EXERCISE SET 2, TMA4160 - KRYPTOGRAFI

Ch. 1

21.

b) We are given the following ciphertext encrypted with a vigenère cipher:

KCCPKBGUFDPHQTYAVINRRTMVGRKDNBVFDETDGILTXRGUD
DKOTFMBPVGEGLTGCKQRACQCWDNAWCRXIZAKFTLEWRPTYC
QKYVXCHKFTPONCQQRHJVAJUWETMCMSPKQDYHJVDAHCTRL
SVSKCGCZQQDZXGSFRLSWCWSJTBHAFSIASPRJAHKJRJUMV
GKMITZHFPDISPZLVLGWTFPLKKEBDPGCEBSHCTJRWXBAFS
PEZQNRWXCVYCGAONWDDKACKAWBBIKFTIOVKCGGHJVLNHI
FFSQESVYCLACNVRWBBIREPBBVFEXOSCDYGZWPFDTKFQIY
CWHJVLNHIQIBTKHJVNPIST

We try the Kasiski test. By using a script we see that the string HJV occurs five times in the text. We count the positions started to be 108, 126, 264, 318, 330. From this we see that the distances between the starting point of each of these substrings are: 18, 138, 54 and 12 for which the gcd is 6. This is therefore a likely length of the keyword. By using the index of coincidence method we can get further evidence for this. We divide the text into 6 substrings $y_1, y_2, y_3, y_4, y_5, y_6$. For example we have for $i = 1$:

y_1 = KBPARRVDXDMECCN$(\cdots)$EFCWKCLIJ

We compute the index of coincidence for $m = 6$ to be: $0.0630, 0.0838, 0.0494, 0.0649, 0.0429, 0.0734$. Which is a fairly good result; It certainly fits better than for other choices of m. We can therefore assume that the length of the keyword is 6. We now turn to the probability distribution of each y_i. Each of these vectors is obtained by a shift encryption and therefore with a sufficient amount of data one would expect that the shifted distributions are close to the naturally occuring one.

I.e. if the length of y_i is n, $f_0, f_2, \cdots f_{25}$ is the letter frequencies in y_i and k_i is the value that y_i is shifted by, then $\frac{f_{k_i+j \bmod 25}}{n} \approx p_j$, where p_j is the ideal probability of hitting letter j in a text. We define for $0 \leq g \leq 25$ the sum $M_g = \sum_{i=0}^{25} p_i f_{i+g}/n$. Note that if $g = k_i$, then $M_g \approx \sum_{i=0}^{25} p_i^2 \approx 0.065$ which makes the particular choice of g highly probable of being in the key. We now compute this value for the different y_i:

i	$M_g(y_i)$												
1	0.031	0.036	**0.065**	0.041	0.035	0.043	0.041	0.035	0.046	0.047	0.027	0.037	0.041
	0.044	0.041	0.045	0.036	0.044	0.046	0.034	0.032	0.047	0.047	0.036	0.043	0.038
2	0.038	0.039	0.049	0.041	0.039	0.036	0.044	0.029	0.026	0.036	0.045	0.029	0.034
	0.047	0.040	0.033	0.036	**0.069**	0.035	0.030	0.029	0.033	0.029	0.037	0.046	0.035
3	0.035	0.037	0.034	0.038	0.036	0.042	0.029	0.039	0.035	0.042	0.042	0.047	0.041
	0.043	0.037	0.032	0.035	0.040	0.044	0.032	0.039	0.035	0.036	0.044	**0.059**	0.047
4	0.045	0.039	0.044	0.038	0.037	0.039	0.036	0.037	0.042	0.038	0.038	0.055	0.044
	0.033	0.037	**0.066**	0.037	0.033	0.042	0.042	0.027	0.042	0.044	0.035	0.036	0.048
5	0.040	0.034	0.034	0.041	0.045	0.035	0.046	0.048	0.049	0.034	0.035	0.039	0.037
	0.036	0.035	0.046	0.034	0.038	0.037	**0.057**	0.042	0.040	0.045	0.046	0.031	0.035
6	0.042	0.038	0.037	0.042	0.039	0.026	0.032	0.038	0.036	0.034	0.047	0.033	0.024
	0.036	**0.070**	0.042	0.032	0.031	0.038	0.032	0.040	0.039	0.035	0.037	0.039	0.047

This corresponds to $K = [2, 17, 24, 15, 19, 14]$ that gives the keyword $CRYPTO$. This keyword gives the plaintext (edited to be readable):

> I learned how to calculate the amount of paper needed for a room when I was at school. You multiply the square footage of the walls by the cubic contents of the floor and ceiling combined, and double it. You then allow half the total for openings such as windows and doors. Then you allow the other half for matching the pattern. Then you double the whole thing again to give a margin of error, and then you order the paper.

d) We are given the following ciphertext encrypted with an unspecified cipher:

BNVSNSIHQCEELSSKKYERIFJKXUMBGYKAMQLJTYAVFBKVT
DVBPVVRJYYLAOKYMPQSCGDLFSRLLPROYGESEBUUALRWXM
MASAZLGLEDFJBZAVVPXWICGJXASCBYEHOSNMULKCEAHTQ
OKMFLEBKFXLRRFDTZXCIWBJSICBGAWDVYDHAVFJXZIBKC
GJIWEAHTTOEWTUHKRQVVRGZBXYIREMMASCSPBNLHJMBLR
FFJELHWEYLWISTFVVYFJCMHYUYRUFSFMGESIGRLWALSWM
NUHSIMYYITCCQPZSICEHBCCMZFEGVJYOCDEMMPGHVAAUM
ELCMOEHVLTIPSUYILVGFLMVWDVYDBTHFRAYISYSGKVSUU
HYHGGCKTMBLRX

We run this cipthertext through a Friedman test with $m = 1$ and get the index 0.0414. This suggests that the text is encrypted using a polyalphabetic cipher. We assume that it is a Vigenière cipher and see where we end up with this assumption. We run the same test with different values for m and get:

i					
1	0.0414				
2	0.0443	0.0462			
3	0.0442	0.0479	0.0484		
4	0.0430	0.0563	0.0465	0.0479	
5	0.0444	0.0418	0.0404	0.0433	0.0370
6	0.0497	0.0613	0.0550	0.0709	0.0555	0.0698

From which we assume that $m = 6$. Using the same method as above we get the key $K = [19, 7, 4, 14, 17, 24]$ corresponding to $THEORY$. With this key the ciphertext decrypts to the plaintext (edited to be readable):

> I grew up among slow talkers, men in particular, who dropped words a few at a time like beans in a hill, and when I got to Minneapolis where people took a Lake Wobegon comma to mean the end of a story, I couldn't speak a whole sentence in company and was considered not too bright. So I enrolled in a speech course taught by Orville Sand, the founder of reflexive relaxology, a self-hypnotic technique that enabled a person to speak up to three hundred words per minute.

Ch. 2

2. We study a Latin square of order n. That is an $n \times n$ array L such that the integers from 1 to n occur once in each row and column. Let $\mathcal{P} = \mathcal{K} = \mathcal{C} = \{1, 2, \cdots, n\} = A$ and $e_i(j) = L(i, j)$ for $1 \leq i \leq n$. Assume the keys are equidistributed. Then this cryptosystem achieves perfect secrecy. We need to show that $P[x|y] = P[x]$ for all $x \in \mathcal{P}, y \in \mathcal{C}$. So let $y \in \mathcal{C}, x \in \mathcal{P}$, then:

$$P[\mathbf{y} = y] = \sum_{K:y \in C(K)} P[\mathbf{K} = K] P[\mathbf{x} = d_K(y)]$$

Note that $P[\mathbf{K} = K] = \frac{1}{n}$, since $\mathcal{K}$ is equidistributed. If $y \in C(K) = \{e_K(x) : x \in \mathcal{P}\}$ then there is some x and K such that $y = e_K(x) = L(K, x) \Rightarrow d_K(y) = x$. Furthermore $C(K) = \mathcal{C} = \mathcal{K} = \mathcal{P}$ and therefore:

$$P[\mathbf{y} = y] = \frac{1}{n} \sum_{\mathcal{K}} P[\mathbf{x} = x] = \frac{1}{n} \sum_{\mathcal{P}} P[\mathbf{x} = x] = \frac{1}{n}$$

Hence $P[y] = 1/n$. We also have:

$$P[\mathbf{y} = y | \mathbf{x} = x] = \sum_{\{K : x = d_K(y)\}} P[\mathbf{K} = K]$$

That is, given x, y sum over all K such that $x = d_K(y)$, but since every number in the column is different there is really only one such K and the probability for choosing this K is $\frac{1}{n}$. This implies by Bayes rule that

$$P[x|y] = \frac{P[x]P[y|x]}{P[y]} = P[x]$$

so this cryptosystem achieves perfect security.

5. If a cryptosystem has perfect secrecy and $|\mathcal{K}| = |\mathcal{C}| = |\mathcal{P}| = N$, then every ciphertext is equally probable.

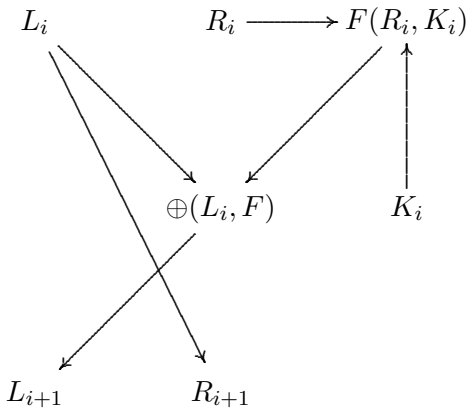
Proof. Note that the assumptions implies by Shannon's theorem that the keyspace is equidistributed. Furthermore for every $x \in \mathcal{P}$ and every $y \in \mathcal{C}$ there is a unique key K such that $e_K(x) = y$. Since the system has perfect secrecy we have that $P[x|y] = P[x]$. By Bayes rule we also therefore must have $P[y|x] = P[y]$. Now given x, y there is a unique key K such that $e_K(x) = y$ and so $P[y] = P[y|x] = P[K : y = e_K(x)|x, y] = 1/N$. Hence every ciphertext is equally probable. □

Ch. 3

2. The Feistel cipher can be decrypted by applying the encryption algorithm with reversing the key schedule. Let $K_0, K_1, \cdots K_n$ be the keyscheme, F the round function and $X = (L_0, R_0)$ be the plaintext splitted into a left and a right half. The Feistel cipher is defined by setting:

$$L_{i+1} = R_i \qquad 0 \leq i \leq n$$
$$R_{i+1} = L_i \oplus F(R_i, K_i) \qquad 0 \leq i \leq n$$

Or schematically:



Consider the last step $i = n+1$. We are given (L_{n+1}, R_{n+1}). By the encryption function we know that $L_{n+1} = R_n$ and therefore we have retrieved the right side of the round n encrypted message. Since $\oplus$ is an involution ($K \oplus M = C \Leftrightarrow (K \oplus M) \oplus M = K \oplus (M \oplus M) = K \oplus 0 = K = C \oplus M$) we also

have that $R_{n+1} \oplus F(R_n, K_n) = L_n$ and since $R_n = L_{n+1}$ we obtain:

$$R_n = L_{n+1}$$
$$L_n = R_{n+1} \oplus F(L_{n+1}, K_n)$$

By induction it now follows that for every $1 \leq i \leq n$:

$$R_i = L_{i+1}$$
$$L_i = R_{i+1} \oplus F(L_{i+1}, K_i)$$

Which essentially is the encryption algorithm with reversed indexation i.e. reversed key schedule.

Ch. 4

11. The CFB-MAC is described like this: Given plaintext blocks $x_1, x_2, \cdots, x_n$ and define $y_0 = \text{IV} = x_1$. We use the key K in CFB mode and get cipherblocks $y_i = x_{i+1} \oplus e_K(y_{i-1})$ for $i = 1, \cdots, n-1$. Finally define the MAC to be $e_K(y_{n-1})$. We shall prove that this MAC is identical to CBC-MAC. For reference the CBC-MAC is given by $y_0 = \text{IV} = 00\cdots0$, $y_i = e_K(y_{i-1} \oplus x_i)$ for $i = 1, \cdots, n$ where $i = n$ gives the MAC. We prove that they are identical by following the computations at each step in a table:

i	CBC-MAC	CFB-MAC
0	$00\cdots0$	x_1
1	$e_K(y_0 \oplus x_1)$	$x_2 \oplus e_K(y_0)$
2	$e_K(y_1 \oplus x_2)$	$x_3 \oplus e_K(y_1)$
$\vdots$	$\vdots$	$\vdots$
$n-1$	$e_K(y_{n-2} \oplus x_{n-1})$	$x_{n+1} \oplus e_K(y_{n-2})$
n	$e_K(y_{n-1} \oplus x_n)$	$e_K(y_{n-1})$

For $i = 1$ we note that $y_0 \oplus x_1 = x_1$ in CBC as $y_0 = 00\cdots0$. Hence we have $e_K(x_1)$ which is almost the same as $x_2 \oplus e_K(x_1)$ in the CFB-mode. In fact we see that this patter continues throughout the whole computations and for every $i = 1, \cdots n-1$ the CBC "lag" behind with a XOR-operation. This is remedied in the last step, however, and therefore they are identical.