3. Homophone Substitutionschiffrierung

Eine einfache Substitutionschiffrierung wird gegeben durch eine injektive Abbildung

$$f : \text{Klartextalphabet} \longrightarrow \text{Chiffretextalphabet},$$

wobei ein Ausgangstext $a_1a_2a_3\dots$ dann zu $f(a_1)f(a_2)f(a_3)\dots$ verschlüsselt wird. Man spricht auch von monoalphabetischer Chiffrierung. Die Caesar-Verschlüsselung ist ein Beispiel dafür. Durch Häufigkeitsanalyse kann man oft einen Chiffretext entschlüsseln.

Homophone Substitutionschiffrierung:

(1) Sei Σ_1 das Klartextalphabet, Σ_2 das Chiffretextalphabet.
(2) Für jedes $a \in \Sigma_1$ wählt man eine Teilmenge $F(a) \subseteq \Sigma_2$, wobei die Mengen $F(a)$ für verschiedene a's disjunkt sind, d.h. $F(a) \cap F(b) = \emptyset$ für $a \neq b$.
(3) Ein Ausgangstext $a_1a_2a_3\dots$ wird nun wie folgt verschlüsselt:
   - Für jedes i wählt man zufällig ein $b_i \in F(a_i)$.
   - Der Chiffretext ist $b_1b_2b_3\dots$.

Beispiel: Ungefähr folgende Tabelle wurde im 2. Weltkrieg auf amerikanischer Seite (Brigadier General Leslie R. Groves, [**Kahn**, S.546] benutzt:

	1	2	3	4	5	6	7	8	9	0
1	I	P	I		O	U	O		P	N
2	W	E	U	T	E	K		L	O	
3	E	U	G	N	B	T	N		S	T
4	T	A	Z	M	D		I	O	E	
5	S	V	T	J		E		Y		H
6	N	A	O	L	N	S	U	G	O	E
7		C	B	A	F	R	S		I	R
8	I	C	W	Y	R	U	A	M		
9	M	V	T		H	P	D	I	X	Q
0	L	S	R	E	T	D	E	A	H	E

Um einen Buchstaben zu verschlüsseln, sucht man sich zufällig eine Stelle in der Tafel, wo er auftritt, und notiert sich dann Zeilen- und Spaltennummer. So wird TODAY zu 24 69 06 42 58 oder auch zu 41 63 97 62 84 verschlüsselt.

Überlegung: Wir denken hier an $\Sigma_1 = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \dots, \mathrm{Z}\}$. Sei $h(a)$ die Häufigkeit, mit der das Zeichen a in einem 'durchschnittlichen' Text auftritt. Wählen wir bei der Verschlüsselung $b \in F(a)$ (gleichverteilt) zufällig, so wird die Häufigkeit von $b \in F(a)$ im verschlüsselten Text

$$h(b) \approx \frac{h(a)}{\#F(a)}.$$

Wollen wir, dass alle Zeichen im verschlüsselten Text gleichhäufig auftreten, so müssen wir verlangen, dass

$$\frac{h(a)}{\#F(a)}$$

unabhängig von a ist, d.h. es gibt eine Konstante c mit

$$\#F(a) \approx c \cdot h(a) \quad \text{für alle} \quad a \in \Sigma_1.$$

Ist $N = \sum_{a\in\Sigma_1} \#F(a)$, so gilt also

$$N = \sum_{a\in\Sigma_1} \#F(a) \approx \sum_{a\in\Sigma_1} c \cdot h(a) = c \cdot \sum_{a\in\Sigma_1} h(a) = c,$$

d.h.

$$\#F(a) \approx N \cdot h(a).$$

Bei dieser Wahl von $F(a)$ sollten die Zeichen des Chiffretexts alle ungefähr gleichhäufig auftreten. Daher nützt hier eine Häufigkeitsanalyse der Zeichen nichts mehr.

Beispiel: Wir wählen $\Sigma_1 = \{\mathrm{A}, \ldots, \mathrm{Z}\}$ und $\sigma_2 = \{00, 01, \ldots, 99\}$. In der folgenden Tabelle ist $h(a)$ eine hypothetische Zeichenwahrscheinlichkeit des Zeichens a, wie sie sich bei [**?**, S.286] findet. Dazu wählen wir $F(a) \approx 100 \cdot h(a)$, sodass $\#F(a) \geq 1$ (Jedes Zeichen muss sich verschlüsseln lassen.) und $\sum_{a \in \Sigma_1} \#F(a) = 100$ gilt.

	$h(a)$	$\#F(a)$
A	6.47	6
B	1.93	2
C	2.68	* 2
D	4.83	5
E	17.48	17
F	1.65	2
G	3.06	3
H	4.23	4
I	7.73	8
J	0.27	* 1
K	1.46	1
L	3.49	3
M	2.58	* 2

	$h(a)$	$\#F(a)$
N	9.84	10
O	2.98	3
P	0.96	1
Q	0.02	* 1
R	7.54	* 7
S	6.83	7
T	6.13	6
U	4.17	4
V	0.94	1
W	1.48	1
X	0.04	* 1
Y	0.08	* 1
Z	1.14	1

(Das Zeichen * in der dritten Spalte bedeutet, dass $\#F(a)$ nicht als die $100 \cdot h(a)$ nächstgelegene ganze Zahl gewählt wurde.)

Beispiel:

a	$F(a)$
A	$\{11, 20, 30, 68, 77, 85\}$
B	$\{91, 94\}$
C	$\{10, 39\}$
D	$\{02, 08, 41, 51, 93\}$
E	$\{04, 07, 17, 18, 22, 32, 35, 44, 49, 53, 56, 57, 60, 64, 73, 74, 90\}$
F	$\{24, 36\}$
G	$\{09, 25, 65\}$
H	$\{14, 40, 47, 66\}$
I	$\{03, 12, 15, 43, 76, 95, 97, 98\}$
J	$\{42\}$
K	$\{27\}$
L	$\{79, 81, 92\}$
M	$\{88, 89\}$
N	$\{05, 06, 19, 38, 54, 58, 59, 61, 69, 83\}$
O	$\{16, 55, 84\}$
P	$\{71\}$
Q	$\{34\}$
R	$\{00, 01, 23, 28, 37, 46, 48\}$
S	$\{21, 29, 52, 63, 67, 86, 87\}$
T	$\{26, 33, 70, 72, 75, 82\}$
U	$\{45, 62, 96, 99\}$
V	$\{80\}$
W	$\{78\}$
X	$\{50\}$
Y	$\{13\}$
Z	$\{31\}$

Damit verschlüsselt sich ‘KRYPTOGRAPHIEUNDMATHEMATIK’ zu

```
27 01 13 71 26 55 25 46 85 71 47 43 90 96 05 02 89 30 26 40 73 89 85 33 95 27
```

oder auch zu

```
27 37 13 71 82 55 25 23 68 71 47 43 18 99 38 08 88 11 75 14 07 88 68 72 97 27
```