CODES AND CRYPTOGRAPHY – Example Sheet 4

TKC Lent 2015

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

   (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 of (i) the substitution cipher, and (ii) the Vigenère cipher.

2. I encrypt a binary sequence $a_1, a_2, a_3, \dots, a_N$ using a one-time pad with key sequence $k_1, k_2, k_3, \dots$. So I send $a_1 + k_1, a_2 + k_2, a_3 + k_3, \dots, a_N + k_N$. Then , by mistake, I transmit $a_1 + k_2, a_2 + k_3, a_3 + k_4, \dots, a_N + k_{N+1}$. Assuming that you know I have made this error and that my message makes sense, how would you find the message? Can you now decipher other messages sent using the same key sequence?

3. 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?

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

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

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

5. I decide to use an RSA cipher to encode a message by first converting each letter of the message to an integer (space = 0, a=1, b=2, *etc*) and then encrypting this integer k as $k^e \pmod N$. Explain why this is foolish.

   You intercept the ciphertext 02940 00365 18718 18718 01759 02940 02940 and know that the public key for the cipher is $(18721, 25)$. Decipher the message.

   [*Hint: Use a spreadsheet to calculate a^e, or ask a crossword solver.*]

6. The Foolish Internet Service Provider plc. decided to provide each of its customers with their own RSA ciphers using a common modulus N. Customer j is given the public key (N, e_j) and sent secretly their decrypting exponent d_j. The company then sends out the same message, suitably encrypted, to each of its customers. You intercept two of these messages to customers i and j with $(e_i, e_j) = 1$. Explain how to decipher the message.

   You are one of the customers, and so also know your own decrypting exponent, explain how you could decipher any message sent to another customer?

7. 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.

8. Describe the Elgamal signature scheme. Explain what rôle the hash function h plays by considering the scheme where the hash function is taken as the identity. Show that given one signed message I can construct other validly signed messages (existential forgeries), although those messages may not make sense.

   Alice uses the Elgamal signature scheme to sign a sequence of messages, incrementing the value of k by 2 for each new message. Show how to determine Alice's private key from any two successive signed messages.

9. Consider a binary message $A_1A_2A_3\dots$. Suppose that the probability that A_n is 0 is $\frac{1}{2}$ and the probability that $A_{n+1} = A_n$ is $p \leqslant \frac{1}{2}$. Find the entropy of A_1, and of the block A_1A_2. Find the Shannon-Fano coding for A_1A_2 and the Huffman encoding of A_1A_2. Explain how this allows us to compress the initial binary code for certain values of p.

10. Look up (or calculate) the probability for each letter in English prose. What is the entropy of a random letter with these probabilities? Find the optimal binary coding for this random variable.

11. (a) Suppose that x_n is a stream which is periodic with period M and y_n is a stream which is periodic with period N. Show that the streams $x_n + y_n$ and $x_n y_n$ are periodic with periods dividing the lowest common multiple of M and N.

    (b) One of the most confidential German codes (called FISH by the British) 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 stream of period 1497, what is the longest possible period of k_n? How many consecutive values of k_n do you need to specify the sequence completely?

12. Criticise the following authentication procedure. Alice chooses N as the public key for a 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 $r \equiv m_1^2 \pmod{N}$.

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

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

    (b) Show that the sequence $x_n = T(\alpha^n)$ is the output from a LFSR of length d.

    (c) 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 (b) has period $2^d - 1$.

    This shows that we can find LFSR of length d that achieve the maximum possible period.

14. Suppose that $N = pq$ where p and q are distinct primes with the same number of binary digits. We will 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.

Please send any comments or corrections to me at: t.k.carne@dpmms.cam.ac.uk .

Decipher:

```
klqhikg ip pl bawrqifre wp pmoikg

                                    gaowox jwkeat hlmcik
```

[*Hint: It is a substitution cipher.*]