MATHEMATICAL TRIPOS PART II (2022–2023)
CODING AND CRYPTOGRAPHY
EXAMPLE SHEET 4 OF 4

1 Show that, subject to a suitable non-degeneracy condition, any output stream $x_0, x_1, x_2, \ldots$ produced by a linear feedback shift register is *purely periodic*, i.e. there exists r such that $x_{n+r} = x_n$ for all $n \geq 0$.

2 A linear feedback shift register was used to generate the stream $110001110001\ldots$.

Recover the feedback polynomial by the Berlekamp-Massey method. (The LFSR has length 4 but you should work through the trials for length d for $1 \leqslant d \leqslant 4$.)

3 We model English text by a sequence of random variables $(X_n)_{n\geq 1}$ taking values in $\Sigma = \{A, B, \ldots, Z, \texttt{space}\}$. 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 $0 \leq H_E \leq \log 27$.

(b) Taking $H_E \approx \log 3 \approx 1.58$, estimate the unicity distance of (i) the substitution cipher, and (ii) the Vigenère cipher.

4 We work with streams of symbols in $\mathbb{F}_2$. I have a key 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 key sequence?

5 A 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 4 cycles of lengths 1, 2, 4 and 9.

6 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 $r = m_1^2 \pmod N$. [Consider what happens when many mutually distrusting parties communicate with Alice in this way.]

7 I announce that I shall be using the Rabin code 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 cat 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 code 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?

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

(ii) A popular choice for the RSA encryption exponent is $e = 65537$. Using this exponent how many multiplications are required to encrypt a message?

(iii) Why might it be a bad idea to use an RSA modulus $N = pq$ with $|p - q|$ small?

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 in at most $n^2 - n$ communications of 'Diffie–Hellman type numbers'. The numbers p and g of our original Diffie-Hellman system are nown 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.

10 Describe briefly the Elgamal[1] signature scheme and indicate how it defeats a homomorphism attack.

(i) 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).

(ii) Suppose we drop the requirement that $1 \leqslant r \leqslant p - 1$ from the Elgamal signatuire scheme. How might we then be able to forge signatures from old? [Hint: use the Chinese remainder theorem for the coprime moduli p and $p - 1$.

11 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[2] 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?

12 Let K be the finite field with 2^d elements. We recall that K^* is a cyclic group, generated by α say. Let $T : K \to \mathbb{F}_2$ be any non-zero $\mathbb{F}_2$-linear map.

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

(ii) Show that the sequence $x_n = T(\alpha^n)$ is the output from a linear feedback shirt register of length d.

(iii) The period of $(x_n)_{n \geq 0}$ 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 $2^d - 1$.

Further Problems

[1] Tahir Elgamal is an Egyptian cryptographer, known as the 'father of SSL'. 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.

[2] 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).

13 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?

14 Confident in the unbreakability of RSA, I write the following.

0000001 0000000 0002048 0000001 1391142

0000000 0177147 1033288 1391142 1174371

What mistakes have I made? Advise me on how to increase the security of messages[4]

15 NASA's Mariner 9 mission to Mars 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 $x_{3r+i} + y_i \bmod 3$ $(0 \leqslant i \leqslant 2)$ 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 by NASA has picked up the following message from the Martians:

$$1200220102111210010010210020 21.$$

Although nobody in NASA reads Martian, it is believed that the last letter of the message will be B if the Martian Battlefleet has launched. NASA are desperate to know the last letter and send a representative to your rooms in College to ask your advice. Give it.

[3]The first Chief of SIS was Captain Sir George Mansfield Smith-Cumming, who held office from 1909 until his death in 1923. 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. He is referred to as 'Control' in the John le Carré novel *The spy who came in from the cold*, and in other novels. In the original Bond novels, Ian Fleming refers to the chief of SIS as M. A rather workmanlike biography *The Quest for C: Mansfield Cumming and the Founding of the Secret Service* by Alan Judd portrays an extraordinary man as somewhat ordinary.

[4]This code is really no more sophisticated than the ones appearing in Edgar Allan Poe's *The Gold-Bug* or the Sherlock Holmes novel *The Adventure of the Dancing Men* by Arthur Conan Doyle.

16 Implement Shamir's (k, n)-threshold scheme (i.e. the 'secret sharing') of Chapter 20, taking $k = 2$, $n = 3$, $x_j = j + 1$, $p = 7$, $a_0 = S = 2$ and $a_1 = 3$. Check directly that any two people can find the secret S but that no single individual can.

Take $k = 3$, $n = 4$ and suppose that I foolishly try to implement Shamir's $(3, 4)$-threshold scheme by choosing $p = 6$. By considering systems of congruences mod 6, show that if I take $x_j = j$ 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.

My final message to you:

```
klqhikg ip pl bawrqifre wp pmoikg
                                  gaowox jwkeat hlmcikp
```

[Hint: it is a substitution cipher.]

SM, Lent Term 2023
Comments on and corrections to this sheet may be emailed to sm137@cam.ac.uk