0.0.1 Simple Substitution

The System and its Main Weakness

With the method of a *simple substitution* one chooses a fixed permutation π of the alphabet $\{a, b, \ldots, z\}$ and applies that to all letters in the plaintext.

Example 0.1

In the following example we only give that part of the substitution π that is relevant for the given plaintext. We use the Mathematica function StringReplace.

```
StringReplace["plaintext",
 {"a" -> "k", "e" -> "z", "i" -> "b", "l" -> "r",
  "n" -> "a", "p" -> "v", "t" -> "q", "x" -> "d"}]
```

```
vrkbaqzdq
```

A more formal description of the simple substitution system is as follows: the key space $\mathcal{K}$ is the set S_q of all permutations of $\{0, 1, \ldots, q-1\}$ and the cryptosystem $\mathcal{E}$ is given by

$$\mathcal{E} = \{E_\pi \mid \pi \in S_q\},$$

where

$$E_\pi(m) = \pi(m),\ 0 \le m < q.$$

The decryption function D_π is given by $D_\pi = E_{\pi^{-1}}$, as follows from

$$D_\pi(E_\pi(m)) = D(\pi(m)) = E_{\pi^{-1}}(\pi(m)) = \pi^{-1}(\pi(m)) = m,\quad 0 \le m < q.$$

Unlike Caesar's cipher, this system does not have the drawback of a small key space. Indeed, $|\mathcal{K}| = |S_{26}| = 26! \approx 4.03\ 10^{26}$. This system however does demonstrate very well that a large key space should not fool one into believing that a system is secure! On the contrary, by simply counting the letter frequencies in the ciphertexts and comparing these with the letter frequencies in Table 1.1, one very quickly finds the images under π of the most frequent letters in the plaintext. Indeed, the most frequent letter in the ciphertext will very likely be the image under π of the letter *e*. The next one is the image of the letter *n*, etc. After having found the encryptions of the most frequent letters in the plaintext, it is not difficult to fill in the rest. Of course, the longer the cipher text, the easier the cryptanalysis becomes. In Chapter 5, we come back to the cryptanalysis of the system, in particular how long the same key can be used safely.

Cryptanalysis by The Method of a Probable Word

In the following example we have knowledge of a very long ciphertext. This is not necessary at all for the cryptanalysis of the ciphertext, but it takes that long to know the full key. Indeed, as long as two letters are missing in the plaintext, one does not know the full key, but the system is of course broken much earlier than that.

Apart from the ciphertext, given in Table 2.2, we shall assume in this example that the plaintext discusses the concept of "bidirectional communication theory". Cryptanalysis will turn out to be very easy.

```
zhjeo ndize hicle osiol digic  lmhzq zolyi zehdp zhjeo ndize
hycdh hlpvs uczyc dhzhj eondi  zehge moylk zhjpm lhylg gidiz
gizyd ppsdo lylzr losye nnmhz  ydize hicle osceu lrloq lgyoz
vlgic lneol flhlo dpydg lzhuc  zyciu eeone olzhj eondi zehge
moylg zhjpm lhyll dycei clogi  dizgi zydpp siclq zolyi zehej
iczgz hjpml hylzg lkaol gglqv  sqzol yilqi odhgj eondi zehxm
dhizi zlguc zycyd hehps vlqlo  zrlqz jiclp duejy dmgdp ziszg
evglo rlqqz gizhf mzgcz hficl  ldopz loydm gljoe niclp dilol
jjlyi zhvze pefsd hqgey zepef  syenn mhzyd izehi cleos gllng
iecdr luzql daapz ydize hgqml  ieicl jdyii cdipz rzhfv lzhfg
dolvs iclzo dyize hggem oylge  jzhje ondiz ehucz yczhj pmlhy
lldyc eiclo zhdpp aeggz vplqz  olyiz ehgic laolg lhiad aloql
gyzvl gicly dglej vzqzo lyize  hdpye nnmhz ydize hicle osdaa
pzlqi eiclg eyzdp vlcdr zemoe  jneht lsg…
```

Ciphertext obtained with a simple substitution

Table 0.1

Assuming that the word "communication" will occur in the plaintext, we look for strings of 13 consecutive letters, in which letter 1 = letter 8, letter 2 = letter 12, letter 3 = letter 4, letter 6 = letter 13 and letter 7 = letter 11.

Indeed, we find the string "yennmhzydizeh" three times in the ciphertext. This gives the following information about π.

c	*o*	*m*	*u*	*n*	*i*	*a*	*t*
↓	↓	↓	↓	↓	↓	↓	↓
y	*e*	*n*	*m*	*h*	*z*	*d*	*i*

Assuming that the word "direction" does also occur in the plaintext, we need to look for strings of the form "*z**yizeh" in the ciphertext, because of the information that we already have on π. It turns out that "qzolyizeh" appears four times, giving:

d	*r*	*e*
↓	↓	↓
q	*o*	*l*

If we substitute all this information in the ciphertext one easily obtains π completely. For instance, the text begins like

in*ormationt*eor*treat*t*eunid…,

which obviously comes from

information theory treats the unid(irectional) …,

This gives the π-image of the letters f, h, y and s.

Continuing like this, one readily obtains π completely.

```
a b c d e f g h i j k l m n o p q r s t u v w x y z
↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓
d v y q l j f c z w t p n h e a x o g i m r u k s b
```

Example 0.2

Mathematica makes is quite easy to find a substring with a certain pattern.For instance, to test where in a text one can find a substring of length 6 with letters 1 and 4 equal and also letters 2 and 5 (as in the Latin word "quoque"), one can use the Mathematica functions If, StringTake, StringLength, Do, Print, and the following:

```
ciphertext = "xyuysuyifvyxi";
Do[
 If[StringTake[ciphertext, {i + 1}] == StringTake[ciphertext,
     {i + 4}] ⋀ StringTake[ciphertext, {i + 2}] ==
    StringTake[ciphertext, {i + 5}],
  Print[i + 1, "   ", StringTake[ciphertext, {i + 1, i + 6}]]],
 {i, 0, StringLength[ciphertext] - 6}]
```

```
3   uysuyi
```

This example was taken from Table 2.1.