PMAT 329 Introduction to Cryptography

SOLUTIONS TO ASSIGNMENT 2

---

(1) [12 points] Cryptanalyze the following polyalphabetic ciphertext. Show your work. Note that this ciphertext is the example used in the handout given out in class to illustrate the factoring method for resolving the number of alphabets.

```
SIJYU MNVCA ISPJL RBZEY QWYEU LWMGW ICJCI MTZEI MIBKN
QWBRI VWYIG BWNBQ QCGQH IWJKA GEGXN IDMRU VEZYG QIGVN
CTGYO BPDBL VCGXG BKZZG IVXCU NTZAO BWFEQ QLFCO MTYZT
CCBYQ OPDKA GDGIG VPWMR QIIEW ICGXG BLGQQ VBGRS MYJJY
QVFWY RWNFL GXNFW MCJKX IDDRU OPJQQ ZRHCN VWDYQ RDGDG
BXDBN PXFPU YXNFG MPJEL SANCD SEZZG IBEYU KDHCA MBJJF
KILCJ MFDZT CTJRD MIYZQ ACJRR SBGZN QYAHQ VEDCQ LXNCL
LVVCS QWBII IVJRN WNBRI VPJEL TAGDN IRGQP ATYEW CBYZT
EVGQU VPYHL LRZNQ XINBA IKWJQ RDZYF KWFZL GWFJQ QWJYQ
IBWRX
```

Solution. This ciphertext is the example used in class to illustrate the factoring method for resolving the number of alphabets (5 alphabets were used). Attempting to apply Kerkhoff's shortcut fails, so we conclude that a mixed polyalphabetic cipher (mixed Vigenére) was used. Using frequency data and symmetry of position, one can find the five cipher alphabets:

Plain	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
C 1	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H
C 2	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O
C 3	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K	L	M	O	P	Q
C 4	I	N	G	B	C	D	F	J	K	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T
C 5	L	M	O	P	Q	R	V	W	Y	Z	E	X	H	A	U	S	T	I	N	G	B	C	D	F	J	K

Note that the alphabet was reordered using the (rather appropriate!) key word EXHAUSTING, and that the key indicating which shift of the reordered alphabet to use for which subtext can be found by reading the cipher characters corresponding to plaintext A (keyword=APRIL).

The plaintext is the following:

> CO Troop B:
> Enemy has retired to NEWCHESTER. One troop is reported at HENDERSON MEETING HOUSE; two other troops in orchard at southwest edge of NEWCHESTER. Second Sq is preparing to attack from the south. One troop of third Sq is engaging hostile troop at NEWCHESTER. Rest of third Qs is moving to attack NEWCHESTER from the north. Move your Sq into woods east of crossr. five-three-nine and be prepared to support attack of second and third Sq. Do not advance beyond NEWCHESTER. Messages here.
> Treer, col.

The complete solution to this cipher can be found in "Elements of Cryptanalysis" by William Friedman.

(2) Consider a cryptosystem with key space $\mathcal{K}$, plaintext space $\mathcal{M}$, and ciphertext space $\mathcal{C}$ that provides perfect secrecy. Assume that $p(C) > 0$ for all $C \in \mathcal{C}$.

(a) [5 points] Prove that $|\mathcal{K}| \geq |\mathcal{C}|$.

Solution. Since our system provides perfect secrecy, we have for all $M \in \mathcal{M}$ with $p(M) > 0$ and all $C \in \mathcal{C}$:

$$0 < p(C) = p(C|M) = \sum_{\substack{K \in \mathcal{K} \\ E_K(M) = C}} p(K),$$

so for every $M \in \mathcal{M}$ with $p(M) > 0$ and every $C \in \mathcal{C}$, at least one of the terms $p(K)$ in the above sum is positive. This means that for every $M \in \mathcal{M}$ with $p(M) > 0$ and every $C \in \mathcal{C}$, there exists at least one key $K \in \mathcal{K}$ such that $E_K(M) = C$.
Fix $M \in \mathcal{M}$ with $p(M) > 0$ (such a message always exists, obviously) and consider the map $f_M : \mathcal{K} \to \mathcal{C}$ via $f_M(K) = E_K(M)$. By our above reasoning, for every ciphertext $C \in \mathcal{C}$, there exists a key $K \in \mathcal{K}$ with $C = E_K(M) = f(K)$. This says exactly that the map f is surjective (onto). Since both $\mathcal{K}$ and $\mathcal{C}$ are finite sets, this implies that $|\mathcal{K}| \geq |\mathcal{C}|$.

(b) [5 points] Conclude that if $|\mathcal{K}| \leq |\mathcal{M}|$, then $|\mathcal{K}| = |\mathcal{M}| = |\mathcal{C}|$ and all encryptions are bijections.

Solution. Since every encryption function $E_K : \mathcal{M} \to \mathcal{C}$ is an injection (proved in class), we have $|\mathcal{M}| \leq |\mathcal{C}|$. By (a), we have $|\mathcal{C}| \leq |\mathcal{K}|$. By assumption, we have $|\mathcal{K}| \leq |\mathcal{M}|$. Altogether, we must have $|\mathcal{M}| = |\mathcal{K}| = |\mathcal{C}|$. Since all encryptions are injections and $|\mathcal{M}| = |\mathcal{C}|$ is finite, all encryptions are bijections.

(c) [6 points] Show that under the condition of part (b) every ciphertext is equally probable, i.e. $p(C) = 1/|\mathcal{C}|$ for all $C \in \mathcal{C}$.
(*Hint:* Let $C \in \mathcal{C}$ be any ciphertext. Use the statement on the uniqueness of keys in Shannon's Theorem to show that the function $g_C : \mathcal{K} \to \mathcal{M}$ via $g_C(K) = D_K(C)$ is a bijection. Now use the other statement in Shannon's Theorem, i.e. that every key is used with equal likelihood.)

Solution. By the second statement of Shannon's Theorem, for every plaintext $M \in \mathcal{M}$ and every ciphertext $C \in \mathcal{C}$, there exists a unique key $K \in \mathcal{K}$ with $C = E_K(M)$, or equivalently (since E_K is a bijection by part (b)), $M = D_K(C)$.
As suggested in the hint, let $C \in \mathcal{C}$ be any ciphertext, and consider the map $g_C : \mathcal{K} \to \mathcal{M}$ via $g_C(K) = D_K(C)$. Now by our previous statement, $D_K(C)$ exists for every key K, so g_C is defined on all of $\mathcal{K}$. Furthermore, for every $M \in \mathcal{M}$, there exists a unique key $K \in \mathcal{K}$ with $M = D_K(C) = g_C(K)$, so g_C is a bijection. By the first statement of Shannon's Theorem, every key is used with equal probability $1/|\mathcal{K}|$. Therefore,

$$\begin{aligned} p(C) &= \sum_{K \in \mathcal{K}} p(K)p(D_K(C)) = \frac{1}{|\mathcal{K}|} \sum_{K \in \mathcal{K}} p(D_K(C)) \\ &= \frac{1}{|\mathcal{K}|} \sum_{K \in \mathcal{K}} p(g_C(K)) = \frac{1}{|\mathcal{K}|} \cdot 1 = \frac{1}{|\mathcal{K}|} = \frac{1}{|\mathcal{C}|}. \end{aligned}$$

Here, we use the facts that $|\mathcal{K}| = |\mathcal{C}|$ and that as K runs through $\mathcal{K}$, the values of $g_C(K)$ run through all of $\mathcal{M}$ (since g_C is a bijection), so the sum over the probabilities of all these values is one.

Solution 2. : Here is another way to prove this without the hint. As in the previous solution, you argue that for any messae $M \in \mathcal{M}$ and any ciphertext $C \in \mathcal{C}$, there exists a unique key $K = K_{MC} \in \mathcal{K}$ with $C = E_{K_{MC}}(C)$. Then for any $M \in \mathcal{M}$ and $C \in \mathcal{C}$:

$$p(C) = p(C|M) = \sum_{\substack{K \in \mathcal{K} \\ E_K(M)=C}} p(K) = p(K_{MC}) = \frac{1}{|\mathcal{K}|} = \frac{1}{|\mathcal{C}|}$$

since the sum has only one term $p(K_{MC})$ and $|\mathcal{K}| = |\mathcal{C}|$.

(3) [8 points] Use the characterization $p(C) = p(C|M)$ for all $C \in \mathcal{C}$ and $M \in \mathcal{M}$ to prove that one-time pad provides perfect secrecy under the assumption that each key is chosen with equal likelihood. Can you say anything about the distribution of ciphertexts?

Solution. We have $\mathcal{M} = \mathcal{C} = \mathcal{K} = \mathbb{Z}_2^n$ ($n \in \mathbb{N}$), and for every $M, K, C \in \mathbb{Z}_2^n$, encryption of M under key K is given by $E_K(M) = M \oplus K$, and decryption of C under K is given by $D_K(C) = C \oplus K$. We assume that each key $K \in \mathbb{Z}_2^n$ is chosen with equal likelihood $p(K) = 1/|\mathbb{Z}_2^n| = 2^{-n}$. Let $C \in \mathbb{Z}_2^n$. Then

$$p(C) = \sum_{K \in \mathbb{Z}_2^n} p(K)p(D_K(C)) = 2^{-n} \sum_{K \in \mathbb{Z}_2^n} p(C \oplus K).$$

But the map $f : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ via $f(K) = C \oplus K$ is a bijection whose inverse map is f, since $f(f(K)) = f(C \oplus K) = C \oplus C \oplus K = K$. Hence, the sum in the formula above runs through all the elements in $\mathbb{Z}_2^n$ and can be written as $\sum_{K' \in \mathbb{Z}_2^n} p(K')$ which is equal to 1. This shows that $p(C) = 2^{-n}$ for all $C \in \mathbb{Z}_2^n$; in particular, all ciphertexts occur with equal probability, regardless of the probability distribution on the plaintext space.

On the other hand, it is easy to see that for every message $M \in \mathbb{Z}_2^n$ and every ciphertext $C \in \mathbb{Z}_2^n$, there exists a unique key K such that $C = E_K(M)$, and that key is $K = M \oplus C$. To see this, not that $K = K \oplus \mathbf{0} = K \oplus M \oplus M = C \oplus M$. Also, suppose we have two keys $K_1, K_2 \in \mathbb{Z}_2^n$ such that $C = M \oplus K_1 = M \oplus K_2$, then x-or'ing this identity with M yields $M \oplus C = K_1 = K_2$. Therefore, for all $C, M \in \mathbb{Z}_2^n$ with $p(M) > 0$, we have

$$p(C|M) = \sum_{\substack{K \in \mathbb{Z}_2^n \\ E_K(M)=C}} p(K) = p(M \oplus C) = 2^{-n} = p(C),$$

and the one-time pad provides perfect secrecy under the outlined conditions.

(4) For a bit string $\mathbf{x} \in \mathbb{Z}_2^n$, denote by $\overline{\mathbf{x}}$ the *ones' complement* of $\mathbf{x}$; that is, the i-th bit of $\overline{\mathbf{x}}$ is a '1' if and only if the i-th bit of $\mathbf{x}$ is a '0' for $1 \leq i \leq n$. Note that $\overline{\mathbf{x}} = \mathbf{1} \oplus \mathbf{x}$ where $\mathbf{1} \in \mathbb{Z}_2^n$ is the string consisting of n ones.

(a) [4 points] Let M be a DES plaintext and K a DES key. Suppose $C = E_K(M)$ where E_M denote DES encryption under key K. Show that $\overline{C} = E_{\overline{K}}(\overline{M})$.

Solution. Let L_i and R_i, $i = 0, 1, \ldots, 16$, be the 32-bit blocks occurring during the computation of $E_K(M)$ and let $L_i^{'}$ and $R_i^{'}$, $i = 0, 1, \ldots, 16$, be the 32-bit blocks occurring during the computation of $E_{\overline{K}}(\overline{M})$. Similarly, let K_i, $i = 1, \ldots, 16$, be the subkeys used during the computation of $E_K(M)$ and let K'_i be the subkeys used during the computation of $E_{\overline{K}}(\overline{M})$.
First, consider the DES key schedule. The subkeys are produced by combinations of permutations, expansions, and shifts. All of these operations only rearrange the bits of the input — none of them actually complement or otherwise modify a single input bit. The same sequence of bits are taken for the subkeys irregardless of the input, so the sequences of bits occurring in the key schedules for K and $\overline{K}$ will simply be complements of each other. This implies that $K'_i = \overline{K}_i$.
Second, note that $L'_0 = \overline{L}_0$ and $R'_0 = \overline{R}_0$, since the initial permutation only rearranges the bits of the input, i.e., $IP(\overline{M}) = \overline{IP(M)}$. When the first round is executed, we get $L'_1 = R'_0$, and since $L_1 = R_0$ we obtain

$$R'_0 = \overline{R}_0 = \overline{L}_1 \implies L'_1 = \overline{L}_1 \ .$$

What about R'_1? We have that $R'_1 = L'_0 \oplus f(R'_0, K'_i) = \overline{L}_0 \oplus f(\overline{R}_0, \overline{K}_0)$. Consider the function f. The first step is to compute $E(\overline{R}_0) \oplus \overline{K_0}$, where E is an expansion function. E only rearranges and duplicates bits, so $E(\overline{R}_0) = \overline{E(R_0)}$. Now, for two bits a and b, it can be seen by simply constructing truth tables that $a \oplus b = \overline{a} \oplus \overline{b}$, and hence for bit strings A and B, bit-wise XOR yields $A \oplus B = \overline{A} \oplus \overline{B}$. Hence,

$$E(\overline{R}_0) \oplus \overline{K_1} = \overline{E(R_0)} \oplus \overline{K_1} = E(R_0) \oplus K_1,$$

and this implies that

$$f(R'_0, K'_1) = f(\overline{R}_0, \overline{K}_1) = f(R_0, K_1) \ .$$

By constructing truth tables we can also show that $\overline{A} \oplus B = A \oplus \overline{B} = \overline{A \oplus B}$, yielding

$$R'_1 = L'_0 \oplus f(R'_0, K'_1) = \overline{L}_0 \oplus f(R_0, K_1) = \overline{L_0 \oplus f(R_0, K_1)} = \overline{R}_1 \ .$$

We can carry out the same argument for each of the 16 round to show that executing each round on $L'_i = \overline{L}_i$ and $R'_i = \overline{R}_i$ yields $L'_{i+1} = \overline{L}_{i+1}$ and $R'_{i+1} = \overline{R}_{i+1}$, and in particular we get that $L'_{16} = \overline{L}_{16}$ and $R'_{16} = \overline{R}_{16}$. IP^{-1} rearranges bits without modifying any of them, so $IP^{-1}(\overline{R}_{16}, \overline{L}_{16}) = \overline{IP^{-1}(R_{16}, L_{16})} = \overline{C}$. Thus, we have $E_{\overline{K}}(\overline{M}) = \overline{C}$ as required.

(b) [4 points] Suppose a cryptanalyst knows two plaintext-ciphertext pairs (M_1, C_1) and (M_2, C_2) with $C_i = E_K(M_i)$ $(i = 1, 2)$ for some DES key K (i.e. the same key is used for both encryptions) and $M_2 = \overline{M_1}$ (this scenario amounts to a CTA). How and by how much can this information reduce the effort of an exhaustive key search attack on DES? Explain.

Solution. This complementation property can be used to reduce the search effort for K by half as follows. Suppose that the adversary knows $C_1, C_2, M_1 = M$, and $M_2 = \overline{M}$ such that $C_1 = E_K(M)$ and $C_2 = E_K(\overline{M})$. Further, suppose that during the course of an exhaustive search attack we are testing the key L :

- If L is the correct key, then

$$C_1 = E_L(M) \text{ and } C_2 = E_{\overline{L}}(\overline{M}) \ .$$

- If L is the one's complement of the correct key, then

$$\overline{C}_1 = E_L(\overline{M}) \text{ and } \overline{C_2} = E_L(M) \ .$$

Thus, to test the key L simply compute $C_L = E_L(M)$ and

- compare C_L to C_1 : if they are equal, then L is the correct key;
- compare C_L to $\overline{C}_2$: if they are equal, then L is the one's complement of the correct key.

This allows us to simultaneously test keys L and $\overline{L}$ (by computing only *one* DES encryption), thereby cutting the total number of keys to check in half.

(5) In a cryptographic system, one wishes to avoid keys that provide a poor level of encryption; the worst scenario would obviously be $E_K(M) = M$ for all plaintexts M, but other keys have less drastic weaknesses.

Two DES keys K_1 and K_2 are *dual* or *semi-weak* if $E_{K_1}(M) = D_{K_2}(M)$ for every $M \in \mathbb{Z}_2^{64}$. Such keys are obviously a disaster for double encryption as $E_{K_2}(E_{K_1}(M)) = M$ for all plaintexts M. If in addition, $K_1 = K_2$ (= K say), i.e. $D_K = E_K$, then K is called *self-dual* or *palindromic*[1] or simply *weak*.

(a) [4 points] Let C_0 be the left half and D_0 be the right half of the image of the relevant 56 bits of a DES key K under DES Permuted Choice PC-1. If C_0 is either all 0's or all 1's and D_0 is either all 0's or all 1's, then K is self-dual. Prove this in the case $C_0 = D_0 = 0^{56}$ (the other three cases can be proved analogously).

Solution. Each round key is produced from a combination of C_i and D_i, where C_i and D_i are obtained by a circular shift of the bits of C_{i-1} and D_{i-1}, respectively. If all the bits of C_0 are identical, then any circular shift produces the same result and $C_0 = C_1 = \cdots = C_{16}$. Similarly, if all the bits of D_0 are identical, then $D_0 = D_1 = \cdots = D_{16}$. Thus, if for a key K we obtain C_0 with all bits the same and D_0 with all bits the same, the subkeys produced will all be identical. Since decryption is simply the DES algorithm with the reverse key schedule, encryption and decryption will be the same when all the subkeys identical, and any key producing identical subkeys is a weak key.

(b) [4 points] The following four DES keys (given in hexadecimal, i.e. base 16, notation) are self-dual. Prove this fact for the first of these four keys (again, the proof for the other three is analogous).

```
0101 0101 0101 0101
FEFE FEFE FEFE FEFE
1F1F 1F1F 0E0E 0E0E
E0E0 E0E0 F1F1 F1F1
```

It turns out that these are the only weak keys. It is a fact that each such key K has 2^{32} *fixed points*, i.e. plaintexts M for which $E_K(M) = M$.

[1] A *palindrome* is a sequence of symbols that reads the same forwards as backwards, for example *"never odd or even"* or *"able was I ere I saw elba"*

Solution. Note that the first key consists of the byte 00000001 repeated 8 times. The DES specification (FIPS publication) states that bits $8, 16, \ldots, 64$ of the key are used as parity bits (recall that only 56 bits of the key are actually used as key material). When "PERMUTED CHOICE 1" is applied to the key to produce C_0 and D_0, these bits are ignored. Thus, since all the remaining bits of the first key are 0, C_0 and D_0 will both consist solely of 0's, and by Part (a) this key is weak.
For completion, we discuss the other three weeak keys. The second key consists of 11111110 repeated 8 times, and using the above reasoning C_0 and D_0 will consist solely of 1's, implying that the second key is also weak.
There are two more possible weak keys, the first corresponding to C_0 all 0's and D_0 all 1's, and the second corresponding to C_0 all 1's and D_0 all 0's. Setting all the bits of C_0 to 0 and all the bits of D_0 to 1 yields the third key $1F1F \ldots 1F$ (using the table for "PERMUTED CHOICE 1" from the FIPS document). Similarly, the fourth key yields C_0 all 1's and D_0 all 0's.

(c) [4 points] Let C_0 and D_0 be as in part (a). Prove that $C_0 = 0101 \ldots 01$ (in binary), then $C_i \oplus C_{17-i} = 1111 \ldots 11$ for $1 \leq i \leq 16$. State an analogous property for the D_i's.

Solution. Looking at the table of left shifts in the key schedule, we see that $C_1 = (10)^{14}$, $C_2 = C_3 = \cdots = C_8 = (01)^{14}$, $C_9 = C_{10} = \cdots = C_{15} = (10)^{14}$, and $C_{16} = (01)^{14}$. In other words, $C_i = \overline{C_{17-i}}$, and hence $C_i \oplus C_{17-i} = 1^{28}$ for $1 \leq i \leq 16$.
Simply replace C_i by D_i for $1 \leq i \leq 16$ to obtain the same property and proof for the D_i.

(d) [4 points] The following pairs of keys (given in hexadecimal notation) are dual:

01FE 01FE 01FE 01FE	FE01 FE01 FE01 FE01
1FE0 1FE0 0EF1 0EF1	E01F E01F F10E F10E
01E0 01E0 01F1 01F1	E001 E001 F101 F101
1FFE 1FFE 0EFE 0EFE	FE1F FE1F FE0E FE0E
011F 011F 010E 010E	1F01 1F01 0E01 0E01
E0FE E0FE F1FE F1FE	FEE0 FEE0 FEF1 FEF1

Prove this for the first of these six pairs of keys (again, one can give analogous proofs for the other five). These are the only semi-weak keys.

Solution. It is not hard to show that key $(01FE)^4$ corresponds to $C_0 = D_0 = (01)^{14}$ and key $(FE01)^4$ corresponds to $C_0 = D_0 = (10)^{14}$. In other words, if $K_1, K_2, \ldots, K_{16}$ are the 16 round keys obtained from key $(01FE)^4$, then the 16 round keys obtained from key $(FE01)^4$ are $K_{16}, K_{15}, \ldots, K_1$. Since decryption under any key K reverses the key schedule obtained when encrypting with K, encryption under key $(01FE)^4$ and decryption under $(FE01)^4$ have the same key schedule and thus produce the same ciphertext.

In practice, it is obviously easy to avoid the 16 keys listed above.