MT362/462/5462 Cipher Systems: Sheet 1

Attempt every question. Please remember to write your name or student number. Submit your work through Moodle. Instructions are under 'General Information' on the Moodle page. The eight problem sheets are worth 15% of your final mark.

Question 1(b) is deliberately similar to the first group work problem in Week 2. The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Friday 9th October.

It is helpful if you indicate questions you did but are uncertain about, or would like done in the Q&A session.

The MATHEMATICA notebook `AlphabetCiphers.nb` on Moodle can encrypt and decrypt using substitution ciphers, and compute frequencies and the Index of Coincidence. Remember '**Evaluate Notebook**' to get started.

1. In Example 1.2, Alice agreed to send Bob his exam mark $x \in \{0, 1, \ldots 99\}$ by encrypting it as the ciphertext $(x+k)$ mod 100. Assume that the key $k \in \{0, 1, \ldots 99\}$ is known to only Alice and Bob and is chosen at random. Eve, the eavesdropper, learns all messages that Alice sends to Bob. Malcolm, the man-in-the-middle learns the message *and* can change it. Following Kerckhoff's Principle, everyone knows all this.

The only secret information, known only to Alice and Bob, is the key.

(a) If Eve observes the ciphertext 20, what if anything, can she learn about the plaintext x?

(b) Suppose Eve believes Bob got 0 with probability $\frac{1}{5}$, and that if not, each mark between 40 and 79, is equally likely. Let K be the random key and let Y be the random ciphertext.

(i) If Eve observes the ciphertext 17, so $Y = 17$, what can she learn about the key? [*Hint:* could it be that $K = 30$?]

(ii) What is $\mathbb{P}[K = 85|Y = 17]$? [*Hint:* the answer is *not* $\frac{1}{100}$.]

(iii) What is $\mathbb{P}[K = 45|Y = 17]$?

(iv) Sketch a graph showing $\mathbb{P}[K = k|Y = 17]$ as k varies between 0 and 99.

(c) Malcolm, the man-in-the-middle, can modify the ciphertext. Suppose he is confident that Bob's mark is between 40 and 79. (He does not know the key.) Can he trick Bob into thinking he failed?

(d) Suppose that next year Alice sends Bob her own exam mark $x' \in \{0, 1, \ldots, 99\}$ using the same cryptoscheme, *and using the same key* k. What can Eve learn?

(e) Suppose the scheme is simplified so that Alice sends $x + k$, without reducing modulo 100. What are some problems with this simplified scheme?

(f) Find a way to modify the scheme to avoid the problem in (c). [*Hint:* a cryptosystem may have more ciphertexts than plaintexts.]

2. Decrypt `BYIKVXRYVVYGKI`, assuming it is the ciphertext output by a Caesar cipher. What is the key?

3. In the first week you were assigned to a *block* of four people, identified as Alice, Bob, Alice$'$, Bob$'$. The *pairs* are {Alice, Bob} and {Alice$'$, Bob$'$}. You were then emailed a substitution cipher key. Each person in a pair has the same key.

(a) Write a plaintext message x of at least 75 words on a subject of your choice, and encrypt it using your substitution cipher key π. (Keep the spaces please!) Email the ciphertext $e_\pi(x)$ to all three people in your block.

(b) Decrypt the message from the other person in your pair. [*Hint:* do **not** use frequency analysis!]

(c) Using frequency analysis, decrypt either of the messages sent to you by a person not in your pair.

(d) Write up (c), explaining your method. (An annotated printout is fine.) Did you learn the entire key? If you only looked at one message, why might using both, but still decrypting only one, have been easier?

4. The ciphertext below is the output of a Vigenère cipher. Each line has length 50.

```
01234567890123456789012345678901234567890123456789
CQUAJHXHVWGJMRTAJHBPIHTLTHIPRKKYTHWBKUKZCUKWGDZLFZ
UYFLTAJHIPRKKYVHDAVKOZKVUMVHTKWHZVVZULSXGSYRXKULTA
JHVSCLTAGAZPPSUZKWOVPVJPHIKYKQMIADSBNWOWNHUMVKKSGQ
MAJRLAJHQLALZPUNTVYQGZMDYPUNOZVHYA
```

(a) Compute the Index of Coincidence on the samples of size 20 (or larger if you prefer) obtained by taking every m-th position in the ciphertext starting with the first letter `C` [**correction, not** `W`]. in position 0, for each $m \in \{2, 3, 4, 5, 6\}$.

For example the sample for $m = 3$ of size 20 is `CAXWMABHTPKHKZKDFYTH`. To get these samples in MATHEMATICA, evaluate `AlphabeticCiphers.nb`; then `StringTake[SplitText[Q5Ciphertext, 3][[1]], {1, 20}]`.

(b) What does this suggest about the key length?

(c) Determine the key. [*Hint:* `AJH` appears starting in positions 3, 15, 55, ...] [**Corrected off-by-one error.**] What are the final two words of the ciphertext?

(d) Explain why the Index of Coincidence is largest for $m = 4$, smallest for $m = 3$ and $m = 5$ and in the middle for $m = 2$ and $m = 4$. Give a detailed answer referring to the Caesar shifts that are relevant to each ciphertext sample.

(e) How could you use the positions of `AJH` to guess the length of the key? [As you will know, if you did a complete decryption, this is the *Kasiski test.*]

MT362/462/5462 Cipher Systems: Sheet 2

Attempt at least Questions 1 to 4, and also 5 and 6 if you are an M.Sc. student. (All students may swap one of the first four questions for a later optional question.) Please remember to write your name or student number. Submit your work through Moodle. Instructions are under 'General Information' on the Moodle page. The eight problem sheets are worth 15% of your final mark.

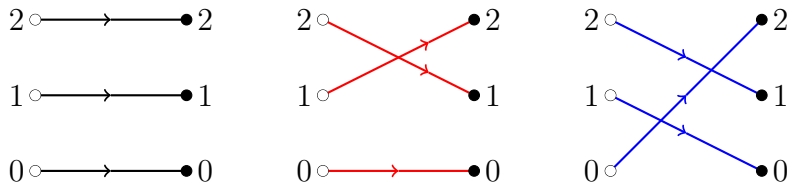
The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Friday 23rd October. Note you have a fortnight to do this sheet.

It is helpful if you indicate questions you did but are uncertain about, or would like done in the Q&A session.

Throughout we use the notation of §3, so $\mathcal{K}$ is the keyspace, $\mathcal{P}$ the plaintexts and $\mathcal{C}$ the ciphertexts in a cryptosystem, with encryption functions $e_k : \mathcal{P} \to \mathcal{C}$ and decryption functions $d_k : \mathcal{C} \to \mathcal{P}$ indexed by keys $k \in \mathcal{K}$.

1. The cryptosystem shown below uses three keys from the affine cipher on $\mathbb{Z}_3$, each with probability $\frac{1}{3}$. Suppose that plaintext 1 is sent with probability p and plaintext 2 is sent with probability $1-p$, so plaintext 0 is never sent.



(a) Recall that $e_{(a,c)}(x) = ax + c \bmod 3$. Which keys (a, c) are used in this cryptosystem?

(b) Find $\mathbf{P}[Y = 1|X = 1]$. (Your answer should not involve p.)

(c) Express $\mathbf{P}[Y = 1]$, $\mathbf{P}[X = 1|Y = 1]$ in terms of p.

(d) When does the cryptosystem have perfect secrecy with respect to the probability distribution $p_0 = 0$, $p_1 = p$, $p_2 = 1 - p$ on plaintexts?

2. Alice and Bob communicate using the numeric one-time pad cryptosystem from Example 3.5, in which $\mathcal{K} = \mathcal{P} = \mathcal{C} = \{0, 1, \ldots, n-1\}$ and the encryption functions are defined by $e_k(x) = (x + k) \bmod n$. Each key $k \in \mathcal{K}$ is chosen with equal probability. Let p_x be the probability distribution on plaintexts: $\mathbf{P}[X = x] = p_x$.

(a) Show that if $x \in \mathcal{P}$ and $p_x > 0$ then $\mathbf{P}[Y = y|X = x] = \frac{1}{n}$ for all $y \in \mathcal{C}$. [*Note:* we need $p_x > 0$ since the conditional probability $\mathbf{P}[Y = y|X = x]$ is only defined when $\mathbf{P}[X = x] > 0$.]

(b) Find $\mathbf{P}[Y = y]$ for each $y \in \mathcal{C}$. [*Hint:* condition on the plaintext: you may use the convention that $\mathbf{P}[Y = y|X = x]\mathbf{P}[X = x]$ is taken as 0 if $\mathbf{P}[X = x] = 0$. The correct answer can be stated without using x or k.]

(c) Hence show that $\mathbf{P}[X = x|Y = y] = p_x$ for all $x \in \mathcal{P}$ with $p_x > 0$.

(d) What is $\mathbf{P}[X = x|Y = y]$ if $p_x = 0$? Deduce from this and (c) that the numeric one-time pad has perfect secrecy.

3. (a) Is there a cryptosystem such that $|\mathcal{C}| < |\mathcal{P}|$?

(b) Is there a practical (see Definition 3.9) cryptosystem with perfect secrecy such that $|\mathcal{K}| < |\mathcal{C}|$?

(c) A student writes: 'since the encryption functions e_k are injective, if $k \neq k'$ then $e_k(x) \neq e_{k'}(x)$'. Is this claim correct? Justify your answer with a proof or counterexample, as appropriate. If the claim is wrong, can you identify the misconception?

(d) Give at least three different examples of cryptosystems with perfect secrecy such that $|\mathcal{P}| = |\mathcal{K}| = |\mathcal{C}| = 4$. [*Hint:* Latin squares. Please make it clear how a Latin square defines a cryptosystem.]

4. In a *chosen plaintext attack*, **you choose** a plaintext x. You are then **given** the corresponding ciphertext $e_k(x)$ for the key k.

Explain how to find the key by a chosen plaintext account when the cipher is (a) a Caesar cipher, (b) a substitution cipher e_π; (c) a Vigenère cipher e_k where k has length exactly 10. Make it clear which plaintexts the attacker should choose. [*Hint:* if you write about frequency analysis in (b) you have missed the point.]

5. (M.Sc.) Work with the Shamir secret sharing scheme over $\mathbb{F}_{11}$ with 5 people and threshold 3 using evaluation points $c_i = i$ for $i \in \{1, 2, 3, 4, 5\}$.

(a) Find the shares for the secret $5 \in \mathbb{F}_{11}$, choosing a polynomial at random.

(b) Alice (Person 1), Bob (Person 2) and Charlie (Person 3) have the shares 7, 5, 3 respectively. The three agree to meet, simultaneously reveal their shares, and together compute the secret.

(i) What is the secret?

(ii) Show, by giving an explicit example, that if Alice lies about her share to Bob and Charlie, then she can both learn the secret and leave Bob and Charlie knowing an incorrect secret.

(iii) Suggest a way to avoid some of the problems in (ii).

6. (M.Sc.) If $x \in \mathbb{N}_0$ and $x < 2^\ell$ then x has a unique expression as $2^{\ell-1}x_{\ell-1} + \cdots + 2x_1 + x_0$ where each x_i is a bit in $\{0, 1\}$. We define the *ℓ-bit binary form* of x to be $x_{\ell-1} \ldots x_1 x_0$. For instance since $11 = 2^3 + 2 + 1$, the 5-bit binary form of 11 is 01101. We write modular addition as $\boxplus$.

For $j \in \{0, 1, \ldots, \ell - 1\}$, let $f_j : \mathbb{F}_2^\ell \to \mathbb{F}_2$ be the Boolean function defined so that $f(x_{\ell-1}, \ldots, x_1, x_0)$ is the bit in position j of $x_{\ell-1} \ldots x_1 x_0 \boxplus 5 \bmod 2^\ell$.

For example, taking $\ell = 4$, since $6 = 0110_2$ $6 \boxplus 5 = 11$ and $11 = 1011_2$, we have $f_3(0110) = 1$, $f_2(0110) = 0$, $f_1(0110) = 1$ and $f_0(0110) = 1$.

Express f_0, f_1, f_2, f_3 as polynomials in x_3, x_2, x_1, x_0. What is the coefficient of the monomial $x_0x_1x_2$ in f_3?

For general j, what is the monomial with the maximum number of variables in f_j?

Optional questions

($\star$) marks those that are harder than average. M.Sc. students are encouraged to look at these.

7. This question asks you to prove the converse result to Shannon's Theorem (Theorem 3.12) stated in Proposition 3.14. Suppose that $|\mathcal{P}| = |\mathcal{C}| = |\mathcal{K}|$. Let n be the common value. Show that if each key is used with equal probability and for all $x \in \mathcal{P}$ and $y \in \mathcal{C}$ there is a unique key k such $e_k(x) = y$, then

(a) $\mathbf{P}[Y = y] > \frac{1}{n}$ for all $y \in \mathcal{C}$;

(b) the cryptosystem has perfect secrecy.

8. To define perfect secrecy for a general cryptosystem, we require that

$$\mathbf{P}[X = x | Y = y] = p_x$$

for all probability distributions p_x on the plaintexts and all $y \in \mathcal{C}$ such that $\mathbf{P}[Y = y] > 0$. Note that for practical cryptosystems $\mathbf{P}[Y = y] > 0$ always holds, so this is the usual definition. The aim of this question is to show that without the practicality assumption, Shannon's Theorem may fail in various ways.

(a) Show that if the hypothesis 'for all $y \in \mathcal{C}$ there exists $x \in \mathcal{P}$ and $k \in \mathcal{K}$ such that $e_k(x) = y$' is dropped then conclusions (a), (b) and (c) of Theorem 3.12 may fail to hold. (Define perfect secrecy by $\mathbf{P}[X = x|Y = y] = p_x$ whenever $P[Y = y] > 0$.)

(b) Show that if the hypothesis '$\mathbf{P}[K = k] > 0$ for each $k \in \mathcal{K}$' is dropped then again (a), (b) and (c) may fail to hold.

(c) Show that there is a (non-practical) cryptosystem with perfect secrecy with $|\mathcal{K}| < |\mathcal{C}|$, so (d) fails to hold.

9. This extends Question 2 on the Group Work for Week 2 on a toy model for the Vigenère Cipher in which the alphabet is `a`, `b`, `c`, `d`, shifts are modulo 4, and letters `a` and `d` are *common*, each with probability $\frac{1}{3}$ and letters `b` and `c` are *rare*, each with probability $\frac{1}{6}$.

(l) In (e) we supposed that in a typical ciphertext, the letters had probabilities q_0, q_1, q_2, q_3. Find with proof a formula for the IOC of a long ciphertext as the length tends to infinity.

(m) What probability distribution q_0, q_1, q_2, q_3 minimizes the IOC just found? What is this IOC? Proof this is the unique minimum. [*Hint:* write $q_j^2 = (q_j - c)^2 + 2cq_j - c^2$ for a suitable c.]

(n) Using (ii) find all keys of length 2 that minimize the IOC of a typical long ciphertext. (This extends (k) in the Group Work.)

(o) What keys maximize the IOC? ($\star$) Prove your answer is correct using any standard inequalities of your choice.

10. ($\star$) Prove or disprove by finding a counterexample: in any practical cryptosystem in which $|\mathcal{P}| = |\mathcal{C}|$, each ciphertext is equally likely.

Remark. This is true when $|\mathcal{K}| = |\mathcal{P}| = |\mathcal{C}|$ by Theorem 3.12(d).

11. ($\star$) Consider the cryptosystem obtained from the numeric one-time pad on $\{0, 1, \ldots, n-1\}$ by removing the key 0 corresponding to the identity permutation.

Give a necessary and sufficient condition on the distribution p_x for $x \in \{0, 1, \ldots, n-1\}$ on plaintexts for there to exist a probability distribution r_k for $k \in \{1, \ldots, n-1\}$ on the $n-1$ keys such that each ciphertext is equally likely.

(Suggested by a question in the Q&A session in 2020 Teaching Week 1.)

MT362/462/5462 Cipher Systems: Sheet 3

Attempt at least questions 1 to 4. Question 6 is compulsory for **M.Sc.** students. Please remember to write your name or student number. Submit your work through Moodle. Instructions are under 'General Information' on the Moodle page. The eight problem sheets are worth 15% of your final mark.

The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Friday 30th October.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in the plenary session.

1. Consider the affine cipher (see Example 4.2) with $q = 151$.

(a) Decrypt the ciphertext 138 sent using the key $(12, 10)$.

(b) In a known plaintext and ciphertext Mark learns that $e_{(a,c)}(21) = 18$. Find all the possibilities for the key (a, c). Suppose that later he learns that $e_{(a,c)}(18) = 21$. What is the key?

2. Let q be prime. Suppose that Alice and Bob communicate using the affine cipher on $\mathbb{Z}_q$ with keyspace $\mathcal{K} = \{(a, c) : a, c \in \mathbb{Z}_q, a \neq 0\}$, and that Alice's plaintext is $x \in \mathbb{Z}_q$ with probability p_x.

(a) What is the size $|\mathcal{K}|$ of the keyspace?

(b) Show that for each x, $y \in \mathbb{Z}_q$ there are exactly $q-1$ keys k such that $e_k(x) = y$.

(c) Show that if each key is equally probable then the cryptosystem has perfect secrecy. Can Eve learn anything about the plaintext from a known ciphertext?

(d) Show that the key can be determined by a chosen plaintext attack using two plaintexts. Does this contradict perfect secrecy? Does a single plaintext suffice?

(e) Can Malcolm, the man-in-the-middle, modify a ciphertext without Bob noticing? How might this problem be reduced? [*Hint:* change $\mathcal{P}$ to a subset of $\mathbb{Z}_q$.]

3. Show that if K and X are independent random variables, taking values in sets $\mathcal{K}$ and $\mathcal{P}$ respectively, then

$$H(K, X) = -\sum_{k \in \mathcal{K}} \sum_{x \in \mathcal{P}} \mathbb{P}[K = k]\mathbb{P}[X = x](\log_2 \mathbb{P}[K = k] + \log_2 \mathbb{P}[X = x]).$$

Deduce that $H(K, X) = H(K) + H(X)$. [*Hint:* please explain your steps, taking care to use sigma notation correctly. The joint entropy $H(K, X)$ is defined in Definition 5.6.]

4. Alice the Spy Master has just learned that Bob, her field agent, has been uncovered. She warns him using the one-time pad kept strictly for emergency use.

(a) Eve the Eavesdropper observes the ciphertext `xioneh`. Find the key if the plaintext is (i) `report`, (ii) `status` (iii) `fluffy`. Can Eve learn anything about the plaintext?

(b) Eve then observes `yadhpu` also sent by Alice.

(i) What, in fact, is the key?

(ii) What are Alice's two messages?

(c) Finally Eve observes `albusa` sent by Bob. What port should she guard?

[*Hint:* the code used in the video is online at `repl.it/@mwildon/OneTimePad2`. (Try Google Chrome if it doesn't work in your first choice of browser.) You can also use the MATHEMATICA notebook `AlphabetCiphers` to add and subtract strings: skip to end for Example 4.8 and this question.]

5. Shannon's estimate for unicity distance (the length of ciphertext needed to determine a key when the plaintext is an English message) is $H(K)/R$, where R is the per-character redundancy in English.

Given that $R \approx 3.2$ bits, what is his estimate for the substitution cipher, when all keys are equally likely? [*Hint:* you need to count the number of different permutations of $\{\mathtt{a}, \ldots, \mathtt{z}\}$; then use Example 5.5(2) to find $H(K)$.]

6. (a) **(All M.Sc.)** Let $f : \mathbb{F}_2^3 \to \mathbb{F}_2$ be defined by $f(x_0, x_1, x_2) = x_1x_2$. Find all the correlations $\mathrm{corr}(f, L_T)$ for $T \subseteq \{0, 1, 2\}$ and hence check Theorem 4.7(c) that

$$(-1)^f = \sum_{T \subseteq \{0,1,2\}} \mathrm{corr}(f, L_T)(-1)^{L_T}$$

(b) Let $S \triangle T = \{u \in S \cup T : u \notin S \cap T\}$. Show that if f is an n-variable Boolean function then $\mathrm{corr}(f + L_S, L_T) = \mathrm{corr}(f, L_{S \triangle T})$.

(c) Let $g(x_0, x_1, x_2) = x_0 + x_1x_2$. Express $(-1)^g$ in the form in (a). [*Hint:* use (b) and Theorem 4.7(c).]

7. Let X and Y be random variables taking values in sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $f : \mathcal{X} \to \mathcal{Y}$ be a function.

(a) Prove the inequality $H\big(f(X) \,|\, Y\big) \leq H(X|Y)$ used in the 'extras' for Part A. [*Hint:* one proof uses the Chaining Rule, Lemma 5.9.]

(b) Show that if f is injective then equality holds. Does the converse hold?

8. What is Shannon's estimate of unicity distance for the Vigenère Cipher with equiprobable keys of length ℓ? Is it a reliable estimate in this case? Discuss.

MT362/462/5462 Cipher Systems: Sheet 4

Attempt at least questions 1 to 3. Question 4 is compulsory for **M.Sc.** students. Please remember to write your name or student number. Submit your work through Moodle. Instructions are under 'General Information' on the Moodle page. The eight problem sheets are worth 15% of your final mark.

The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Friday 13th November.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in the plenary session.

The MATHEMATICA notebook `LFSRs.nb` used is available from Moodle. By definition, the LFSR of width ℓ with taps T, where $T \subseteq \{1, 2, \ldots, \ell\}$, has keystream $k_0k_1k_2\ldots$ such that $k_s = \sum_{t\in T} k_{s-t}$ for all $s \geq \ell$.

1. By Definition 6.8, the *period* of a keystream is its length until its first repeat. For instance $00110011\ldots$ has period 4. Let G be the LFSR of width 5 with taps $\{4, 5\}$.

(a) (i) Let $k = 10000$. Calculate the keystream $k_0k_1k_2, \ldots$, defined by G for k. What is the period of this keystream?

(ii) Find s such that $(k_s, k_{s+1}, k_{s+2}, k_{s+3}, k_{s+4}) = 11100$.

(iii) Do all binary words of length 5 appear somewhere in the keystream $k_0k_1k_2\ldots$?

(iv) How would your answer to (i) change if the key was 11100?

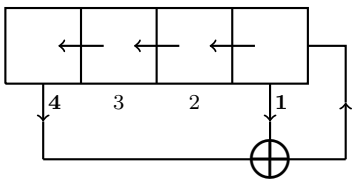
(b) Find a key k' such that the keystream defined by G for k' has period 7.

(c) Find all the periods of keystreams for G. What is the lowest common multiple of the periods?

(d) By Definition 5.8, the *period* of G is the least m such that $G^m = \mathrm{id}$, the identity function. What is the period of G?

[You can use your answer to (a) in Question 3.]

2. Let F be the LFSR of width 4 with taps $\{1, 4\}$, as shown in the circuit diagram below; the numbers correspond to the four positions that may be tapped.



(a) Solve the equation $F\big((x_0, x_1, x_2, x_3)\big) = (y_0, y_1, y_2, y_3)$ and hence find a formula for F^{-1}.

(b) Draw a circuit diagram for F^{-1}. [*Hint:* be careful with the directions of the wires.] According to the strict wording of Definition 6.2, is F^{-1} an LFSR?

3. Let F be an LFSR of width ℓ with taps T, so by definition

$$F\big((x_0, x_1, \ldots, x_{\ell-2}, x_{\ell-1})\big) = (x_1, x_2, \ldots, x_{\ell-1}, \sum_{t\in T} x_{\ell-t}).$$

(a) Show that: F is invertible $\implies \ell \in T$. [*Hint:* use the contrapositive.]

(b) Show conversely that if $\ell \in T$ then F is invertible and give a formula for F^{-1}.

4. (**M.Sc.**) Let $k_0k_1k_2\ldots$ and $k'_0k'_1k'_2\ldots$ be keystreams of LFSRs with taps T and T' and widths ℓ and ℓ', respectively. Let $u_s = k_s + k'_s$ for $s \in \mathbb{N}_0$.

(a) Use annihilators to show that $u_0u_1u_2\ldots$ is a keystream of the LFSR of width $\ell + \ell'$ and feedback polynomial $g_T(z)g_{T'}(z)$, as claimed in Corollary 5.5.

[*Hint:* let $\kappa(z) = k_0 + k_1z + k_2z^2 + \cdots$ and $\kappa'(z) = k'_0 + k'_1z + k'_2z^2 + \cdots$ be the power series representing these keystreams. Show that $\kappa(z) + \kappa'(z)$ annihilated by $g_T(z)g_{T'}(z)$.]

(b) Give an example where $u_0u_1u_2\ldots$ is also the keystream of an LFSR of strictly smaller width than $\ell + \ell'$.

5. This question generalizes the result in Question 1(d). Suppose that an invertible LFSR $F : \mathbb{F}_2^\ell \to \mathbb{F}_2^\ell$ has keystreams of periods $p^{(1)}, \ldots, p^{(r)}$.

(a) Suppose that $F^m = \mathrm{id}$. Show using **(VUP)** that m is divisible by all of $p^{(1)}, \ldots, p^{(r)}$. [*Hint:* if $k \in \mathbb{F}_2^\ell$ has keystream of length p, what does **(VUP)** say about $F^s(k)$ for $1 \le s < p$?]

(b) Let P be the lowest common multiple of $p^{(1)}, \ldots, p^{(r)}$. Show that F^P is the identity function id.

(c) Deduce that, as claimed after Definition 6.8, the period of F is the lowest common multiple of the periods of its keystreams.

6. A *de Bruijn sequence* of *order* ℓ is a cyclic sequence containing every element of $\mathbb{F}_2^\ell$ exactly once. Thus 00010111 is a de Bruijn sequence of order 3; for instance, to find 110, take the final two 1s and the initial 0.

(a) Use the LFSR in Example 6.3 to construct a de Bruijn sequence of order 4.

(b) Prove that there exist de Bruijn sequences of every order. (You may assume there exists an invertible LFSR of period $2^\ell - 1$ for every $\ell \in \mathbb{N}$.)

7. Show that if F is an invertible LFSR then there is a keystream of F whose period is equal to the period of F and hence that the period of F is the maximum of the periods of its keystreams, strengthening the result of Q5(c). [*Hint:* use Lemma 5.4 in the M.Sc. add-on notes; interested people doing MT362/462 will be able to read §5 without earlier M.Sc. extras.]

MT362/462/5462 Cipher Systems: Sheet 5

Attempt questions 1 to 3. Question 4 is compulsory for **M.Sc.** students. Please remember to write your name or student number. Submit your work through Moodle. Instructions are under 'General Information' on the Moodle page. The eight problem sheets are worth 15% of your final mark.

The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Monday 23rd November. Note you have an extra weekend to do this sheet.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in the plenary session.

The MATHEMATICA notebook `LFSRs.nb` used is available from Moodle. By definition, the LFSR of width ℓ with taps T, where $T \subseteq \{1, 2, \ldots, \ell\}$, has keystream $k_0k_1k_2\ldots$ such that $k_s = \sum_{t\in T} k_{s-t}$ for all $s \geq \ell$.

1. In 8-bit ASCII, '`a`' is encoded as the 8-bit binary form of 97, namely 01100001, '`b`' as the binary form of 98, namely 01100010, and so on.

Fix $n \in \mathbb{N}$ and consider the cryptosystem with plaintexts $\mathcal{P} = \{\mathtt{a}, \ldots, \mathtt{z}\}^n$ and ciphertexts $\mathcal{C} = \mathbb{F}_2^{8n}$, in which a message of n characters is first converted to 8-bit ASCII, and then encrypted using the cryptosystem defined in Definition 6.4 with the LFSR F of width 5 with taps $\{3, 5\}$.

Your key is the first 5 bits of the binary key in your email from the lecturer.

(a) Let $k_0k_1k_2\ldots$ be the keystream for your key. What is its period? Show that $k_{31+s} = k_s$ for all $s \in \mathbb{N}_0$ and deduce that $k_{32m} = k_m$ for each $m \in \mathbb{N}_0$.

(b) Encrypt a message (lower-case, no spaces) of at least 25 characters. Send the sequence of bits to everyone in your block.

[*Hint:* to do this in MATHEMATICA, after loading and evaluating `LFSRs.nb` use `EncryptString[{3, 5}, {k0, k1, k2, k3, k4}, "message"]` You can also use `StringToASCIIBits["x"]` to get the 8 bits for `x`, and so on.

(c) Decrypt the message from your partner.

(d) Decrypt either of the messages from the other two people in your block. You **must** explain your method. [*Hint:* start by looking at bits 0 and 32 in the ciphertext. If you do not have a ciphertext to decrypt, use the one in the MATHEMATICA notebook.]

(e) What is the minimum length of ciphertext needed to determine the key?

2. (a) Suppose that $k_0k_1k_2k_3k_4k_5k_6k_7$ is the keystream of an LFSR of width 4. Let $T \subseteq \{1, 2, 3, 4\}$ be the taps. Show that

$$k_s = k_{s-1}\begin{cases}1 & \text{if } 1 \in T\\ 0 & \text{otherwise}\end{cases} + k_{s-2}\begin{cases}1 & \text{if } 2 \in T\\ 0 & \text{otherwise}\end{cases} + k_{s-3}\begin{cases}1 & \text{if } 3 \in T\\ 0 & \text{otherwise}\end{cases} + k_{s-4}\begin{cases}1 & \text{if } 4 \in T\\ 0 & \text{otherwise}\end{cases}$$

for each $s \geq 4$. For instance if the taps are $\{3, 4\}$, this says $k_s = k_{s-1} \times 0 + k_{s-2} \times 0 + k_{s-3} \times 1 + k_{s-4} \times 1$.

Deduce that the matrix equation with unknowns b_1, b_2, b_3, b_4

$$\begin{pmatrix} k_0 & k_1 & k_2 & k_3 \\ k_1 & k_2 & k_3 & k_4 \\ k_2 & k_3 & k_4 & k_5 \\ k_3 & k_4 & k_5 & k_6 \end{pmatrix} \begin{pmatrix} b_4 \\ b_3 \\ b_2 \\ b_1 \end{pmatrix} = \begin{pmatrix} k_4 \\ k_5 \\ k_6 \\ k_7 \end{pmatrix}$$

has a solution for b_1, b_2, b_3, b_4.

(b) The hypothesis in (a) is '$k_0k_1k_2k_3k_4k_5k_6k_7$ is the keystream of an LFSR of width 4'. Call this statement P. Let Q be the conclusion, 'the matrix equation above has a solution'. By (a), $P \implies Q$. Is the converse true?

(c) Which of the bit sequences 00100011, 00100010, 11100001 and 1101100 is a keystream of an LFSR of width 4? (In the last you are only given $k_0k_1 \dots k_6$.) Justify your answers. Do they change if the LFSR is required to be invertible?

3. Let $B_0, B_1, \dots, B_{n-1}$ be a sequence of bits, each 0 or 1 independently with probability $\frac{1}{2}$. For b, $b' \in \{0, 1\}$, let $M_{bb'}$ be the number of $s \in \{0, \dots, n-2\}$ such that $(B_s, B_{s+1}) = (b, b')$.

(a) Show that the expected value of M_{00} is $\mathbb{E}[M_{00}] = (n-1)/4$ and find $\mathbb{E}[M_{01}], \mathbb{E}[M_{10}], \mathbb{E}[M_{11}]$.

(b) For the sequence below $M_{00} = 4$. Write down the statistics M_{10}, M_{01}, M_{11}.

$$\begin{array}{c} (0,1,0,1,1,0,0,1,0,1,0,1,0,1,0,1,0,0,1,1,0,0,1,1,0,0,1,0,1,0,1,1,0) \\ 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 0\ 1\ 2 \end{array}$$

(c) Perform a χ^2-test on $M_{00}, M_{01}, M_{10}, M_{11}$ to test the sequence in (b) for randomness on pairs of bits. [*Hint:* use $M_{00} + M_{01} + M_{10} + M_{11} = 32$ to find the number of degrees of freedom. Tables of the χ^2 distribution are on the web.]

(d) Does the sequence in (b) pass the Monobit Test in Exercise 7.4?

4. (a) Let $u_0u_1 \dots u_{n-1} \in \mathbb{F}_2^n$. Let $U_n(z) = u_0 + u_1z + \dots + u_{n-1}z^{n-1}$ be the corresponding polynomial. Prove that $u_0u_1 \dots u_{n-1}$ is the output of the LFSR with width ℓ and taps $T \subseteq \{1, \dots, \ell\}$ if and only if $U_n(z)g_T(z) = h(z) + z^n r(z)$ for some polynomials $h(z)$ and $r(z)$ with $\deg h < \ell$.

[*Hint:* adapt the proof of Lemma 5.4. The proof given in the M.Sc. Week 7 Plenary Session is on Moodle.]

(b) The LFSR of width 3 with taps $\{2, 3\}$ and the LFSR of width 5 with taps $\{2, 4, 5\}$ generate the top two keystreams below. Let $u_0u_1 \dots u_9u_{10}$ be the bottom sequence.

$$\begin{array}{c} 1,0,1,1,1,0,0,1,0,1,1 \\ 1,0,1,1,1,0,0,0,0,1,0 \\ 1,0,1,1,1,0,0,0,0,1,1 \\ 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 0 \end{array}$$

(i) Write down the feedback polynomials $f_{\{2,3\}}(z)$ and $f_{\{2,4,5\}}(z)$.

(ii) What is the smallest m such that the LFSR with taps $\{2, 3\}$ generates $u_0u_1 \dots \overline{u}_m$?

(iii) The LFSR with taps $\{2, 4, 5\}$ generates $u_0u_1 \dots u_9\overline{u}_{10}$. Use Proposition 6.5 to find the feedback polynomial and hence the taps of an LFSR generating $u_0u_1 \dots u_9u_{10}$.

(iv) Is there an LFSR of width 5 generating $u_0u_1 \dots u_9u_{10}$? Justify your answer.

MT362/462/5462 Cipher Systems: Sheet 6

Attempt questions 1 and 2. Questions 3 and 4 are compulsory for **M.Sc.** students. Please remember to write your name or student number. Submit your work through Moodle. Instructions are under 'General Information' on the Moodle page. The eight problem sheets are worth 15% of your final mark.

The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Monday 30th November.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in the plenary session.

The MATHEMATICA notebook `LFSRs.nb` used is available from Moodle under Teaching Week 8, as is `QuadraticStreamCipher.nb`. By definition, the LFSR of width ℓ with taps T, where $T \subseteq \{1, 2, \ldots, \ell\}$, has keystream $k_0k_1k_2\ldots$ such that $k_s = \sum_{t\in T} k_{s-t}$ for all $s \geq \ell$.

1. Let $(k_0, k_1, k_2, \ldots)$ be a keystream of the LFSR F of width 2 with taps $\{1, 2\}$. Let $(k'_0, k'_1, k'_2, \ldots)$ be a keystream of an LFSR G of width 3 with unknown taps. The keystreams are multiplied to give $(k_0k'_0, k_1k'_1, k_2k'_2, \ldots)$. Suppose you know the product begins 101100000101.

(a) Explain why the keystreams of F and G have the form $1\star11\star\star\star\star1\star1$, where $\star$ denotes an unknown bit. By considering the possible keystreams produced by F, deduce the key for F.

(b) By considering the keystream for F explain why the keystream of G is of the form $1\star11\star00\star01\star1$. Hence find a possible set of taps and the unique key for G.

(c) Are the taps you found in (b) unique? Justify your answer.

2. Let F be the Feistel Network for the function $f : \mathbb{F}_2^m \to \mathbb{F}_2^m$ so, by definition, $F\big((v, w)\big) = (w, v + f(w))$ for $(v, w) \in \mathbb{F}_2^{2m}$.

(a) Compute $F\big((0001, 0001)\big)$ in the special case when $m = 4$ and $f(x_0, x_1, x_2, x_3) = (x_2, x_3, x_0 + x_1x_2, x_1 + x_2x_3) + (1, 1, 1, 1)$.

(b) In this part your argument should work for general m and f. The Feistel network is used to encrypt a plaintext (v, w) to a ciphertext $(v', w') = F\big((v, w)\big)$.

(i) Show that $w = v'$ and express v in terms of v' and w'.

(ii) Show that (w, v) is the encryption of (w', v').

(iii) You have a black box that implements F. That is: given any (v, w), the box will output the encryption $(v', w') = F\big((v, w)\big)$. Suppose you are given (v', w'). Can the black box be used to find (v, w)? That is, can you use the black box to decrypt?

3. (**M.Sc.**) The table below shows the first 14 steps in the Berlekamp–Massey algorithm applied to the sequence

$$(u_0, u_1, \ldots, u_{14}) = (1,0,1,1,1,1,1,1,1,0,1,1,1,1,0)$$
$$\phantom{(u_0, u_1, \ldots, u_{14}) =} \ 0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 0\ 1\ 2\ 3\ 4$$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ℓ_n	1	1	2	2	3	3	3	3	3	7	7	7	7	7	7
T_n	$\varnothing$	$\varnothing$	$\{2\}$	$\{1,2\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\star$	$\{1,5,6,7\}$	$\star$	$\star$	$\star$	$\star$
m	0	0	2	2	4	4	4	4	4	9	9	9	$\star$	9	

For instance, the LFSR F_9 has length $\ell_9 = 3$ and taps $T_9 = \{1\}$. Performing step 9 of the algorithm using $m = 4$ gives the LFSR F_{10} of length $\ell_{10} = 7$ and taps T_{10} that you are asked to find in (i). Since the length goes up, m is updated to 9. You should find that the final LFSR F_{15} generating all 15 bits has taps $\{1,2,3,4,6,7\}$.

(i) Verify that case (a) applies for steps $5,6,7,8$ and perform step 9 to obtain the entry marked $\star$ in the column for $n = 10$.

(ii) Find the five remaining entries marked $\star$.

(iii) Given that the entire sequence $u_0, u_1, u_2, \ldots$ is generated by an LFSR of width 7, will the taps change in further steps of the Berlekamp–Massey algorithm? Justify your answer.

4. The 2-quadratic stream cipher was defined in Example 8.5. Recall that F is the LFSR of width 5 with taps $\{3,5\}$ and F' is the LFSR of width 6 with taps $\{2,3,5,6\}$. Given keys $k \in \mathbb{F}_2^5$ and $k' \in \mathbb{F}_2^6$, the keystream $u_0u_1u_2\ldots$ is defined by $u_0 = 0$ and $u_s = k_s k'_s + k_{s-1}k'_{s-1}$ for each $s \in \mathbb{N}$.

Using the attack in this example, the attacker guesses that k is $v_0v_1v_2v_3v_4$ and computes the correlation between the keystream $v_0v_1\ldots v_{1023}$ and $u_0u_1\ldots u_{1023}$. (Here $u_0u_1\ldots u_{1023}$ is obtained via a chosen plaintext attack, as in Exercise 7.1.)

The keystream for $k = 00001$ has period $2^5 - 1 = 31$ and has the form **00001**00101$\ldots\star$**00001**.

(a) What is the bit in position 30, marked $\star$ above, just before the first repeat of the key **00001**?

(b) What are the bits $u_{26}u_{27}u_{28}u_{29}u_{30}$? [*Hint:* rather than compute 30 bits, you could try applying the LFSR 'in reverse', thinking of k'_{30} as k'_{-1}, and so on.]

The table below shows the four guessed keys $v_0v_1v_2v_3v_4$ with the highest correlations for several different k and k'. In each case the correlations for the other 32 guessed keys are close to 0. (Use the MATHEMATICA notebook `QuadraticStreamCipher.nb` under Teaching Week 8 if you want to check this.)

k	k'	guessed key, correlation			
00001	000001	00000, 0.223	00001, 0.242	10000, 0.230	10001, 0.203
00001	000011	00000, 0.230	00001, 0.215	10000, 0.219	10001, 0.211
00111	000001	00000, 0.238	00111, 0.199	10011, 0.199	10100, 0.254
00111	000011	00000, 0.199	00111, 0.219	10011, 0.234	10100, 0.254

(c) Explain why in each case there are three 'fake keys', with correlation about $\frac{1}{4}$, as well as the correct key $k_0k_1k_2k_3k_4$. Predict the three fake keys when $k = 01000$ and k' is unknown.

[*Hint:* for $\frac{1}{4}$ of the positions in the F' keystream, $k'_s = 0$ and $k'_{s-1} = 1$ and so $u_s = k_{s-1}$. What keystream for F should $u_0u_1 \dots u_{1023}$ then be compared with? You know the key for this keystream from (a). This should give you one 'fake' key.]

5. The stream cipher Trivium has an 80 bit key $k_0k_1 \dots k_{79} \in \mathbb{F}_2^{80}$. The key is used, together with an 80 bit *initialization vector* v, to generate a keystream $u_0u_1u_2 \dots$. The encryption functions $e_k : \mathbb{F}_2^n \to \mathbb{F}_2^n$ for $k \in \mathbb{F}_2^{80}$ are then

$$e_k(x) = (v, y)$$

where $v \in \mathbb{F}_2^{80}$ is the initialization vector and $y_0 \dots y_{n-1}$ is obtained in the usual way (see Example 6.4) by adding the keystream to the plaintext: thus $y_s = x_s + u_s$ for each s.

(a) Show that anyone knowing the key k can decrypt a Trivium ciphertext (v, y).

In a hypothetical application, a 'test-and-trace' app uses an 80 bit Trivium key to secure its communications. If a user Alice notifies the app that she is infected, the app encrypts the plaintext message 'One of your contacts has been infected: you should self-isolate' and sends it to all of Alice's close contacts known to the app.

Assume that it is not possible to extract the key from the app.

(b) Suppose that the app always uses the all-zeros initialization vector. Bob receives the warning message from the app. Should he believe it?

(c) Suppose that the app chooses a random initialization vector for each message. Should Bob believe the warning message now?

(d) Suppose that when sending a message to Bob, the app uses the initialization vector obtained by converting Bob's unique identifier into an 80-bit vector. Should Bob believe the warning message now?

As a further feature, and only with the user's consent, the app may send a message to a central authority confirming that the user has been required to self-isolate.

(e) How might this be implemented securely?

Remark. The assumption that users cannot extract the key from the app is realistic: smartphones are highly locked down and Android and iOS have specific measures to protect cryptographic keys. Some aspects of this question were discussed in the Q&A Session in Week 8: see the answer posted to the Moodle forum.

6. In an attack on a stream cipher using a key of length ℓ, a correct guess at the first m bits of the key gives a keystream $v_0v_1v_2\ldots$ having average correlation $c > 0$ with the correct keystream $u_0u_1u_2\ldots$.

(a) Assuming that different bits are independent, what is $\mathbb{P}[v_s = u_s]$?

(b) Let $n \in \mathbb{N}$. If the correlation is computed by comparing $v_0v_1\ldots v_{n-1}$ and $u_0u_1\ldots u_{n-1}$ then what is the distribution of the correlation statistic?

(c) Show that if c is nearly 0 (as is typical) then n has to be at least $1/c^2$ for this correlation attack to be effective.

7. The m-quadratic stream cipher in Example 8.5 has keyspace $\{(k, k') : k \in \mathbb{F}_2^5, k' \in \mathbb{F}_2^6\}$. It is used to encrypt plaintexts in $\mathbb{F}_2^{1024}$.

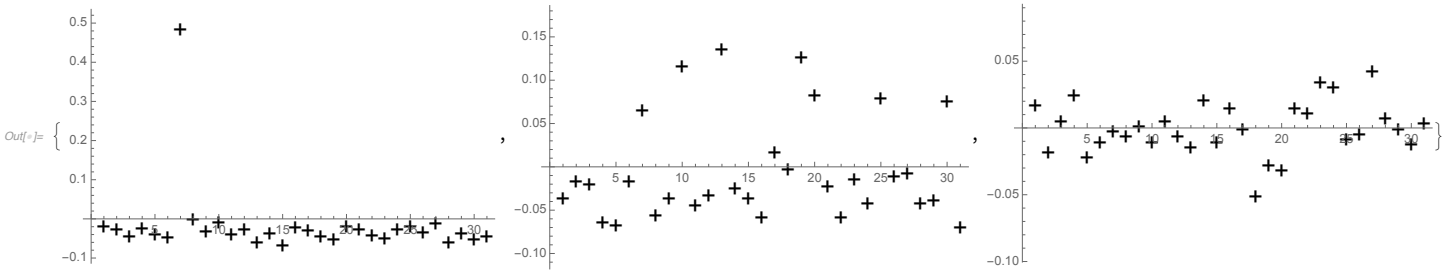
(a) Show that in the attack in Example 8.5, the expected correlation for a correct guess of k is $\frac{1}{2^m}$. [*Hint:* use the Piling Up Lemma, Lemma 4.11 in the (**M.Sc.**) notes.] How many fake keys are there? [*Hint:* an upper bound is enough to know if the attack will work.]

(b) Is the attack effective when $m = 3$? If so, is it subexhaustive? [*Hint:* use Question 4(d).]

(c) Is the attack effective when $m = 5$? If so, is it subexhaustive?

(d) Find a different correlation attack that breaks the 5-quadratic stream cipher.

Remark. The graphs below show correlations for all 31 non-zero keys taking $m = 1$, $m = 3$ and $m = 5$. They were produced using the MATHEMATICA notebook `QuadraticStreamCipher.nb` available from Moodle.



8. Prove the Piling-Up Lemma, Lemma 4.11 in the (**M.Sc.**) notes.

MT362/462/5462 Cipher Systems: Sheet 7

Attempt questions 1 to 3. Question 3(c) is optional. **M.Sc.** students should also attempt question 4. Please remember to write your name or student number. Submit your work through Moodle. Instructions are under 'General Information' on the Moodle page. The eight problem sheets are worth 15% of your final mark.

The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Monday 7th December.

It is helpful if you indicate questions you did but are uncertain about, or would like seen done in the plenary session or Q&A.

The MATHEMATICA notebook `BlockCiphers.nb` is available from Moodle and can be used to check answers to the questions on the Q-block cipher.

1. Let $S : \mathbb{F}_2^4 \to \mathbb{F}_2^4$ be the S-box in the Q-block cipher, defined by $S\big((x_0, x_1, x_2, x_3)\big) = (x_2, x_3, x_0 + x_1x_2, x_1 + x_2x_3)$. Recall from Example 9.5 that the Feistel network in round i of this cipher is

$$(v^{(i-1)}, v^{(i)}) \mapsto \big(v^{(i)}, v^{(i-1)} + S(v^{(i)} + k^{(i)})\big)$$

where $k^{(i)} \in \mathbb{F}_2^4$ is the round key. The encryption functions e_k for $k = (k^{(1)}, k^{(2)}, k^{(3)}) \in \mathbb{F}_2^{12}$ are defined by $e_k\big((v^{(0)}, v^{(1)})\big) = (v^{(3)}, v^{(4)})$.

(a) Encrypt $0000\,0000 \in \mathbb{F}_2^8$ using the key $0011\,0011\,0011$.

(b) Decrypt the ciphertext $0111\,0111$ using the key in (a)

(c) Find a key $k \in \mathbb{F}_2^{12}$ such that $e_k(0001\,0001) = 0000\,0000$.

(d) Does the Q-block cipher have the 'confusion' property: i.e. does it mix up nearby bits in a non-linear way?

(e) Does the Q-block cipher have the 'diffusion' property: i.e. does it mix up bits so that *every* bit of the ciphertext depends on *every* bit of the key?

2. Let $S : \mathbb{F}_2^4 \to \mathbb{F}_2^4$ be the S-box in the Q-block cipher from Question 1.

(a) Let $\boldsymbol{\Delta} \in \mathbb{F}_2^4$. Show that if $\boldsymbol{\Delta}_2 = 0$, i.e. $\boldsymbol{\Delta}$ is of the form $(\star, \star, 0, \star)$ then

$$S(x + \boldsymbol{\Delta}) + S(x) = \begin{cases} (0, \boldsymbol{\Delta}_3, \boldsymbol{\Delta}_0, \boldsymbol{\Delta}_1) & \text{if } x_2 = 0 \\ (0, \boldsymbol{\Delta}_3, \boldsymbol{\Delta}_0 + \boldsymbol{\Delta}_1, \boldsymbol{\Delta}_1 + \boldsymbol{\Delta}_3) & \text{if } x_2 = 1. \end{cases}$$

(b) Deduce Lemma 10.1(i), that $S(x + 1000) = S(x) + 0010$ for all $x \in \mathbb{F}_2^4$.

(c) Using (b) show that $e_k(x) = e_{k+1000\,0010\,1000}(x)$ for all $x \in \mathbb{F}_2^8$.

(d) **change to four rounds so can be consistent with Feistel** In Attack 9.5 the attacker encrypts $x = 1111\,0000$ and $x_\Delta = 1111\,1000$ to ciphertexts z and z_Δ. She then guesses the final two rounds keys $k^{(2)}$ and $k^{(3)}$ and decrypts z and z_Δ over the final two rounds to $w = 0000\,1110$ and $w_{\boldsymbol{\Delta}} = 1000\,1101$. Is the attacker's guess correct?

3. 3DES is the block cipher of block size 64 and keyspace $\mathbb{F}_2^{56} \times \mathbb{F}_2^{56} \times \mathbb{F}_2^{56}$ with encryption functions defined by

$$e_{(k,k',k'')}(x) = e_{k''}\big(d_{k'}\big(e_k(x)\big)\big)$$

where e_k and d_k are the encryption and decryption functions for DES.

(a) Show that there is a meet-in-the-middle attack using multiple chosen plaintexts that finds the key using about 2^{112} encryptions/decryptions.

[*Hint:* see Attack 9.8. A small example of the meet-in-the-middle attack was seen in the Group Work from Week 9; answers are on Moodle.]

(b) Assume no attack better than (a) exists. Is 3DES secure?

(c) ($\star$) Suggest why the middle map is decryption rather than encryption.

4. (**M.Sc.**) Again let $S : \mathbb{F}_2^4 \to \mathbb{F}_2^4$ be the S-box in the Q-block cipher.

(a) Find all possibilities for $S(x + 0010) + S(x)$ where $x \in \mathbb{F}_2^4$.

(b) Let $\mathbf{\Gamma} = 0000\,1000$. Let $(v, w) \in \mathbb{F}_2^8$ be chosen uniformly at random. Let (v', w') and $(v'_{\mathbf{\Gamma}}, w'_{\mathbf{\Gamma}})$ be the encryptions of (v, w) and $(v, w) + \mathbf{\Gamma}$, respectively **over the first two rounds** of the Q-block cipher.

Show that no matter what the key is, $(v', w') + (v'_{\mathbf{\Gamma}}, w'_{\mathbf{\Gamma}})$ is equally likely to be each of the four differences $\{0010\,0000, 0010\,0001, 0010\,0010, 0010\,0011\}$.

(c) Suggest a subexhaustive attack on the Q-block cipher in which the attacker first guesses $k^{(3)}$, and then $(k^{(1)}, k^{(2)})$. Make it clear why your attack is subexhaustive.

(d) Is the attack in (c) an improvement on Attack 10.2, where the attacker guessed $(k^{(2)}, k^{(3)})$, and then with 16 possibilities for this pair, guessed $k^{(1)}$?

5. In a cryptosystem based on the AES S-box, a plaintext $x \in \mathbb{F}_2^8$ is encrypted by a key $(k, k') \in \mathbb{F}_2^{16}$ to the ciphertext $S(x + k) + k'$.

(a) Let $x = 0000\,0000 \in \mathbb{F}_2^8$ and let $\mathbf{\Delta} = \mathbf{0000\,0001}$. Let $y = x + k$ and let $y_{\mathbf{\Delta}} = x_{\mathbf{\Delta}} + k$. What is a simple form for $y + y_{\mathbf{\Delta}}$?

(b) Let $z = S(x + k) + k'$ and let $z_{\mathbf{\Delta}} = S(x_{\mathbf{\Delta}}) + k'$. What is a simple form for $z + z_{\mathbf{\Delta}}$?

(c) Lemma 10.8 immediately implies that unless $\mathbf{\Gamma} = \mathbf{0000\,0001}$, the equation $S(w) + S(w + \mathbf{\Delta}) = \Gamma$ has exactly two solutions $w \in \mathbb{F}_2^8$. What are these solutions when $\mathbf{\Gamma} = z + z_{\mathbf{\Delta}}$?

(d) Hence show that, with one exceptional case, the ciphertexts z and $z_{\mathbf{\Delta}}$ determine $\{k, k + \mathbf{\Delta}\}$.

(e) Is this attack on the cryptosystem a sign that the AES S-box is weak? Justify your answer.

MT362/462/5462 Cipher Systems: Sheet 8

Attempt questions 1 to 4 and 6. **M.Sc.** students should also attempt question 7: the final part ($\star$) is optional. Submit your work through Moodle. The eight problem sheets are worth 15% of your final mark.

The lecturer will be happy to discuss any of the questions in the office hour or the live Q&A session.

To be submitted by midnight on Wednesday 16th December.

Private keys, and other private information, are written in red.

1. (a) Compute 2^{257} mod 10007. [*Hint:* to do this by hand, first compute 2^2, 2^4, 2^8, 2^{16}, …, 2^{128} mod 10007 by repeated squaring: note that $(2^m)^2 = 2^{2m}$.]

(b) Find $\operatorname{dlog}_2 45 \bmod 139$; that is, find m such that $2^m \equiv 45 \bmod 139$. [*Hint:* you could use the MATHEMATICA notebook `PKC.nb` to do a brute-force search.]

Remark. Despite the smaller numbers, you should find that (b) is a harder problem than (a). Correspondingly the exponentiation function $m \mapsto 2^m \bmod p$ is 'one-way': easy to compute but hard to invert.

2. Suppose that Bob's RSA public key is $(2279, 17)$. As Eve you observe the RSA ciphertext 37 sent to Bob. Find Bob's private key and hence find the plaintext.

3. Generate an RSA public key (n, a) with $n > 2^{128}$ and private key (p, q, r). Use the MATHEMATICA notebook `PKC.nb` on Moodle and the `PowerMod` function.

(a) Email your public key to your partner in your block.

(b) Email a message x of your choice, using the RSA Cryptosystem, to your partner in your block. [*Hint:* you know their public key when you receive their email from (a). Your message can be a number between 0 and $n-1$, or if you use the functions in the notebook, an English string.]

(c) Decrypt the message from your partner. [If your partner is uncooperative, you may use the lecturer as a substitute in (a) and (c).]

(d) Suppose all emails are observed by Eve. What, if anything, can she learn?

(e) Suppose all emails can be modified by Malcolm. What, if anything, can he learn?

4. Consider the cryptoscheme in which English plaintexts are converted to 8-bit ASCII ('`a`' ↔ 01100001, '`b`' ↔ 01100010, and so on, as on Problem Sheet 5) and then encrypted using RSA with the appropriate public key.

For example 'hi' becomes 1101000 1101001 which is the binary form of 13409. If Alice's public key is (n, a) then she is sent $13409^a \bmod n$. Assume that $n \approx 2^{2048}$.

(a) Alice is expected an important message 'yes' or 'no' from Bob. Show that Eve can decrypt Bob's ciphertext without knowing Alice's private key.

(b) Can the problem in (a) occur if Alice and Bob use a symmetric cipher such as AES where the key is entirely private? How can it be avoided while still using the RSA cryptosystem?

5. Let (n, a) be Alice's RSA public key. Suppose that $n = pq$. Let $t = (p-1)(q-1)$. Show that an attacker who knows n and t can easily find p and q. [*Hint:* find a quadratic equation for p with coefficients expressed in terms of n and t.]

6. In Diffie–Hellman Key Exchange, we saw that the eavesdropper Eve knows the prime p, the base g and $g^a \bmod p$. Only Alice knows her exponent a. (We write $g^a \bmod p$ entirely in black because although a is private, $g^a \bmod p$ is public.)

Bob wants to send a message $x \in \{1, \ldots, p-1\}$ to Alice.

(a) Suppose Bob sends $xg^a \bmod p$. Show that Eve can find x.

(b) Explain why Bob can send $xg^{ar} \bmod p$ for any private r of his choice. (This is not entirely obvious because Bob knows $g^a \bmod p$ but he does not g^a and he does not know a.) Can Alice find x?

(c) Suppose Bob sends $xg^{ar} \bmod p$ and then sends r. Can Alice find x? Can Eve find x?

(d) Suppose Bob sends $xg^{ar} \bmod p$ and then sends $g^r \bmod p$. Can Alice find x? Can Eve find x?

Remark: (d) is the ElGamal cryptoscheme: Alice publishes (g, g^a, p) as her public key, and keeps (g, a, p) as her private key.

7. (**M.Sc.**) Let $e_k : \mathbb{F}_2^8 \to \mathbb{F}_2^8$ for $k \in \mathbb{F}_2^{12}$ be the encryption maps in the Q-block cipher. Find $\mathrm{corr}(L_{\{0\}} \circ e_k, L_{\{2,5\}})$ and $\mathrm{corr}(L_{\{0\}} \circ e_k, L_{\{2,6\}})$. Assuming you have good estimates for these statistics, and for $\mathrm{corr}(L_{\{0\}} \circ e_k, L_{\{2\}}) = \frac{1}{2}(-1)^{k_0+k_6}$, how many possibilities are there for k? ($\star$) Find some further high correlations that give more information about the key.

8. (**M.Sc.**) Let $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$. Suppose that $\mathrm{corr}(L_U \circ F, L_T) = c > 0$. Let $k \in \mathbb{F}_2^n$ and define $G : \mathbb{F}_2^n \to \mathbb{F}_2^n$ by $G(x) = F(x+k)$.

(a) Show that $\mathrm{corr}(L_U \circ G, L_T) = (-1)^{L_T(k)} c$.

An attacker has a collection $\{(v^{(j)}, v'^{(j)}) : 1 \le j \le q\}$ of chosen plaintext/ciphertext pairs for a cryptosystem defined by $e_k(x) = F(x+k)$. She estimates the correlation by computing $S_j = L_U(v'^{(j)}) + L_T(v^{(j)})$ for each j, and taking $C = \frac{1}{q}\sum_{j=1}^{q}(-1)^{S_j}$.

(b) Let $Z_j = (-1)^{S_j}$. Find $\mathbb{P}[Z_j = 1]$ and $\mathbb{P}[Z_j = -1]$.

(c) Show that if q is large then the distribution of C is approximately normal with mean c and variance $\frac{1-c^2}{q}$. [*Hint:* use the Central Limit Theorem.]

(d) How large must q be for the attacker to be confident of learning $L_T(k)$?

9. A draft of this year's examination paper (the form examinations will take is still to be decided) has been posted to Moodle. It is encrypted using AES. The key is the first 128 bits (or 32 hexadecimal characters) of the SHA-512 hash of the lecturer's password. The SHA-256 hash of this password is

```
170972f840215582a876e057f7b22ff662d77e94526df8e1f57c854ccd29c6c5
```

What cryptographic assumptions are needed, on AES, the SHA-256 hash function, the SHA-512 hash function, and the lecturer's password, to prove that the examination paper is secure? Which is likely to be the weakest link?