CS6222 Grad Cryptography, Homework 1

Response by: Your Name, (Computing ID)

Total points: 40 awarded maximum. 50 available. Points are noted after each problem.

Instructions. For each problem, typeset your solution in the `answer` environment, and if there are sub-problems, mark them clearly. Feel free to use as much space as you require, and be sure to update your name and computing ID above, and the acknowledgements box at the end.

Policies. In short, you are encouraged to think about the problems on your own, and then discuss them and work toward solutions with your classmates. You must write and submit your own solutions. You may also read any published material that helps you come to an understanding of the problems, but you must acknowledge and/or reference any discussion or published material, with the exception of lecture notes, in-class and in-office-hours discussions, textbook sections we have covered, and basic LaTeX help or dictionary lookups. It is a violation of the honor code if any of the following occur:

- You copy text directly from any source.
- You use any material or discussion without acknowledgment or citation, excluding the above special cases.
- You are unable to explain your work orally.

See https://jackdoerner.net/teaching/2026/Fall/CS6222/#policies for more details.

This homework recaps a few preliminary concepts that we will use in later lectures, and asks a few simple questions about perfect security.

Notation. For any $n \in \mathbb{N}$, let $[n] = \{1, ..., n\}$. For any finite set $\mathcal{S}$, let $|\mathcal{S}|$ be the cardinality of $\mathcal{S}$.

Definition 1 (Statistical Distance)**.** *Let X and Y be probability distributions over the finite domain $\mathcal{U}$. Their* statistical distance *(also known as statistical* difference *or* total variation distance*) can be defined in two equivalent ways*

$$\mathrm{SD}(X, Y) = \max_{\mathcal{E} \subseteq \mathcal{U}} |\Pr[x \in \mathcal{E} : x \leftarrow X] - \Pr[y \in \mathcal{E} : y \leftarrow Y]| \tag{1}$$

$$\mathrm{SD}(X, Y) = \frac{1}{2} \sum_{e \in \mathcal{U}} |\Pr[x = e : x \leftarrow X] - \Pr[y = e : y \leftarrow Y]| \tag{2}$$

Remark: *We might abuse notation by applying statistical distance to random variables directly, rather than their distributions.*

Remark: *You may use whichever definition of statistical distance suits you.*

Problem 1 (Statistically Close Distributions, 4pts)**.** Let $\mathcal{X} = \{X_1, X_2, \ldots\}$ and $\mathcal{Y} = \{Y_1, Y_2, \ldots\}$ be two *ensembles* of distributions.[1] Suppose that there exists a function $\varepsilon(n)$ such that for all $n \in \mathbb{N}$, $\mathrm{SD}(X_n, Y_n) \leq \varepsilon(n)$. Prove that for any deterministic algorithm D, for all $n \in \mathbb{N}$,

$$|\Pr[D(t) = 1 : t \leftarrow X_n] - \Pr[D(t) = 1 : t \leftarrow Y_n]| \leq \varepsilon(n).$$

[1] That is, each of them is a set of distributions indexed by the natural numbers.

Remark: We say that $\mathcal{X}$ and $\mathcal{Y}$ are *statistically close* if and only if ε is a negligible function. This proves that statistically close ensembles are always indistinguishable.

Problem 2 (Some Call it Δ, 4pts)**.** Prove that statistical distance obeys the triangle inequality. That is, prove that for any distributions X, Y, Z over some domain $\mathcal{U}$ it holds that

$$\mathrm{SD}(X, Z) \leq \mathrm{SD}(X, Y) + \mathrm{SD}(Y, Z)$$

Problem 3 (Statistically Far Distributions, 4pts)**.** Let U_n be the uniform distribution over $\{0,1\}^n$ and let $G : \{0,1\}^n \to \{0,1\}^{n+1}$ be any deterministic function[2] on n-bit inputs. Show that

$$\mathrm{SD}(U_{n+1}, G(U_n)) \geq \frac{1}{2}.$$

Problem 4 (Equivalence of Definitions, 6pts each)**.** Let $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ be an encryption scheme on message-spaces $\{\mathcal{M}_n\}_{n\in\mathbb{N}}$ and ciphertext-spaces $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$. In class, we established that such a scheme is *perfectly secret* if for every $n \in \mathbb{N}$, every $m_0, m_1 \in \mathcal{M}_n$, and every $c \in \mathcal{C}_n$, we have

$$\Pr\left[\mathsf{Enc}_k(m_0) = c : k \leftarrow \mathsf{Gen}(1^n)\right] = \Pr\left[\mathsf{Enc}_k(m_1) = c : k \leftarrow \mathsf{Gen}(1^n)\right] \tag{3}$$

We also established that it is *perfectly indistinguishable* if for every $n \in \mathbb{N}$ and every two-part algorithm $D = (D_1, D_2)$,[3] we have

$$\Pr\left[D_2(c, s) = b : (m_0, m_1, s) \leftarrow D_1(1^n), b \leftarrow \{0,1\}, k \leftarrow \mathsf{Gen}(1^n), c \leftarrow \mathsf{Enc}_k(m_b)\right] = \frac{1}{2} \tag{4}$$

(a) Prove that $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is perfectly indistinguishable if it is perfectly secret.

(b) Prove that $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ is perfectly indistinguishable *only* if it is perfectly secret.

Hint: For part (b), consider that an encryption scheme is *not* perfectly indistinguishable if you can construct an algorithm D that violates Equation 4.

Problem 5 (Deterministic WLOG, 3pts each)**.** This question explores the role of randomness in the context of perfect security. Recall that we define *randomized* algorithms to have access to a string of *uniform* bits.

(a) Show that for any encryption scheme satisfying Equation 3, regardless of the distribution from which Gen samples k, there is another scheme (possibly with a different keyspace $\mathcal{K}'$) that samples k from $\mathcal{K}'$ *uniformly*, and produces the same distribution of ciphertexts for any given message as the original scheme did. You may assume that there is a concrete polynomial upper-bounding the runtime of the original Gen.

(b) Show that for any randomized encryption scheme satisfying Equation 3, there is another scheme (possibly with a different keyspace $\mathcal{K}'$) that has a deterministic encryption procedure, and produces the same distribution of ciphertexts for any given message as the original scheme did. You may assume that there are concrete polynomials upper-bounding the runtimes of the original Gen and Enc.

[2] The function need not be polynomial-time.

[3] The algorithm need not be polynomial-time, but you can assume that it always outputs $m_0, m_1 \in \mathcal{M}_n$.

(c) Prove that in Equation 4 it is sufficient to quantify over deterministic distinguishers. In other words, prove that if an encryption scheme satisfies Equation 4 with respect to deterministic distinguishers only, then it also satisfies the equation with respect to randomized ones. You may assume that every D terminates with output in a finite amount of time. Specifically, assume that for every D there is some finite function that upper bounds the runtime of D, relative to the length of its input.

Problem 6 (Slightly-Imperfect Secrecy, 5pts)**.** Imagine that we weakened the definition of perfect secrecy just a little bit. In particular, suppose we define the property of "ε-imperfect secrecy" to mean that for all $m_1, m_2 \in \mathcal{M}$ and every $c \in \mathcal{C}$,

$$|\Pr[\mathsf{Enc}_k(m_1) = c : k \leftarrow \mathsf{Gen}] - \Pr[\mathsf{Enc}_k(m_2) = c : k \leftarrow \mathsf{Gen}]| \leq \varepsilon$$

for some very small ε. Construct an encryption scheme and prove that it has the following properties:

- Perfect correctness.
- "ε-imperfect secrecy" with $\varepsilon = 2^{-1000}$.
- Ciphertexts completely reveal the plaintext with probability 1. Your scheme should be so broken that even a child could recover the plaintext from the ciphertext without using any computational tools.

Problem 7 (Two-Time Perfect Encryption, 6pts each)**.** In class, we explored what might happen if we use one-time pad to encrypt multiple messages under the same key, and observed that our definition of perfect secrecy considers only a single encrypted message. Consider a seemingly-natural definition of security for *two* messages that are encrypted under the same key: For all $m_1, m_2, m_1', m_2' \in \mathcal{M}$ and all $c_1, c_2 \in \mathcal{C}$,

$$\Pr[\mathsf{Enc}_k(m_1) = c_1 \wedge \mathsf{Enc}_k(m_2) = c_2 : k \leftarrow \mathsf{Gen}] = \Pr\left[\mathsf{Enc}_k(m_1') = c_1 \wedge \mathsf{Enc}_k(m_2') = c_2 : k \leftarrow \mathsf{Gen}\right]$$

(a) Show that no perfectly-correct encryption scheme can achieve the above definition. Your proof should hold even if the encryption algorithm is randomized.

(b) To overcome the limitation we have just proved, let us modify our encryption and decryption procedures so that they both take the *index* of the message to be encrypted. That is, for $i \in \{1, 2\}$, the i^{th} message m_i is encrypted by computing $\mathsf{Enc}_k(m_i, i)$ and the resulting ciphertext is decrypted by computing $\mathsf{Dec}_k(c_i, i)$. Next, suppose that we modify our definition of two-time perfect encryption to insist that for all $m_1, m_2, m_1', m_2' \in \mathcal{M}$ and all $c_1, c_2 \in \mathcal{C}$,

$$\begin{aligned}&\Pr[\mathsf{Enc}_k(m_1, 1) = c_1 \wedge \mathsf{Enc}_k(m_2, 2) = c_2 : k \leftarrow \mathsf{Gen}] \\ &= \Pr\left[\mathsf{Enc}_k(m_1', 1) = c_1 \wedge \mathsf{Enc}_k(m_2', 2) = c_2 : k \leftarrow \mathsf{Gen}\right].\end{aligned}$$

Construct a simple scheme that meets this notion of security. For an optional extra challenge, can you do this with $|\mathcal{K}| \leq 2 \cdot |\mathcal{M}|$?

Hint: For part (b), your scheme need not be efficient.

Acknowledgments

In this box, you should acknowledge your collaborators and the resources you used, if any. For example:

Problem 1: I discussed this problem with Alice and Bob. In addition, I asked Carol for help understanding Conditional Probability, but we did not discuss the problem further.

Problem 2: I asked ChatGPT "What is a turing machine?", and it gave me the following transcript: https://chatgpt.com/share/68b4dba7-e19c-8013-a137-e8db901493b7.

Problem 3: It helped me to read the proof of [Vad12, Theorem X, Page Y].

Instructor's Acknowledgments

Problems in this homework have been borrowed from or inspired by a number of sources. Citations can be provided on request, after the homework is completed.

References

[Vad12] Salil P. Vadhan. *Pseudorandomness*. 2012. https://people.seas.harvard.edu/~salil/pseudorandomness/pseudorandomness-published-Dec12.pdf.