2 Monographische, monoalphabetische Substitution

2.1 Lineare Substitution

- ROT13 ist definiert durch eine Funktion $\chi : V \longrightarrow W$, die zeichenweise angewandt wird
  - $\mathbf{X}$ wird durch χ unter Konkatenation induziert: $\mathbf{X}(x_1x_2x_3\ldots) = y_1y_2y_3\ldots$ mit $x_i \in V, y_i = \chi(x_i) \in W$
- zur Erinnerung:

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

- definiere Isomorphismus $V \cong W \cong \mathbb{Z}_{26}$ gemäß $\mathtt{a} \equiv \mathtt{A} \equiv 0, \mathtt{b} \equiv \mathtt{B} \equiv 1, \ldots, \mathtt{z} \equiv \mathtt{Z} \equiv 25$
- damit $\chi(x) = x + 13$
  - $\chi(\mathtt{i}) \equiv \chi(8) = 21 \equiv \mathtt{V}$
  - $\chi(\mathtt{s}) \equiv \chi(18) = 31 \simeq 5 \equiv \mathtt{F}$

- einer der Nachteile von ROT13: einmal bekannt, immer bekannt
- wünschenswert: Verfahren, das über einen Schlüssel parametriert werden kann
  - Wechsel des Schlüssels einfacher als Wechsel des Verfahrens
- CAESAR-Addition $\chi(x) = x + t$ mit Schlüssel t als Verallgemeinerung von ROT13
- Dechiffrierung mit $\chi^{-1}(x) = x - t$
- Beispiel für $t = 3$:

a	l	e	a	i	a	c	t	a	e	s	t
D	O	H	D	L	D	F	W	D	H	V	W

Angriffsmöglichkeiten:

- exhaustiv (bei nur 26 möglichen Schlüsseln)
- über Buchstabenhäufigkeiten

```
PDQGD UIVLF KGDKH UZHGH UYHUU HQQHQ QRFKD EVFKU HFNHQ ODVVH QXQGP XVVHV
PDFKH QZLHL QGHUS ROLWL NXPVF KZHQN HQHWL HQQHE DCHUL HV
```

C	D	E	F	G	H	I	K	L	N	O	P	Q	R	S	U	V	W	X	Y	Z
1	7	2	6	5	20	1	6	7	3	2	4	14	2	1	8	9	2	3	1	3

Häufigkeiten im Deutschen:

Große Ähnlichkeit der Häufigkeitsgebirge für $t = 3$.


```
PDQGD UIVLF KGDKH UZHGH UYHUU HQQHQ QRFKD EVFKU HFNHQ ODVVH QXQGP XVVHV
PDFKH QZLHL QGHUS ROLWL NXPVF KZHQN HQHWL HQQHE DCHUL HV
```

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

```
manda rfsic hdahe rwede rverr ennen nocha bschr ecken lasse nundm usses
mache nwiei nderp oliti kumsc hwenk eneti enneb azeri es
```

Man darf sich daher weder verrennen noch abschrecken lassen und muss es machen wie in der Politik: umschwenken. — Étienne Bazeries[1]

[1] Étienne Bazeries (1846–1931), französischer Kryptologe, Verfasser des Buches „Les chiffres secrets dévoilés“

2.2 Bestimmung der Verschiebung t mittels Chi

Seien T, T' Texte über demselben Alphabet V, $|V| = N$, p_i und p'_i jeweils die relativen Häufigkeiten des i-ten Zeichens des Alphabets; definiere:

$$\mathrm{Chi}(T, T') = \sum_{i=1}^{N} p_i \cdot p'_i$$

Betrachte Chi für T langen deutschen Text (p_i Buchstabenhäufigkeiten im Deutschen), T' entschlüsselten Text für verschiedene $t \Rightarrow$ Korrelation der Histogramme

$t = 0$:

T	a	b	c	d	e	f	g	h	i	j	k	l	m
p_i in %	6.47	1.93	2.68	4.83	17.48	1.65	3.06	4.23	7.74	0.27	1.46	3.49	2.58
T'	A	B	C	D	E	F	G	H	I	J	K	L	M
p'_i in %	0.00	0.00	0.93	6.54	1.87	5.61	4.67	18.69	0.93	0.00	5.61	6.54	0.00
$p_i \cdot p'_i/10^2$	0.00	0.00	0.03	0.32	0.33	0.09	0.14	0.79	0.07	0.00	0.08	0.23	0.00

T	n	o	p	q	r	s	t	u	v	w	x	y	z
p_i in %	9.84	2.98	0.96	0.02	7.54	6.83	6.13	4.17	0.94	1.48	0.04	0.08	1.14
T'	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
p'_i in %	2.80	1.87	3.74	13.08	1.87	0.93	0.00	7.48	8.41	1.87	2.80	0.93	2.80
$p_i \cdot p'_i/10^2$	0.28	0.06	0.04	0.00	0.14	0.06	0.00	0.31	0.08	0.03	0.00	0.00	0.03

$$\mathrm{Chi}(T, T') = 3.10 \cdot 10^{-2}$$

$t = 1$:

T	a	b	c	d	e	f	g	h	i	j	k	l	m
p_i in %	6.47	1.93	2.68	4.83	17.48	1.65	3.06	4.23	7.74	0.27	1.46	3.49	2.58
T'	B	C	D	E	F	G	H	I	J	K	L	M	N
p'_i in %	0.00	0.93	6.54	1.87	5.61	4.67	18.69	0.93	0.00	5.61	6.54	0.00	2.80
$p_i \cdot p'_i/10^2$	0.00	0.02	0.18	0.09	0.98	0.08	0.57	0.04	0.00	0.02	0.10	0.00	0.07

T	n	o	p	q	r	s	t	u	v	w	x	y	z
p_i in %	9.84	2.98	0.96	0.02	7.54	6.83	6.13	4.17	0.94	1.48	0.04	0.08	1.14
T'	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
p'_i in %	1.87	3.74	13.08	1.87	0.93	0.00	7.48	8.41	1.87	2.80	0.93	2.80	0.00
$p_i \cdot p'_i/10^2$	0.18	0.11	0.13	0.00	0.07	0.00	0.46	0.35	0.02	0.04	0.00	0.00	0.00

$$\mathrm{Chi}(T, T') = 3.50 \cdot 10^{-2}$$

$t = 2$:

T	a	b	c	d	e	f	g	h	i	j	k	l	m
p_i in %	6.47	1.93	2.68	4.83	17.48	1.65	3.06	4.23	7.74	0.27	1.46	3.49	2.58
T'	C	D	E	F	G	H	I	J	K	L	M	N	O
p'_i in %	0.93	6.54	1.87	5.61	4.67	18.69	0.93	0.00	5.61	6.54	0.00	2.80	1.87
$p_i \cdot p'_i/10^2$	0.06	0.13	0.05	0.27	0.82	0.31	0.03	0.00	0.43	0.02	0.00	0.10	0.05

T	n	o	p	q	r	s	t	u	v	w	x	y	z
p_i in %	9.84	2.98	0.96	0.02	7.54	6.83	6.13	4.17	0.94	1.48	0.04	0.08	1.14
T'	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
p'_i in %	3.74	13.08	1.87	0.93	0.00	7.48	8.41	1.87	2.80	0.93	2.80	0.00	0.00
$p_i \cdot p'_i/10^2$	0.37	0.39	0.02	0.00	0.00	0.51	0.52	0.08	0.03	0.01	0.00	0.00	0.00

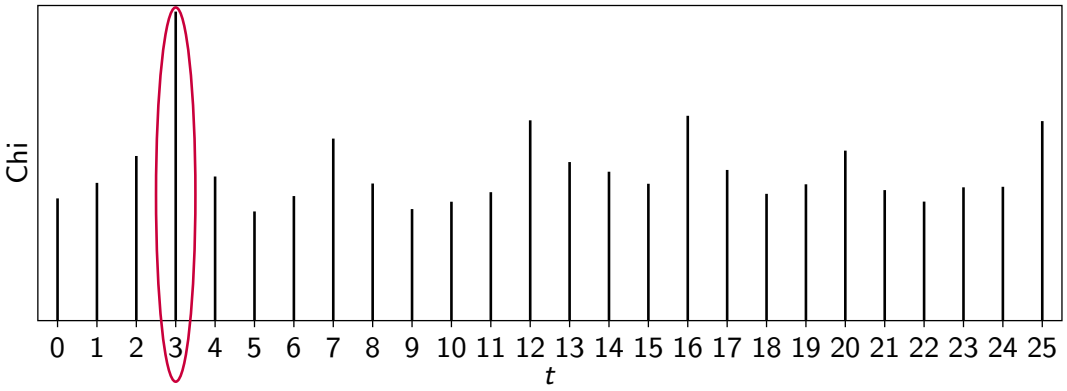
$$\mathrm{Chi}(T, T') = 4.18 \cdot 10^{-2}$$

$t = 3$:

T	a	b	c	d	e	f	g	h	i	j	k	l	m
p_i in %	6.47	1.93	2.68	4.83	17.48	1.65	3.06	4.23	7.74	0.27	1.46	3.49	2.58
T'	D	E	F	G	H	I	J	K	L	M	N	O	P
p_i' in %	6.54	1.87	5.61	4.67	18.69	0.93	0.00	5.61	6.54	0.00	2.80	1.87	3.74
$p_i \cdot p_i'/10^2$	0.42	0.04	0.15	0.23	3.27	0.02	0.00	0.24	0.51	0.00	0.04	0.07	0.10

T	n	o	p	q	r	s	t	u	v	w	x	y	z
p_i in %	9.84	2.98	0.96	0.02	7.54	6.83	6.13	4.17	0.94	1.48	0.04	0.08	1.14
T'	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
p_i' in %	13.08	1.87	0.93	0.00	7.48	8.41	1.87	2.80	0.93	2.80	0.00	0.00	0.93
$p_i \cdot p_i'/10^2$	1.29	0.06	0.01	0.00	0.56	0.57	0.11	0.12	0.01	0.04	0.00	0.00	0.01

$$\text{Chi}(T, T') = 7.85 \cdot 10^{-2}$$



Maximum bei $t = 3$ deutlich erkennbar

2.3 Permutation

- bijektive Funktion $\chi : V \longrightarrow V$, V endlich heißt Permutation
- entspricht einfacher Substitution, wenn $V = W$
  - hier auch $V = W$, falls V Kleinbuchstaben, W Großbuchstaben
- Darstellung durch Tabelle

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
S	E	C	U	R	I	T	Y	A	B	D	F	G	H	J	K	L	M	N	O	P	Q	V	W	X	Z

- oder Zyklenschreibweise

(asnhyxwvqlfi)(bermgtoj)(c)(dupk)(z)

- CAESAR für $t = 1$ in Zyklenschreibweise:

(abcdefghijklmnopqrstuvwxyz)

- ROT13 in Zyklenschreibweise:

(an)(bo)(cp)(dq)(er)(fs)(gt)(hu)(iv)(jw)(kx)(ly)(mz)

- voll zyklisch, falls nur ein Zyklus
- involutorisch, falls nur Einer- und Zweier-Zyklen
- *echt involutorisch*, falls nur Zweier-Zyklen

Angriffsmöglichkeiten:

- Buchstabenhäufigkeiten – schwieriger als bei linearer Substitution
- Häufigkeiten von Buchstabengruppen (Bigramme, evtl. Trigramme) – relativ viel Text nötig, um gute Annäherung an Wahrscheinlichkeiten zu erreichen
- Mustersuche, falls wahrscheinliches Wort bekannt:
  `QJBOY BETCT EYXTU JYTHC IBMEX TEECO FCOOX TEECO TEYDC HICOS Y`
  Wir vermuten, der Klartext enthält "Wissen", Muster 123345.
  - `wissen` ↔ `XTEECO` oder `wissen` ↔ `FCOOXT`
  - erste Möglichkeit:
    `QJBnYBsieisYwiUJYiHeIBMswissenFennwissenisYDeHIenSY`
  - zweite Möglichkeit:
    `QJBsYBEninEYenUJYnHiIBMEenEEiswissenEEisnEYDiHIisSY`
  - verfolge erste Möglichkeit mit `F=d`, `Y=t`:
    `QJBntBsieistwiUJtiHeIBMswissendennwissenistDeHIenSt`
  - usw. . .
    Phantasie ist wichtiger als Wissen, denn Wissen ist begrenzt. — Albert Einstein
- falls kein Buchstabe auf sich selbst abgebildet werden kann: zusätzlich negative Mustersuche
- ohne wahrscheinliches Wort: Ausnutzen markanter Muster im Geheimtext, z.B. 12132435 im Englischen nur "fiftieth"

2.4 Schlüsselgewinnung aus Kennwörtern

- Aus praktischen Gründen ist häufig ein gut memorierbarer Schlüssel, ein Kennwort oder Kennsatz, wünschenswert.
- Beispiel: `KREIDETAFEL`
- häufiges Verfahren:
  1. mehrfach vorkommende Buchstaben streichen ⇒ `KREIDTAFL`
  2. fehlende Buchstaben ergänzen ⇒ `KREIDTAFLBCGHJMNOPQSUVWXYZ`
  3. in Substitutionstabelle

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
K	R	E	I	D	T	A	F	L	B	C	G	H	J	M	N	O	P	Q	S	U	V	W	X	Y	Z

  oder als Zyklus (kreidtaflbcghjmnopqsuvwxyz) verwenden
- Nachteile:
  - Nach erstem Einstieg kann die Rekonstruktion des Kennworts die Entschlüsselung vereinfachen.
  - mit obiger Vorschrift: zu häufig z = Z, y = Y, ...

Um die Nachteile eines Kennworts zumindest zu lindern, sollte wenigstens eine Umstellung z.B. folgender Art erfolgen:

K	R	E	I	D	T	A	F	L
B	C	G	H	J	M	N	O	P
Q	S	U	V	W	X	Y	Z	

spaltenweises Auslesen: `KBQRCSEGUIHVDJWTMXANYFOZLP`

Überblick über die kombinatorische Komplexität:

Anzahl der Permutationen	$\|V\| = N$	$\|V\| = 26$
alle	$N!$	$4{,}03 \cdot 10^{26}$
voll zyklische	$(N-1)!$	$1{,}55 \cdot 10^{25}$
involutorische	$\approx N \cdot \sqrt{N!}$	$5{,}33 \cdot 10^{14}$
echt involutorische	$(N-1)!!$	$7{,}91 \cdot 10^{12}$
aus sinnvollen Kennwörtern gewonnene		$10^4 \ldots 10^6$

2.5 Multipartite Substitution

- $\chi : V \longrightarrow W^m, m > 1$
  - $|W|^m \geq |V|$
  - in der Regel $|W| < |V|$
- Beispiel für $W = \{1, 2, 3, 4, 5\}, m = 2$ (V enthält kein `j`):

	1	2	3	4	5
1	a	b	c	d	e
2	f	g	h	i	k
3	l	m	n	o	p
4	q	r	s	t	u
5	v	w	x	y	z

  `nachricht` $\Rightarrow$ 331131322442313244
  - Nutzung als Klopfcode in Haftanstalten
- Einsatz (in Kombination mit einer Transposition) an der deutschen Westfront 1918 mit $W = \{\mathtt{A}, \mathtt{D}, \mathtt{F}, \mathtt{G}, \mathtt{V}, \mathtt{X}\}, m = 2$ und $V = \{\mathtt{a}, \ldots, \mathtt{z}, 0, \ldots, 9\}$
- falls $|W|^m > |V|$ Homophone und/oder Blender möglich

3 Monographische, polyalphabetische Substitution

- monographisch: Verschlüsselung wird auf einzelne Zeichen angewendet
- monoalphabetisch: jedes Zeichen wird gleich verschlüsselt
- polyalphabetisch: Zeichen werden unterschiedlich verschlüsselt
  - $\mathbf{X}(x_1 x_2 x_3 \ldots) = y_1 y_2 y_3 \ldots$ mit $x_i \in V, y_i = \chi_i(x_i) \in W$
- Vorteil: Angriff über Zeichenhäufigkeiten nicht (ohne weiteres) möglich
- praktisches Problem: Bestimmung der χ_i durch Schlüssel
  - sicherste Variante: jedes χ_i ist eine von $N!$ mögichen Substitutionen – aber Schlüssel schwer handhabbar
  - begleitende Alphabete als systematische Ableitung zusätzlicher Chiffrierschritte