MT362/462/5462 Cipher Systems: Preliminary Problem Sheet

Attempt every question. Binary forms, Bayes' Theorem and conditional probability will be used a lot in this course. This sheet gives you a chance to revise them.

This sheet need not be handed in. You are welcome to discuss the questions with the lecturer after lectures or in office hours: Tuesday 3pm, Wednesday 10am, Thursday 11am, or by appointment.

Binary Form. Given $x \in \mathbb{N}_0$ and $n \in \mathbb{N}_0$ such that $2^n > x$, there exist unique *bits* $x_0, x_1, \ldots, x_{n-1} \in \{0, 1\}$ such that $x = 2^{n-1}x_{n-1} + \cdots + 2x_1 + x_0$. We say x has *binary form* $x_{n-1} \ldots x_1 x_0$. For example 19 has binary form 10011; in symbols we write $19 = 10011_2 = 010011_2 = \ldots$.

1. (a) What is the binary form of 43?

(b) What number has binary form 01010101?

(c) If x has binary form $x_{n-1} \ldots x_1 x_0$ with $x_{n-1} = 1$ then we say x has *length* n. Which natural numbers have

(i) length 3?

(ii) length at most 3?

(iii) length n, for a general $n \in \mathbb{N}_0$?

2. (a) A friend knows a number between 0 and 15 (inclusive). How many yes/no questions do you need to guess it? What is the maximum length of the number in binary? Explain why the answers are the same.

(b) How would your answers to (a) change if instead the number is between 0 and 26 (inclusive)?

(c) Now suppose the number is between 0 and 15 (inclusive) but your friend is permitted to lie in the answer to at most one question.

(i) Show that no strategy can guarantee to find the number by asking exactly six questions. [*Hint:* as well as learning four bits of information about the number, you learn about the lie.]

(ii) (Optional, but instructive.) Suggest a good strategy.

3. Let $x \in \mathbb{N}$ have binary form $x_{n-1} \ldots x_1 x_0$.

(a) What is the binary form of $2x$?

(b) Explain how to obtain the binary form of $x + 1$ from $x_{n-1} \ldots x_1 x_0$.

(c) What is the binary form of x mod 8? (Assume that $n \geq 3$.)

(d) What is the binary form of $2^n - 1 - x$? **Introduce bit flip notation here**

(e) Explain how to obtain the binary form of $-x$ mod 2^n from $x_{n-1} \ldots x_1 x_0$. Check your answer works when $x = 43$ and $n = 7$, so -43 mod 128 is 85.

Probability reminder. Let Ω be a probability space. Recall that events $A, B \subseteq \Omega$ are *independent* if $\mathbf{P}[A \cap B] = \mathbf{P}[A]\mathbf{P}[B]$. If $\mathbf{P}[B] > 0$ we define the *conditional probability* of A given B by $\mathbf{P}[A|B] = \mathbf{P}[A \cap B]/\mathbf{P}[B]$. A *random variable* is a function $X : \Omega \to \mathbb{R}$. The *expectation* of X is defined by $\mathbf{E}[X] = \sum_x x\mathbf{P}[X = x]$. See the revision notes on probability on Moodle for more background. You can look up Bayes' Theorem on the web if necessary.

4. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ where $p_i = \frac{1}{6}$ for each $i \in \Omega$ be the probability space for rolls of a fair die.

(a) Let $A = \{2, 4, 6\}$ be the event an even number if rolled. Let $B = \{3, 6\}$ be the event that a multiple of 3 is rolled.

Find $\mathbf{P}[A]$, $\mathbf{P}[B]$, $\mathbf{P}[A \cup B]$, $\mathbf{P}[A \cap B]$. Are A and B independent?

(b) Given an example of events C, $D \subseteq \Omega$ such that $\mathbf{P}[C]$, $\mathbf{P}[D] > 0$ and $\mathbf{P}[C|D] \neq \mathbf{P}[D|C]$.

5. People may have illness F. In its early stages, F has no symptoms. A certain test satisfies $\mathbf{P}[\text{test positive}|F] = 1$ and $\mathbf{P}[\text{test positive}|\text{not } F] = p$.

In words: the test always works on people that have F, but has a false-positive probability of p.

Suppose that one in a thousand people have F.

(a) You have an envelope with the test results in your hands. Which probability is more important to you: $\mathbf{P}[\text{test positive}|F]$ or $\mathbf{P}[F|\text{test positive}]$?

(b) Using Bayes' Theorem (or, equivalently, the definition of conditional probability) calculate $\mathbf{P}[F|\text{test positive}]$.

(c) Comment on your answer to (b) in the cases $p = 0.002$ and $p = 0.0001$. When would you recommend the test be widely used?

6. Let $\Omega = \{\text{HHH}, \text{HHT}, \text{HTH}, \ldots, \text{TTT}\}$ be the probability space for flips of three independent coins, each biased to land heads with probability p. Let $S : \Omega \to \mathbb{R}$ be the total number of heads, so $S(\text{HHH}) = 3$, $S(\text{HHT}) = 2$, and so on.

(a) What is the probability of the outcome $\text{HTH} \in \Omega$?

(b) Find $\mathbf{P}[S = 2]$ in terms of p.

(c) Write down $\mathbf{E}[S]$.

(d) Find $\mathbf{E}[S^2]$ and $\text{Var}[S]$.

MT362/462/5462 Cipher Systems: Sheet 1

Attempt at least questions 1 to 4. Question 5 is compulsory for **Msc** students. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 10am, Thursday 11am, or by appointment.

To be handed in by noon on Wednesday 11th October, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

The MATHEMATICA notebook `AlphabeticCiphers` on Moodle can be used to find frequencies and compute the Index of Coincidence. Please use it!

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

2. You know that Alice and Bob are communicating using a substitution cipher. As Eve you intercept the following ciphertext

```
XNKWBMOW KWH JKXKRJKRZJ RA KWRJ ZWXCKHI XIH IHNRXYNH EBI THZRCWHIRAO
DHJJXOHJ JHAK RA HAONRJW KWH IHXTHI JWBMNT ABK EBIOHK KWXK KWH JKXKRJKRZJ
EBI XABKWHI NXAOMXOH XIH GMRKH NRLHNU KB YH TREEHIHAK WBQHPHI HGMRPXNHAK
JKXKRJKRZJ XIH XPXRNXYNH EBI BKWHI NXAOMXOHJ RE KWH ZIUCKXAXNUJK TBHJ ABK
LABQ KWH NXAOMXOH RA QWRZW KWH DHJJXOH QXJ QIRKKHA KWHA BAH BE WRJ ERIJK
CIBYNHDJ RJ KB KIU KB THKHIDRAH RK KWRJ RJ X TREERZMNK CIBYNHD
```

The table below shows the frequencies of the most common letters, as percentages.

H	K	R	X	J	I
13.2	12.1	8.7	7.9	7.9	6.4

(a) Find the plaintext, explaining your method.

(b) Do you now know the key? Will you have any difficulty decrypting further ciphertexts sent from Alice to Bob using the same cipher?

3. In a *chosen plaintext attack*, the attacker chooses a plaintext x, and is given the corresponding ciphertext $e_k(x)$ for the key k.

Explain how to recover the key by a chosen plaintext account when the cipher is (a) a substitution cipher e_π; (b) the Vigenère cipher e_k where k has length at most 10.

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

```
12345678901234567890123456789012345678901234567890
WKMSDBPZPQYBGLLSDBTHCBLDNBAHLECQNBOTEOCRWOCOAXRDZT
MQZFLSDBAHLECQPBVSPEGREPMEPBLCQBRNPTMDMRYKSLPCOFLS
DBNKWFLSAURHJMMREQGNJPBHBCCQEKEAUXKTHQGOHBMEPECKAK
ESDLDSDBIDUFRHOHLNSKYRGXQHOHGRPBQS
```

(a) Find all positions in which `SDB` appear in the ciphertext.

(b) Compute the Index of Coincidence on the samples of size 20 obtained by taking every m-th position in the ciphertext starting with the `W` in position 1, for each $m \in \{2, 3, 4\}$.

(c) What do (a) and (b) suggest about the key length?

(d) Determine the key: start by guessing the plaintext corresponding to each `SDB`.

(e) Why is the Index of Coincidence least for $m = 3$ and in the middle for $m = 2$?

5. Let R be defined on plaintexts by $R(x)_i = x_i + i \bmod 26$. For example $R(\text{bead}) =$ `CGDH` since bead $\longleftrightarrow (1, 4, 0, 3)$, $R\big((1, 4, 0, 3)\big) = (2, 6, 3, 7)$ and $(2, 6, 3, 7) \longleftrightarrow$ `CGDH`. Similarly $R^2(\text{aaaa}) =$ `CEGI`.

Let e_π denote the substitution cipher with key π.

Propose known ciphertext attacks on the two ciphers (a) $x \mapsto R^j\big(e_\pi(x)\big)$ and (b) $x \mapsto e_\pi\big(R(x)\big)$. In (a) the key is (π, j) for some $j \in \{0, \ldots, 25\}$; in (b) the key is simply π. Assume the plaintext is a lengthy English message.

Which cipher has more possible keys? Which appears harder to break?

6. Let y be every m-th position in a ciphertext output by the Vigenère cipher. What statistic would you compute to perform a χ^2-test with null hypothesis that the letters in y are distributed uniformly? How is this statistic related to the Index of Coincidence?

7. Let k be a key of length ℓ. For $q \in \{1, \ldots, \ell\}$, let $y^{(q)}$ be every ℓ-th position, starting at position q, in a ciphertext output by the Vigenère cipher with key k. Let x be a plaintext chosen to have the frequency distribution of typical English and roughly the same length as each $y^{(q)}$.

(a) Let c_s denote the Caesar cipher with shift s. Let $x : c_s(y)$ denote the concatenation of x and $c_s(y)$.
Explain why $I\big(x : c_s(y^{(i)})\big)$ should be maximized when $s = 26 - k_1$.

(b) Using (a) we can hope to find $x^{(1)}$, the plaintext corresponding to $y^{(1)}$. Which statistic would you expect to work best for decrypting the rest of the ciphertext: $I\big(x : c_s(y^{(i)})\big)$ or $I\big(x^{(1)} : c_s(y^{(i)})\big)$?

8. Which of the ciphertexts `XXXXX` and `VWXYZ` could be the output of (a) a substitution cipher, (b) a Vigenère cipher with key of length 3? Assume that the plaintext is a single English word. (The Vigenère key need not be an English word.)

MT362/462/5462 Cipher Systems: Sheet 2

Attempt at least questions 1 to 4. Question 5 is compulsory for **Msc** students. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 10am, Thursday 11am, or by appointment.

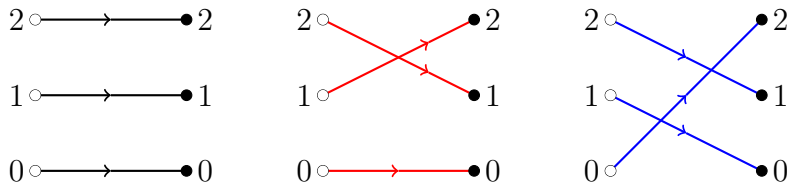
To be handed in by noon on Wednesday 18th October, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

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 $1 - p$.



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

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

(c) When does the cryptosystem have perfect secrecy?

2. Let q be prime. Suppose that Alice and Bob communicate using the affine cipher on $\mathbb{Z}_q$, and that Alice sends plaintext $x \in \mathbb{Z}_q$ with probability p_x.

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

(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 used with equal probability then the cryptosystem has perfect secrecy.

(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?

3. One of the encryption functions in the affine cipher on $\mathbb{Z}_{103}$ is $e_{(23,39)}$, defined by $x \mapsto 23x + 39$. What is $d_{(23,29)}(1)$?

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

(b) Is there a cryptosystem with perfect secrecy such that $|\mathcal{K}| < |\mathcal{C}|$? [**typo** $\leq$]

5. (MSc.) 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 an appropriate 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. This question proves the converse of Corollary 3.13. Suppose that $|\mathcal{K}| = |\mathcal{C}|$ and plaintext $x \in \mathcal{P}$ is sent with probability $p_x > 0$ for all $x \in \mathcal{P}$. 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) $\mathbb{P}[Y = y] > 0$ for all $y \in \mathcal{C}$;

(b) the cryptosystem has perfect secrecy.

7. Let $X : \Omega \to \mathcal{X}$ be a random variable taking values in a finite set $\mathcal{X}$. Let $f : \mathcal{X} \to \mathcal{X}$ be a function.

(a) Show that if f is bijective then $H(f(X)) = H(X)$. [*Hint:* the entropy of a random variable depends only on its probability distribution.]

(b) Show that $H(f(X)) \leq H(X)$. [*Hint:* use the chaining rule.]

8. Let $p_1, \ldots, p_n$ and $q_1, \ldots, q_n$ be probability distributions with $q_i > 0$ for all i. Gibbs' inequality states that

$$\sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log_2 \frac{1}{q_i}.$$

(a) Deduce from Gibbs' inequality that $H(X|Y) \leq H(X)$ [**typo:** $H(X)$ **was wrongly** $H(Y)$] for any two random variables X and Y. (The quantity $H(X) - H(X|Y)$ is known as the *mutual information* of X and Y.)

(b) Prove Gibbs' inequality. [*Hint:* the inequality $\log(1 + x) \leq x$ can be used. Another proof uses the concavity of the logarithm function.]

MT362/462/5462 Cipher Systems: Sheet 3

Attempt at least questions 1 to 4. Question 5 is compulsory for **Msc** students. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Wednesday 10am, Thursday 11am, or by appointment.

To be handed in by noon on Wednesday 25th October, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

1. Prove that if K and X are independent random variables, taking values in sets $\mathcal{K}$ and $\mathcal{P}$, respectively, then $H(K, X) = H(K) + H(X)$.

2. Eve intercepts the three ciphertexts `cqhk`, `wqvj`, `bqsq` [**typo: was** `bpsq`] encrypted using the same one-time pad. Find all three plaintexts and the key.

[*Hint:* the code used in the lecture is online at `https://repl.it/M78M/3`. If you do it by hand, it will be helpful to know that the plaintexts are four letter words related to cryptography.]

3. Alice and Bob communicate using the one-time pad cryptosystem of length n, in which $\mathcal{K} = \mathcal{P} = \mathcal{C} = \{\mathrm{a}, \ldots, \mathrm{z}\}^n$. Each key $k \in \mathcal{K}$ is chosen with equal probability. Let p_x be the probability that $x \in \mathcal{P}$ is Alice's message.

(a) Show that if $x \in \mathcal{P}$ and $p_x > 0$ then $\mathbb{P}[Y_n = y | X_n = x] = \frac{1}{26^n}$ for all $y \in \mathcal{C}$.

(b) Find $\mathbb{P}[Y_n = y]$ for each $y \in \mathcal{C}$.

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

(d) Deduce that the one-time pad has perfect secrecy.

(e) Is there a contradiction with the results of Question 2?

4. Let $\mathcal{A} = \{\mathrm{a}, \ldots, \mathrm{z}\}$

(a) Estimate the unicity distance (see Definition 4.13) of the Vigenère cipher using keys of length 10, chosen with equal probability from $\mathcal{A}^{10}$.

(b) Given bijections $\pi, \sigma : \mathcal{A} \to \mathcal{A}$ define $e_{(\pi,\sigma)} : \{\mathrm{a}, \ldots, \mathrm{z}\}^n \to \{\mathrm{a}, \ldots, \mathrm{z}\}^n$ by

$$e_{(\pi,\sigma)}(x)_i = \begin{cases} \pi(x_i) & \text{if } i \text{ is odd} \\ \sigma(x_i) & \text{if } i \text{ is even.} \end{cases}$$

For example, if $n = 3$ then $e_{(\pi,\sigma)}(x_1, x_2, x_3) = (\pi(x_1), \sigma(x_2), \pi(x_3))$.

(i) Estimate the unicity distance of the cryptosystem with keys all $e_{(\pi,\sigma)}$, supposing that keys are chosen with equal probability.

(ii) Propose a chosen plaintext attack on this cryptosystem. [*Hint:* please read the definition on page 13 of the printed lecture notes.]

5. TTTT (Totally Trusted Transmission Technologies) decides to use the affine cipher over $\mathbb{Z}_{26}$ as part of a cryptosystem with $\mathcal{P} = \mathcal{C} = \mathbb{Z}_{26}$.

(a) To maximize the size of the keyspace, they take $\mathcal{K} = \{(a, c) : a \in \mathbb{Z}_{26}, a \neq 0, c \in \mathbb{Z}_{26}\}$. What goes wrong?

(b) In the revised version, only those keys (a, c) such that $\mathrm{hcf}(a, 26) = 1$ are used. How many keys are there? Supposing that each key is used with equal probability, does the cryptosystem have perfect secrecy?

6. A computer is programmed to guess English plaintexts. It is told as soon as it guesses a character correctly, and then moves on to the next. For example if the plaintext is '`information`', the computer might guess

- `e, t, a, i` (correct, plaintext starts `i`),
- `t, s, n` (correct, plaintext starts `in`),
- `t, f` (correct, plaintext starts `inf`)

and then get every subsequent character right immediately, except for a misguess of ␣ (space) rather than `a` on character 7.

(a) Explain why given the sequence $(4, 3, 2, 1, 1, 1, 2, 1, 1, 1, 1)$ and access to the computer, you can reconstruct the plaintext.

(b) Calculate the entropy of the sequence above. Why is this entropy an upper bound on the entropy of the plaintext?

(c) Play the game online at `https://repl.it/LX00/2` and hence estimate the redundancy of English.

This game is one of Shannon's many brilliant discoveries.

7. Eve intercepts the ciphertexts `jaekbwwoswoppljoeow` and `eszxzagrhaofvquwkhj` encrypted using the same one-time pad. Decrypt and find the key.

[*Remark:* During the Second World War, the NSA decrypted many highly sensitive messages encrypted by Soviet Intelligence by re-used one-time pads. Project Venona helped expose the double agent Klaus Fuchs at Los Alamos.]

MT362/462/5462 Cipher Systems: Sheet 4

Attempt at least questions 1 to 4. Question 5 is compulsory for **Msc** students. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 10am, Thursday 10am, or by appointment.

To be handed in by noon on Wednesday 1st November, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

You are welcome to use the code from lectures at `https://repl.it/NE32`. The comments at the top show how to use it for Question 1.

1. Define the *period* of a keystream to be its length until first repeat. For instance $001100110\ldots$ has period 4. Let G be the LFSR of width 5 with taps $\{0, 1\}$.

(a) (i) Let $k = 00001$. Calculate the keystream $k_0, k_1, k_2, \ldots$, defined by G for k. What is the period of this keystream?

(ii) Find s such that $(k_s, k_{s+1}, \ldots, k_{s+4}) = 10001$.

(iii) How would your answer to (i) change if the key was 10001?

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

(c) Find all the possible periods of keystreams for G.

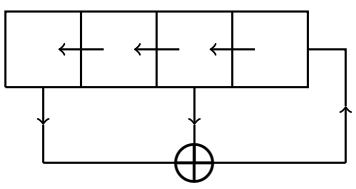
(d) What is the period of G?

2. Let F be an LFSR of width ℓ. Let $k \in \mathbb{F}_2^\ell$ and let $k_0, k_1, \ldots$ be the keystream. Show that, as claimed after Definition 5.6,

$$F^s(k) = (k_s, k_{s+1}, \ldots, k_{s+\ell-1}).$$

[*Hint:* the case $s = 0$ was Exercise 5.5.]

3. Let F be the LFSR of width 4 with taps $\{0, 2\}$, as shown in the circuit diagram below.



(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}. Is F^{-1} an LFSR?

(c) Draw a picture for F, as in Example 4.2′ from lectures, showing every cycle made by iterating F.

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

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

(a) Show that if F is invertible then $0 \in T$. [*Hint:* use the contrapositive.]

(b) Show conversely that if $0 \in T$ then F is invertible. Give a formula for F^{-1}.

5. (MSc.) Recall from the Preliminary Problem Sheet that $x_{\ell-1} \ldots x_1 x_0$ is the binary form of $2^{\ell-1}x_{\ell-1} + \cdots + 2x_1 + x_0$.

For $j \in \{0, 1, 2, 3\}$, let $f_j : \mathbb{F}_2^4 \to \mathbb{F}_2$ be the Boolean function defined so that $f(x_3, x_2, x_1, x_0)$ is the bit in position j of $x_3x_2x_1x_0 + 5 \bmod 16$.

For example, since 0110 is the binary form of 6, $6 + 5 = 11$ and 11 has binary form 1011, we have $f_3\big((0, 1, 1, 0)\big) = 1$, $f_2\big((0, 1, 1, 0)\big) = 0$, $f_1\big((0, 1, 1, 0)\big) = 1$ and $f_0\big((0, 1, 1, 0)\big) = 1$.

Express each f_j as a polynomial in x_3, x_2, x_1, x_0.

6. Show that if F is an invertible LFSR with an odd number of taps then F has at least three cycles (counting fixed points as cycles of length 1).

7. Let F be an invertible LFSR of width ℓ and let $k = (0, \ldots, 0, 1) \in \mathbb{F}_2^\ell$. Let M be the matrix representing F as in Proposition 5.9.

(a) Show that the vectors $k, kM, \ldots, kM^{\ell-1}$ form a basis for $\mathbb{F}_2^\ell$.

(b) Deduce that the keystream for k has period equal to the period of F.

8. A *de Bruijn sequence* of *order* ℓ is a circular 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 5.1 to construct a de Bruijn sequence of order 4.

(b) Prove that there exist de Bruijn sequences of every order. (The proof using LFSRs needs some finite field theory.)

9. Let M be the matrix, as in Proposition 5.9, representing an LFSR of width ℓ with taps T. Show that the characteristic polynomial of M is $X^\ell + \sum_{t\in T} X^t$. [*Hint:* use the Cayley–Hamilton Theorem and Lemma 5.10. Alternatively, expand $\det(M + XI)$ on the final column as a sum of ℓ determinants of minor matrices.]

MT362/462/5462 Cipher Systems: Sheet 5

Attempt at least questions 1 to 4. Question 5 is compulsory for **Msc** students. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 10am, Thursday 11am, or by appointment.

To be handed in by noon on Wednesday 8th November, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

A MATHEMATICA notebook for generating keystreams and solving the matrix equation in Question 1 is available from Moodle.

1. Show that the keystream $(k_0, k_1, \ldots, k_{n-1})$ is the output of a LFSR of width ℓ if and only if the $(n-\ell) \times \ell$ system of equations

$$\begin{pmatrix} k_0 & k_1 & \ldots & k_{\ell-1} \\ k_1 & k_2 & \ldots & k_\ell \\ k_2 & k_3 & \ldots & k_{\ell+1} \\ \vdots & \vdots & \ddots & \vdots \\ k_{n-\ell-1} & k_{n-\ell} & \ldots & k_{n-2} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{\ell-1} \end{pmatrix} = \begin{pmatrix} k_\ell \\ k_{\ell+1} \\ k_{\ell+2} \\ \vdots \\ k_{n-1} \end{pmatrix}$$

has a solution. [*Hint:* this is the general version of Example 5.13. Remember that 'if and only if' means you have to prove two implications.]

2. (a) Find an LFSR of minimal possible width that generates the keystream

$$\underset{k_0\ k_1\ k_2\ k_3\ k_4\ k_5\ k_6\ k_7\ k_8\ \ldots}{(0,0,1,0,0,0,1,0,1,\ldots)}$$

making clear its width and taps. Does your answer change if the LFSR is required to be invertible?

(b) Find with proof a keystream that is generated by an invertible LFSR of width 6 but not by any invertible LFSR of smaller width.

[*Hint:* the proof can be as short as one line. The final question in the quiz at the end of Lecture 15 has a relevant idea. Slides are on Moodle.]

3. Let $B_0, B_1, \ldots, B_{n-1}$ 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 $i \in \{0, \ldots, n-2\}$ such that $(B_i, B_{i+1}) = (b, b')$.

(a) Find the statistics M_{00}, M_{01}, M_{10}, M_{11} when the sequence is

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

(b) What, in general, is $M_{00} + M_{01} + M_{10} + M_{11}$?

(c) For $i \in \{0, 1, \ldots, n-2\}$, let

$$X_i = \begin{cases} 1 & \text{if } B_i = B_{i+1} = 0 \\ 0 & \text{otherwise.} \end{cases}$$

(i) Fix i. Find $\mathbb{P}[X_i = 1]$ and hence write down $\mathbb{E}[X_i]$.

(ii) Explain why $M_{00} = X_0 + X_1 + \cdots + X_{n-2}$.

(iii) Hence show that $\mathbb{E}[M_{00}] = (n-1)/4$.

(d) What are $\mathbb{E}[M_{01}], \mathbb{E}[M_{10}], \mathbb{E}[M_{11}]$?

(e) Use a χ^2-test on $M_{00}, M_{01}, M_{10}, M_{11}$ to test the sequence in (a) for randomness. [*Hint:* use (b) to determine the degrees of freedom.]

(f) Does the sequence in (a) pass the monobit test in Exercise 6.4?

4. In 7-bit ASCII, 'a' is encoded as the binary form of 97, namely 1100001, 'b' as the binary form of 98, namely 1100010, and so on.

Fix $n \in \mathbb{N}$ and consider the cryptosystem with plaintexts $\{\mathrm{a}, \ldots, \mathrm{z}\}^n$ and ciphertexts $\mathbb{F}_2^{7n}$, in which a message of n characters is first converted to 7-bit ASCII, and then encrypted by the LFSR cryptosystem of width 4 with taps $\{0, 1\}$, as defined in Definition 5.6(b).

(a) Convert 'bob' to a sequence of 21 bits using 7-bit ASCII.

(b) Hence encrypt 'bob' using the key 0101.

(c) Show that from a known ciphertext $y_0 y_1 \ldots$ an attacker can deduce $k_0, k_7, k_{14}, \ldots$. How long a ciphertext does the attacker need to learn the key (k_0, k_1, k_2, k_3)?

5. (**M.Sc.**) Define a boolean function $f : \mathbb{F}_2^5 \to \mathbb{F}$ by

$$f(x_1, x_2, x_3, x_4, x_5) = \begin{cases} 1 & \text{if at least 3 of the } x_i \text{ are 1} \\ 0 & \text{otherwise.} \end{cases}$$

Express f in (i) algebraic normal form; (ii) disjunctive normal form; (iii) conjunctive normal form.

Prove a generalization of at least one of (i), (ii), (iii) in which 3 and 5 are replaced with arbitrary t and n. What is the connection with secret sharing schemes?

MT362/462/5462 Cipher Systems: Sheet 6

Attempt at least questions 1 and 2. Question 3 is compulsory for **Msc** students. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 10am, Thursday 11am, or by appointment.

To be handed in by noon on Wednesday 15th November, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

1. Consider the cryptoscheme in which English plaintexts are converted to 7-bit ASCII ('a' $\leftrightarrow$ 1100001, 'b' $\leftrightarrow$ 1100010, and so on) and then encrypted by adding a keystream from the LFSR of width 7 with taps $\{0, 3\}$.

Suppose Malcolm observes the ciphertext

$$(0,1,1,1,0,0,0,0,0,1,1,1,1,1,1,0,0,1,1,0,1,0,0,1,1,0,1,0).$$

He knows that the second letter in the plaintext was 'e'. Find the plaintext.

2. Let $(k_0, k_1, k_2, \ldots)$ be a keystream of the LFSR F of width 2 with taps $\{0, 1\}$. Let $(k'_0, k'_1, k'_2, \ldots)$ be a keystream of an LFSR G of width 3. The keystreams are multiplied to give $(k_0k'_0, k_1k'_1, k_2k'_2, \ldots)$. Suppose you know the product is 101100000101

(a) Explain why the keystreams of F and G have the form $1\star11\star\star\star\star\star1\star1$, 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 calculate the taps and the key for G.

3. (**M.Sc.**) The table below shows the first 14 steps in the Berlekamp–Massey algorithm applied to the keystream [**corrected, see below**]

$$(u_0, u_1, \ldots, u_{14}) = (1,0,1,1,1,1,1,1,1,0,1,1,1,1,0,0)$$

As in Example 4.5 in the M.Sc. notes, m is shown only case (b) applies.

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
$\widetilde{T}_n$	$\varnothing$	$\varnothing$	$\{2\}$	$\{1,2\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\{1\}$	$\star$	$\{1,5,6,7\}$	$\star$	$\star$	$\star$	$\star$
m		0	2	2					4		9		$\star$		

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

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

(c) Will the LFSR change in further steps of the Berlekamp–Massey algorithm? Justify your answer. [**Omit please.**]

Correction. As printed, the keystream was the generating cycle of the LFSR of width 7 with taps $\{0, 1, 5, 6\}$ in Example 7.2(b):

$$(u_0, u_1, \ldots, u_{20}) = (1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0)$$

For this keystream the table should have been:

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	5	5	5	6	6	7	7	7
$\widetilde{T}_n$	$\varnothing$	$\varnothing$	$\{2\}$	$\{2\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{3,5\}$	$\{3,5\}$	$\{3,5\}$	$\{5\}$	$\{5\}$	$\{2,7\}$	$\{1,2,6,7\}$	$\{1,2,6,7\}$
m		0		2			4			7		10	12		

If you prefer (as agreed on Monday), please instead verify these entries. Part (c) above now can be answered, since the sequence is generated by a known LFSR.

4. A keystream $k_0, k_1, k_2, \ldots$ can be encoded as a formal power series with coefficients in $\mathbb{F}_2$,

$$\sum_{s=0}^{\infty} k_s X^s = k_0 + k_1 X + k_2 X^2 + \cdots .$$

Given an LFSR F of width ℓ with taps T, define $\widetilde{g}_F(X) = 1 + \sum_{t \in T} X^{\ell - t}$.

(a) Show that the keystream $k_0, k_1, k_2, \ldots$ is the output of the LFSR F if and only if

$$\widetilde{g}_F(X) \sum_{s=0}^{\infty} k_s X^s = a(X)$$

for some polynomial $a(X)$ of degree at most $\ell - 1$.

(b) Let $k_0, k_1, k_2, \ldots$ and $k'_0, k'_1, k'_2, \ldots$ be keystreams of LFSRs F and F' of widths ℓ and ℓ' respectively. Show that $k_0 + k'_0, k_1 + k'_1, k_2 + k'_2, \cdots$ is the keystream of an LFSR G of width $\ell + \ell'$ with taps defined by

$$\widetilde{g}_G(X) = \widetilde{g}_F(x)\widetilde{g}_{F'}(x).$$

(c) Let $(u_0, u_1, u_2, \ldots)$ be the sym of the keystreams for the LFSRs of width 3 and 4 with taps $\{0, 1\}$ and $\{0, 3\}$ for keys k and k', respectively.

(i) Show, as claimed in Example 7.1(b) that $(u_0, u_1, u_2, \ldots)$ is a keystream of the LFSR H of width 7 with taps $\{0, 1, 5, 6\}$.

(ii) Show that the map $(k, k') \to (u_0, u_1, u_2, u_3, u_4, u_5, u_6)$ is injective. [*Hint:* show that if $(k, k') \to (0, 0, 0, 0, 0, 0, 0)$ then $(u_0, u_1, u_2, \ldots)$ is a keystream of both the LFSRs F and G, and use (a).]

(iii) Suppose you know s and $(u_s, u_{s+1}, \ldots, u_{s+6})$. Explain how to find the keys k and k'.

(d) How is $\widetilde{g}_F(X)$ related to the minimal polynomial $g_F(X)$ of F?

MT362/462/5462 Cipher Systems: Sheet 7

Attempt at least questions 1 to 4. M.Sc. students should also attempt question 5 or 6. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 10am, Thursday 11am, or by appointment.

To be handed in by noon on Wednesday 29th November, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

1. 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 $(v,w) \in \mathbb{F}_2^{2m}$.

Let $(v',w') = (w, v + f(w))$. Noting the order w', v' carefully, express $F\big((w',v')\big)$ in terms of v and w.

2. Consider the Q-block cipher as defined in Example 8.4, consisting of three rounds of the Feistel Network

$$F\big((v,w)\big) = (w, v + S(w + k^{(i)}))$$

where $v, w \in \mathbb{F}_2^4$ and $S\big(x_0,x_1,x_2,x_3\big) = (x_2, x_3, x_0 + x_1x_2, x_1 + x_2x_3)$. Given a key $k \in \mathbb{F}_2^{12}$, the three round keys are $k^{(1)} = (k_0,k_1,k_2,k_3)$, $k^{(2)} = (k_4,k_5,k_6,k_7)$ and $k^{(3)} = (k_8,k_9,k_{10},k_{11})$.

(a) Encrypt $(0,0,0,0,0,0,0,0) \in \mathbb{F}_2^8$ using the key $(0,0,1,1,0,0,1,1,0,0,1,1)$.

(b) Decrypt the ciphertext $(0,1,1,1,0,1,1,1)$ using the key in (a).

(c) Find a key $k \in \mathbb{F}_2^{12}$ such that $e_k\big((0,0,0,1,0,0,0,1)\big) = (0,0,0,0,0,0,0,0)$.

(d) Show that given $(v,w) \in \mathbb{F}_2^8$ and $w' \in \mathbb{F}_2^8$ there is a unique round key $k_{\text{round}} \in \mathbb{F}_2^4$ such that $(w, v + S(w + k_{\text{round}})) = (w,w')$.

(e) How many keys $k \in \mathbb{F}_2^{12}$ have the property in (c)?

(f) Would your answer to (e) change if $(0,0,0,1,0,0,0,1)$ and $(0,0,0,0,0,0,0,0)$ were replaced with different plaintexts and ciphertexts?

3. You have a black box implementing a round of a Feistel block cipher with block size $2m$. Thus, given $(v, w) \in \mathbb{F}_2^{2m}$ and a round key k_{round}, the box will output $\big(w, v + S(w + k_{\text{round}})\big)$. You do not know S.

Explain how to use the box to decrypt a ciphertext $(x, y) \in \mathbb{F}_2^{2m}$ encrypted using the key k. [*Hint:* Question 1 is relevant.]

4. 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 that finds the key using about 2^{112} operations. (You may use multiple chosen plaintexts.)

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

(c) (Optional.) Why is the middle map decryption rather than encryption?

5. The *affine block cipher* of block size n has keyspace all pairs (A, b), where A is an invertible $n \times n$ matrix with entries in $\mathbb{F}_2$ and $b \in \mathbb{F}_2^n$. The encryption functions $e_{(A,b)} : \mathbb{F}_2^n \to \mathbb{F}_2^n$ are defined by

$$e_{(A,b)}(x) = xA + b.$$

(a) Define the decryption functions $d_{(A,b)} : \mathbb{F}_2^n \to \mathbb{F}_2^n$.

(b) How can the key be recovered in a chosen plaintext attack? How many plaintext / ciphertext pairs are required?

(c) Does repeating the cipher (as in the example of 2DES, so two potentially different keys are used) make this cipher any more secure?

(d) Does this cipher have the 'confusion' property?

(e) Does this cipher have the 'diffusion' property?

6. The University of Erewhon has, at fabulous expense, purchased an examination database from TTTT (Totally Trusted Transmission Technologies) in which the grades, which must be numbers between 0 and 100, are encrypted using AES with a secret key $k \in \mathbb{F}_2^{128}$. A typical table is a list of ordered pairs

$$\big(\text{Alice}, e_k(75)\big), \big(\text{Bob}, e_k(40)\big), \big(\text{Charlie}, e_k(65)\big), \ldots.$$

Criticize the security of this system. How could it be improved?

MT362/462/5462 Cipher Systems: Sheet 8

Attempt at least questions 1 to 4. M.Sc. students should also attempt question 5. Please staple your answers together and put your name and student number on this sheet.

The lecturer will be happy to discuss any of the questions in drop-in sessions: Tuesday 3.30pm, Wednesday 10am, Thursday 11am, or by appointment.

To be handed in by noon on Wednesday 6th December, or at the Monday lecture.

☐ Tick this box if you *do not* want written feedback on your solutions.

Your feedback to the lecturer: what question, if any, do you most want solved in lectures? What was easy, hard, interesting this week?

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 8.4 that the Feistel network in round i of this cipher is

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

where $k^{(i)} \in \mathbb{F}_2^4$ is the round key.

(a) Let $\Delta \in \mathbb{F}_2^4$. Show that if $\Delta_2 = 0$, i.e. Δ is of the form $(\star, \star, 0, \star)$ then

$$S(x + \Delta) + S(x) = \begin{cases} (0, \Delta_3, \Delta_0, \Delta_1) & \text{if } x_2 = 0 \\ (0, \Delta_3, \Delta_0 + \Delta_1, \Delta_1 + \Delta_3) & \text{if } x_2 = 1. \end{cases}$$

(b) Deduce Lemma 9.7(i), namely that $S(x + 1000) = S(x) + 0010$ for all $x \in \mathbb{F}_2^4$.

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

(d) Show that after two rounds of the Q-block cipher, the input difference $0000\,1000$ is sent to one of $0010\,0000, 0010\,0001, 0010\,0010, 0010\,0011$, each with equal probability.

2. Let e_k for $k \in \mathbb{F}_2^{12}$ be the encryption maps in the Q-block cipher. Show using Lemma 9.7(i) that $e_k(x) = e_{k+1000\,0010\,1000}(x)$ for all $x \in \mathbb{F}_2^8$.

3. (a) Compute $2^{131} \bmod 3023$. [*Hint:* if $2^m \equiv y \bmod 3023$ then $2^{2m} \equiv y^2 \bmod 3023$ and $2^{m+1} \equiv 2y \bmod 3023$. You should not need more than a pocket calculator or its computational equivalent.]

(b) Find x such that $2^x \equiv 35 \bmod 37$.

4. For $k \in \mathbb{F}_2^8$ define $e_k : \mathbb{F}_2^8 \to \mathbb{F}_2^8$ by $e_k(x) = P(x) + k$ where $P : \mathbb{F}_2^8 \to \mathbb{F}_2^8$ is the pseudo-inverse function defined by identifying $\mathbb{F}_2^8$ with the finite field $\mathbb{F}_{2^8}$. In Example 9.3 we attacked the cipher with key space $\mathbb{F}_2^8 \times \mathbb{F}_2^8$ defined by

$$E_{(k_{\text{otp}},k)}(x) = e_k(x + k_{\text{otp}}).$$

By Exercise 9.6 (see optional Question 6 below), the attack typically finds k and one false key $k + \Gamma$ using 2^9 decryptions. You know $x \in \mathbb{F}_2^8$ and $E_{(k_{\text{otp}},k)}(x)$.

(a) Explain why computing $E_{(k_{\text{otpguess}},k)}(x)$ and $E_{(k_{\text{otpguess}},k+\Gamma)}(x)$ for all guesses k_{otpguess} will typically leave just two possibilities for the key (k_{otp}, k).

(b) How many encryptions/decryptions are required in total? Is the attack subexhaustive?

5. (M.Sc.) By Theorem 5.8(c) in the M.Sc. notes, if $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a Boolean function then

$$(-1)^f = \sum_{T \subseteq \{1,\ldots,n|} \text{corr}(f, L_T)(-1)^{L_T}$$

where $L_T(x_1, \ldots, x_n) = \sum_{t \in T} x_t$.

(a) Define $\neg f : \mathbb{F}_2^n \to \mathbb{F}$ by $(\neg f)(x) = \neg f(x)$. Show that

$$\text{corr}(\neg f, L_T) = -\text{corr}(f, L_T).$$

(b) Show that $\sum_{T \subseteq \{1,\ldots,n\}} \text{corr}(f, L_T)^2 = 1$.

(c) Define $g : \mathbb{F}_2^3 \to \mathbb{F}_2$ by $g(x_1, x_2, x_3) = x_1x_2x_3$. Compute the eight correlations $\text{corr}(g, L_T)$ for $T \subseteq \{1, 2, 3\}$ and so check that part (b) and Theorem 5.8(c) holds for g.

(d) (Optional: far longer than I realised.) What are the possible values for the eight correlations $\text{corr}(f, L_T)$ when T is a 3-variable Boolean function? [*Hint:* Example 5.6 and (c) show two possibilities. Don't forget the linear functions.]

6. Let $p : \mathbb{F}_{2^8} \to \mathbb{F}_{2^8}$ be the pseudo-inversion function, defined by

$$p(\beta) = \begin{cases} 0 & \text{if } \beta = 0 \\ \beta^{-1} & \text{otherwise.} \end{cases}.$$

Let $\gamma \in \mathbb{F}_{2^8}$ be non-zero. Show that for each non-zero $\delta \in \mathbb{F}_2^8$

$$\{\beta : \beta \in \mathbb{F}_2^8 : p(\beta) + p(\beta + \gamma) = \delta\}$$

has size 0 or 2, except when $\delta^{-1} = \gamma$, when it has size 4. [*Hint:* quadratic equations over any field have at most two roots.]

MT362/462/5462 Cipher Systems: Sheet 9

Attempt questions 1 to 5. **M.Sc.** students should also attempt question 6.

This sheet need not be handed in. Model answers will be posted on Moodle as usual. You are welcome to email the lecturer `mark.wildon@rhul.ac.uk` with any questions.

Private keys, and other private information, are written in red. You are welcome to follow this convention or ignore it, as you prefer.

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

2. 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 $x \in \{1, \dots, p-1\}$ to Alice.

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

(b) Suppose Bob sends $x(g^a)^r \bmod p$ for some private r of his choice. Can Alice find x?

(c) Suppose Bob sends $x(g^a)^r \bmod p$ and then sends r. Can Alice find x? Can Eve find x?

(d) Suppose Bob sends $x(g^a)^r \bmod p$ 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.

3. Let (a, n) 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.]

4. Alice's RSA public key (a_{pub}, n) can be downloaded from her website. Alice's private key (a_{priv}, n) is secure on Alice's computer.

Any data entering or leaving Bob's computer may be modified by Malcolm, the man-in-the-middle. Bob goes to Alice's website and reads what he thinks is her RSA public key. He uses it to send her a message x using the RSA cryptoscheme.

(a) Show that Malcolm can read x, and if he wishes change x to another message x', without Alice or Bob noticing.

(b) Is the problem avoided if Alice signs her public key?

(c) Is the problem avoided if Trevor, a person trusted by Bob, signs Alice's key? Assume Trevor's public key is already on Bob's computer.

5. Consider the cryptoscheme in which English plaintexts are converted to 7-bit ASCII ('a' $\leftrightarrow$ 1100001, 'b' $\leftrightarrow$ 1100010, and so on) 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 (a, n) then she is sent $e_a(13409) = 13409^a \bmod n$. 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 non-public key cipher such as AES? How can it be avoided while still using the RSA cryptosystem?

6. (**M.Sc.**) Let $e_k : \mathbb{F}_2^8 \to \mathbb{F}_2^8$ 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?

7. (**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 randomly chosen plaintext/ciphertext pairs. She estimates the correlation in (a) by computing $Z_j = (-1)^{L_U(v'^{(j)}) + L_T(v^{(j)})}$ for each j, and taking the mean $C = \frac{1}{q}\sum_{j=1}^{q} Z_j$.

(b) 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 $L_T(k)$?

8. (**M.Sc.**)

(a) Let $G, F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be functions. Show using the Discrete Fourier Inversion Theorem that if V, $T \subseteq \{1, \ldots, n\}$ then

$$\mathrm{corr}(L_V \circ G \circ F, L_T) = \sum_{U \subseteq \{1,\ldots,n\}} \mathrm{corr}(L_V \circ G, L_U)\,\mathrm{corr}(L_U \circ F, L_T).$$

(b) Let e_k for $k \in \mathbb{F}_2^\ell$ be the encryption functions in a block cipher of blocksize n and key length ℓ, defined by iterating some number of rounds. Suppose, as in Example 8.4 and the DES cipher,

(i) each round key is given by taking n specified bits from the key k,

(ii) each round is a function of the form $x \mapsto F(x + k_{\mathrm{round}})$.

Let T, $U \subseteq \{1, \ldots, n\}$. Show that there exist $c_V \in \mathbb{R}$ for $V \subseteq \{1, \ldots, n\}$ such that

$$\mathrm{corr}(L_U \circ e_k, L_T) = \sum_{V \subseteq \{1,\ldots,\ell\}} c_V (-1)^{L_V(k)}.$$

Remark: in practice, one c_V is often much bigger in absolute value than the rest; in this case the attack in Question 7 can be applied.