsf{Vfy}(k, m, t)$ if $f(k, m) \stackrel{?}{=} t$ output 1, otherwise output 0.

Show if f is a good PRF then $(\mathsf{Gen}, \mathsf{Mac}, \mathsf{Vfy})$ is a good MAC scheme for multiple messages against computational adversaries as defined in part b).

(a) **5 pts** State the contrapositive of the above statement.

(b) **10 pts** Describe an algorithm $\mathcal{A}'$ that is able to distinguish the function f from a truly random function (this algorithm may use an $\mathcal{A}$ breaks security as defined in part b)). Be clear about what $\mathcal{A}'$ may receive as input and how it prepares inputs to $\mathcal{A}$.

(c) **10 pts** Show that $\mathcal{A}'$ distinguishes f from a truly random function.