6.875: Cryptography and Cryptanalysis **September 8, 2021**

Problem Set 1

Instructor: Vinod Vaikuntanathan **TAs:** Lali Devadas and Sacha Servan-Schreiber

Instructions.

- **When:** This problem set is due on **September 22, 2021** before **11pm ET**.
- **How:** You should use LaTeX to type up your solutions (you can use our LaTeX template from the course webpage). Solutions should be uploaded on Gradescope as a single pdf file.
- **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*.
- **Reference your sources:** If you use material from outside the lectures, you must reference your sources (papers, websites, wikipedia, ...).
- **When in doubt, ask questions:** Use Piazza or the TA office hours for questions about the problem set. See the course webpage for the timings.

Problem 1. Perfect vs. Statistical Secrecy

Shannon's Perfect Secrecy. Recall the definition of an encryption scheme satisfying *prefect correctness* and *perfect secrecy*. An encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ with message space $\mathcal{M}$ and ciphertext space $\mathcal{C}$ is said to be:

- **perfectly correct** if for all messages $m \in \mathcal{M}$,

$$\Pr\left[\ k \leftarrow \mathsf{Gen}(1^\lambda) : \mathsf{Dec}(k, \mathsf{Enc}(k, m)) = m\ \right] = 1$$

  That is, the encryption scheme always correctly decrypts an encrypted message under every secret key k.

- **perfectly secret** if for all messages $m_0, m_1 \in \mathcal{M}$, and for all $c \in \mathcal{C}$

$$\Pr\left[\ k \leftarrow \mathsf{Gen}(1^\lambda) : \mathsf{Enc}(k, m_0) = c\ \right] = \Pr\left[\ k \leftarrow \mathsf{Gen}(1^\lambda) : \mathsf{Enc}(k, m_1) = c\ \right],$$

  where the probability is over the randomness of Gen and Enc.

Note that Gen and Enc could be randomized, and Dec is deterministic. It was shown in lecture that, in order to achieve these requirements, the size of the key space has to be at least as large as the message space. In this problem, we will consider variations and relaxations of this definition. Before diving in, we will define the notions of *statistical distance* and *statistical secrecy*.

Statistical Distance. Let X and Y be two random variables over a finite set S with distributions D_X and D_Y, respectively. The **statistical distance** (also called the total variation distance) between X and Y, denoted $\Delta(X, Y)$ is defined as

$$\Delta(X, Y) := \frac{1}{2} \sum_{z \in S} \Big| \Pr[X = z] - \Pr[Y = z] \Big|.$$

Note that $0 \leq \Delta(X, Y) \leq 1$. $\Delta(X, Y) = 1$ precisely when the supports of the distributions D_X and D_Y are disjoint, and $\Delta(X, Y) = 0$ precisely when the two distributions are identical.

Statistical Secrecy. Let λ be a security parameter. An encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ with message space $\mathcal{M}$ and ciphertext space $\mathcal{C}$ is said to be **statistically secret** if for all messages $m_0, m_1 \in \mathcal{M}$, and for all $c \in \mathcal{C}$, there exists a negligible function $\epsilon = \epsilon(\lambda)$ such that

$$\Big| \Pr\big[\ k \leftarrow \mathsf{Gen}(1^\lambda) : \mathsf{Enc}(k, m_0) = c\ \big] - \Pr\big[\ k \leftarrow \mathsf{Gen}(1^\lambda) : \mathsf{Enc}(k, m_1) = c\ \big] \Big| \leq \epsilon,$$

where the probability is over the randomness of Gen and Enc.

(a) Suppose Alice and Bob want to communicate securely without an evil three-letter agency E.V.E. decrypting their communication. However, Alice and Bob live in a world where the MOD operator has been banned by E.V.E., who does not want people using Shannon's one-time pad. That is, using $\text{XOR}(b_1, b_2) = b_1 + b_2 \pmod 2$ is no longer possible, even though computing $\text{ADD}(b_1, b_2) = b_1 + b_2$ *is* possible.

Alice proposes using the following encryption scheme, which she hopes will be secure against even the most powerful computers that E.V.E. has at their disposal.

Alice's One-time Pad without MOD:
Let $\mathbb{Z}^+$ be the set of non-negative integers (i.e., $\{0, 1, 2, \dots\}$) and let λ be a security parameter. Let $\mathcal{M} = \mathcal{K} = \{i \in \mathbb{Z}^+ \mid i \leq 2^\lambda\}$, and define $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ as follows.

- $k \leftarrow \mathsf{Gen}(1^\lambda)$: sample a uniformly random k from $\mathcal{K}$.
- $c \leftarrow \mathsf{Enc}(k, m)$: Output $k + m$.
- $m \leftarrow \mathsf{Dec}(k, c)$: Output $c - k$.

Prove that Alice's scheme satisfies perfect correctness but does not satisfy perfect secrecy. Compute the statistical distance ϵ between encryptions of (any given) m_0 and m_1 under Alice's scheme. Does Alice's scheme satisfy *statistical secrecy*? If not, modify her scheme to make it statistically secret. (Hint: you do not need to modify Enc and Dec.)

(b) Alice and Bob find a classified document claiming that E.V.E. can only read *partially-corrupted* ciphertexts sent over the internet (even though Alice and Bob receive the *uncorrupted* ciphertext). Both Alice and Bob grow tired of having to trek across the MIT campus to generate a new random key for each message they encrypt, and they wonder if they can use E.V.E.'s lack of error-correction to their advantage.

Formally, E.V.E.'s lack of error-correction can be modeled by her receiving ciphertexts $\tilde{c}$ where the i^{th} bit of $\tilde{c}$ is the i^{th} bit of c with probability $1 - p$ and is the opposite bit with probability p. Observe that if $p = 0$, E.V.E. receives the uncorrupted ciphertext, i.e. $\tilde{c} = c$.

Let (Gen, Enc, Dec) be Alice's OTP-without-MOD from part (a) (without your modifications). What is the smallest value of p that would make (Gen, Enc, Dec) *statistically secret*? How about *perfectly secret*?

Problem 2. Pseudorandom Generators (PRG)

Notation. In this problem, $||$ denotes concatenation of strings, and $\bar{x}$ denotes the bit-wise complement of x, e.g., if $x = 01011$ then $\bar{x} = 10100$. A substring of x is denoted $x_{[a:b]}$, where indices a and b are inclusive. We will write $x_{[i]}$ to denote just the i^{th} bit of x.

Let $G_1 : \{0,1\}^\lambda \to \{0,1\}^{\lambda+1}$ and $G_2 : \{0,1\}^\lambda \to \{0,1\}^{\lambda+1}$ be PRGs stretching a λ-bit input into a $(\lambda+1)$-bit output.

1. $G'(x) = G_1(x) \,||\, G_2(x)$
2. $G'(x) = G_2(G_1(x)_{[1:\lambda]}) \,||\, G_1(x)_{[\lambda+1]}$
3. $G'(x) = G_1(x) \oplus G_2(\bar{x})$
4. $G'(x||y) = G_1(x||0^{\lambda/2}) \,||\, y$ where $|x| = |y| = \lambda/2$.

For each of the above constructions of G', prove that G' is a PRG OR provide a counterexample and proof showing why it is not a PRG.

Problem 3. Pseudorandom Permutations (PRP)

After the first few lectures of 6.875, Bob feels confident that he can build pseudorandom functions. He puts together a pseudorandom function family $\mathcal{F} = \{f_k : \{0,1\}^\ell \to \{0,1\}^\ell\}_{k\in\{0,1\}^q}$ for $q = q(\lambda), \ell = \ell(\lambda)$.

He brags about this to Alice, who complains that even given the key k, Bob may not be able to *invert* his pseudorandom function; that is, given k and $f_k(x)$, it is unclear how to recover x. She claims that she can one-up him by transforming his pseudorandom function into one that is also invertible (i.e. a permutation) given the key. She defines a **pseudorandom permutation family** as a family of functions $\mathcal{P} = \{p_s : \{0,1\}^n \to \{0,1\}^n\}_{s\in\{0,1\}^m}$ for $m = m(\lambda), n = n(\lambda)$ such that

- **correctness:** for all $s \in \{0,1\}^m$, $p_s : \{0,1\}^n \to \{0,1\}^n$ is an efficiently computable bijection.
- **security:** for all probabilistic polynomial-time (PPT) distinguishers $\mathcal{A}$, there exists some negligible function $\epsilon = \epsilon(\lambda)$ such that

$$\left|\Pr\left[\; s \xleftarrow{R} \{0,1\}^m : \mathcal{A}^{p_s}(1^\lambda) = 1 \;\right] - \Pr\left[\; p \xleftarrow{R} \mathbb{P}_n : \mathcal{A}^{p}(1^\lambda) = 1 \;\right]\right| \leq \epsilon$$

  where $\mathbb{P}_n$ is the set of all permutations (i.e., bijective functions) on $\{0,1\}^n$.

(a) Alice's first idea is to set $n = 3\ell$ and split the input $x \in \{0,1\}^n$ into three equal parts, which we will call $L_0, M_0, R_0 \in \{0,1\}^\ell$. That is,

$$L_0 := x_{[1:\ell]},\;\; M_0 := x_{[\ell+1:2\ell]},\;\; R_0 := x_{[2\ell+1:n]}.$$

She defines $p_s(\cdot)$ to be:

$p_s(x)$

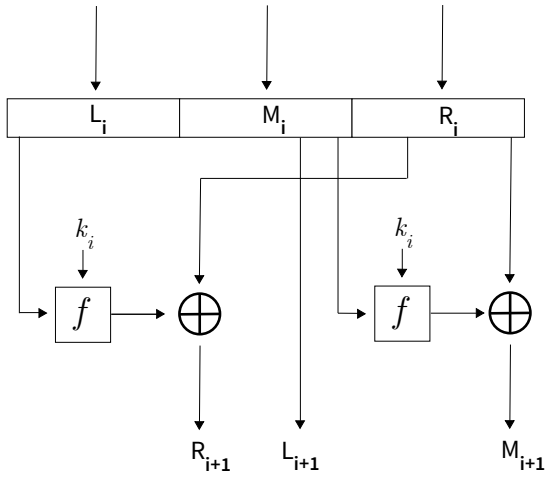
1 : **parse** $x = L_0||M_0||R_0$ // as above
2 : **parse** $s = k$
3 : $L_1 := M_0$
4 : $M_1 := R_0 \oplus f_k(M_0)$
5 : $R_1 := L_0 \oplus f_k(R_0)$
6 : **return** $L_1||M_1||R_1$

Prove that this is a permutation.

(b) Since the above is obviously not pseudorandom (convince yourself!), Alice's next idea is to repeat the procedure iteratively multiple times. She picks a number of "rounds" r and sets $m = rq$ so that each round can have a different key $k_i \in \{0,1\}^q$ for the pseudorandom function (i.e. $s = k_0||\cdots||k_{r-1} \in \{0,1\}^m$). This time she defines $p_s(\cdot)$ to be:

$p_s(x)$

1 : **parse** $x = L_0||M_0||R_0$
2 : **parse** $s = k_0||\cdots||k_{r-1}$
3 : $L_1 := M_0$
4 : $M_1 := R_0 \oplus f_{k_1}(M_0)$
5 : $R_1 := L_0 \oplus f_{k_1}(R_0)$
6 : **foreach** $i = 1, \ldots, r-1$ **do**
7 : $\quad L_{i+1} := M_i$
8 : $\quad M_{i+1} := R_i \oplus f_{k_i}(M_i)$
9 : $\quad R_{i+1} := L_i \oplus f_{k_i}(R_i)$
10 : **endfor**
11 : **return** $L_r||M_r||R_r$



Prove that p_s is still a permutation for any $r \geq 1$.

(c) Alice gets excited and claims that the above is a pseudorandom permutation for $r = 3$, but Bob thinks it still isn't pseudorandom.

Who is correct? Prove your answer. (If you take Bob's side, you need to show an attack on the permutation family; if you take Alice's side, you need to show a proof of security that it satisfies the definition above.)

Problem 4. Pseudorandom Functions (PRF)

Alice and Bob have a different problem. They want to use a NIST-approved pseudorandom function family defined as

$$\mathcal{F} = \{f_k : \{0,1\}^\ell \to \{0,1\}^\ell\}_{k \in \{0,1\}^q}$$

where $q = q(\lambda)$, just as before. But they discover that what they need is slightly different, namely, a pseudorandom function that operates on inputs of length 3ℓ.

(a) Bob has a bright idea: he suggests to XOR the function outputs! That is, he defines a function family $\mathcal{G} = \{g_k : \{0,1\}^{3\ell} \to \{0,1\}^{\ell}\}_{k \in \{0,1\}^{3q}}$ by the functions

$$g_{k_0||k_1||k_2}(x_0||x_1||x_2) = f_{k_0}(x_0) \oplus f_{k_1}(x_1) \oplus f_{k_2}(x_2) \ .$$

Is $\mathcal{G}$ a pseudorandom function family? Provide an attack or a proof of that it satisfies the definition of a PRF.

(b) It is now Alice's turn. She proceeds to define the function family $\mathcal{H} = \{h_k : \{0,1\}^{3\ell} \to \{0,1\}^{\ell}\}_{k \in \{0,1\}^{\ell}}$ by the following iterative process:

$$\begin{aligned} y_0 &= f_k(x_0); \\ y_1 &= f_{y_0}(x_1); \\ y_2 &= f_{y_1}(x_2). \end{aligned}$$

Finally, she sets $h_k(x_0||x_1||x_2) = y_2$.

Is $\mathcal{H}$ a pseudorandom function family? Provide an attack or a proof of that it satisfies the definition of a PRF.

(c) Bob thinks Alice's construction is *too iterative*. He optimizes her construction and defines the function family $\mathcal{H}' = \{h'_k : \{0,1\}^{3\ell} \to \{0,1\}^{\ell}\}_{k \in \{0,1\}^{\ell}}$ as follows:

$$\begin{aligned} y_0 = f_k(x_0); \ \ y_1 = f_k(x_1); \\ y_2 = f_{y_0 \oplus y_1}(x_2). \end{aligned}$$

He sets $h'_k(x_0||x_1||x_2) = y_2$.

Is $\mathcal{H}'$ a pseudorandom function family? Provide an attack or a proof of that it satisfies the definition of a PRF.