KAPITEL 2

Klassische Chiffrierverfahren

Im Folgenden werden wir ein paar klassische Chiffrierverfahren vorstellen. Weitergehende historische Anmerkungen findet man unter anderem bei [**Bauer**], [**Kahn**], [**Kippenhahn**], [**Singh**].

1. Vigenère-Verschlüsselung und Kasiski-Test

Die Vigenère-Verschlüsselung ist eine Strom-Verschlüsselung, bei der das Schlüsselwort immer wieder aneinandergehängt wird um den Schlüsselstrom zu erzeugen.

VIGENÈRE-Verschlüsselung:

(1) Als Alphabet liegen die 26 Großbuchstaben A,B,C,...,Z zugrunde, die wir mit den Zahlen 0,1,2,...,25 identifizieren.

(2) **Schlüssel:** Als Schlüssel dient ein Wort (aus Großbuchstaben) $k_1k_2k_3\dots k_n$, das wir auch als Folge von n Zahlen aus $\{0,\dots,25\}$ interpretieren können. Der Schlüssel entsteht aus dem Schlüsselwort durch periodische Fortsetzung:

$$k_1k_2\dots k_nk_1k_2\dots k_nk_1k_2\dots k_nk_1k_2\dots k_n\dots.$$

Den Schlüssel kann man sich auch durch

$$k_i = k_{i \bmod n} \text{ für } i \geq n \quad \text{mit} \quad k_0 = k_n$$

periodisch fortgesetzt denken.

(3) **Verschlüsselung:** Den zu verschlüsselnden Text fassen wir als Zahlenfolge $a_1, a_2, a_3, \dots$ mit $0 \leq a_i \leq 25$ auf. Dieser wird zur Folge $b_1, b_2, b_3, \dots$ verschlüsselt, wobei

$$b_i = a_i + k_{i \bmod n} \bmod 26$$

berechnet wird. (Denkt man sich den Schlüssel periodisch fortgesetzt, so lautet die Verschlüsselungsvorschrift einfach $b_i = a_i + k_i \bmod 26$.)

(4) **Entschlüsselung:** Entschlüsselt wird mit $a_i = b_i - k_{i \bmod n} \bmod 26$.

Beispiel: Mit dem Schlüsselwort „FREITAG“ verschlüsseln wir den Text „AUCH IM OKTOBER KANN ES REGNEN“:

Text	A	U	C	H	I	M	O	K	T	O	B	E	R	K	A	N	N	E	S	R	E	G	N	E	N
Schlüssel	F	R	E	I	T	A	G	F	R	E	I	T	A	G	F	R	E	I	T	A	G	F	R	E	I
Chiffretext	F	L	G	P	B	M	U	P	K	S	J	X	R	Q	F	E	R	M	L	R	K	L	E	I	V

Bemerkung: Häufigkeitsanalyse von Zeichen hilft bei der Vigenère-Verschlüsselung zunächst nicht weiter. Daher galt diese Verschlüsselung längere Zeit als sicher. Mitte des 19. Jahrhunderts hatte Friedrich Wilhelm Kasiski eine Idee, wie man die Vigenère-Verschlüsselung doch erfolgreich angreifen kann (Kasiski-Methode).

Der Kasiski-Test - Überlegungen: Wir nehmen an, wir haben ein Schlüsselwort $k_1, k_2, \dots, k_n$ der Länge n, das wir uns durch $k_i = k_{i \bmod n}$ und $k_0 = k_n$ zum Schlüsselstrom fortgesetzt denken. Ist der Ausgangstext $a_1, a_2, a_3, \dots$, so ist der Vigenère-verschlüsselte Text $b_1, b_2, b_3, \dots$ mit $b_i = a_i + k_i \bmod 26$ oder auch $b_i = a_i + k_{i \bmod n} \bmod 26$.

Datei: ki_klass.tex. Version vom 24.10.2019

(1) Wir schauen im verschlüsselten Text nach, ob kurze Zeichenketten mehrfach auftreten. Wir finden z.B.

$$(b_i b_{i+1} b_{i+2} b_{i+3}) = (b_j b_{j+1} b_{j+2} b_{j+3}) \quad \text{mit} \quad i < j,$$

d.h. beginnend mit dem i-ten bzw. j-ten Buchstaben stimmen 4 Buchstaben überein. Eine Möglichkeit für das Zustandekommen dieses Phänomens ist, dass

$$(a_i a_{i+1} a_{i+2} a_{i+3}) = (a_j a_{j+1} a_{j+2} a_{j+3}) \quad \text{und} \quad i \equiv j \bmod n$$

gilt, denn dann folgt

$$k_i = k_j, \quad k_{i+1} = k_{j+1}, \quad k_{i+2} = k_{j+2}, \quad k_{i+3} = k_{j+3}$$

und damit

$$b_i = b_j, \quad b_{i+1} = b_{j+1}, \quad b_{i+2} = b_{j+2}, \quad b_{i+3} = b_{j+3}.$$

In der Praxis stellt sich nun heraus, dass diese Erklärung sehr oft richtig ist. Aus $i \equiv j \bmod n$ folgt dann

$$n \mid j - i,$$

d.h. die Schlüssellänge n ist ein Teiler der Differenz $j - i$, wobei die erste Zeichenkette an der Stelle i, die zweite an der Stelle j begann.

(2) Wir durchsuchen jetzt unseren verschlüsselten Text nach gleichen Zeichenketten der Länge 4 (oder ähnlich) und merken uns die Abstände $j_1 - i_1$, $j_2 - i_2$, $j_3 - i_3$, ... Trifft die Erklärung aus (1) zu, so erhalten wir

$$n \mid j_1 - i_1, \quad n \mid j_2 - i_2, \quad n \mid j_3 - i_3, \quad \dots,$$

sodass insbesondere

$$n \mid \mathrm{ggT}(j_1 - i_1, j_2 - i_2, j_3 - i_3, \dots)$$

gilt. Bei entsprechend vielen gleichen kurzen Zeichenketten kann man hoffen, dass sogar

$$n = \mathrm{ggT}(j_1 - i_1, j_2 - i_2, j_3 - i_3, \dots)$$

gilt. (Natürlich kann es auch „Ausrutscher“ $j - i$ geben, die man dann weglassen sollte.) Dieses Vorgehen ist unter der Bezeichnung **Kasiski-Test** bekannt. Mit dem Kasiski-Test kann man also die Schlüssellänge n bestimmen, wenn alles gut geht.

(3) Wir nehmen jetzt an, dass wir die Schlüssellänge n kennen. Wir schreiben den Chiffretext zeilenweise in eine Tabelle mit n Spalten:

$$\begin{array}{cccccc} b_1 & b_2 & b_3 & \dots & b_{n-1} & b_n \\ b_{n+1} & b_{n+2} & b_{n+3} & \dots & b_{2n-1} & b_{2n} \\ b_{2n+1} & b_{2n+2} & b_{2n+3} & \dots & b_{3n-1} & b_{3n} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \end{array}$$

Dies ist einfach (unter Weglassung von mod26):

$$\begin{array}{cccccc} a_1 + k_1 & a_2 + k_2 & a_3 + k_3 & \dots & a_{n-1} + k_{n-1} & a_n + k_n \\ a_{n+1} + k_1 & a_{n+2} + k_2 & a_{n+3} + k_3 & \dots & a_{2n-1} + k_{n-1} & a_{2n} + k_n \\ a_{2n+1} + k_1 & a_{2n+2} + k_2 & a_{2n+3} + k_3 & \dots & a_{3n-1} + k_{n-1} & a_{3n} + k_n \\ \vdots & \vdots & \vdots & & \vdots & \vdots \end{array}$$

Die erste Spalte ist also CAESAR-k_1-chiffriert, die zweite CAESAR-k_2-chiffriert, ..., die j-te Spalte CAESAR-k_j-chiffriert. Wir bestimmen nun das häufigste Zeichen in der j-ten Spalte: c_j. Ist E$\simeq 4$ das häufigste Zeichen in den zugehörigen Klartextspalten, so gilt $4 + k_j = c_j$ (modulo 26), womit wir den Schlüssel $k_1, k_2, \dots, k_n$ bestimmt haben.

(4) Hat man nun n und $k_1, k_2, \dots, k_n$ auf diese Weise bestimmt, kann man leicht entschlüsseln und testen, ob die Annahmen alle richtig waren.

Beispiel: Der folgende Text wurde Vigenère-verschlüsselt:

```
NHVMVWEIEWSAVVASEWZZKIIJFSKXUVWPIXRFHGYFUAPVPZMYEQFBGYGKHLRRMYMXNDLNVGRMFVSSAVVA
OTVVARHKVARPIZWMGKVOMKWTUWGHLEUAHZRKXRNNDWHVEWBRVZGVLKRSVLKMWARANZKIRYLLIZAMGHNN
JXMEMAKOVYKVLDVVMHVESGHVEWXMETGEHRETXMKRJDSEALXRRPZLIZAWFELFKXLVACTYDFWVLQRZGNRU
JXLROWGYEQLTXNBZENVGRMISRFLIZAWXVJGWKIZFWBRUVWPICGVXVROWGHNNJLGYBFAIINMYKVXGFQVA
MGHUVWLGYNLMIEQWKVZRKXRXEGLWVAXBGYGWGYEQTNGYRFOIISAGWKRJMIEQWGWTUETPVAOXKRHXWIDQ
AXFVVVXRNNFWIIGWGHVERBVBRDLGYZAWXJPZKMKGOTGBRJOSIJSXVKFMGHGSAYJVVFEMVQKVLNNLSXVN
MVLQHOXMCRFFMKZMGXVEKXMEREAYEQMGHJPZBIEFAVLEVUAXMVWEHREMFDLXMXQDRJGHRFKWMVASVLKA
AVLKZWAVWRJGHVFLHJVEFXVROWKHZRFTITUKMIYRJUIITWLIZNTXVWRDBBURJZSCQSKFVVLXVJNZLMTU
GYXRRFZWKYAVLLZOXREQWKAZAVWYIPZWMVOSXYDRJTYJPZMIJBOTVVFAAQRYKASVEWXVKEAMXVUAGXVE
KBGYJWGRUNKZIJGJTILPZTQNRYXLZAMGHYRJPEEXLXYEQKBGYGWBPKRYEELOLXIITWLMTULXVYVFMIIQ
WGFLRKVLVADTYVEFSYJRZXR
```

(1) Wir durchsuchen den Text nach gleichen Zeichenketten der Länge ≥ 4 und notieren die Stellen i und j, an denen die Zeichenfolgen auftreten, sowie die Differenz $j - i$

Zeichenfolge	i	j	Differenz $j-i$
MVWE	4	529	525
SAVVA	11	76	65
UVWPI	29	284	255
BGYG	52	352	300
BGYG	52	767	715
NVGRM	68	258	190
OTVV	81	691	610
ZRKXR	114	339	225
XRNN	117	407	290
HVEW	123	183	60
SVLK	136	556	420
LIZA	152	212	60
LIZA	152	267	115
ZAMGH	154	749	595
AMGH	155	320	165
GHNNJ	157	297	140
GHVE	182	417	235

Zeichenfolge	i	j	Differenz $j-i$
VEWX	184	704	520
EHRE	192	532	340
LIZAW	212	267	55
ROWG	244	294	50
WGYEQ	246	356	110
ZAWX	269	430	161 $*$
XVROW	292	582	290
AMGH	320	750	430
MIEQW	332	377	45
EQWK	334	659	325
BGYGW	352	767	415
GXVEK	497	717	220
RJGH	545	570	25
IITWL	603	783	180
LMTU	637	787	150
KBGY	721	766	45

Es fällt auf, dass alle Differenzen $j-i$ durch 5 teilbar sind, bis auf die zu ZAWX gehörige Differenz 161. Lässt man 161 weg, so ergibt sich als ggT der Differenzen 5. Dies legt die Vermutung nahe, dass die Schlüssellänge 5 ist.

(2) Wir machen jetzt eine Häufigkeitsanalyse der 5 Teilfolgen:

Zeichenfolge	Häufigste Zeichen
$(b_{5i+1})_{i\geq 0}$	W (17.58), K (9.09)
$(b_{5i+2})_{i\geq 0}$	X (16.97), G (12.73)
$(b_{5i+3})_{i\geq 0}$	I (16.36), H (10.30)
$(b_{5i+4})_{i\geq 0}$	V (21.95), E (9.76)
$(b_{5i+5})_{i\geq 0}$	R (16.46), A (10.37)

Ist jeweils E das häufigste Zeichen des Ausgangstextes, so erhalten wir wegen

$$\mathrm{E} \overset{x\mapsto x+18}{\longrightarrow} \mathrm{W}, \quad \mathrm{E} \overset{x\mapsto x+19}{\longrightarrow} \mathrm{X}, \quad \mathrm{E} \overset{x\mapsto x+4}{\longrightarrow} \mathrm{I}, \quad \mathrm{E} \overset{x\mapsto x+17}{\longrightarrow} \mathrm{V}, \quad \mathrm{E} \overset{x\mapsto x+13}{\longrightarrow} \mathrm{R}$$

$$k_1 = 18, \quad k_2 = 19, \quad k_3 = 4, \quad k_4 = 17, \quad k_5 = k_0 = 13,$$

was dem Wort „STERN“ entspricht.

(3) Entschlüsselt man jetzt mit Schlüsselwort „STERN“, so erhält man

```
VORVIELENJAHRENALSIMSPESSARTDIEWEGENOCHSCHLECHTUNDNICHTSOHAEUFIGALSJETZTBEFAHREN
WARENZOGENZWEIJUNGEBURSCHENDURCHDIESENWALDDEREINEMOCHTEACHTZEHNJAHREALTSEINUNDWA
REINZIRKELSCHMIDTDERANDEREEINGOLDARBEITERKONNTENACHSEINEMAUSSEHENKAUMSECHZEHNJAH
REHABENUNDTATWOHLJETZTEBENSEINEERSTEREISEINDIEWELTDERABENDWARSCHONHERAUFGEKOMMEN
UNDDIESCHATTENDERRIESENGROSSENFICHTENUNDBUCHENVERFINSTERTENDENSCHMALENWEGAUFDEMD
```

```
IEBEIDENWANDERTENDERZIRKELSCHMIDTSCHRITTWACKERVORWAERTSUNDPFIFFEINLIEDSCHWATZTEA
UCHZUWEILENMITMUNTERSEINEMHUNDUNDSCHIENSICHNICHTVIELDARUMZUKUEMMERNDASSDIENACHTN
ICHTMEHRFERNDESTOFERNERABERDIENAECHSTEHERBERGESEIABERFELIXDERGOLDARBEITERSAHSICH
OFTAENGSTLICHUMWENNDERWINDDURCHDIEBAEUMERAUSCHTESOWARESIHMALSHOEREERTRITTEHINTER
SICHWENNDASGESTRAEUCHAMWEGEHINUNDHERWANKTEUNDSICHTEILTEGLAUBTEERGESICHTERHINTERD
ENBUESCHENLAUERNZUSEHEN
```

(aus: Wilhelm Hauff, Das Wirtshaus im Spessart)

(4) Die Häufigkeitsverteilung des Chiffretexts (823 Großbuchstaben wurden ausgewertet, Zahlen in %) sieht so aus:

A	B	C	D	E	F	G	H	I	J	K	L	M
4.98	1.94	0.49	1.46	5.10	3.77	5.71	2.92	4.37	3.16	4.98	5.22	4.86
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
3.04	1.82	1.70	2.19	6.68	2.43	2.67	1.94	9.48	5.59	5.47	4.01	4.01

Die Häufigkeitsverteilung des Ausgangstexts (823 Großbuchstaben wurden ausgewertet, Zahlen in %) sieht so aus:

A	B	C	D	E	F	G	H	I	J	K	L	M
6.08	1.82	4.37	5.47	17.86	1.58	1.82	6.93	6.93	0.73	0.97	2.67	2.55
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
9.36	2.07	0.24	0.00	7.41	6.20	6.20	4.13	0.61	2.43	0.12	0.00	1.46

Bemerkung: Eine andere Angriffsmethode liefert der sogenannte Koinzidenz-Index von Friedman.