MATHEMATICAL TRIPOS PART II (2025-2026)

Coding and Cryptography - Example Sheet 4 of 4

1 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 $1 \leqslant d \leqslant 4$.)

2 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.

3 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?

4 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.

5 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 $r = m_1^2 \pmod N$.

6 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 cipher with public 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?

7 (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) Show that if the primes p and q are chosen to be close then RSA is insecure. (By 'close' we mean $(p+q)/2$ is not much larger than $\sqrt{pq}$: we know that it is always at least as big).

8 Alice and Bob are issued with RSA public keys (N, e_1) and (N, e_2), and corresponding private keys (N, d_1) and (N, d_2).
(i) The same message m is sent to both Alice and Bob. Assuming e_1 and e_2 are coprime, how can we recover m from the intercepted cipher texts c_1 and c_2?
(ii) How can Alice read messages sent to Bob?

9 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.[1]

10 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 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.

11 Recall the Elgamal signature scheme and briefly 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 signature scheme. How might we then be able to forge signatures from old? [Hint: use the Chinese remainder theorem for the coprime moduli p and $p - 1$.]

12 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?

---

[1] 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.

[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).

Further Problems

13 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.

14 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 shift register of length at most 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 $2^d - 1$.

15 Konstantin uses a one-time pad to communicate with the notorious spy Villanelle. The messages are coded in the obvious way, namely, if the pad has C, the third letter of the alphabet and the message has I, the ninth, then the encrypted message has L as the (3+9)th. We will work modulo 26. Unknown to them, the person whom they employ to carry the messages is actually the MI6 agent Eve in disguise. MI6 are on the verge of arresting Villanelle when Eve is given the message

$$LRPFOJQLCUD.$$

Eve knows that the actual message is

$$FLYXATXONCE,$$

and suggests that Q[3] 'changes things a little' so that Villanelle deciphers the message as

$$REMAINXHERE.$$

How will Q achieve this?

. *Comments & corrections should be sent to Rachel Camina (rdc26).*

[3]Actually the character Q (code for Quartermaster) never appears in the novels by Ian Fleming, where Q and 'Q Branch' are merely mentioned. Charles Fraser-Smith is widely credited as the inspiration for Q, for which see the New Scientist article 'Careful Carruthers That Paper Clip Is Loaded' from August 1993.