CMPT 404 — Cryptography and Protocols
Spring 2012

Exercises on Probability and Perfect Security. Due: Thursday, February 2nd (at the beginning of the class)

Reminder: the work you submit must be your own. Any collaboration and consulting outside resources must be explicitly mentioned on your submission.

1. The following text was encrypted using a substitution cipher. Find the plaintext and explain the steps you took to find it. Punctuation marks are removed from the text. Tables of English letters frequencies can be found on the web.

   GHDG XLBHSFH IHBB LS YKH CLGGBH TI D OTWG TZY OHSY DBB BLJKYX YKH ILWH BHDUHG ZU LS EBDFA XCTAHX DXKHX DSG FLSGHWX OHWH LS YKH HQHX TI YKH GODWNHX DSG YKH OTTG ODX ILBBHG DJDLS OLYK YKHLW FBDCTZW DSG YKHLW FWLHX ELBET ITZSG KLCXHBI WZSSLSJ WTZSG DSG WTZSG DX KH YKTZJKY DSG FDBBLSJ DSG FDBBLSJ GTWL STWL TWL TLS JBTLS ILBL ALBL ETCEZW ETIZW GODBLS EDBLS YKTWLS TDAHSXKLHBG OKLBH UHTUBH KH FTZBG STY XHH TW IHHB OHWH GTLSJ YKH XDCH DBB WTZSG KLC OLYK DS TFFDXLTSDB ELBET YKWTOS LS EZY YKH FWLHX TI YKH TYKHWX JTY XYHDGLBQ IZWYKHW DSG IDLSYHW DSG YKTZJK DIYHW D OKLBH LY XHHCHG YT KLC YKHQ FKDSJHG YT QHBBX DSG FWLHX ITW KHBU LS YKH GLXYDSFH DBB STLXH DY BDXY GLHG WLJKY DODQ DSG KH ODX BHIY DBTSH LS FTCUBHYH XLBHSFH DSG GDWASHXX

2. One method to improve the security of substitution ciphers is to compress the plaintext (say, using gzip) before applying the encryption algorithm. Explain why this simple procedure improves security. You may think of two explanations, linguistic and mathematical. Explain, how you need to change your substitution cipher so that it can still be used.

3. Prove that the statistical distance between random variables X and Y satisfy the following properties:

   (a) $\Delta(X,Y) = \frac{1}{2}\sum_v |\Pr[X=v] - \Pr[Y=v]|$;

   (b) *(triangle inequality)* for any random variables Z, $\Delta(X,Y) \leq \Delta(X,Z) + \Delta(Z,Y)$.

4. Suppose that (K,E,D) is a SES with key 100 bit shorter than the plaintext, that is $E\colon \{0,1\}^{m-100} \times \{0,1\}^m \to \{0,1\}^*$.

   (a) Prove that there is an algorithm ADV for the Eavesdropper (not necessarily efficient), and plaintexts P_1, P_2 such that

   $$\Pr[\mathrm{ADV}(E_k(P_i)) = i] > 0.99,$$

   where the probability is over random choices of the key and $i \in \{1,2\}$.

   (b) Prove that for each plaintext P_1 there are at least 2^{99} plaintexts P_2 such that the disequality above holds.

Hint: Review the proof of the theorem on statistical security and short keys.

5. Let (K, E, D) be a SES that uses keys at least 1 bit shorter than messages. Give two algorithms breaking the encryption scheme that work in the following settings:

   (a) The algorithm must distinguish two plaintext P_1, P_2 chosen by the algorithm. As in the game model of security, it requests an encryption of one of them (Alice chooses which one to encrypt), however, it receives several encryptions (of the same plaintext) using different keys. How many encryptions do you need to distinguish plaintexts with probability 0.9999? (One of the possible solutions uses Chernoff bound in probability. However, there are solutions — although less elegant — that do not use it.) Model Alice and the encryption scheme as black boxes.

   (b) The algorithm must recover the key Alice uses. It is allowed to request encryptions of several (well, possibly, many) plaintext in either non-adaptive (all at once), or adaptive (use the encryptions of previous plaintexts to choose the next one) mode. You can choose which mode you like. Model Alice and the encryption scheme as black boxes.

   (c) (*optional*) Write codes for the above algorithms and run them for key length 5 and message length 6. Use yourself as Alice. Use the following encryption scheme:
   – a key is a random prime number p with $2^{m-2} \le p < 2^{m-1}$, that is of length $m-1$ bits;
   – to encrypt a single plaintext P we first compute $k = 2^{m-1} \pmod p$, and then set $C_i = P_i \oplus k_i \pmod 2$ for $i < m$, and $C_m = P_m \oplus k_1$;
   – to encrypt multiple plaintexts, we encrypt the first one as before, and to encrypt every consequent one we replace k with $k^2 \pmod p$.

   Hint: The previous exercise can be helpful.

6. (*optional*) A collection $\{X_n\}$ of random variables with $X_n \in \{0,1\}^n$ is called (T, ε)-unpredictable, if for any probabilistic algorithm Alg with time complexity at most $T(n)$

$$\Pr[\mathrm{Alg}(X_n^1 X_n^2 \dots X_n^i)) = X_n^{i+1}] < \frac{1}{2} + \varepsilon(n),$$

   for any $i < n$, where $X_n = X_n^1 X_n^2 \dots X_n^n$. Proof that collection $\{X_n\}$ is (T, ε)-unpredictable if and only if it is (T, ε)-pseudorandom.
   Hint: Difficult, proving that a pseudorandom collection is unpredictable is extremely difficult. To prove that an unpredictable collection is pseudorandom, you need to assume that it is not pseudorandom, take an algorithm that distinguishes it from U_n, and convert it into an algorithm that predicts X_n.

7. (*optional*) All our definitions of security are given assuming that the adversary deals with only one pair of plaintexts. What if she allowed to use many pairs? Give a definition of computational security if Eve has access to many ciphertexts.