COM S 6830 – Cryptography Due Feb 3, 2016

Homework 1

Instructor: Rafael Pass *TA: TBD*

You may collaborate with other students on the homework but you must submit your own individually written solution and you must identify your collaborators. If you make use of any other external source, you must acknowledge it. You are not allowed to submit a problem solution which you cannot explain orally to the course staff.

Problem 1. *Expectation and Variance*

Let X and Y be two independent random variables. Prove the following facts.

(a) $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$

(b) $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$

Give examples when X and Y are not independent and equalities (a) and (b) do not hold.

Problem 2. *Pairwise Independence*

Let $r_1, r_2, \ldots, r_k$ be n-bit strings picked uniformly at random. For any subset S of $\{1, 2, \ldots, k\}$, define a random variable $z_S = \oplus_{i \in S} r_i$. Prove that the set of random variables $\{z_S \mid S \subseteq \{1, 2, \ldots, k\}\}$ are pairwise independent.

Problem 3. *Sum of Pairwise Independent Variables*

Let $X_1, X_2, \ldots, X_n$ be random variables that are pairwise independent. Further, for all i, let $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2$

(a) Show that,

$$\Pr\left[\left|\frac{\sum X_i}{n} - \mu\right| \geq \epsilon\right] \leq \frac{\sigma^2}{n\epsilon^2}$$

Note that this is a Chernoff like bound when the random variables are only pairwise independent. (Hint: Use Chebyshev's inequality)

(b) Suppose further that the random variables assume only the values 1 and -1. Show that the inequality can be simplified to,

$$\Pr\left[\left|\frac{\sum X_i}{n} - \mu\right| \geq \epsilon\right] \leq \frac{1-\mu^2}{n\epsilon^2}$$

Problem 4. *Secure Encryption*

The cryptographers at secure-encryption.com have come up with an encryption scheme using secret keys. This scheme has a known vulnerability that an attacker, given a random message $m \in \mathcal{M}$ and its encryption, can guess for *any* i, the i^{th} bit of the secret key with probability $\frac{1}{2} + \epsilon$.

Let us suppose the attacker gains access to k such messages $m_1, m_2, \ldots, m_k$ with its encryption.

(a) If the messages $m_1, \ldots, m_k$ are known to be independent, show that the attacker can find any bit of the secret key with very high probability (Hint: Use Chernoff bound). Using the union bound, find a lower bound on the probability of the attacker guessing the entire secret key, if the key is made up of n bits.

(b) Find the number of messages required to guess the entire key with 99% probability when $\epsilon = 0.0001$ and $n = 1024$.

(c) In real life, it is hard for the messages to be independent. Repeat parts (a) and (b) if the messages are known to be only pairwise independent (Hint: Use Problem 3).

Problem 5. *Book Codes*

A Book-code is like the One-time pad, except that instead choosing a key uniformly at random, the key is drawn from the same distribution as the message. For example, if the messages are texts in English dictionary, so are the keys.

1. Show that breaking the one-time pad, when the same key is reused twice (this is sometimes called the "two-time pad" is no harder than breaking the Book-code. That is, assume there exists an attacker A that "breaks" the Book-code when both the message and the key come from the probability distribution D; by "breaking" the code we mean that A is able to fully recover the key used to encrypt the message (and can thus also recover the message). Show how to construct an attacker B, which uses A, to break the one-time pad when used to encrypt two messages coming from the distribution D, with the same uniformly random key.

2. Prove the converse to the above question. That is, if you have an attacker B that can break the two-time pad when messages come from the distribution D, then show how to use B to construct an attacker A

that breaks the Book-code when both the message and the key come from D. Together with the first part, this shows that breaking the two time pad is equivalent to breaking the Book-code.

Problem 6. *Breaking the Nazi Code*

In this problem you will be asked to break the Geheimschreiber (a.k.a. the G-writer, or the Sturgeon). (We thank Johan Hastad for the following description of the Geheimschreiber.) The Geheimschreiber had 10 wheels of lengths, 47,53,59,61,64,65,67,69,71,and 73 respectively. Each position on each wheel contained a bit. The plaintext and crypto-text alphabets were given by the string

2T3O4HNM5LRGIPCVEZDBSYFXAWJ6UQK7

in that 2 corresponds to the integer 0, T to 1 etc.until 7 the corresponds to 31.

Each day a fixed permutation of the wheels was chosen e.g. it could be that the wheel of length 71 was put in position one, the wheel of length 59 was set in position two etc. This remained fixed for the day. Each wheel was then rotated to a starting position that was different for each message. We here assume that the starting position was agreed between the sender and receiver. In some situations some of the positions were transmitted as part of the message but this is not the case for our messages. Encryption was performed as follows.

- Read one bit from each of the ten wheels getting bits b_0, b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8 and b_9.
- Use the given alphabet to convert the clear-text to a number between 0 and 31 and let c_0, c_1, c_2 c_3 and c_4 be the bits of this written in binary.
- For $i = 0 \ldots 4$, $c_i = c_i \oplus b_i$ (take xor)
- If b_5 is 1 interchange c_0 and c_4.
- If b_6 is 1 interchange c_0 and c_1.
- If b_7 is 1 interchange c_1 and c_2.
- If b_8 is 1 interchange c_2 and c_3.
- If b_9 is 1 interchange c_3 and c_4.

- Output the character written as $c_0 \dots c_4$ in binary

Before the next character was encrypted each wheel is stepped one step.

1. Implement the encryption and decryption algorithm. Check your implementation with the plaintext, wheel configuration and ciphertext given on the course web page.

2. On the course homepage you will find a plaintext and corresponding ciphertext under some wheel configuration. You are given that the order of the wheels is 47,53,59,61,64,65,67,69,71,and 73 (i.e. b_0 comes from the 47-size wheel). Recover the bits on all wheels (you may assume without loss of generality that the offset is 0 for all wheels).

3. On the course homepage you will find another plaintext and corresponding ciphertext under some wheel configuration. For this part of the problem, you no longer know the wheel order. Recover the bits on all the wheels and additionally the order of the wheels (again, you may assume without loss of generality that the offset is 0 for all wheels).

4. **(Bonus)** You have a number of intercepted G-writer ciphertexts in the directory called "bonus"; decode them! Hint: The plaintext follows the same structure as the plaintext in the earlier questions on the HW (e.g., they have the same header).

For all the above questions, you need to explain how you obtained your solutions: what method did you use, and why. It is perfectly reasonable to perform a "brute-force" attack (i.e., try all possibilities) on parts of the scheme if you so desire. Additionally, attempt to provide a formal justification for why your attack works, and try to provide bounds on the length of the ciphertexts needed to recover the keys. (Hint: use the Chernoff bound.)