Queen Mary
University of London

MTH6115 **Cryptography**

Notes 1 **Winter/Spring 2011**

Introduction

Cryptography refers to the art of protecting transmitted information from unauthorised interception or tampering. The other side of the coin, *cryptanalysis*, is the art of breaking such secret ciphers and reading the information, or perhaps replacing it with different information. Sometimes the term *cryptology* is used to include both of these aspects. However, the term cryptography is often used colloquially to include both cryptography and cryptanalysis. In these notes I will use the term *cryptography* exclusively.

Cryptography is closely related to another part of communication theory, namely *coding theory*. This involves translating information of any kind (text, scientific data, pictures, sound, and so on) into a standard form for transmission, and protecting this information against distortion by random noise. There is a big difference, though, between interference by random noise, and interference by a purposeful enemy, and the techniques used are quite different.

The need for both coding theory and cryptography has been recognised for a long time. Here, from “The Tale of Lludd and Llevelys” in *The Mabinogion* (a collection of ancient Welsh stories), is a tale that illustrates both subjects.

> When Lludd told his brother the purpose of his errand Llevelys said that he already knew why Lludd had come. Then they sought some different way to discuss the problem, so that the wind would not carry it off and the Corannyeid learn of their conversation. Llevelys ordered a long horn of bronze to be made, and they spoke through that, but whatever one said to the other came out as hateful and contrary. When Llevelys perceived there was a devil frustrating them and causing trouble he ordered wine to be poured through the horn to wash it out, and the power of the wine drove the devil out.

Here the horn is a cryptographic device, preventing the message from being intercepted by the enemy (the Corannyeid); this is an example of a *secure channel*, which we will discuss later. Pouring wine down the horn is a bizarre form of error-correction.

Steganography and cryptography

There are two principal ways to keep a message out of the enemy's hands:

- You can conceal the message and hope that the enemy can't find it: this is known as *steganography*. [From the Greek words στεγανός 'steganos' meaning 'cover, roof' and γράφειν 'graphein' meaning 'to write'.]

- You can scramble the message, and hope that (assuming that it is intercepted) the enemy is unable to unscramble it: this is what is properly known as *cryptography*. [From the Greek word κρυπτός 'kryptos' meaning 'hidden'. Cryptology also has a part λόγος 'logos' meaning 'word'.]

We are mainly concerned with cryptography; but here are a few of the many methods of steganography that have been used or proposed.

- Herodotus relates that one Histauaeus shaved the head of his messenger, wrote the message on his scalp, and waited for the hair to re-grow. On reaching his destination, the messenger shaved his head again and the recipient, Aristogoras, read the message. Not to be recommended if you are in a hurry!

- Invisible ink comes into this category; the recipient develops the message by applying heat or chemicals to it.

- A message can be concealed in a much longer, innocent-looking piece of text; the long text is composed so that a subsequence of the letters (chosen by some rule known to the recipient) forms the message. For example, taking every fifth letter of

  ```
  The prepared letters bring news of amounts
  ```

  gives the message "Retreat".

- The message can be photographed and reduced to a tiny speck called a *microdot*, which can be concealed in a full stop in an ordinary letter.

- A recent proposal uses the fact that a molecule of DNA (the genetic material in all living things) can be regarded as a very long word in an alphabet of four letters A, C, G, T (the bases adenine, cytosine, guanine and thymine). Now that the technology exists to modify DNA very freely, it is possible to encode the message in this four-letter alphabet and insert this sequence into a DNA molecule. A small amount of DNA can then be concealed in a letter, in the same way as a microdot. (This method may or may not have been used.)

- Recently there has been a lot of discussion of the idea of "watermarks" in files. A large computer file such as a picture or a piece of music contains millions of bytes of information. The basic idea is to change a few bytes to include a copyright message. This can only work if the change in the look of the picture, or the sound of the music, is so small that it cannot be noticed, and if the changed bytes are in apparently random positions which cannot easily be found by someone attempting to copy the file illegally. No completely satisfactory method has yet been found.

Of course, steganography can be combined with cryptography: the message can be scrambled and then hidden, for extra security.

Some terms defined

Figure 1 shows the general scheme of cryptography. Traditionally, the two parties who want to communicate are called Alice and Bob, and the eavesdropper who is trying to read their message is Eve. Alice and Bob both have access to the key, but Eve doesn't. The black boxes input plaintext and key and output ciphertext (in Alice's case), or input ciphertext and key and output plaintext (in Bob's).

The terms in the figure have the following meanings.

Plaintext: The plaintext is not quite the same as the message being sent. The message probably has to be translated into some standard form to be encrypted; for example, this might be leaving out the punctuation, turning it into ASCII code or a sequence of numbers, etc. But there is nothing secret about this stage; knowing the plaintext is equivalent to knowing the message.

Ciphertext: The ciphertext is what is actually transmitted. In general Alice and Bob must assume that Eve can get her hands on the ciphertext, and they must design the system so that this will not enable her to recover the plaintext.

Key: The encryption uses some extra information, known as the key, which can be varied from one transmission to another. Both Alice and Bob must have information about the key, in order to perform the encryption and decryption.

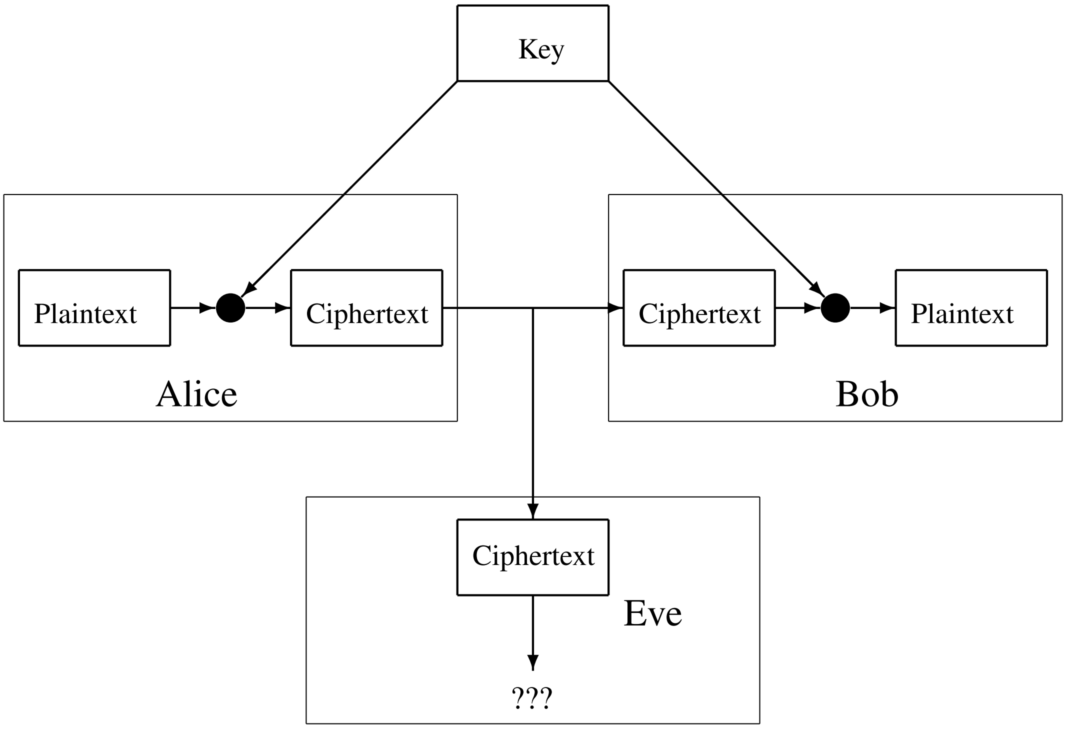

Figure 1: The set-up

There are three main types of encryption method:

Transposition: The order of the letters in the plaintext is rearranged in some systematic way. The key is the permutation applied to the positions.

Substitution: Individual letters are replaced by different letters in a systematic way. (We do allow a letter to 'replaced' by itself.) This may be more complicated than just a single permutation; we may apply different permutations to the letters in different positions. The key is the sequence of applied permutations.

Codebook: Complete words in the message are replaced by other words with quite different meanings. The key is the codebook, the list of words and their replacements.

Of course, the types are not completely separate, and some or all of them can be used together.

Note on the word "code": This word is used with many different meanings in communication theory. Often it just means a scheme for translating information from one format to another. Thus, for example, the Morse code (used in early telegraph and radio communication) would translate the word “Code” into the sequence

−·−· −−− −·· ·

of dots and dashes, while seven-bit ASCII (used in computer communication and representation of data) would translate it into the four numbers 67, 111, 100, 101, or, in binary notation,

$$1000011110111111001001100101.$$

An error-correcting code translates a string of symbols into a different string for the purposes of error correction. For example, a $[7,4,3]$-code might translate 1010 into 1010101, and more generally, translate the bit-string $abcd$ into the bit-string $abcdefg$ where $e = b+c+d$, $f = a+c+d$ and $g = a+b+d$ with addition being modulo 2.

The term “secret code” might mean what we have called a cipher system, or perhaps a cryptogram (the result of encrypting a message using a cipher system).

Within cryptography, a code replaces certain key words in the message by other words or combinations of symbols, as specified in the code book. This is sometimes contrasted with a cipher, which operates on the individual letters or symbols.

Pig-Latin

Pig-Latin is a simple form of transposition cipher with a “null” character. These rules are taken from the Pig-Latin homepage at

`http://www.idioma-software.com/pig/home.htm`.

(I have extended and clarified the rules in order to deal with ambiguous cases not dealt with by the rules from the said website.)

For words which begin with a single consonant followed by a vowel take the consonant off the front of the word and add it to the end of the word. Then add ‘ay’ after the consonant. Here are some examples:

cat	↦	atcay
dog	↦	ogday
simply	↦	implysay
noise	↦	oisnay

For words which began with double or multiple consonants take the group of consonants off the front of the word and add them to the end, adding ‘ay’ at the very end of the word. The encrypted word should commence with a vowel. Here are some examples:

scratch	↦	atchscray
thick	↦	ickthay
flight	↦	ightflay
grime	↦	imegray

For words that begin with a vowel, just add ‘yay’ at the end. For example:

is	↦	isyay
apple	↦	appleyay
under	↦	underyay
octopus	↦	octopusyay

For words consisting solely of consonants, just add ‘ay’ at the end. This is the only case when the encrypted word does not begin with a vowel.

rhythm	↦	rhythmay
cwm	↦	cwmay
nth	↦	nthay

In the above, by consonant and vowel, I mean orthographic consonant and vowel, rather their sounds. So the only vowels are ‘a’, ‘e’, ‘i’, ‘o’, ‘u’, with the remaining letters being consonants, despite the fact that ‘y’ often represents a vowel in English.

A sample of pig-Latin:

> Igpay-Atinlay opensyay upyay ayay ewnay orldway atthay ouyay evernay ouldway avehay oughtthay ossiblepay. Byay usingyay Igpay-Atinlay, ouyay ootay ancay ulfillfay ouryay ascinatingfay uturefay unctionsfay otay ethay ullestfay ullnessfay astfay. Ouyay illway ebay ayay etterbay ersonpay, avehay ayay etterbay exsay ifelay, andyay ebay etterbay anthay ouryay eighborsnay.