Tutorial 2 Solutions

1. *Define the basic structure shared by all running key encryption algorithms. Agent Alice has to communicate with various operatives around the world, sending them long text messages. She decides to communicate with them using a literature based version of the running key algorithm, where the reference document is the Unix man page for the sh command, available to all of her operatives. How does this algorithm function and what is the key? This algorithm is not secure. Briefly describe how it can be attacked.*

In the running key algorithm the key is the same length as the plain text. Each letter of key and plaintext is combined to produce the code letter. A typical way of combining them is to add them modulo the alphabet size. In the literature based version the key consists of consecutive letters of a standard text, with the actual key the standard text and the starting character position in that text. This algorithm can be attacked using the redundancy in English (Friedman's method). The approach is to assume that the code letters are produced by a combination of high frequency letters in both the plain text and the key. This reduces the number of possible values for both the plain text and key text, and will be true often enough for digram and trigram analysis to guess a lot of the key and plain text.

2. *Describe how a rotor machine could be used to encrypt a text document, explaining the advantages of using it. Describe the Enigma machine variant of the rotor algorithm, pointing out the weakness incorporated by Enigma. What factors led to the breaking of the Enigma code?*

Each disk in a rotor machine has electrical contacts arranged in circles on both sides. Each contact on one each side is wired to a contact on the other side in a way that is unique to each rotor. There is the same number of contacts as letters in the encryption alphabet, and each pair of contacts is labelled with a letter on the circumference of the rotor. In this way, each rotor is a letter substitution machine. Several rotors are mounted on the same shaft so that their contacts touch and the substitutions are combined. The front rotor is connected to a keyboard while the back one is connected to a display. The rotors are then rotated with each key press using an odometer mechanism. The key is the initial letter positions at the top of the rotor. The advantage is the very long period before the settings repeat. The Enigma machine reflected the signal back through the rotors for a second time, which meant that corresponding plain text and cipher text letters could not be the same. The Enigma code was broken by lax security procedures which allowed a known plaintext attack, coupled with the large number of messages sent with each key. The weakness mentioned above allowed the position of the known plain text in the document to be determined.

3. *It is proposed that a modern mechanical rotor machine would make a useful encryption machine since it does not rely on computer technology, which can be*

compromised. How many rotors would be needed for secure encryption? Justify your answer.

The number of rotors determines the key length, and although the encryption devices would be mechanical, computers could be used to break the code. If we assume an alphabet of 32 letters to make the calculations easier, then each rotor would provide 5 key bits. We would need a key length of around 100 bits, leading to 20 rotors.

4. *Examine the following description of a single key encryption algorithm used to encrypt English text and answer the following questions.*

The plain text and cipher text both use an alphabet of 32 characters, the 26 letters of the alphabet and 6 punctuation characters. Encryption consists of taking the characters in blocks of 5 and rearranging them to form the cipher text. The rearrangement permutation is the key, and the inverse permutation is used to decrypt.

Why is this not a two key system, since different permutations are used to encrypt and decrypt? If a cipher text only attack were used to try and break the encryption, how many letters of cipher text would be needed to be confident that the code had been broken? Suggest a mechanism for dealing with messages that are not an exact multiple of 5 letters long. You may assume that the number of different ways of permuting n letters is $n \times (n-1) \times \ldots \times 1$.

This is not a two key system because each key can be derived from the other. Now we need a calculation of the unicity distance. $N = H(Key) / D$. There are 120 different keys, and so H(Key) is approximately 7 ($2^7 = 128$). R = 5 and we can assume r = 1.5, so that D = 3.5 for English, and so the answer is about 2 letters. Any sensible block padding mechanism is fine. For example, terminating the message with an impossible letter pair and then adding junk to the next block boundary.

5. *An isolated civilisation has developed a written language based on an alphabet with just 4 letters: α, β, γ and δ. Their written documents are very long. The letters do not occur with equal frequency: α occurs 3/8 of the time ; β and γ 1/4 of the time each and δ 1/8 of the time. The probability of any two letter combination occurring is however just the product of the probability of each letter occurring independently and there are no special digrams or trigrams. Calculate the redundancy of his language. You do not need to calculate an exact number but can leave terms like $\log_2(3)$ in your answer.*

R, the maximum rate of the language = 2
r = entropy per character = $3/8 \log_2(8/3) + 2 * ¼ * \log_2(4) + 1/8 * \log_2(8)$
$= 3/8 \log_2(8) – 3/8 * \log_2(3) + ½ * \log_2(4) + 1/8 * \log_2(8)$
$= 9/8 + 1 + 3/8 – 3/8 * \log_2(3)$

= 2 ½ - 3/8 * $\log_2(3)$
D = R – r = 3/8 $\log_2(3)$ – 1/2

6. *This civilisation is aware of the English language and has decided to encrypt their secret documents by using some English language letters for the cipher text. Show how they can hide the redundancy in their language by using the alphabet {A, B, C, D, E, F, G, H} as the cipher text alphabet.*

I am looking for a homophonic cipher. A suitable encoding would be:
α → ABC (or any three letters)
β → DE
γ → FG
δ → H

7. *This civilisation is also investigating the possibility of hiding two different messages in the cipher text, each with its own key. How many letters from the English language would be need? Give an example of such an encoding. Explain why this code would be easier to break than the previous one.*

This is a second order homophonic cipher. A matrix consisting of α β γ δ for both rows and columns is constructed. It will have 16 elements and so 16 letters would be needed. Letters from both messages are paired up and identify an element in the matrix, which is the cipher text letter. One key will associate all cipher text letters in each row with a plain text symbol, while the other would locate all letters in a column. Any example with this structure will do. This code does not destroy letter frequencies and high frequency letter pairs from both messages will make it easier to break.

