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?

Solution The coicidence index for each of the ciphertext samples is given in the table below.

1.	0.04387225548902195608782435129
2.	0.06570233203879852160372225661
3.	0.04472396925227113906359189377
4.	0.06295453108202112186175931195
5.	0.04128012438455558434827675563
6.	0.06552250683626455449665258769
7.	0.04123391156371732030776195996
8.	0.06749180610660686562014835259
9.	0.06855734826766228859701757894
10.	0.06573417580940245377974433462
11.	0.06658477799932729884470080593

All but ciphertexts 1, 3, 5 and 7 are consistent with output from a simple substitution or transposition cipher. It is likely that the exceptional ones employ a polyalphabetic cipher. In order to distinguish substitution and transposition ciphers, it is necessary to look at character distributions, e.g. a close match with the frequency distribution of English (or a modern language) suggests a transposition cipher.

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?

Solution Using the average coincidence index of the ciphertext decimations, we find that the periods of the ciphertexts 1, 3, 5, and 7 are 11, 6, 14, and 9, respectively. The code to verify this is:

```
CI := CoincidenceIndex;
for m in [1..9] do
    print m, &+[ CI(Decimation(c05,i,m)) : i in [1..m] ]/m;
end for;
```

with output (for ciphertext 5):

```
1 0.04128012438455558434827675563
2 0.04252125629012667013708138122
3 0.04116798565358075708549374293
4 0.04333449356073790462930734423
5 0.04067538126361655773420479302
6 0.04311525495153813737884534344
7 0.05476804123711340206185567010
8 0.04296218487394957983193277310
9 0.04264264264264264264264264264
```

Noting the minor peak at $m = 7$, we continue with the loop:

```
for m in [10..18] do
    print m, &+[ CI(Decimation(c05,i,m)) : i in [1..m] ]/m;
end for;
```

to find the true period is $m = 14$:

```
10 0.04205443371378402107111501316
11 0.04187779433681073025335320417
12 0.04415584415584415584415584415
13 0.03927369764473836872026917275
14 0.07231509625126646403242147922
15 0.03939393939393939393939393938
16 0.04478803716608594657375145179
17 0.04185520361990950226244343891
18 0.04479479479479479479479479479
```

Note that the coincidence index for each even test period is slightly higher, since these involve an averaging over only 7, rather than 14, distinct substitutions.

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.

Solution The first step in recovering the keys and plaintext is to determine the type of cipher; further techniques are studied in the next tutorial. Note that the ciphertexts 1, 3, 5, and 7 are the result of Vigenère cryptosystems, and can be deciphered by statistical analysis of the each of the decimations with respect to their periods. A javascript program from the course web page can be used for this purpose.