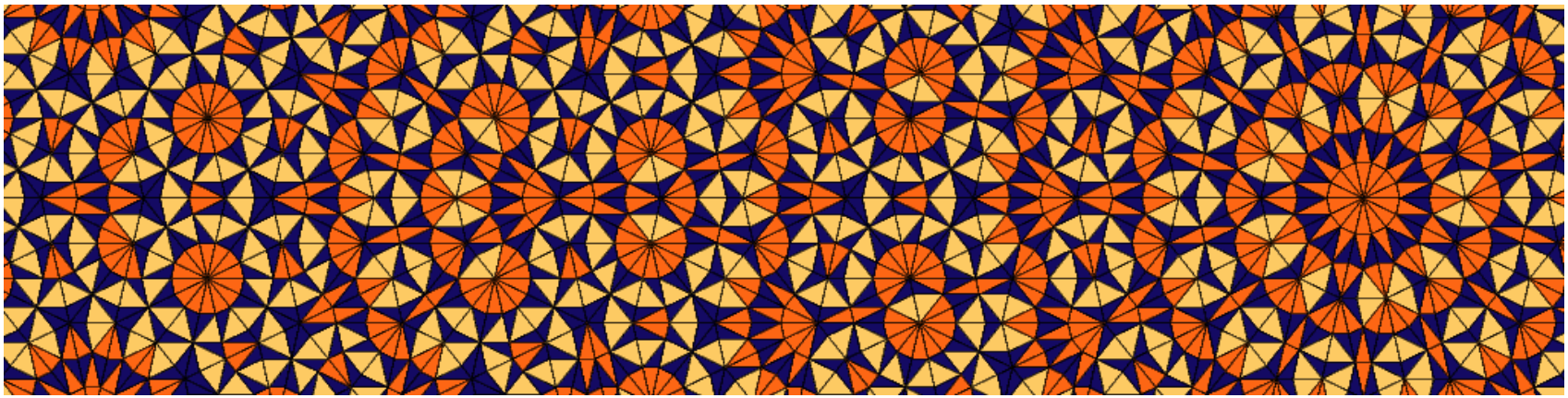
Panorama der Mathematik und Informatik

16: Kryptographie I

Dirk Frettlöh
Technische Fakultät / richtig einsteigen



Kryptographie: Verschlüsselung (Geheimschrift, Code, *cipher*...)

Kryptoanalyse: Entschlüsselung

(Simon Singh: Geheime Botschaften, 1999)

Schon in der Antike: Z.B. Cäsar-Code.

a	b	c	d	e	f	g	h	i	j	k	l	m
D	E	F	G	H	I	J	K	L	M	N	O	P

n	o	p	q	r	s	t	u	v	w	x	y	z
Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Also z.B. julius caesar $\rightarrow$ MXOLXV FDHVDU

Bezeichnung: C_*, wobei $C_A : a \rightarrow A$, $b \rightarrow B$ usw., oder $C_B : a \rightarrow B$, $b \rightarrow C$ usw. (das oben ist also C_D)

Im Folgenden immer **Klartext**: unverschlüsselt, kleine Buchstaben, **Geheimtext**: verschlüsselt, große Buchstaben.


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

Wurde wirklich früher genutzt. Ist aber nicht sicher, auch wenn wilder codiert wird (statt verschobenes Alphabet zufällige Permutation, etwa $a \to X, b \to M, c \to F, \ldots$).

Substitutionscode: Ein Code, in dem jeder Buchstabe mit genau einem Zeichen codiert wird (allgemeiner auch: jedes Buchstabenpaar durch ein Zeichen bzw Buchstabenpaar,...)

Schon islamische Wissenschaftler entdeckten die *Häufigkeitsanalyse*.

In einem längeren verschlüsselten Text (hier: englisch), zähle die Anzahl der verschiedenen Buchstaben.

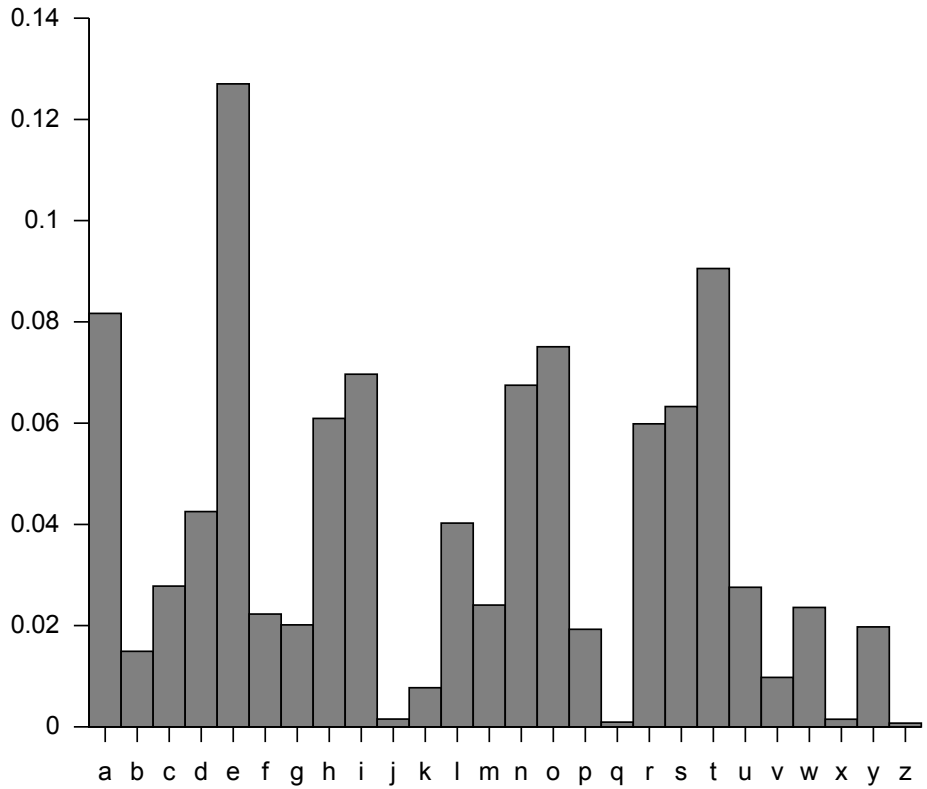
BT JPX RMLX PCUV AMLX ICVJP IBTWXVR CI M LMT'R PMTN, MTN YVCJX CDXV MWMBTRJ JPX AMTNGXRJBAH UQCT JPX QGMRJXV CI JPX YMGG CI JPX HBTW'R QMGMAX; MTN JPX HBTW RMY JPX QMVJ CI JPX PMTN JPMJ YVCJX. JPXT JPX HBTW'R ACUTJXTMTAX YMR APMTWXN, MTN PBR JPCUWPJR JVCUFGXN PBL, RC JPMJ JPX SCBTJR CI PBR GCBTR YXVX GCCRXN, MTN PBR HTXXR RLCJX CTX MWMBTRJ MTCJPXV. JPX HBTW AVBXN MGCUN JC FVBTW BT JPX MRJVCGCWXVR, JPX APMGNXMTR, MTN JPX RCCJPRMEXVR. MTN JPX HBTW RQMHX, MTN RMBN JC JPX YBRX LXT CI FMFEGCT, YPCRCXDXV RPMGG VXMN JPBR YVBJBTW, MTN RPCY LX JPX BTJXVQVXJMJBCT JPXVXCI, RPMGG FX AGCJPXN YBJP RAM

Häufigkeiten im Text oben:

A	B	C	D	E	F	G	H	I	J	K	L	M
11	27	37	2	2	5	17	8	9	49	0	7	44

N	O	P	Q	R	S	T	U	V	W	X	Y	Z
20	0	42	6	36	1	40	6	22	13	56	11	0

Häufigkeiten in echten englischen Texten allgemein:



Also probiere X=e, J=t: dann ist das häufige Wort JPX = tPe.

Also JPX = the. Im Text kommt das Wort JPXT vor, das wäre dann sicher then (oder they, aber T ist sehr häufig, also eher n).

Probieren: M=a, C=o, dann wohl CI = of; also I=f. Dann JPXVXCI = the?eof, also V=r.

MWMBTRJ = a?ain?t, also against. W=g, R=s.

MTN (8mal im Text!) = an?, also N=d; usw.

"in the same tour came forth fingers of a man's then, and wrote over against the candlestick upon the plaister of the wall of the king's palace..." (Bibel, Daniel 5,5)

In Mittelalter und Renaissance wurden zunächst einfache Substitutionscodes benutzt; evtl aufgebrezelt mit weiteren Zeichen für häufige Worte (das, ich, ist, da...)

Aufwendigere Codes (homophone Codes) benutzten z.B. 7 verschiedene Zeichen für e, 5 Zeichen für a und n, 4 Zeichen für s und t usw. Damit wurde Häufgkeitsanalyse erschwert.

Bsp: *Grand Chiffre* für franz. Könige, u.a. Louis XIV (Chef-Kryptographen und -analytiker: Familie Rossignol). Benutzt homophone Codes für Silben (häufige Silbe: viele Zeichen), einige Zeichen stehen für ganze Wörter, einige Zeichen stehen für nichts, einige Zeichen heben das letzte Zeichen auf (!)

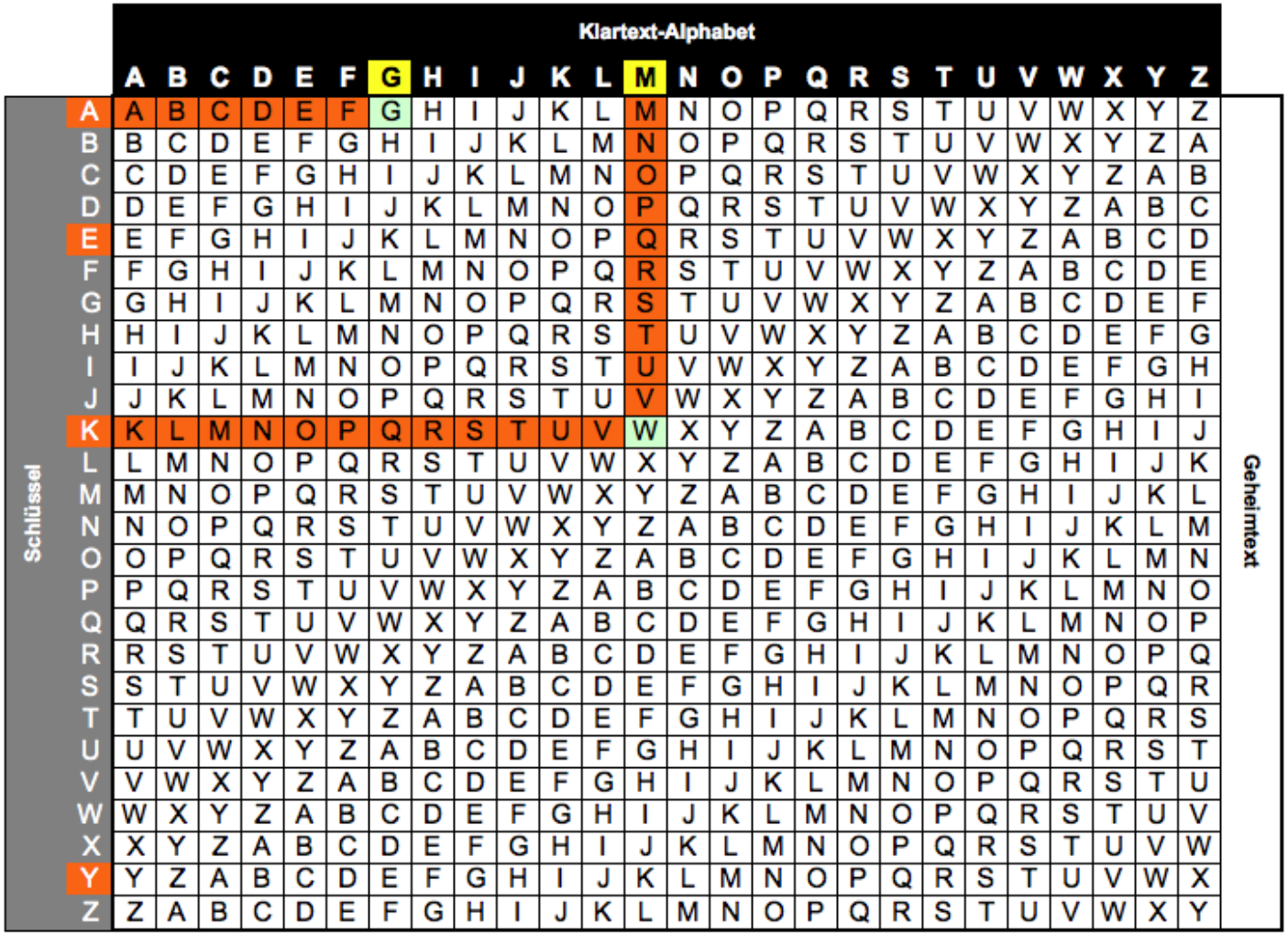
Wurde erst 1893 geknackt. Nachteil: dicke Codebücher, lange Ver- und Entschlüsselungsdauer.

Anderes kurioses Beispiel: *the Beal Ciphers* (USA 1885). Drei verschlüsselte Texte berichten von der Lage eines Goldschatzes, vergraben um 1820 in Virginia. Text 2 wurde entschlüsselt, die anderen bis heute nicht.

Eine sehr gute Methode, die lange nicht geknackt werden konnte: **Vigenère-Verschlüsselung** (le chiffre indéchiffrable)

Idee: benutze mehrere Cäsar-Substitutionen. Ein Schlüsselwort (hier: a key = AKEY)

Benutze dann abwechselnd die Cäsar-Substitutionen
$C_A, C_K, C_E, C_Y, C_A, C_K, C_E, C_Y, C_A, \ldots$.



(Bild: Log(z)equalsY, wikipedia)

Entwickelt bereits 1553 (G.B. Bellaso), einige Ideen noch eher.

Zugeschrieben Blaise Vigenère (1583). Hat sich aber erst spät durchgesetzt. Galt lange als unknackbar. Häufigkeitsanalyse versagt hier.

Bis um 1854 C. Babbage (unpubliziert) und F.W. Kasiski (1863) eine Methode zum Entschlüsseln fanden.

Idee:

- Suche Wortwiederholungen
- Errate daraus die Länge n des Schlüsselworts (z.B. 6) (Kandidaten: gemeinsame Teiler der Wortabstände)
- Führe dann n Häufigkeitsanalysen für die Buchstaben $1, 7, 13, 19, \ldots$ durch, für $2, 8, 14, 20, \ldots$ usw.

Wir gehen davon aus, dass gleiche Worte im Geheimtext mit der gleichen Folge von Cäsarsubstitutionen verschlüsselt wurden.

WUBEFIQLZURMVOFEHMYMWT
IXCGTMPIFKRZUPMVOIRQMM
WOZMPULMBNYVQQQMVMVJLE
YMHFEFNZPSDLPPSDLPEVQM
WCXYMDAVQEEFIQCAYTQOWC
XYMWMSEMEFCFWYEYQETRLI
QYCGMTWCWFBSMYFPLRXTQY
EEXMRULUKSGWFPTLRQAERL
UVPMVYQYCXTWFQLMTELSFJ
PQEHMOZCIWCIWFPZSLMAEZ
IQVLQMZVPPXAWCSMZMORVG
VVQSZETRLQZPBJAZVQIYXE
WWOICCGDWHQMMVOWSGNTJP
FPPAYBIYBJUTWRLQKLLLMD
PYVACDCFQNZPIFPPKSDVPT
IDGXMQQVEBMQALKEZMGCVK
UZKIZBZLIUAMMVZ

WUBEFIQLZURMVOFEHMYMWT
IXCGTMPIFKRZUPMVOIRQMM
WOZMPULMBNYVQQQMVMVJLE
YMHFEFNZPSDLPPSDLPEVQM
WCXYMDAVQEEFIQCAYTQOWC
XYMWMSEMEFCFWYEYQETRLI
QYCGMTWCWFBSMYFPLRXTQY
EEXMRULUKSGWFPTLRQAERL
UVPMVYQYCXTWFQLMTELSFJ
PQEHMOZCIWCIWFPZSLMAEZ
IQVLQMZVPPXAWCSMZMORVG
VVQSZETRLQZPBJAZVQIYXE
WWOICCGDWHQMMVOWSGNTJP
FPPAYBIYBJUTWRLQKLLLMD
PYVACDCFQNZPIFPPKSDVPT
IDGXMQQVEBMQALKEZMGCVK
UZKIZBZLIUAMMVZ

Abstände (der jeweils ersten Buchstaben):

- EFIQ: 95
- PSDLP: 5
- WCXYM: 20

Gemeinsamer Teiler: 5. (Hier ist's deutlich. Es gibt im Allg. auch Ausnahmen: zwei gleiche Worte, die *nicht* nach demselben Schlüssel codiert sind)

Also führen wir Häufigkeitsanalyse für Buchstaben $1, 6, 11, 16, \ldots$ durch:

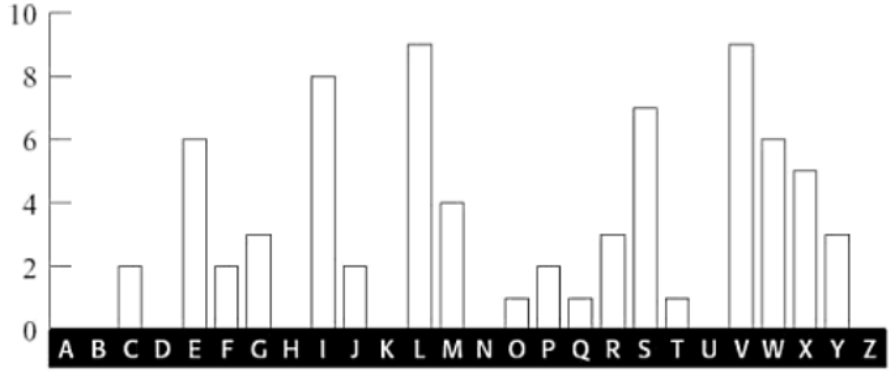


Figure 14 Frequency distribution for letters in the ciphertext encrypted using the L_1 cipher alphabet (number of occurrences).

Figure 15 Standard frequency distribution (number of occurrences based on a piece of plaintext containing the same number of letters as in the ciphertext).

Vergleiche häufigste Buchstaben (hm...) oder noch besser: auffällige Muster.
Hier z.B.: Real sind r,s,t besonders häufig (drei Bergspitzen), gefolgt von 6 seltenen Buchstaben (Tal der Breite 6).
Im Geheimtext sind V,W,X besonders häufig, gefolgt von sechs seltenen.
Also wohl C_E.

Für die anderen 4 Gruppen analog. Schlüssel ist “EMILY”.
Decodiere mit $C_E, C_M, C_I, C_L, C_Y, C_E, C_M, C_I, C_L, C_Y \ldots$.

sittheedownandhavenoshamecheekbyjowlandkneebykneewhatcareifor...

bzw.

Sit thee down, and have no shame,
Cheek by jowl, and knee by knee:
What care I for any name?
What for order or degree?

Let me screw thee up a peg:
Let me loose thy tongue with wine:
Callest thou that thing a leg?
Which is thinnest? thine or mine?

Thou shalt not be saved by works:
Thou hast been a sinner too:
Ruined trunks on withered forks,
Empty scarecrows, I and you!

Fill the cup, and fill the can:
Have a rouse before the morn:
Every moment dies a man,
Every moment one is born.