Fundamentals of Cryptography (Fall 2018)
Homework Set 1
By: Guang Yang

1. Due October 15th, 2018.
2. Assignment Submission.
   - Please sent your PDF copy to **1701213987@pku.edu.cn** and cc to **yangguang01@ict.ac.cn**.
   - The file should be named as "hw*-studentnumber-name.pdf" where the "*"stand for the homework number.
   - Or hand in the hard copy to TA.
3. You can give your answer both in Chinese or English.
4. **Important:** Plagiarism is strictly prohibited. Discussion is allowed, but remember:
   - no answer should be written during the discussion;
   - acknowledge everyone involved in the discussion.

Problem 1 (20 + 10 points) [Statistical distance.] Part A. Prove the equivalence of the two definitions of statistical distance, where X and Y are random variables with a finite range S:

$$\Delta(X,Y) = \frac{1}{2}\sum_{s\in S}\big|\Pr[X=s]-\Pr[Y=s]\big|$$

and

$$\Delta(X,Y) = \max_{T\subseteq S}\big(\Pr[X\in T]-\Pr[Y\in T]\big).$$

Part B. Let $D : S \to \{0,1\}$ be a "distinguisher". Assume for simplicity that D is *deterministic*. Prove that

$$\Pr[D(X)=1]-\Pr[D(Y)=1] \le \Delta(X,Y)$$

where "$D(X)=1$" means that D outputs 1 after being given a single sample of X, *etc.*, where X and Y are any two probability distributions[1] over S.
Part C. (*bonus*) Prove that

$$\Delta(X,Y) = \max_{D}\left(\Pr[D(X)=1]-\Pr[D(Y)=1]\right)$$

where the maximum is taken over all *probabilistic* distinguisher $D : S \to \{0,1\}$. You can think that for every $s \in S$, there exists a probability $p_s := \Pr[D(s)=1]$, where $p_s \in [0,1]$. Here we don't care about the running time of D, and such algorithm is sometimes called *information-theoretic*. (*Hint: You can first prove that the optimal D satisfies $p_s \in \{0,1\}$ for every $s \in S$.*)

[1] A *probability distribution* over S is, technically, a function $f : S \to [0,1]$ such that $\sum_{s\in S} f(s) = 1$. However if $X : \Omega \to S$ is a random variable defined over a probability space (Ω,μ) where S is the range of X, then X is sometimes also called a "probability distribution" over S because one can think of X as being specified by the function $f(s) = \Pr[X=s]$ where $\sum_{s\in S} f(s) = 1$ by definition. Note, however, that knowing just the function $f(s) = \Pr[X=s]$ does not tell you everything about X, because as Parts A and B make clear, you cannot compute $\Pr[X \neq Y]$ with only $\Pr[X=s]$ and $\Pr[Y=s]$ for each $s \in S$, where $Y : \Omega \to S$ is some other random variable over (Ω,μ). Note also that $\Pr[X \neq Y]$ *makes no sense* if X and Y are random variables defined over *different* probability spaces, even if X and Y both have range S.

Problem 2 (20 points) Prove the claim in the last slide of Lecture 2 in your own words:
For every symmetric-key encryption scheme $\Pi = (\mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ with message space $\mathcal{M}$ and key space $\mathcal{K}$, there exist $m_0, m_1 \in \mathcal{M}$ such that

$$\Delta(\mathrm{Enc}(K, m_0), \mathrm{Enc}(K, m_1)) \geq 1 - \frac{|\mathcal{K}|}{|\mathcal{M}|}.$$

Problem 3 (20 points) A *secret-sharing scheme* is a way to break a "secret" (say an n-bit value) into two pieces, or "shares" (traditionally called the "left share" and "right share" of the secret) such that (i) the left share doesnt reveal anything about the secret, (ii) the right share doesnt reveal anything about the secret, and (iii) the secret can be reconstructed from the left share and the right share together.

For example, to share a secret $x \in \{0,1\}^n$, select a uniformly random value $y \in \{0,1\}^n$ and let $y \oplus x$ be the left share and y be the right share. Obviously you can reconstruct the secret x from both shares, and you can tell nothing about x if you only know one of the two shares.

You goal: to propose a *formal definition* of a "perfectly secure" secret-sharing scheme. You have to invent the definitions in two separate parts: first, what is a "secret-sharing scheme" (formally); second, what is "perfectly secure". There are two different definitions.

For example, the "first part" and "second part" of a symmetric-key encryption scheme can be defined as follows:

First part. A *symmetric-key encryption scheme* is a 6-tuple $(\mathcal{K}, \mathcal{M}, \mathcal{C}, \mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ where $\mathcal{K}, \mathcal{M}$ and $\mathcal{C}$ are finite sets and where

- $\mathrm{Gen} : \emptyset \to \mathcal{K}$ is a randomized algorithm;
- $\mathrm{Enc} : \mathcal{K} \times \mathcal{M} \to \mathcal{C}$ is a randomized algorithm;
- $\mathrm{Dec} : \mathcal{K} \times \mathcal{C} \to \mathcal{M}$ is a deterministic algorithm.

Moreover, we require that $\mathrm{Dec}(k, \mathrm{Enc}(k, m)) = m$ for every $(k, m) \in \mathcal{K} \times \mathcal{M}$.

Second part. Let $\Pi = (\mathcal{K}, \mathcal{M}, \mathcal{C}, \mathrm{Gen}, \mathrm{Enc}, \mathrm{Dec})$ be a symmetric-key encryption scheme. Let K be the distribution over $\mathcal{K}$ induced by Gen (so K is a random variable of range $\mathcal{K}$). We say that Π is *perfectly secure* if for every two messages $m_0, m_1 \in \mathcal{M}$,

$$\Delta(\mathrm{Enc}(K, m_0), \mathrm{Enc}(K, m_1)) = 0$$

P.S. The "first part" is sometimes called the *syntax* of the scheme (encryption, secret-sharing, etc.), while the "second part" is sometimes called the *semantics* of the scheme.

Problem 4-7 (30+10 points) Do exercises 1.3, 2.3, 2.11 part 1: prove perfect secrecy for a constant $t > 0$ (rather than $t \geq 1$). Part 2 of 2.11 (the lower bound on the size of $\mathcal{K}$) is *bonus*. Note: in 2.11 you should consider a general message space, e.g. $\mathcal{M} = \{0,1\}^n$ or $\mathcal{M} = \{1, 2, \ldots, n\}$, rather than a trivialized case when $\mathcal{M} = \{0,1\}$ or $|\mathcal{M}| = O(1)$. Thus the lower bound should be in terms of both t *and* $|\mathcal{M}|$.