6.5620 (6.875), Fall 2022

Homework # 3 **Due: October 19 2022, 11:59:59pm ET**

- **Typsetting:** You are encouraged to use LaTeX to typeset your solutions. You can use the following template.
- **Submissions:** Solutions should be submitted to Gradescope.
- **Reference your sources:** If you use material outside the class, please reference your sources (including papers, websites, wikipedia).
- **Acknowledge your collaborators:** Collaboration is permitted and encouraged in small groups of at most three. You must write up your solutions entirely on your own and acknowledge your collaborators.

Problems:

1. *(3 points)* **Encryption implies one-way functions.**

   (a) *(3 points)* Assume that IND-CPA secure public-key encryption exists: that is, assume that $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is a perfectly correct IND-CPA secure public-key encryption scheme, where $\mathsf{Gen}(1^\lambda)$ outputs a key pair (pk, sk), $\mathsf{Enc}(pk, m)$ outputs a ciphertext c, and $\mathsf{Dec}(sk, c)$ outputs a message m. Prove that one-way functions exist. (That is, construct a one-way function which is secure given that $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is perfectly correct and IND-CPA secure.)

   (b) *(**Bonus**; 2 points)* **Warning: hard (and subtle)!** Assume that IND-CPA secure *secret*-key encryption exists: that is, assume that $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is a perfectly correct IND-CPA secure secret-key encryption scheme, where $\mathsf{Gen}(1^\lambda)$ outputs a secret key sk, $\mathsf{Enc}(sk, m)$ outputs a ciphertext c, and $\mathsf{Dec}(sk, c)$ outputs a message m. Prove that one-way functions exist.

   You may assume for this part of the problem that the keyspace and the message space of $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ are the same size.

   Hint. We of course do not believe that we can prove that the existence of the one-time-pad implies the existence of one-way functions! As such, your reduction (from the security of your constructed OWF to the IND-CPA security of the secret-key encryption scheme) will somehow need to use many queries to its encryption oracle, which we assume always encrypts messages with the same key.

2. *(10 points)* **Candidate One-Way Functions.** Alice would really like to construct a simple to state one-way function and decides to try a few options. Let S be a multiset of 0's and 1's, and $\mathsf{maj}(S)$ denote the majority element of S. For example, $\mathsf{maj}(\{0,0,1\}) = 0$ and $\mathsf{maj}(\{1,1,1,0\}) = 1$.

   For a set $I = \{i_1, \ldots, i_\ell\}$ and $x \in \{0,1\}^n$, we denote

   $$\mathsf{maj}_I(x) = \mathsf{maj}(x_{i_1}, \ldots, x_{i_\ell}),$$

   i.e. the majority of the bits of x that are indexed by I.

   (a) *(1 point)* Suppose $x, y, z \leftarrow_R \{0,1\}$ is drawn uniformly. What is the probability that $x = \mathsf{maj}(\{x, y, z\})$?

   (b) *(4 points)* As a first try, Alice decides to construct a candidate function f as follows. Let $I_1, I_2, \ldots, I_{\binom{n}{3}}$ denote all possible subsets of triples of [n] (here, $[n]$ denotes the set $\{1, 2, \ldots, n\}$). Let $\ell = \binom{n}{3}$. Consider the function $f_0 : \{0,1\}^n \to \{0,1\}^{\ell(3\log n+1)}$, where

   $$f_0(x) = (I_1, \mathsf{maj}_{I_1}(x), I_2, \mathsf{maj}_{I_2}(x), \ldots, I_\ell, \mathsf{maj}_{I_\ell}(x)),$$

   i.e. the output of f_0 is the list of triples (each takes $3\log n$ bits to denote) and the majority of the corresponding triple of indices of x.

   Show that f_0 is not a one-way function.

   (c) *(5 points)* Alice realises that giving the majority of *every* triple of indices seems to be leaking too much information. She therefore decides to **reduce** the number of triples given in the output of the candidate one-way function in the hopes of leaking less information about x.

   She now picks $\ell = Cn\log n$, where C is a very large constant, and constructs the following function $f_1 : \{0,1\}^{n+\ell} \to \{0,1\}^{\ell(3\log n+1)}$ given by

   $$f_1(x, I_1, I_2, \ldots, I_\ell) = (I_1, \mathsf{maj}_{I_1}(x), I_2, \mathsf{maj}_{I_2}(x), \ldots, I_\ell, \mathsf{maj}_{I_\ell}(x)).$$

   Intuitively, one can view the function f_1 as mapping x to a list of the majority of ℓ randomly chosen triples of indices of x.

   Show that f_1 is *not* a one-way function, i.e. come up with an efficient p.p.t. algorithm that inverts f_1 with $\frac{1}{\mathrm{poly}(n)}$ probability.

   You might find that proving the following claims might be useful (but not necessary). You can also use the following results without proof for partial credit.

   i. For each index i, show that it appears in $\Omega(\log n)$ triples among $I_1, \ldots, I_\ell$ with probability $1 - \frac{1}{\mathrm{poly}(n)}$.
   ii. Consider the following problem. Suppose you have the following pairs $(a_1, a_2), (a_3, a_4), \ldots, (a_{2t-1}, a_{2t})$ such that $a_i \neq a_j$ for all $i \neq j$. Show that the probability that a uniformly sampled $(x, y) \leftarrow_R [n] \times [n]$ is such that $x \neq a_i$ and $y \neq a_i$ for all i is $1 - O(t/n)$.

iii. Suppose $t = C' \log n$ where C' is some large constant. Show that given uniformly sampled pairs $(a_1, a_2), \ldots, (a_{2t-1}, a_{2t}) \leftarrow_R [n] \times [n]$, the probability that there are at least $\Omega(\log n)$ disjoint pairs is $1 - \frac{1}{\text{poly}(n)}$.

(d) (Optional. We will not grade this.) Show that even if Alice replaces the maj function in f_1 with *any* predicate on three values, the resulting function is still not a one-way function.

You might find union bounds and Chernoff bounds extremely useful for this problem. If you are unfamiliar with these, please refer to the probability recitation handout.

3. *(8 points)* **The power of Decisional Diffie-Hellman.**

Given a cyclic group $G = \langle g \rangle$ of *prime* order q, the DDH assumption says that

$$(g, g^a, g^b, g^{ab}) \stackrel{c}{\approx} (g, g^a, g^b, g^c), \tag{1}$$

where $a, b, c \leftarrow \mathbb{Z}_q$ are uniformly random and independent. By grouping the elements appropriately, we can view this assumption in the following compact form:

$$g^{\begin{pmatrix} 1 & a \\ 1 \cdot b & a \cdot b \end{pmatrix}} \stackrel{c}{\approx} g^{\begin{pmatrix} 1 & a \\ b & c \end{pmatrix}},$$

where g^M (for a matrix M over $\mathbb{Z}_q$) is the matrix over G obtained by raising g to each entry of M. Observe that in the left-hand matrix, the two rows are linearly dependent (over $\mathbb{Z}_q$), while in the right-hand matrix they are very likely not to be.

(a) *(6 points)* Prove that the DDH assumption implies that, for any positive integers $w, h = \text{poly}(n)$,

$$g^{(a_i \cdot b_j)_{i \in [h], j \in [w]}} \stackrel{c}{\approx} g^{(c_{i,j})_{i \in [h], j \in [w]}}, \tag{2}$$

where $a_i, b_j, c_{i,j} \leftarrow \mathbb{Z}_q$ are all uniformly random and independent. In other words, show that any efficient algorithm that distinguishes between the two distributions in (2) can be used to devise an efficient algorithm that distinguishes between the two distributions in (1).

(Note that the left-hand matrix (in the exponent) has rank 1, while the right-hand matrix is very likely to be full-rank!)

(b) *(2 points)* Conclude that, under the DDH assumption, there is a PRG family expanding about $2n \log q$ bits to about $n^2 \log q$ bits. (The output need not literally be made up of bits, though.)

4. *(12 points)* **Square encryption.**

(a) *(2 points)* Let N be an integer, and consider the following subset of $\mathbb{Z}_{N^2}^*$:

$$\mathbb{G}_N := \{\, aN + 1 : a \in \{0, \ldots, N-1\} \,\}.$$

Show that $\mathbb{G}_N$ is a multiplicative subgroup of $\mathbb{Z}_{N^2}^*$ of order N.

(b) *(2 points)* Which elements of $\mathbb{G}_N$ are generators? Why?

(c) *(2 points)* Show, for any generator $g \in \mathbb{G}_N$, that the discrete log problem of finding x such that $g^x = y$ given $y \in \mathbb{G}_N$ can be solved efficiently. (Which step breaks down if $g \in \mathbb{G}_N$ is not a generator?)

(d) *(4 points)* Consider the following public-key encryption scheme.

- $\mathsf{SquareGen}(1^\ell)$ generates two distinct ℓ-bit primes p and q, sets $N = pq$, $d = (p-1)(q-1)$, and outputs $pk = N, sk = d$.
- Let $g := N+1$. $\mathsf{SquareEnc}(pk, m)$ generates $h \leftarrow_R \mathbb{Z}^*_{N^2}$, and outputs $c = g^m h^N \mod \mathbb{Z}^*_{N^2}$.

Give a decryption algorithm $\mathsf{SquareDec}(sk, c)$ to complete this public-key encryption scheme, and show that your decryption algorithm satisfies perfect correctness. You may assume that $m \in \mathbb{Z}_N$, and that p and q are such that $(p-1)(q-1)$ is relatively prime to N.

(e) *(2 points)* Show that $(\mathsf{SquareGen}, \mathsf{SquareEnc}, \mathsf{SquareDec})$ is IND-CPA secure under the following assumption. If N is generated as $\mathsf{SquareGen}$ generates it, u is sampled uniformly at random from $\mathbb{Z}^*_{N^2}$, and v is sampled uniformly at random from the subgroup

$$(\mathbb{Z}^*_{N^2})^N := \{h^N : h \in \mathbb{Z}^*_{N^2}\},$$

then the distribution (N, u) is computationally indistinguishable from the distribution (N, v).

5. *(16 points)* **Pseudoentropy.** In class, we learned about pseudorandomness, the computational analogue of information-theoretic randomness. In this question, we will explore the notion of *pseudoentropy*, the computational analogue of information-theoretic (min-)entropy.

Informally speaking, a distribution over bitstrings has high (information-theoretic) *min-entropy* if the highest probability string in the distribution is still fairly unlikely to occur. This notion captures the idea that a sample from the distribution is difficult for any adversary to predict. Formally, the min-entropy of a random variable X with probability mass function given by $p : \{0,1\}^n \to [0,1]$ is

$$H^\infty(X) := \min_x(-\log(p(x))).$$

We can also define the notion of *pseudoentropy*, in analogy with the notion of pseudorandomness: informally speaking, a distribution has high *pseudo-(min-)entropy* if it is computationally indistinguishable from a distribution with high information-theoretic min-entropy. In this question, we will take the following to be the definition of pseudoentropy:

Definition 1. Let X be a random variable[1] over $\{0,1\}^n$. We say that X has *pseudoentropy* at least $k(n)$, denoted $H^{PE}(X) \geq k(n)$, if there exists a random variable Y with

[1]Implicitly, a family of random variables, one for each n, but we will omit this for notational convenience.

information-theoretic min-entropy at least $k(n)$ such that, for any PPT computable function $f : \{0,1\}^n \to \{0,1\}$,

$$\left| \Pr_{x \leftarrow X}[f(x) = 1] - \Pr_{y \leftarrow Y}[f(y) = 1] \right| \leq \text{negl}(n).$$

Perhaps surprisingly, pseudoentropy does not always behave like information-theoretic entropy translated into the computational setting. In this problem, we will see two ways in which pseudoentropy is different from its information-theoretic counterpart.

(a) *(2 points)* A basic fact in information theory is the *concatenation lemma*, which says that, for any two (potentially correlated) random variables X and Y, the min-entropy of (X, Y) is at least as large as the min-entropy of X. Prove the concatenation lemma. (It's a short proof, but make sure you justify all your steps!)

(b) *(3 points)* Show that, assuming that one-way functions exist, the concatenation lemma does not hold for pseudoentropy. That is, give (families of) random variables X, Y for which $H^{PE}(X, Y) < H^{PE}(X)$, and prove that this is the case.

(c) *(4 points)* Another fact in information theory is that random variables which are *bit-wise unpredictable* have high min-entropy. A bit-wise unpredictable random variable X over $\{0,1\}^n$ is one where, for *most* $i \in [n]$, X_i is difficult to predict given $X_1, \ldots, X_{i-1}$.[2] Formally,

Definition 2. Let X be a random variable over $\{0,1\}^n$. We say that X is *ε–unpredictable* in index i if, for any function $f : \{0,1\}^{i-1} \to \{0,1,\perp\}$[3],

$$\Pr_{x \leftarrow X}[f(x_1, \ldots, x_{i-1}) = x_i] < \frac{1}{2} + \varepsilon.$$

In this problem, for simplicity, we will restrict ourselves to *flat sources.* We will define a *flat source* to be a random variable X such that every string in the distribution of X either occurs with some fixed positive probability p, or with probability zero.

Our goal in this subproblem and the next will be to prove the following theorem, which formalises the idea that, in the information-theoretic world, flat sources with many unpredictable coordinates have high min-entropy.

Lemma 3. *For any constant ε such that $0 < \varepsilon < \frac{1}{2}$, if X (a flat source over $\{0,1\}^n$) is $\frac{\varepsilon}{2}$-unpredictable in $\left(1 - \frac{\varepsilon}{2}\right) n$ of its coordinates, then $H^{\infty}(X) \geq kn$ for some constant $k > 0$.*

Start by proving that the following lemma implies Lemma 3.

[2] If X is unpredictable for *all* $i \in [n]$, then X is close to truly random.

[3] $\perp$ is like an 'I give up' symbol that the predictor function f can output; we allow it purely for the sake of notational clarity later.

Lemma 4. *Let X be a random variable over $\{0,1\}^n$. Let $\mathcal{F} = \{f_1, \ldots, f_n\}$ be any family of predictor functions such that $f_i : \{0,1\}^{i-1} \to \{0,1,\perp\}$. For any x in the support of X, define* $\mathsf{wrong}(\mathcal{F}, x)$ *to be the set of indices $i \in [n]$ such that $f_i(x_1, \ldots, x_{i-1})$ is different from x_i.*
There exists a family $\mathcal{F} = \{f_1, \ldots, f_n\}$ of functions such that, for all $x \in X$,

$$|\mathrm{Supp}(X)| \geq c^{|\mathsf{wrong}(\mathcal{F},x)|},$$

for some constant $c > 1$, where $\mathrm{Supp}(X)$ is the support of X.

(Hint: show the contrapositive, i.e. start with the hypothesis that X has min-entropy less than kn. Try to derive from this a statement of the form

$$\Pr_{i \leftarrow_R [n], x \leftarrow X}[f_i(x_1, \ldots, x_{i-1}) = x_i] \geq 1 - \frac{k}{\log c},$$

and then use a Markov bound or a counting argument to 'peel off' the outer probability and prove the desired conclusion.)

(d) *(4 points)* Now we will prove Lemma 4. For any $y \in \{0,1\}^i$ for some $i \leq n$, define the *continuations function* $C(y)$ to be:

$$C(y) = \text{number of } x \in \mathrm{Supp}(X) \text{ s.t. } x_1, \ldots, x_i = y$$

Define a family of predictor functions $\mathcal{F} = \{f_1, \ldots, f_n\}$ as follows:

$$f_i(x_1, \ldots, x_{i-1}) = \begin{cases} 1 & \frac{C(x_1,\ldots,x_{i-1},1)}{C(x_1,\ldots,x_{i-1})} > \frac{2}{3} \\ 0 & \frac{C(x_1,\ldots,x_{i-1},1)}{C(x_1,\ldots,x_{i-1})} < \frac{1}{3} \\ \perp & \text{else.} \end{cases}$$

In other words, f_i given $x_1, \ldots, x_{i-1}$ calculates the number of continuations of the string $(x_1, \ldots, x_{i-1}, 1)$ vs. the number of continuations of $(x_1, \ldots, x_{i-1}, 0)$ in the support of X, and makes its decision based on that.

Prove Lemma 4 using this family of predictor functions. (Hint: Lemma 4 requires us to prove its conclusion for all $x \in \mathrm{Supp}(X)$, so we should start by fixing an arbitrary x in the support of X. For any i such that $0 \leq i \leq n$, define $N_i := C(x_1, \ldots, x_i)$. Note that $N_0 = C(0) + C(1)$ is exactly $\mathrm{Supp}(X)$. Try to prove that, for every $i \in \mathsf{wrong}(\mathcal{F}, x)$, $N_{i-1} \geq \frac{3}{2} N_i$.)

(e) *(3 points)* We have now established that bit-wise unpredictability implies high min-entropy in the information-theoretic setting. Show that, assuming that one-way permutations exist, the *computational* bit-wise unpredictability of a flat source X in a constant fraction of its coordinates does not imply that X has pseudoentropy $\Omega(n)$. That is, give a flat source X over $\{0,1\}^n$ which is ε-unpredictable for PPT computable functions in $\Omega(n)$ of its coordinates with $\varepsilon = \mathrm{negl}(n)$, but which has pseudoentropy $o(n)$, and prove that this is the case. (You can cite your proof from part (b).)