MATHEMATICAL TRIPOS PART II 2013
CODING AND CRYPTOGRAPHY
EXAMPLE SHEET 4

The first 18 examples are minimal to cover the course; you are also encouraged to try questions 19–20.

1 A (binary) linear feedback shift register was used to generate the stream 110001110001 ... Recover the feedback polynomial by the Berlekamp–Massey method. [Hint: the LFSR has length 4 but you should work through the trials for length r for $1 \leqslant r \leqslant 4$.]

2 A (binary) non-linear feedback register of length 4 has defining relation

$$x_{n+1} = x_n x_{n-1} + x_{n-3}.$$

Show that the state space contains four cycles of lengths 1, 2, 4 and 9.

3 Implement the Secret Sharing method of (13.7) with $k = 2$, $n = 3$, $x_j = j + 1$, $p = 7$, $a_0 = S = 2$, $a_1 = 3$. Check directly that any two people can find $S = a_0$ but no single individual can.

Take $k = 3$, $n = 4$ and $p = 6$. Show that if $x_j = j$ then the first two members and the fourth member will be unable to determine a_0 uniquely, whereas if $x_j = j + 1$ then these members can determine a_0 uniquely.

4 Let $C(X) = c_0 + c_1 X + \cdots + c_{n-1} X^{n-1} + X^n$ be the feedback polynomial for a linear feedback shift register over $\mathbb{F}_q$ with $c_0 \neq 0$. Show that the output stream (x_j) is periodic. Show that there is a $n \times n$ matrix M such that

$$\begin{pmatrix} x_j \\ x_{j+1} \\ \vdots \\ x_{j+n-1} \end{pmatrix} = M^j \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{pmatrix}$$

for $j = 0, 1, 2, \ldots$. Prove that $M^N = I$ for some integer $N > 0$. Find the characteristic and minimal polynomials for M. What happens when the coefficient c_0 is allowed to be 0? Do we still get periodicity when the feedback function is not assumed to be a polynomial?

5 Recall that if x_n is a stream which is periodic with period M and y_n is a stream which is periodic with period N then the streams $x_n + y_n$ and $x_n y_n$ are periodic with periods dividing the lowest common multiple of M and N. One of the most confidential German codes[1] involved a complex mechanism which the British found could be simulated by two loops of paper tape of length 1501 and 1497. If $k_n = x_n + y_n$ where x_n is a stream of period 1501 and y_n is a stream of period 1497, what is the longest possible period of k_n? How many consecutive values of k_n would you need to find the underlying linear feedback register using the Berlekamp–Massey method if you did not have the information given in the question? If you had all the information given in the question how many values of k_n would you need? [Hint: look at $x_{n+1497} - x_n$.]

You have shown that, given k_n for sufficiently many consecutive n we can find k_n for all n. Can you find x_n for all n?

6 We work with streams of symbols in $\mathbb{F}_2$. I have a secret sequence $k_1, k_2, \ldots$ and a message $p_1, p_2, \ldots, p_N$. I transmit $p_1 + k_1$, $p_2 + k_2$, $\ldots p_N + k_N$ and then, by error, transmit $p_1 + k_2$, $p_2 + k_3$, $\ldots p_N + k_{N+1}$. Assuming that you know this and that my message makes sense, how would you go about finding my message? Can you now decipher other messages sent using the same part of my secret sequence?

7 Describe briefly the Elgamal[2] signature scheme and indicate how it defeats a homomorphism attack.

Alice signs a sequence of messages, incrementing the value of k by 2 each time. Assuming Bob knows this, show that in most cases he can determine Alice's private key from two consecutive signed messages (without having to solve the discrete logarithm problem).

8 I announce that I shall be using the Rabin scheme with modulus N. My agent in X'Dofro sends me a message m (with $1 \leqslant m \leqslant N - 1$) encoded in the requisite form. Unfortunately, my parrot eats the piece of paper on which the prime factors of N are recorded, so I am unable to decipher it. I therefore find a new pair of primes and announce that I shall be using the Rabin scheme with modulus $N' > N$. My agent now recodes the message and sends it to me again.

The dreaded SNDO of X'Dofro intercept both code messages. Show that they can find m. Can they decipher any other messages sent to me using only one of the coding schemes?

9 Extend the Diffie–Hellman key exchange system to cover three participants in a way that is likely to be as secure as the two party scheme.

Extend the system to n parties in such a way that they can compute their common secret key by at most $n^2 - n$ communications of 'Diffie–Hellman type numbers'. (The numbers p and g of our original Diffie-Hellman system are known by everybody in advance.) Show that this can be done using at most $2n - 2$ communications by including several 'Diffie–Hellman type numbers' in one message.

---

[1] codenamed FISH by the allies; this is because Bletchley Park had revealed that the Germans called their wireless teleprinter transmission systems 'Sägefisch' (sawfish). The marvellous tale of how FISH was broken is in Part 3 of the book 'Code Breakers: the inside story of Bletchley Park' edited by Harry Hinsley and Alan Stripp (OUP, 1993).

[2] Tahir Elgamal is an Egyptian cryptographer. His surname has been spelled as two words (the Arabic "El" part is equivalent to "the" in English), and as a single word with an intra-capital. The cryptographic literature is full of both spellings. Dr Elgamal himself spells it as a singly capitalised surname, as this is less likely to be mangled in English.

10 The Mariner 9 mission established that written Martian used an alphabet containing only three letters (which we'll call A, B and C) and avoided the use of spaces (thus a Martian book consists of single word). In the written Martian language, the letter A has frequency .5 and the letters B and C both have frequency .25. In order to disguise this, the Martian Battlefleet uses codes in which the $(3r+i)$th number is $x_{3r+i}+y_i$ modulo 3 $(0 \leqslant i \leqslant 2)$ where $x_j = 0$ if the jth letter of the message is A, $x_j = 1$ if the jth letter of the message is B, $x_j = 2$ if the jth letter of the message is C and y_0, y_1 and y_2 are the numbers 0, 1, 2 in some order.

Radio interception on Venus has picked up the following message from the Martians:

$$120022010211121001001021002021.$$

Although nobody in Venusian Starfleet Intelligence (called MIV) reads Martian, it is believed that the last letter of the message will be B if the Martian Battlefleet has launched. MIV are desperate to know the last letter and send a representative to Torchwood to ask Captain Jack's advice. What does Jack advise?

11 (i) A careless user of RSA accidentally chooses a large prime for his modulus N. Explain why this system is not secure.

(ii) In textbook examples of the RSA code we frequently see $e = 65537$. Using this exponent, how many multiplications are needed to encrypt a message?

(iii) Why is it unwise to choose primes p and q with $|p-q|$ small when forming $N = pq$ for the RSA method?

12 The Secret Intelligence Service (known since WWII as MI6) is proud of the excellence of its privacy system LoTex. To advertise this fact to the world, the Chief of MI6 (known as C)[3] decrees that the internal telephone directory should bear on its cover a number N (a product of two very large secret primes) and each name in the directory should be followed by their personal encryption number e_i. Now C knows all the secret decryption numbers d_i but gives these out on a need to know basis only. (Of course each member of staff must know their personal decryption number but they are instructed to keep it secret.) Messages a from C to members of staff are encrypted in the standard manner as a^{e_i} modulo N and decrypted as b^{d_i} modulo N.

(i) C sends a message to all members of MI6. An outsider intercepts the encrypted message to individuals i and j where e_i and e_j are coprime. How can the outsider read the message? [This is known as the 'common modulus' attack.] Can she read other messages sent from C to the ith member of staff only?

(ii) By means of a phone tapping device, Agent 007 (number u in the directory) has intercepted messages from C to his hated rival, Agent 006 (number v in the directory). Explain why he can decode them.

What moral should be drawn?

[3]The first Chief was Captain Sir George Mansfield Smith-Cumming, KCMG, CB. He typically signed correspondence with his final initial C in green ink. This usage evolved as a code name, and has been adhered to by all subsequent chiefs of MI6 when signing documents to retain anonymity.

13 In his secret base inside an extinct volcano, Dr Evil uses a one-time pad to communicate with the notorious international spy known as SilverNose. The messages are coded in the obvious way. (If the pad has C the 3rd letter of the alphabet and the message has I the 9th then the encrypted message has L the $(3+9)$th. Work modulo 26.) Unknown to Dr Evil, the person whom he employs to carry the messages is actually the MI6 agent Austin Powers in disguise. MI6 are on the verge of arresting SilverNose when Austin is given the message

$$LRPFOJQLCUD.$$

Powers knows that the actual message is

$$FLYXATXONCE$$

and suggests that 'the boffins change things a little' so that SilverNose deciphers the message as

$$REMAINXHERE.$$

The only boffin available is you. Advise MI6.

14 We model English text by a sequence of random variables $(X_n)_{n\geqslant 1}$ taking values in $\mathcal{A} = \{a, b, \ldots, z, \texttt{space}\}$. The entropy of English is $H_E = \lim_{n\to\infty} H(X_1, \ldots, X_n)/n$.

(a) Assuming H_E exists, show that $0 \leqslant H_E \leqslant \log_2 27$.

(b) Taking $H_E = \log_2 3 \approx 1.58$, estimate the unicity distance of (i) the substitution cipher, and (ii) the Vigenère cipher with keyword of length d.

15 Suppose that X and Y are independent random variables taking values in $\mathbb{Z}_n$. Show that

$$H(X+Y) \geq \max\{H(X), H(Y)\}.$$

Why is this remark of interest in the context of one-time pads?

Does this result remain true if X and Y need not be independent? Give a proof or counterexample.

16 Confident in the unbreakability of RSA, I write the following. What mistakes have I made?

0000001 0000000 0002048 0000001 1391142
0000000 0177147 1033288 1391142 1174371.

Advise me on how to increase the security of messages.

17 Criticise the following authentication procedure. Alice chooses N as the public key for the Rabin cryptosystem. To be sure we are in communication with Alice we send her a 'random item' $r \equiv m^2 \pmod N$. On receiving r, Alice proceeds to decode using her knowledge of the factorisation of N, and finds a square root m_1 of r. She returns m_1 to us and we check that $r = m_1^2 \pmod N$.

18 Let K be the finite field with 2^d elements. We recall that $K^\times = K \setminus \{0\}$ is a cyclic group under multiplication, generated by a primitive root α say. Let $T : K \to \mathbb{F}_2$ be any non-zero $\mathbb{F}_2$-linear map. (Here we treat K as a vector space over $\mathbb{F}_2$.)

(i) Show that the $\mathbb{F}_2$-bilinear form $S : K \times K \to \mathbb{F}_2$ given by $S(x, y) = T(xy)$ is a symmetric (bilinear) form. Show further that S is non-degenerate (i.e. $S(x, y) = 0$ for all $y \in K$ implies $x = 0$).

(ii) Show that the sequence $x_n = T(\alpha^n)$ is the output from a LFSR of length at most d. (Part (iii) shows that it must be exactly d.)

(iii) The period of $(x_n)_{n\geqslant 0}$ is the least integer $r \geqslant 1$ such that $x_{n+r} = x_n$ for all sufficiently large n. Show that the sequence in (ii) has period $2^d - 1$. Explain briefly why this is best possible.

19 Suppose that $N = pq$ where p and q are distinct primes with the same number of binary digits. We use an RSA cipher with modulus N, encrypting exponent e and decrypting exponent d, with $0 < d, e < \varphi(N)$.

(a) Show that $N - \varphi(N) < 3\sqrt{N}$.

(b) Let $k = (de - 1)/\varphi(N)$. Show that k is an integer less than d.

(c) Show that if $d < \frac{1}{3}N^{1/4}$ then

$$\left|\frac{k}{d} - \frac{e}{N}\right| < \frac{1}{3d^2}.$$

(d) It is known that if x is a positive real number and a, b are integers with

$$\left|x - \frac{a}{b}\right| < \frac{1}{2b^2}$$

then a/b arises as one of the convergents of the continued fraction expansion of x. Explain how this observation may be used to attack RSA.

20 Consider the linear recurrence

★ $$x_n = a_0 x_{n-d} + a_1 x_{n-d+1} + \ldots + a_{d-1} x_{n-1}$$

with $a_j \in \mathbb{F}_2$ and $a_0 \neq 0$.

(i) Suppose K is a field containing $\mathbb{F}_2$ such that the auxiliary polynomial C has a root α in K. Show that $x_n = \alpha^n$ is a solution of ★ in K.

(ii) Suppose K is a field containing $\mathbb{F}_2$ such that the auxiliary polynomial C has d distinct roots $\alpha_1, \alpha_2, \ldots, \alpha_d$ in K. Show that the general solution of ★ in K is

$$x_n = \sum_{j=1}^{d} b_j \alpha_j^n$$

for some $b_j \in K$. If $x_0, x_1, \ldots, x_{d-1} \in \mathbb{F}_2$, show that $x_n \in \mathbb{F}_2$ for all n.

(iii) Work out the first few lines of Pascal's triangle modulo 2. Show that the functions $f_j : \mathbb{Z} \to \mathbb{F}_2$

$$f_j(n) = \binom{n}{j}$$

are linearly independent in the sense that

$$\sum_{j=0}^{m} b_j f_j(n) = 0$$

for all n implies $b_j = 0$ for $1 \leq j \leq m$.

(iv) Suppose K is a field containing $\mathbb{F}_2$ such that the auxiliary polynomial C factorises completely into linear factors. If the root α_u has multiplicity $m(u)$ $[1 \leq u \leq q]$, show that the general solution of ★ in K is

$$x_n = \sum_{u=1}^{q} \sum_{v=0}^{m(u)-1} b_{u,v} \binom{n}{v} \alpha_u^n$$

for some $b_{u,v} \in K$. If $x_0, x_1, \ldots, x_{d-1} \in \mathbb{F}_2$, show that $x_n \in \mathbb{F}_2$ for all n.

SM, Lent Term 2013

Comments on and corrections to this sheet may be emailed to `sm@dpmms.cam.ac.uk`