Semester 1 **Exercises for Week 3** 2004

Objectives: We review the facilities for cryptosystems in `Magma`, then investigate cryptanalytic methods for ciphertext. We will use these methods to crypanalyze ciphertext samples from the course web page.

Magma Cryptosystems

A cryptosystem can be created in `Magma` with the following commands:

```
> C := SubstitutionCryptosystem();
> V := VigenereCryptosystem(7);
> V;
Vigenere cryptosystem
```

The latter constructor (of the Vigenère cryptosystem) specifies a key length of 7 characters. Recall that the `Magma` type of a cryptosystem is `Crypt`, and that a cryptosystem key, of type `CryptKey` is viewed as an element of the cryptosystem. It specifies an particular enciphering map in the cryptosystem. A key can be created either by *coercion* (the `!` operator) of a valid keystring into the cryptosystem, or by the `RandomKey` function:

```
> RandomKey(C);
HYTIKQRWXUPZBJSCANEFODLVMG
> K := V!"PRNXJAM";
> K;
PRNXJAM
```

Finally we will need cryptographic text (plaintext and ciphertext). This will be of type `CryptTxt`, and can be created by the `Encoding` and `Enciphering` functions. To extract the underlying string (of `Magma` type `MonStgElt`), use the function `String`.

```
> M := "This is sample message text to be encoded.";
> PT := Encoding(V,M);
> PT;
THISISSAMPLEMESSAGETEXTTOBEENCODED
> CT := Enciphering(K,PT);
IYVPRSEPDCINMQHJNDNTQMKGLKEQCTBAND
> L := InverseKey(K);
> Enciphering(L,CT);
THISISSAMPLEMESSAGETEXTTOBEENCODED
> Type($1);
CryptTxt
> String($1);
THISISSAMPLEMESSAGETEXTTOBEENCODED
> Type($1);
MonStgElt
```

Cryptanalysis

One important measure of a cryptographic text is the *coincidence index*. If we define p_i to be the probability of occurrence of the i-th character of the codomain alphabet in a sample plaintext of ciphertext, then the coincidence index of the cryptosystem is the value:

$$\sum_{i=1}^{n} p_i^2,$$

where n is the size of the alphabet. Recall that the sum of all probabilities is 1, therefore the coincidence index is at most 1. The coincidence index is a measure of the probability that two randomly chosen characters are the same.

To compute the coincidence index for a particular string of length N, we first compute the number of occurrences n_i of each character in the alphabet. Then the probability of equality for two randomly chosen characters from that string is given by

$$\frac{\sum_{i=1}^{n} n_i(n_i - 1)}{N(N-1)},$$

which we define to be the coincidence index of a string.

For random text (uniformly distributed characters) in an alphabet of size 26, the coincidence index is approximately 0.0385. For English text, this value is 0.0661. Therefore we should be able to pick out text which is a simple substitution or a transposition of English text. Other languages will have an associated coincidence index, which

Another important tool for cryptanalysis *Kasiski test*. This is useful for determining periodicity in ciphertext. Ciphertext of a Vigenère cipher has a period m such that for each fixed j the $mi+j$-th characters are given simple substitution of the corresponding $mi+j$-th plaintext. If a frequently occurring pattern, such as `THE` is aligned at the same position with respect to this period, then the same three characters will appear in the ciphertext, at a distance which is an exact multiple of m. By looking for frequently occurring strings in the ciphertext, and measuring the most frequent divisors of the displacements of these strings, it is often possible to identify the period, hence to reduce to a simple substitution.

Exercises

The `Magma` crypto package functions

`CoincidenceIndex`, `Decimation`, and `FrequencyDistribution`

provide functionality for analysis of the ciphertexts in the exercises.

1. For each of the cryptographic texts from the course web page, compute the coincidence index of the ciphertexts. Can you tell which come from simple substitution or transposition ciphers? How could you distinguish the two?

2. For each of the cryptographic texts from the course web page, for various periods extract the substrings of $im + j$-th characters. For those which are not simple substitutions, can you identify a period?

3. For each of the ciphertexts which you have reduced to simple substitutions, consider the frequency distribution of the simple substitution texts. Now recover the keys and original plaintext.