COS 433 — Cryptography — Homework 1.

Boaz Barak

Total of 125 points. Due February 10, 2010.
(Email or hand to Sushant by the beginning of class on Wednesday.)

Important note: In all the exercises where you are asked to prove something you need to give a *well written* and *fully rigorous* proof. This does not mean the proofs have to be overly formal or long — a two-line proof is often enough as long as it does not contain any logical gaps. If a proof is made up of several steps, consider encapsulating each step as a separate claim or lemma.

I prefer you type up your solutions using LaTeX. To make this easier, the LaTeX source of the exercises are available on the course's website.

Exercise 0 (10 points)**.** Send email to Boaz (`boaz@cs.princeton.edu`) with subject `COS433 student` containing **(1)** a couple of sentences about yourself, your background, and what you hope to learn in this course and **(2)** your level of comfort with the following mathematical concepts: mathematical proofs, elementary probability theory, big-Oh notation and analysis of algorithms, Turing machines and NP-completeness. Please also describe any courses you've taken covering these topics.

Exercise 1 (20 points)**.** In the following exercise X, Y denote finite random variables. That is, there are finite sets of real numbers $\mathcal{X}, \mathcal{Y}$ such that $\Pr[X = x] = 0$ and $\Pr[Y = y] = 0$ for every $x \notin \mathcal{X}$ and $y \notin \mathcal{Y}$. We denote by $\mathbb{E}[X]$ the expectation of X (i.e., $\sum_{x\in\mathcal{X}} x \Pr[X = x]$), and by $Var[X]$ the variance of X (i.e., $\mathbb{E}[(X - \mu)^2]$ where $\mu = \mathbb{E}[X]$). The standard deviation of X is defined to be $\sqrt{Var[X]}$.

1. Prove that $Var[X]$ is always non-negative.
2. Prove that $Var[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
3. Prove that always $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$.
4. Give an example for a random variable X such that $\mathbb{E}[X^2] \neq \mathbb{E}[X]^2$.
5. Give an example for a random variable X such that its standard deviation is *not equal* to $\mathbb{E}[|X - \mathbb{E}[X]|]$.
6. Give an example for two random variables X, Y such that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
7. Give an example for two random variables X, Y such that $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$.
8. Prove that if X and Y are independent random variables (i.e., for every $x \in \mathcal{X}, y \in \mathcal{Y}$, $\Pr[X = x \wedge Y = y] = \Pr[X = x]\Pr[Y = Y]$) then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ and $Var[X + Y] = Var[X] + Var[Y]$.

Exercise 2 (20 points)**.** Recall that two distributions X and Y that range over some set S are *identical* if for every s in S, $\Pr[X = s] = \Pr[Y = s]$. Below n is some integer $n \geq 3$. (You can get partial credit for solving the questions below for the special case that $n = 3$ and z (in Question 2) is the string 111.)

1. Let $X_1, ..., X_n$ be random variables where $X_i \in \{0, 1\}$ chosen such that each X_i is chosen to equal 0 with probability 1/2 and equal 1 with probability 1/2, and all of the X_i's are independent. Let $Y_1, ..., Y_n$ be random variables where $Y_i \in \{0, 1\}$ chosen as follows: first an n bit 0/1 string y is chosen uniformly at random from the set $\{0, 1\}^n$ of all possible n-bit 0/1 strings, and then Y_i is set to be the i^{th} coordinate of y. Prove that the distributions $(X_1, ..., X_n)$ and $(Y_1, ..., Y_n)$ are identical.

2. Let z be a fixed string in $\{0, 1\}^n$, and let $Z_1, ..., Z_n$ be random variables chosen as follows: first a string $w \in \{0, 1\}^n$ is chosen uniformly from $\{0, 1\}^n$, and then Z_i is set to $z_i \oplus w_i$, where $\oplus$ is the XOR operation (i.e., $0 \oplus 1 = 1 \oplus 0 = 1$ and $0 \oplus 0 = 1 \oplus 1 = 0$). Prove that the distribution $(Z_1, ..., Z_n)$ is identically distributed to $(X_1, ..., X_n)$ and $(Y_1, ..., Y_n)$ above.

3. Let $W_1, ..., W_n$ be random variables where $W_i \in \{0, 1\}$ chosen as follows: first a string w is chosen uniformly at random from the set of all n-bit 0/1 strings satisfying $w_1 \oplus w_2 \oplus \cdots \oplus w_n = 0$, and then W_i is set to be w_i. **(a)** Prove that W_1 and W_2 are independent. **(b)** Prove or disprove that the random variables $W_1, ..., W_n$ mutually independent.

Exercise 3 (25 points)**.** Show formally that the following schemes do *not* satisfy the definition of perfect security given in class (if it's more convenient you can use Definitions 2.1 or 2.4 from the Katz-Lindell book instead). (Below we use $\mathbb{Z}_n$ to denote the set of numbers $\{0, \ldots, n-1\}$ and identify the letters of the English alphabet with $\mathbb{Z}_{26}$ in the obvious way.)

1. *(Caesar cipher)* Key: a random $k \leftarrow_{\text{R}} \mathbb{Z}_{26}$. Encrypt a length-2 string $x \in \mathbb{Z}_{26}^2$ to the pair $\langle x_1 + k \pmod{26}, x_2 + k \pmod{26} \rangle$

2. *("Two-time pad")* Key: $k \leftarrow_{\text{R}} \{0, 1\}^n$. Encrypt $x \in \{0, 1\}^{2n}$ by $x_{1..n} \oplus k$, $x_{n+1..2n} \oplus k$, where $\oplus$ denotes bitwise XOR.

3. *(Substitution cipher)* Key: a random permutation $\pi : \mathbb{Z}_{26} \to \mathbb{Z}_{26}$. Encrypt $x \in \mathbb{Z}_{26}^2$ by $\pi(x_1), \pi(x_2)$.

Exercise 4 (25 points)**.** Give examples (with proofs) for

1. A scheme such that it is possible to efficiently recover 90% of the bits of the key given the ciphertext, and yet it is still perfectly secure. Do you think there is a security issue in using such a scheme in practice?

2. An encryption scheme that is *insecure* but yet it provably hides the first 20% bits of the key. That is, if the key is of length n then the probability that a computationally unbounded adversary guesses the first $n/5$ bits of the key is at most $2^{-n/5}$.

You can use the results proven in class and above. Also the examples need not be natural schemes but can be "contrived" schemes specifically tailored to obtain a counter-example.

Exercise 5 (Bonus 25 points)**.** In class we saw that any perfectly (and even imperfectly) secure private key encryption scheme needs to use a key as large as the message. But we actually made an implicit subtle assumption: that the encryption process is *deterministic*. In a *probabilistic encryption scheme*, the encryption function E may be probabilistic: that is, given a message x and a key k, the value $\mathsf{E}_k(x)$ is not fixed but is distributed according to some distribution $Y_{x,k}$. Of course, because the decryption function is only given the key k and not the internal randomness used by E, we need to require that $\mathsf{D}_k(y) = x$ for *every* y in the support of $Y_{k,x}$ (i.e., $\mathsf{D}_k(y) = x$ for every y such that $\Pr[\mathsf{E}_k(x) = y] > 0$).

Prove that even a probabilistic encryption scheme cannot have key that's significantly shorter than the message. That is, show that for every probabilistic encryption scheme (D, E) using n-length keys and $n+10$-length messages, there exist two messages $x, x' \in \{0,1\}^{n+10}$ such that the distributions $\mathsf{E}_{U_n}(x)$ and $\mathsf{E}_{U_n}(x')$ are of statistical distance at least $1/10$. See footnote for hint[1]

[1] **Hint:** Define $\mathcal{D}$ to be the following distribution over $\{0,1\}^{n+10}$: choose y at random from $\mathsf{E}_{U_n}(0^{n+5})$, choose k at random in $\{0,1\}^n$, and let $x = \mathsf{D}_k(y)$. Prove that if (E, D) is $1/10$-statistically indistinguishable then for every $x \in \{0,1\}^{n+10}$, $\Pr[\mathcal{D} = x] \geq 2^{-n-1}$. Derive from this a contradiction.