Kapitel 2
Einige klassische Kryptoverfahren

In diesem Kapitel geht es nicht um eine Darstellung der Geschichte der Kryptographie; die vorgestellten Verfahren werden daher nicht in chronologischer Reihenfolge vorgestellt, sondern mehrere Gruppen miteinander verwandter Verfahren werden jeweils in logischer Reihenfolge vom einfachsten zum schwierigsten der behandelten Verfahren präsentiert – auch wenn eine ganze Reihe von zum Teil heute noch gebräuchlichen Verfahren schon zum Zeitpunkt ihrer Einführung hoffnungslos veraltet waren.

§ 1: Monoalphabetische Substitutionen

Monoalphabetischen Substitutionen permutieren das Alphabet, aus dessen Buchstaben die Nachricht zusammengesetzt ist. Klassisch waren dies die Buchstaben des Alphabets; heute sind es, zumindest wenn man mit Computern arbeitet, eher Bytes.

a) Die Nullchiffre

Im Internet kann jeder tun und lassen was er will – mit nur ganz wenigen Ausnahmen. Zu diesen Ausnahmen gehört definitiv nicht, daß irgend jemand dazu gezwungen würde, sich vernünftig zu verhalten und auf Sicherheit zu achten. Dort, wo Verschlüsselung vorgesehen ist, sind die Anwender frei in der Wahl der einzusetzenden Verfahren, und fast überall ist als eine Variante auch die sogenannte Nullchiffre vorgesehen, die einfach darin besteht, überhaupt nicht zu verschlüsseln. Da im Internet zwar praktisch nichts vorgeschrieben, aber alles normiert

ist, hat auch dieses Nichtstun eine rund sechs Seiten lange Beschreibung in RFC 2410; beispielsweise wird dort darauf hingewiesen, daß man auch bei diesem „Verfahren“ mit Schlüsseln arbeiten kann, daß eine Erhöhung der Schlüssellänge die Sicherheit aber nicht wesentlich steigert. Für Einzelheiten siehe `www.faqs.org/rfcs/rfc2410.html`.

b) Die Caesar-Chiffre

Bei CAESAR selbst lesen wir, daß er eine Nachricht an CICERO mit griechischen statt lateinischen Buchstaben schrieb um zu verhindern, daß feindliche Soldaten (die im Gegensatz zu CICERO keine griechischen Buchstaben lesen konnten) den Inhalt verstehen konnten. Der Bote kam zwar nicht bis zu CICERO durch, aber, wie ihm CAESAR für diesen Fall geraten hatte, steckte er die Nachricht auf einen Speer, den er in CICEROs Lager warf. Nach zwei Tagen wurde der Speer gefunden und die unverständliche Nachricht zu CICERO gebracht, der sie natürlich lesen konnte.

Von einem geringfügig komplexeren Verfahren berichtet einige Jahrzehnte später SUETON in Kapitel 56 des ersten Buchs DIVUS IULIUS *(der göttliche Julius)* seines Werks DE VITA CAESARUM:

> extant et ad ciceronem, item ad familiares domesticis de rebus, in quibus, si qua occultius perferenda erant, per notas scripsit, id est sic structo litterarum ordine, ut nullum verbum effici posset: quae si qui investigare et persequi velit, quartam elementorum litteram, id est d pro a et perinde reliquas commutet.
>
> Erhalten sind auch seine Briefe an CICERO, ebenso an seine engeren Freunde über private Angelegenheiten, in denen er, was etwa geheim zu überbringen war, in verschlüsselter Form schrieb, nämlich in einer solchen Anordnung der Buchstaben, daß kein einziges Wort herauskam. Falls hier jemand nachforschen und der Sache nachgehen will, möge er den vierten Buchstaben des Alphabets, d.h. D für A und so fort setzen.

(zitiert nach SUETON: Kaiserbiographien, *Akademie Verlag Berlin,* 1993

Das war zwar weit hinter der Kryptographie, die in anderen Weltgegenden schon Jahrhunderte früher praktiziert wurde, aber wie die meisten Römer war eben auch deren Kryptographie recht primitiv.

GAIUS JULIUS CAESAR (100-44) und MARCUS TULLIUS CICERO (106–43) dürften wohl den meisten bekannt sein. GAIUS SUETONAUIS TRANQUILLUS war ein römischer Geschichtsschreiber, der um das Jahr 70 geboren wurde; er starb wahrscheinlich zwischen 130 und 140. Bekannt ist vor allem seine Lebensgeschichte der Caesaren, die aus je einem Buch für jeden der zwölf Caesaren besteht.

CAESAR verschob also das Alphabet zyklisch um drei Positionen; aus GALLIA EST OMNIS DIVISA IN PARTES TRES wurde das auch für Römer unverständliche JDOOL DHVWR PQLVG LYLVD LQSDU WHVWU HV.

Später soll ein ähnliches Verfahren auch von AUGUSTUS verwendet worden sein, allerdings verschob dieser nur um einen Buchstaben. Dafür soll CAESAR gelegentlich auch um eine andere Anzahl als drei verschoben haben. Für jemand, der seine Allianzen so häufig wechselte wie CAESAR, bot dies natürlich schon damals keine große Sicherheit, und heute kann man ein solches Verfahren erst recht vergessen: Zu CAESARs Zeiten, als das Alphabet aus nur 21 Buchstaben bestand und mangels Internet die Nullchiffre noch nicht als Verschlüsselungsverfahren betrachtet wurde, gab es schließlich nur zwanzig Möglichkeiten. (Die Buchstaben K, Y, Z existieren damals noch nicht, U und V sowie I und J wurden nicht unterschieden, und überflüssiger Komfort wie Kleinbuchstaben oder Satzzeichen wurden frühestens im achten oder neunten Jahrhundert gebräuchlich.)

Heute, egal ob wir mit 26 Buchstaben, 256 ASCII-Zeichen oder irgend etwas dazwischen liegendem arbeiten, kann ein Computer in Sekundenbruchteilen alle Möglichkeiten durchprobieren: Um etwa den Chiffretext

XYXFS DKOCO NCMRY VKONS CMSWE CCOXO

MKOZS CDEVK OWYBK VOCKN VEMSV SEWMF S

zu entschlüsseln, betrachten wir einfach alle 26 Möglichkeiten:

```
 1 YZYGT ELPDP ODNSZ WLPOT DNTXF DDPYP NLPAT DEFWL PXZCL WPDLO WFNTW TFXNG T
 2 ZAZHU FMQEQ PEOTA XMQPU EOUYG EEQZQ OMQBU EFGXM QYADM XQEMP XGOUX UGYOH U
 3 ABAIV GNRFR QFPUB YNRQV FPVZH FFRAR PNRCV FGHYN RZBEN YRFNQ YHPVY VHZPI V
 4 BCBJW HOSGS RGQVC ZOSRW GQWAI GGSBS QOSDW GHIZO SACFO ZSGOR ZIQWZ WIAQJ W
 5 CDCKX IPTHT SHRWD APTSX HRXBJ HHTCT RPTEX HIJAP TBDGP ATHPS AJRXA XJBRK X
 6 DEDLY JQUIU TISXE BQUTY ISYCK IIUDU SQUFY IJKBQ UCEHQ BUIQT BKSYB YKCSL Y
 7 EFEMZ KRVJV UJTYF CRVUZ JTZDL JJVEV TRVGZ JKLCR VDFIR CVJRU CLTZC ZLDTM Z
 8 FGFNA LSWKW VKUZG DSWVA KUAEM KKWFW USWHA KLMDS WEGJS DWKSV DMUAD AMEUN A
 9 GHGOB MTXLX WLVAH ETXWB LVBFN LLXGX VTXIB LMNET XFHKT EXLTW ENVBE BNFVO B
10 HIHPC NUYMY XMWBI FUYXC MWCGO MMYHY WUYJC MNOFU YGILU FYMUX FOWCF COGWP C
11 IJIQD OVZNZ YNXCJ GVZYD NXDHP NNZIZ XVZKD NOPGV ZHJMV GZNVY GPXDG DPHXQ D
12 JKJRE PWAOA ZOYDK HWAZE OYEIQ OOAJA YWALE OPQHW AIKNW HAOWZ HQYEH EQIYR E
13 KLKSF QXBPB APZEL IXBAF PZFJR PPBKB ZXBMF PQRIX BJLOX IBPXA IRZFI FRJZS F
14 LMLTG RYCQC BQAFM JYCBG QAGKS QQCLC AYCNG QRSJY CKMPY JCQYB JSAGJ GSKAT G
15 MNMUH SZDRD CRBGN KZDCH RBHLT RRDMD BZDOH RSTKZ DLNQZ KDRZC KTBHK HTLBU H
16 NONVI TAESE DSCHO LAEDI SCIMU SSENE CAEPI STULA EMORA LESAD LUCIL IUMCV I
17 OPOWJ UBFTF ETDIP MBFEJ TDJNV TTFOF DBFQJ TUVMB FNPSB MFTBE MVDJM JVNDW J
18 PQPXK VCGUG FUEJQ NCGFK UEKOW UUGPG ECGRK UVWNC GOQTC NGUCF NWEKN KWOEX K
19 QRQYL WDHVH GVFKR ODHGL VFLPX VVHQH FDHSL VWXOD HPRUD OHVDG OXFLO LXPFY L
20 RSRZM XEIWI HWGLS PEIHM WGMQY WWIRI GEITM WXYPE IQSVE PIWEH PYGMP MYQGZ M
21 STSAN YFJXJ IXHMT QFJIN XHNRZ XXJSJ HFJUN XYZQF JRTWF QJXFI QZHNQ NZRHA N
22 TUTBO ZGKYK JYINU RGKJO YIOSA YYKTK IGKVO YZARG KSUXG RKYGJ RAIOR OASIB O
23 UVUCP AHLZL KZJOV SHLKP ZJPTB ZZLUL JHLWP ZABSH LTVYH SLZHK SBJPS PBTJC P
24 VWVDQ BIMAM LAKPW TIMLQ AKQUC AAMVM KIMXQ ABCTI MUWZI TMAIL TCKQT QCUKD Q
25 WXWER CJNBN MBLQX UJNMR BLRVD BBNWN LJNYR BCDUJ NVXAJ UNBJM UDLRU RDVLE R
26 XYXFS DKOCO NCMRY VKONS CMSWE CCOXO MKOZS CDEVK OWYBK VOCKN VEMSV SEWMF S
```


Auch ohne Lateinkenntnisse sieht man leicht, daß die korrekte Entschlüsselung nur Zeile sechzehn sein kann: NON VITAE SED SCHOLAE DISCIMUS – EPISTULAE MORALES AD LUCILIUM CVI: *Wir lernen nicht für das Leben, sondern für die Schule*, wie LUCIUS ANAEUS SENECA (∼1–65) im 106. seiner Briefe an LUCILIUS die Schulmeister seiner Zeit verspottete.

Da sie so einfach zu entschlüsseln sind, sollten CAESAR-Chiffren heute völlig vergessen sein. Die kryptographische Realität sieht aber leider anders aus: Selbst dümmste Verfahren sind anscheinend unausrottbar. Der letzte bekannt gewordene prominente Anwender einer (leicht variierten) CAESAR-Chiffre war Mafia-Boss BERNARDO PROVENZANO, der bereits im Alter von acht Jahren die Schule verließ: Er ersetzte den Buchstaben „A“ durch die Zahl „4“ und so weiter bis zur Zahl „29“ für „Z“. Da alle hier benutzten zweistelligen Zahlen mit einer Eins oder Zwei beginnen, die einstelligen aber mindestens gleich vier sind, konnte er diese Zahlen ohne Zwischenraum hintereinander schreiben: 612418221215251218 etwa kann nur interpretiert werden als

$$6, 12, 4, 18, 22, 12, 15, 25, 12, 18 \quad = \quad \textit{Ciao Silvio} \ .$$

Da die meisten Strafverfolger länger zur Schule gegangen waren, stellte sie diese Modifikation vor kein unüberwindbares Problem, und so konnte er nach rund vierzig Jahren am 11. April 2006 endlich gefaßt werden.

Auch im Internet benutzt man in einigen Foren die CAESAR-Chiffre mit Verschiebung um 13 (ROT-13), allerdings nicht zur Geheimhaltung: Hier geht darum, daß Texte, die nicht nach jedermanns Geschmack sind, nur von denen gelesen werden, die das wirklich wollen.

c) Allgemeine monoalphabetische Substitutionen

Schwieriger wird es, wenn die Anzahl möglicher Verschlüsselungen zu groß wird, als daß man alle ausprobieren könnte. Solche Verfahren lassen sich leicht konstruieren: Wie bereits im vorigen Kapitel erwähnt, schlug GIROLAMO CARDANO vor, das Alphabet nicht nur wie bei den CAESAR-Substitutionen zyklisch zu verschieben, sondern es auf irgendeine *beliebige* Weise durcheinander zu bringen, so daß es $26! \approx 4 \cdot 10^{26}$ Möglichkeiten gibt – auch für Supercomputer zu viele, um alle auszuprobieren.

Trotzdem konnten schon die Militärkryptographen zur Zeit des ersten Weltkriegs jede so verschlüsselte Nachricht ab etwa einer Länge von fünfzig Zeichen problemlos entschlüsseln, und das ohne jede Maschine mit einem Aufwand von deutlich unter einer Stunde. Bei den meisten Verfahren, die heute beispielsweise als Teil von Office-Software angeboten werden, geht die Entschlüsselung zwar nicht ganz so einfach, dafür aber wenigstens sehr billig: Bei Anbietern wie `www.pwcrack.com` oder `www.lostpassword.com` kann man nachlesen, für welch geringe Beträge (ab etwa 40 US-$) die Verschlüsselungssysteme der meisten heute üblichen Softwarepakete geknackt werden können – natürlich nur zur Rekonstruktion „vergessener“ Passwörter. Wenn man bedenkt, daß manche interne Dokumente oder Kundendatenbanken für ein Unternehmen Werte im Bereich von Tausenden wenn nicht gar Millionen Euro repräsentieren können, wird klar, wie fahrlässig hier mit Kryptographie umgegangen wird.

Der Ansatzpunkt für den Kryptanalytiker ist praktisch immer derselbe: Die Dokumente, die wir schützen wollen, enthalten keine Zufallsfolgen,

sondern sprachliche Texte oder sonstige strukturierte Information. Diese Struktur muß der Angreifer ausnutzen.

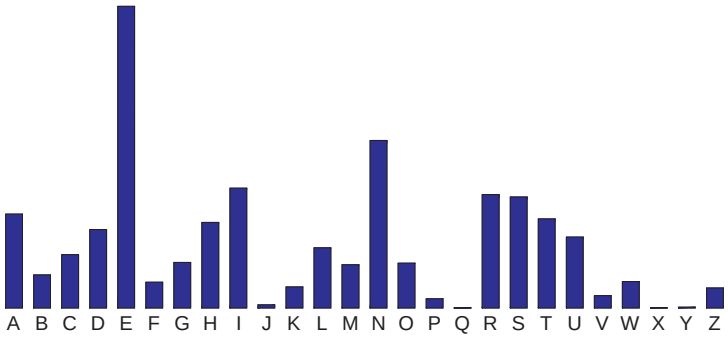
Wir wollen uns anhand eines Beispiels anschauen, wie er etwa im Fall von CARDANOs System vorgehen kann. Angenommen, wir haben die Nachricht

```
MBKFB MPLNL NIEAN KXRBP KKUMK KUNJC NEKXR SMKBN
JENEC PKKEI XRLMB BNIEU MKMAX AJIEO LUNEC NEKXR
NEIEU INRFN REIXR LMBBN IEICK XRJNI ANEBN KNEPN
ALKIX RNIEQ NJEPN EDLIO SNKNE EIXRL MBBNI EIEJN
XREPE OKKMX RNEKF BBUNJ CNEKX RKIXR CPNRN CMXRN
EKFEU NJEMP XRUNJ SNIKR NILBN RJNEC PKKCM ECILQ
NJOEP NONER FNJNE UMKKU INKCI LQNJK LMEUO NKXRM
RSMJR NJJBN RJNJB MNCGN BUMCM VPEUC FJILY UINKN
ANIUN ECFXR LNEIR EUMJP CEIXR LBNIU NEUNE ESNJA
FNKNK LJNIX RNCMX RLOIA LEIXR LMPDU NEBNR JNJMX
RL
```

aufgefangen und wollen sie entschlüsseln. Ein erster Ansatz könnte darin bestehen, daß wir den häufigsten Buchstaben des Kryptogramms als E identifizieren, den zweithäufigsten als den nach E nächsthäufigsten Buchstaben in deutschem Klartext, *usw.*

Dazu müssen wir als erstes wissen, wie häufig welcher Buchstabe in einem typischen deutschen Text ist. Wie die Jahrhunderte alte Erfahrung der Kryptanalytiker (und auch die heutige Linguistik) zeigt, gibt es hier keine nennenswerten Unterschiede zwischen verschiedenen (hinreichend langen) Texten; wir können also einfach irgendeinen deutschen Klartext hernehmen und die Buchstaben zählen. Das folgende Diagramm beruht auf der Auszählung der 260 238 Buchstaben aus JEAN PAULs Novelle *Dr. Katzenbeisers Badereise.* (Wie in der der klassischen Kryptographie üblich, wurden Zwischenräume und Satzzeichen ignoriert und Umlaute sowie „ß“ umschrieben, so daß nur 26 Buchstaben verwendet werden. Dies macht das Lesen der entschlüsselten Nachrichten zwar etwas unbequemer, erhöht aber die Sicherheit.)

Ordnen wir die Buchstaben nach ihrer Häufigkeit in diesem Text, erhal-



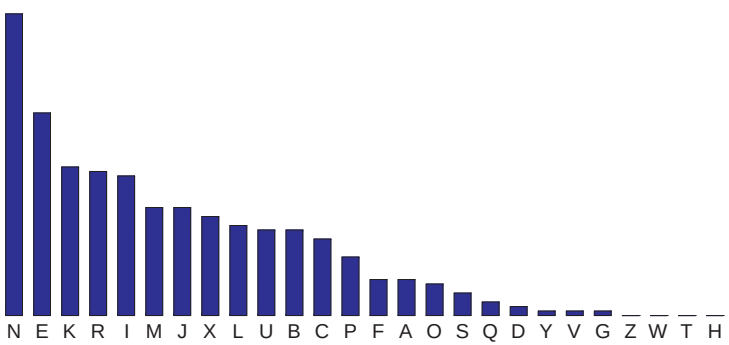
Buchstabenhäufigkeiten in deutschem Klartext

ten wir die Folge

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

Das gleiche können wir auch mit dem Kryptogramm machen; hier erhalten eine ähnliche Abbildung (die uns, falls wir *a priori* nichts über das verwendete Verfahren wissen, auch zeigt, daß wir es wahrscheinlich mit einer monoalphabetischen Substitution zu tun haben), jetzt aber mit der Folge

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



Buchstabenhäufigkeiten im Kryptogramm

Ein naheliegender erster Versuch ist also, daß wir N als E entschlüsseln,

E als N, K als I, R als R, I als S, und so weiter. Leider ist das Ergebnis nicht sehr vielversprechend:

```
ALIOL AGDED ESNME IHRLG IIUAI IUETC ENIHR WAILE
TNENC GIINS HRDAL LESNU AIAMH MTSNB DUENC ENIHR
ENSNU SEROE RNSHR DALLE SNSCI HRTES MENLE IENGE
MDISH RESNF ETNGE NKDSB WEIEN NSHRD ALLES NSNTE
HRNGN BIIAH RENIO LLUET CENIH RISHR CGERE CAHRE
NIONU ETNAG HRUET WESIR ESDLE RTENC GIICA NCSDF
ETBNG EBENR OETEN UAIIU SEICS DFETI DANUB EIHRA
RWATR ETTLE RTETL AECPE LUACA VGNUC OTSDZ USEIE
MESUE NCOHR DENSR NUATG CNSHR DLESU ENUEN NWETM
OEIEI DTESH RECAH RDBSM DNSHR DAGKU ENLER TETAH
RD
```

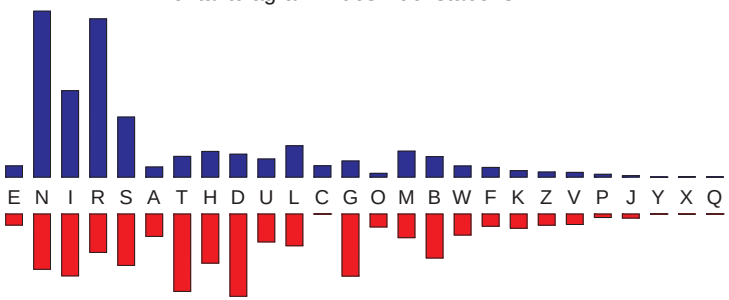
Denkt man etwas nach, sollte dieses Ergebnis eigentlich niemanden verwundern: Die Häufigkeitsunterschiede zwischen ähnlich häufigen Buchstaben sind teilweise so gering, daß sie gerade bei einem relativ kurzen Texten von nur 402 Buchstaben noch im Bereich von Zufallsschwankungen liegen. Ein erfolgreicher Ansatz muß daher mehr von der Struktur der deutschen Sprache ausnützen.

Innerhalb eines Texts ist die Wahrscheinlichkeit für das Auftreten eines Buchstabens stark vom Kontext abhängig: Auch wenn „E“ insgesamt gesehen der häufigste Buchstabe ist, wird man nach einem „C“ oder „Q“ nur selten eines finden und nach einem anderen „E“ auch nicht. Die Diagramme auf den folgenden Seiten zeigen, welche Buchstaben häufig vor (roter unterer Balken) *bzw.* nach (blauer oberer Balken) einem gegebenen Buchstaben auftreten.

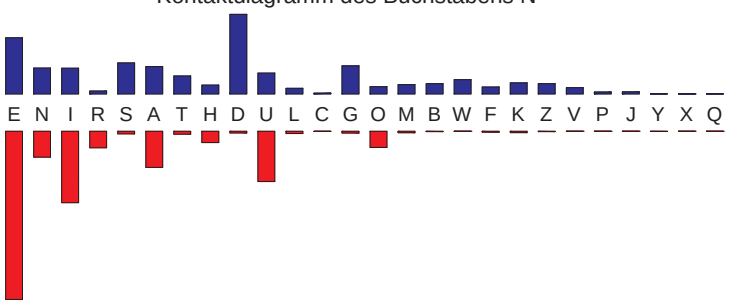
Auch wenn in einem kurzen Text gelegentlich „N“ häufiger vorkommt als „E“: Mit diesen Diagrammen können die beiden Buchstaben leicht unterschieden werden: Beispielsweise kommt vor „N“ häufig „E“ oder „I“ vor, aber fast nie ein seltener Buchstabe, während die Verteilung vor „E“ sehr viel homogener ist. Dafür ist die Verteilung *nach* „N“ homogener als die nach „E“ mit ihren vier großen Zacken bei den häufigen Buchstaben. Zwar sind zu Beginn einer Entschlüsselung *alle* Buchstaben unbekannt, aber wenn man die Buchstaben in den Diagrammen nach

Häufigkeit ordnet, kann man einigermaßen sicher sein, daß sich kein Buchstabe gar zu weit von seiner Position entfernt und damit zumindest die häufigen Buchstaben leicht identifizieren.

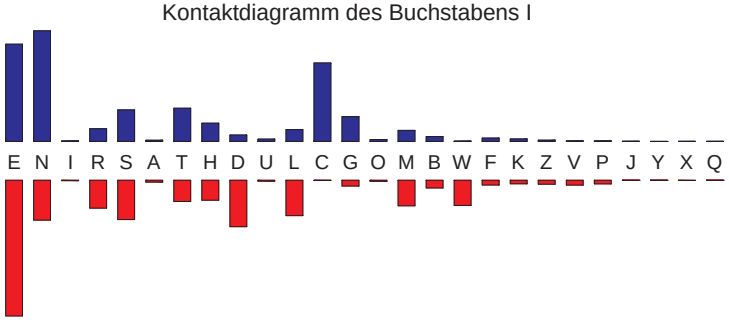
Kontaktdiagramm des Buchstabens E



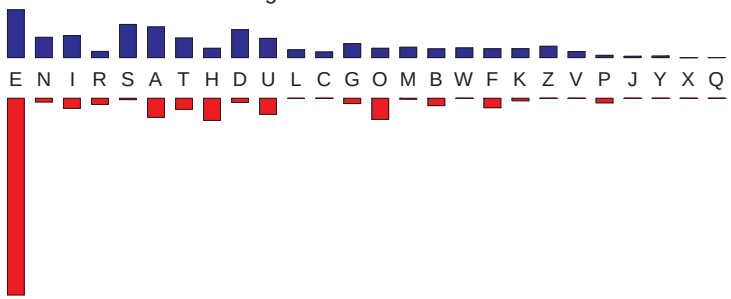
Kontaktdiagramm des Buchstabens N



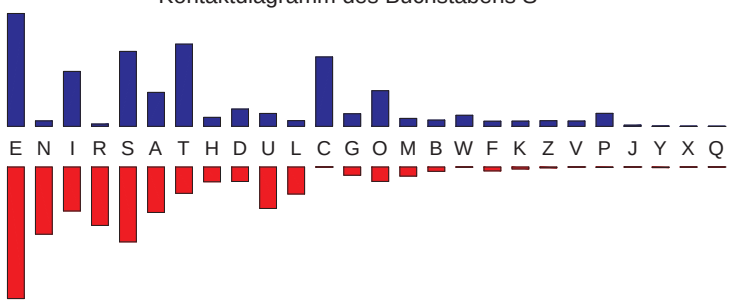
Kontaktdiagramm des Buchstabens I



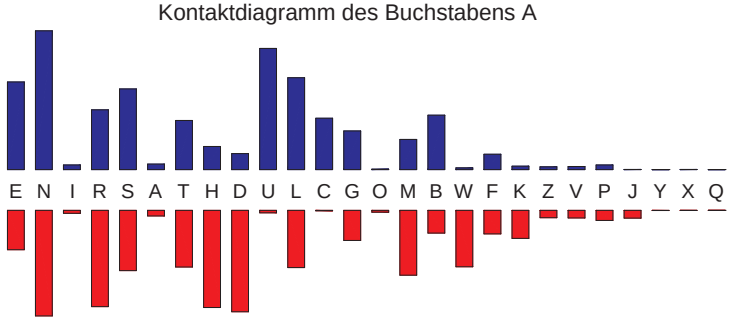
Kontaktdiagramm des Buchstabens R



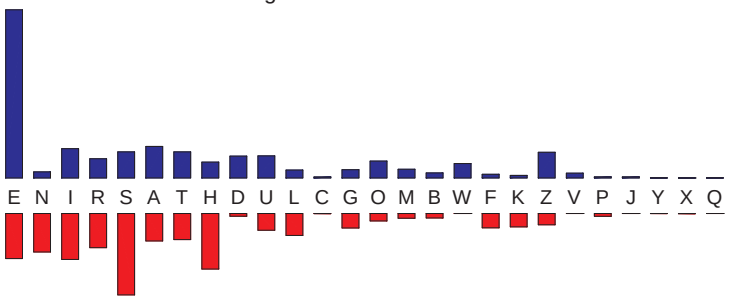
Kontaktdiagramm des Buchstabens S



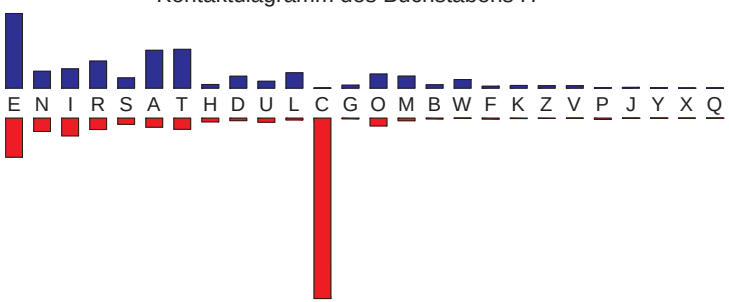
Kontaktdiagramm des Buchstabens A



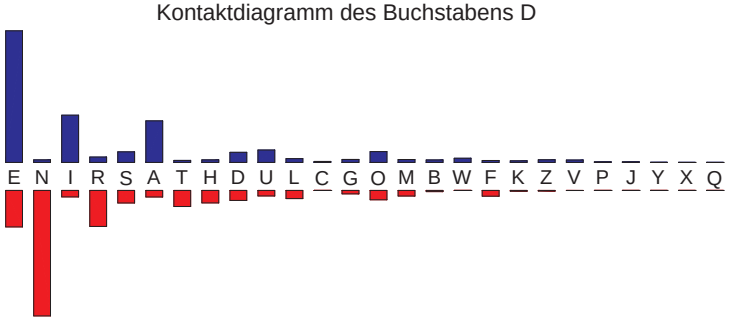
Kontaktdiagramm des Buchstabens T



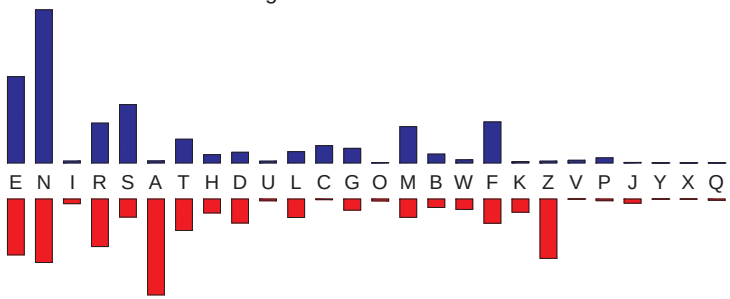
Kontaktdiagramm des Buchstabens H



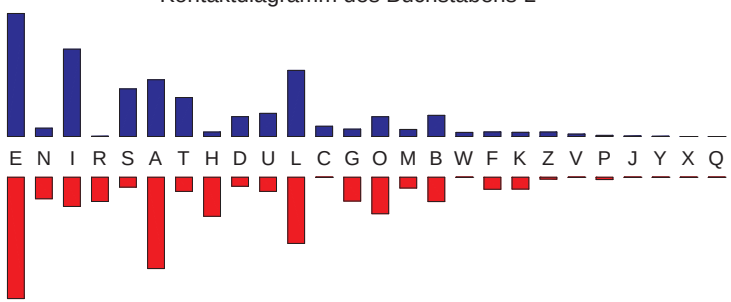
Kontaktdiagramm des Buchstabens D



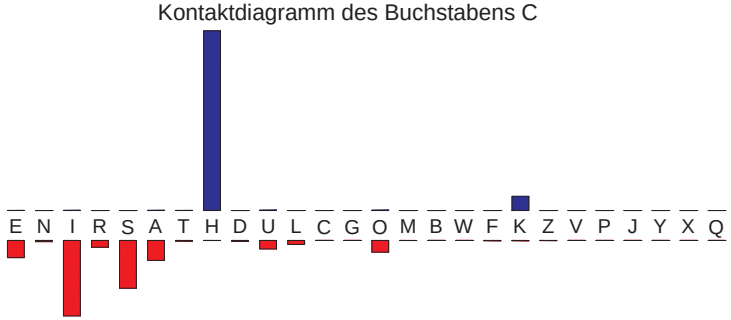
Kontaktdiagramm des Buchstabens U



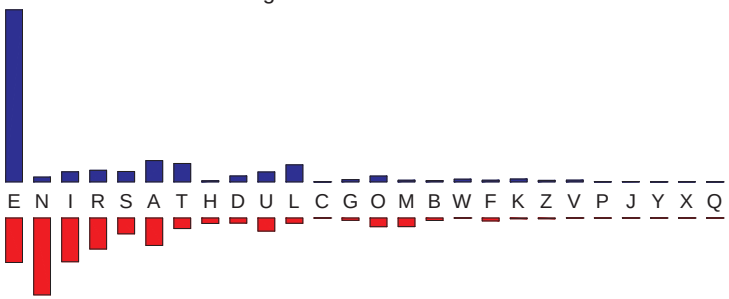
Kontaktdiagramm des Buchstabens L



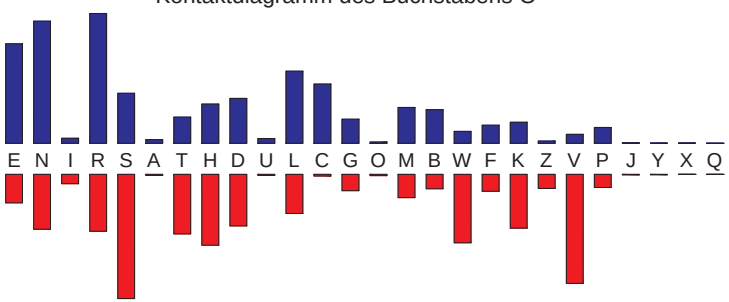
Kontaktdiagramm des Buchstabens C



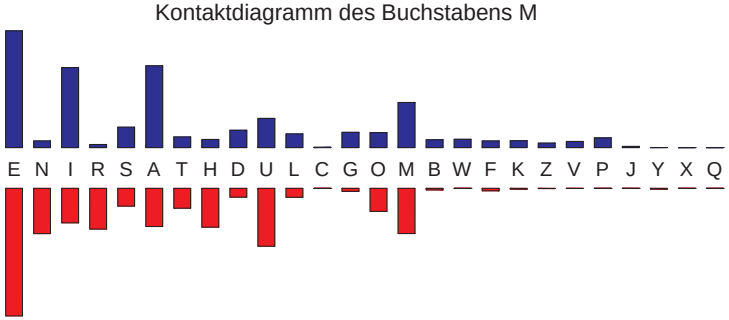
Kontaktdiagramm des Buchstabens G



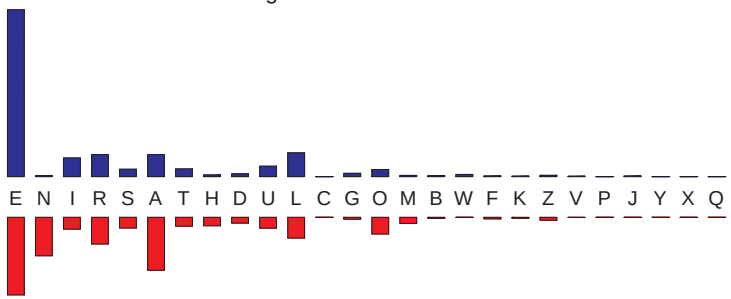
Kontaktdiagramm des Buchstabens O



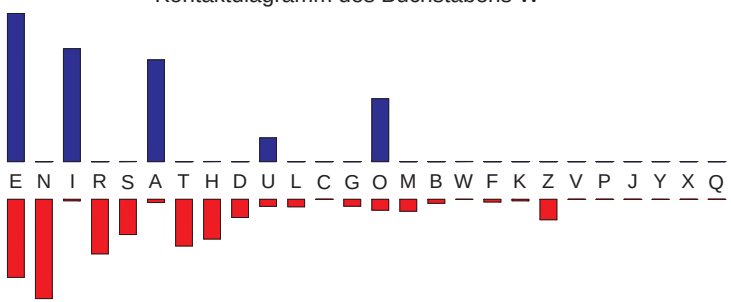
Kontaktdiagramm des Buchstabens M



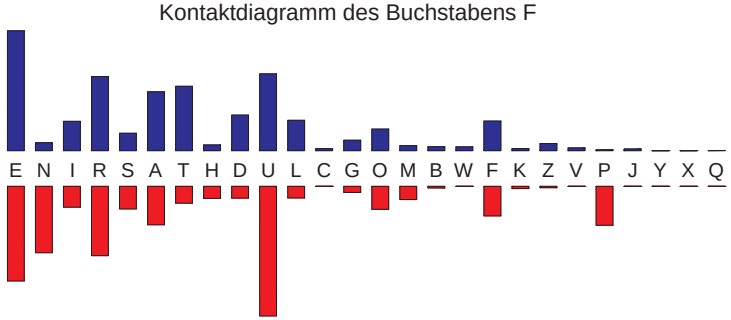
Kontaktdiagramm des Buchstabens B



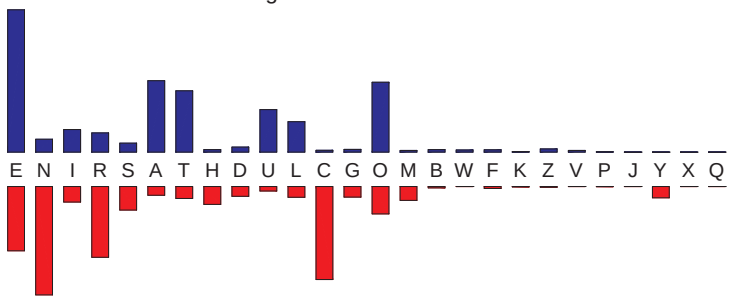
Kontaktdiagramm des Buchstabens W



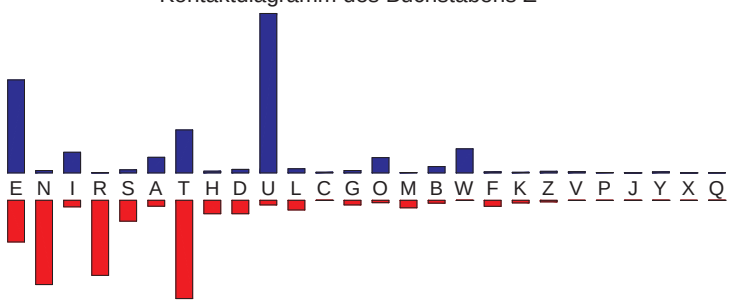
Kontaktdiagramm des Buchstabens F



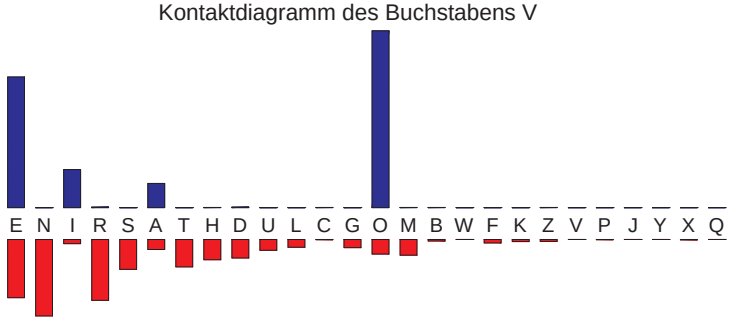
Kontaktdiagramm des Buchstabens K



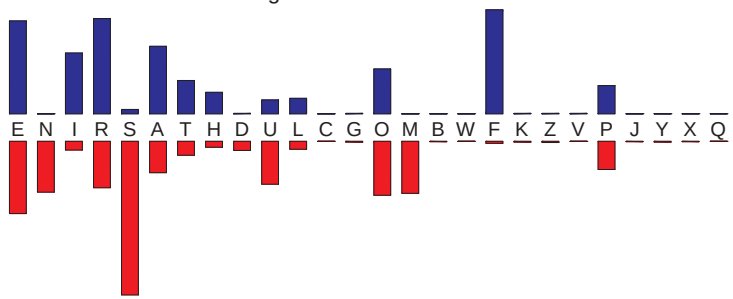
Kontaktdiagramm des Buchstabens Z



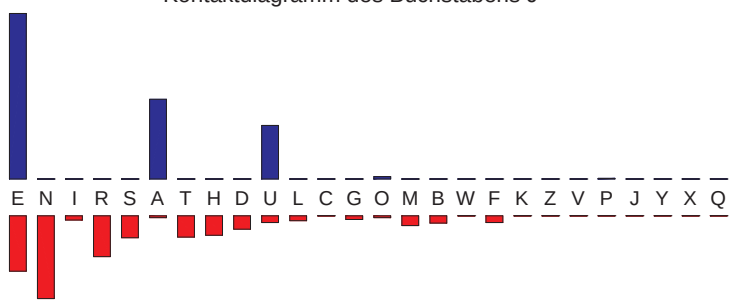
Kontaktdiagramm des Buchstabens V



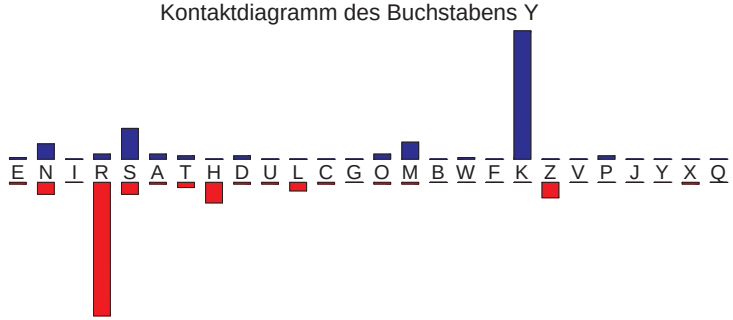
Kontaktdiagramm des Buchstabens P

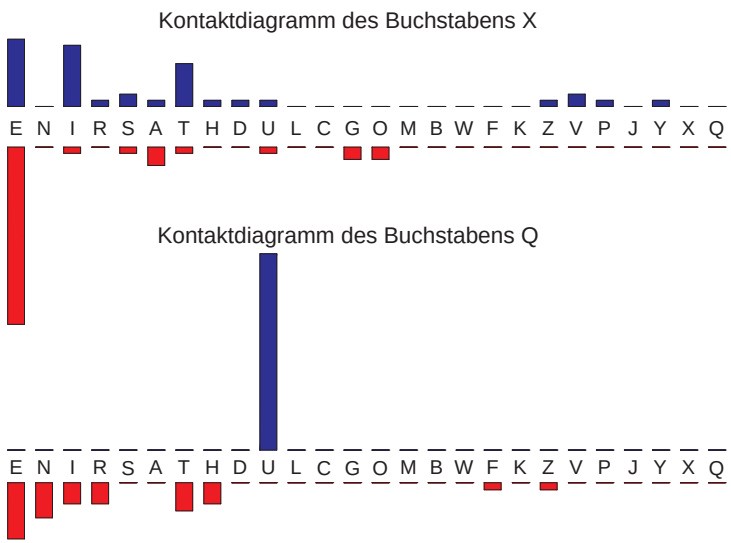


Kontaktdiagramm des Buchstabens J

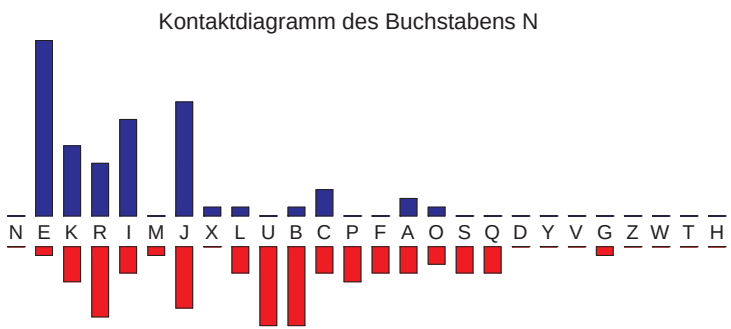


Kontaktdiagramm des Buchstabens Y

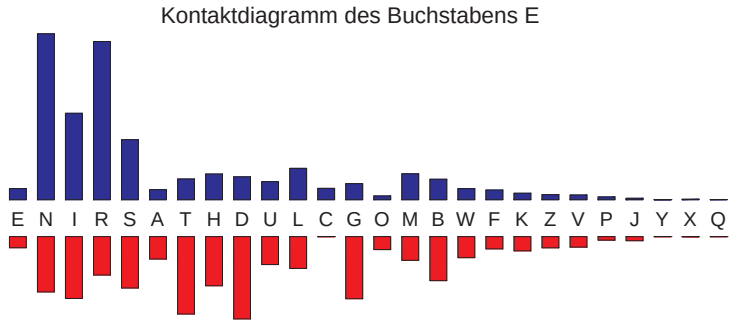




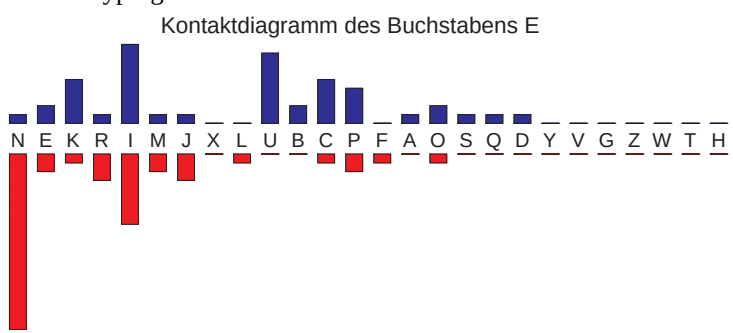
Wenden wir dies an auf unser Kryptogramm! Der häufigste Buchstabe dort ist N, was die Vermutung nahelegt, daß dies für ein Klartext-E stehen könnte. Hier ist das Kontaktdiagramm:



Ein Vergleich mit dem Kontaktdiagramm des Klartext-E zeigt deutliche Unterschiede:



Beispielsweise sehen wir nicht die auf-ab-auf-ab Struktur gleich am Anfang des Balkendiagramm über dem E. Andererseits müssen wir nach der Erfahrung mit unserer ersten ProbeEntschlüsselung damit rechnen, daß eine ganze Reihe von Buchstaben ihre Position verändert haben, und vor diesem Hintergrund sind die vier großen Balken im Anfangsbereich eigentlich alles, was wir realistischerweise erwarten können. Auch sonst ist die qualitative Übereinstimmung recht gut: Im unteren Teil des Diagramms ist die Verteilung deutlich homogener als im oberen, seltene Buchstaben kommen häufiger vor als in den Diagrammen zu den Klartextbuchstaben N, I, R, S und A, so daß wir ziemlich sicher sein können, daß der Kryptogrammbuchstabe N für ein Klartext-E steht.

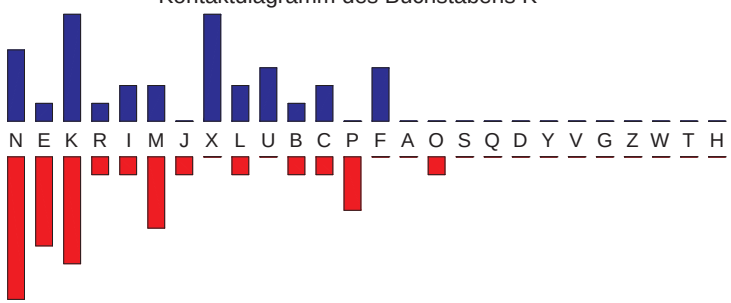


Der zweithäufigste Buchstabe im Kryptogramm ist das E. Auffällig an seinem Kontaktdiagramm ist der lange rote Balken unter dem bereits als Klartext-E erkannten häufigsten Kryptogrammbuchstaben N, der

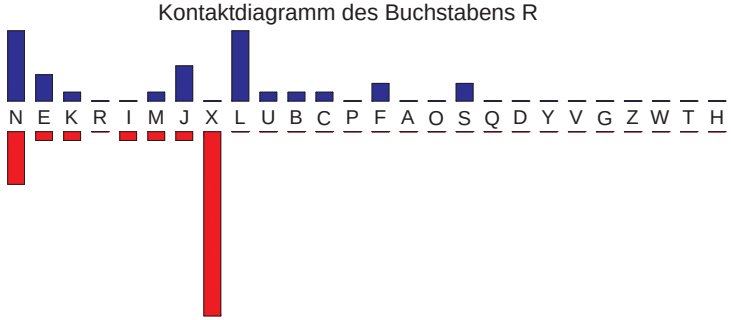
auch das Klartext-N auszeichnet; der zweite lange rote Balken unter dem I könnte dem unter Klartext-I oder eventuell auch U entsprechen. Auch die relativ homogene Verteilung der blauen Balken und die kaum vorhandenen roten Balken im Bereich der seltenen Buchstaben sprechen deutlich für ein N, so daß der Kryptogrammbuchstabe E wohl für ein Klartext-N steht.

Der dritthäufigste Buchstabe des Kryptogramms ist das K. Es zeichnet sich dadurch aus, daß die seltenen Buchstaben weder davor noch dahinter mit nennenswerter Häufigkeit vorkommen, daß häufig ein Klartext-E davorsteht, während N weder davor noch dahinter sonderlich oft vertreten ist. Dafür tritt K häufig als Doppelbuchstabe auf. Unter den in deutschen Klartexten häufigen Buchstaben verhält sich offenbar das S am ähnlichsten: Zwar passen die Balken ganz links besser zu einem R, aber die vielen weiteren langen Balken nach unten und die Häufigkeit von KK schließen dieses mit ziemlicher Sicherheit aus.

Kontaktdiagramm des Buchstabens K



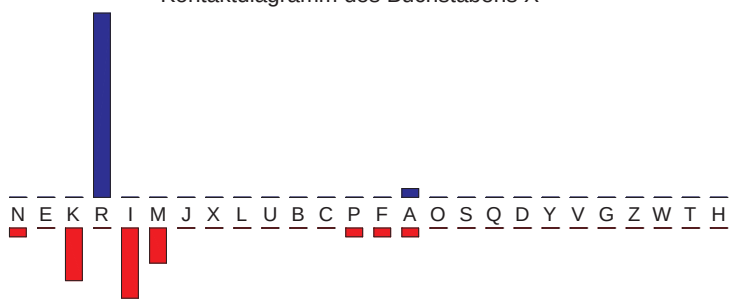
Kontaktdiagramm des Buchstabens R



Als nächstes müssen wir den Kryptogrammbuchstaben R identifizieren. Hier fällt als erstes der lange Balken unter dem X ins Auge; es gibt also einen mittelhäufigen Buchstaben, der das Geschehen vor dem gesuchten Buchstaben deutlich dominiert; ansonsten tritt dort nur noch Klartext-E mit nennenswerter Häufigkeit auf. Nach dem gesuchten Buchstaben ist die Verteilung deutlich homogener.

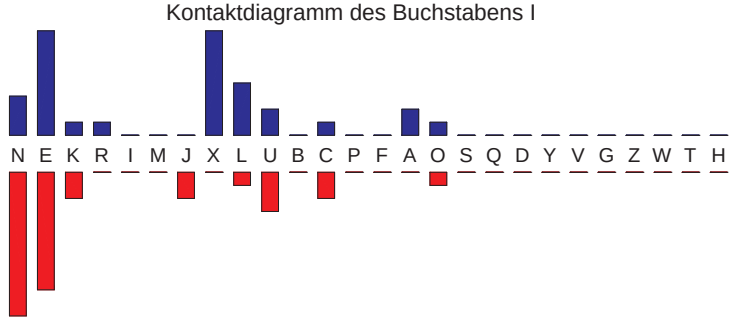
Damit ist eigentlich fast klar, daß R nur für ein H stehen kann, und daß der häufig davor stehende Buchstabe X im Klartext ein C sein muß. Ein direkter Vergleich des Kontaktdiagramms zu X mit dem von Klartext-C stützt diese Vermutung.

Kontaktdiagramm des Buchstabens X



Damit sind gleich zwei Buchstaben identifiziert; wenn wir nun wieder nach der Häufigkeit im Kryptogramm vorgehen, steht als nächstes das I auf dem Programm.

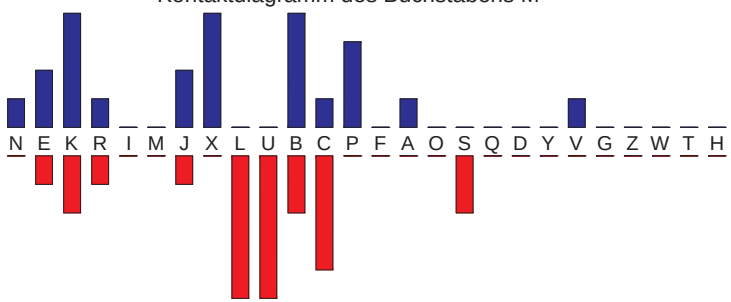
Kontaktdiagramm des Buchstabens I



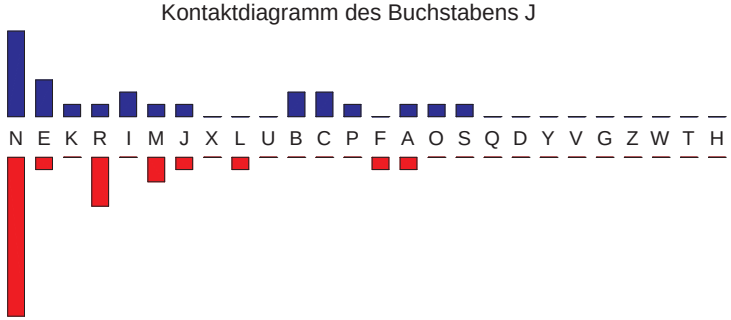
Es kommt häufig nach Klartext-E und N vor; insbesondere beim N auch vorher. Außerdem steht das gerade als C identifizierte X häufig dahinter. Unter den noch unbekannten Buchstaben ist offensichtlich I selbst der beste Kandidat; dieser Buchstabe wird also durch sich selbst verschlüsselt.

Als nächstes wartet das M auf seine Entschlüsselung . Es steht oft nach Buchstaben nur mittlerer Häufigkeit, praktisch nie nach E, aber gelegentlich davor. Zusammen mit I tritt es nie auf. Das deutet auf einen Vokal hin, und in der Tat paßt A sehr gut, wobei die Kombination AE natürlich von der Umschreibung der Umlaute herkommt.

Kontaktdiagramm des Buchstabens M



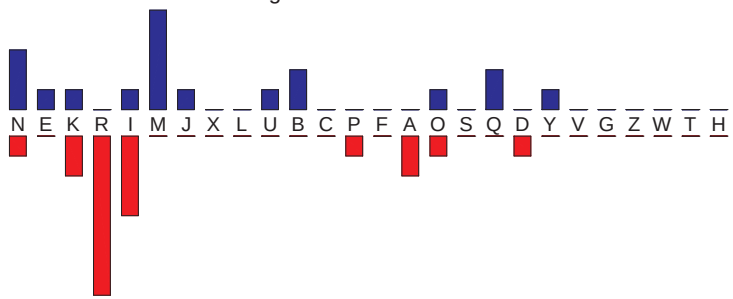
Kontaktdiagramm des Buchstabens J



Unser nächster Kandidat J mag keine seltenen Buchstaben um sich haben und versteht sich am besten mit dem E, insbesondere wenn es vor ihm steht. Die Entschlüsselung R bereitet damit keine Schwierigkeiten.

Das X haben wir bereits als C identifiziert, also kommt nun L an die Reihe.

Kontaktdiagramm des Buchstabens L



Es steht zwar gelegentlich hinter einem E, sehr viel häufiger aber hinter einem H oder I und praktisch nie hinter einem N. Dahinter sind A und E die häufigsten Buchstaben. Teile dieser Beschreibung passen zum T, andere zum D, aber alles zusammen paßt zu keinem der Buchstaben.

Unser Problem ist, daß der Buchstabe L im Kryptogramm nur zwanzig Mal vorkommt, und damit läßt sich keine sonderlich aussagekräftige Statistik über die Verteilung der 26 Buchstaben des Alphabets vor oder nach diesen zwanzig Stellen machen. Ab hier müssen wir also entweder mehrere Möglichkeiten in Betracht ziehen, oder aber zu einer anderen Strategie wechseln.

Mit acht von sechsundzwanzig Buchstaben haben wir zwar nur etwa 30% aller Buchstaben identifiziert, aber da es die acht häufigsten Buchstaben sind, kennen wir deutlich mehr als 30% des Klartexts: Nachzählen zeigt, daß wir von den 402 Buchstaben des Kryptogramms 278 identifiziert haben, also rund 70%. Damit sollte es nicht sehr schwierig sein, zumindest einige weitere Buchstaben aus dem Kontext zu erraten.

Hier ist noch einmal das Kryptogramm, wobei unter jedem bekannten Buchstaben dessen Entschlüsselung steht:

```
MBKFB MPLNL NIEAN KXRBP KKUMK KUNJC NEKXR SMKBN
A S   A  E  EIN E SCH   SS AS S ER  ENSCH  AS E
```

```
JENEC PKKEI XRLMB BNIEU MKMAX AJIEO LUNEC NEKXR
RNEN   SSNI CH A   EIN  ASA C  RIN    EN  ENSCH

NEIEU INRFN REIXR LMBBN IEICK XRJNI ANEBN KNEPN
ENIN  IEH E HNICH  A  E INI S CHREI  EN E SEN E

ALKIX RNIEQ NJEPN EDLIO SNKNE EIXRL MBBNI EIEJN
 SIC HEIN  ERN E N  I   ESEN NICH  A  EI NINRE

XREPE OKKMX RNEKF BBUNJ CNEKX RKIXR CPNRN CMXRN
CHN N  SSAC HENS     ER  ENSC HSICH   EHE  ACHE

EKFEU NJEMP XRUNJ SNIKR NILBN RJNEC PKKCM ECILQ
NS N  ERNA  CH ER  EISH EI  E HREN   SS A N I

NJOEP NONER FNJNE UMKKU INKCI LQNJK LMEUO NKXRM
ER N  E ENH  EREN  ASS  IES I   ERS  AN   ESCHA

RSMJR NJJBN RJNJB MNCGN BUMCM VPEUC FJILY UINKN
H ARH ERR E HRER  AE  E   A A   N    RI    IESE

ANIUN ECFXR LNEIR EUMJP CEIXR LBNIU NEUNE ESNJA
  EI E N  CH  ENIH N AR   NICH   EI  EN EN N ER

FNKNK LJNIX RNCMX RLOIA LEIXR LMPDU NEBNR JNJMX
ESES  REIC HE AC H  I   NICH  A    EN EH RERAC

RL
H
```

Gegen Ende der ersten Zeile haben wir im Klartext die Stelle

„UassUerCensch“ ,

Hier kann eigentlich nur „dass der Mensch“ gemeint sein, d.h. U steht für D und C für M.

In der zweiten Zeile sehen wir die Folge „ssnichLaB“; hier kann L eigentlich nur für T stehen, insbesondere da wir das L bereits provisorisch als D oder T identifiziert haben.

In der dritten Zeile ist der drittletzte Block entschlüsselt als „chrei“; davor steht ein „s“, dahinter ein unbekannter Buchstabe gefolgt von „en’. Dies spricht für eines der beiden Wörter „schreiben“ oder „schreiten“,

wobei letzteres nicht in Frage kommt, da wir bereits wissen, daß T durch L verschlüsselt wird, unser gesuchter Buchstabe aber durch A. Damit ist wohl A die Verschlüsselung des Buchstaben B.

Damit sind vier weitere Buchstaben identifiziert, und wenn wir sie an *allen* Stellen in die Entschlüsselung einsetzen, füllt sich der bereits entschlüsselte Teil des Texts und wir bekommen neue Kontexte zur Identifikation oder zumindest zum Erraten noch fehlender Buchstaben:

```
MBKFB MPLNL NIEAN KXRBP KKUMK KUNJC NEKXR SMKBN
A S   A TET EINBE SCH   SSDAS SDERM ENSCH  AS E

JENEC PKKEI XRLMB BNIEU MKMAX AJIEO LUNEC NEKXR
RNENM  SSNI CHTA   EIND ASABC BRIN  TDENM ENSCH

NEIEU INRFN REIXR LMBBN IEICK XRJNI ANEBN KNEPN
ENIND IEH E HNICH TA  E INIMS CHREI BEN E SEN E

ALKIX RNIEQ NJEPN EDLIO SNKNE EIXRL MBBNI EIEJN
BTSIC HEIN  ERN E N TI   ESEN NICHT A  EI NINRE

XREPE OKKMX RNEKF BBUNJ CNEKX RKIXR CPNRN CMXRN
CHN N  SSAC HENS    DER MENSC HSICH M EHE MACHE

EKFEU NJEMP XRUNJ SNIKR NILBN RJNEC PKKCM ECILQ
NS ND ERNA  CHDER  EISH EIT E HRENM  SSMA NMIT

NJOEP NONER FNJNE UMKKU INKCI LQNJK LMEUO NKXRM
ER N  E ENH  EREN DASSD IESMI T ERS TAND  ESCHA

RSMJR NJJBN RJNJB MNCGN BUMCM VPEUC FJILY UINKN
H ARH ERR E HRER  AEM E  DAMA   NDM  RIT  DIESE

ANIUN ECFXR LNEIR EUMJP CEIXR LBNIU NEUNE ESNJA
BEIDE NM CH TENIH NDAR  MNICH T EID ENDEN N ERB

FNKNK LJNIX RNCMX RLOIA LEIXR LMPDU NEBNR JNJMX
 ESES TREIC HEMAC HT IB TNICH TA  D EN EH RERAC

RL
HT
```

Ab dem dritten Block in der ersten Zeile steht nun

„einbeschBPssdassdermenschSasBernenmPss“ ,

Selbstverständlich muß P für einen Vokal stehen, und aus dem Zusammenhang ist klar, daß es ein U sein muß. Damit ist auch die Entschlüsselung von P als L klar, und S kann eigentlich nur für W stehen.

Da wir nun einen weiteren Vokal identifiziert haben und auch L ein in deutschen Klartext recht häufiger Buchstabe ist, lohnt es sich wahrscheinlich, vor der Suche nach neuen Identifikationen die drei gerade gefundenen im gesamten Text einzusetzen. Wir erhalten

```
MBKFB MPLNL NIEAN KXRBP KKUMK KUNJC NEKXR SMKBN
ALS L AUTET EINBE SCHLU SSDAS SDERM ENSCH WASLE

JENEC PKKEI XRLMB BNIEU MKMAX AJIEO LUNEC NEKXR
RNENM USSNI CHTAL LEIND ASABC BRIN  TDENM ENSCH

NEIEU INRFN REIXR LMBBN IEICK XRJNI ANEBN KNEPN
ENIND IEH E HNICH TALLE INIMS CHREI BENLE SENUE

ALKIX RNIEQ NJEPN EDLIO SNKNE EIXRL MBBNI EIEJN
BTSIC HEIN  ERNUE N TI  WESEN NICHT ALLEI NINRE

XREPE OKKMX RNEKF BBUNJ CNEKX RKIXR CPNRN CMXRN
CHNUN  SSAC HENS  LLDER MENSC HSICH MUEHE MACHE

EKFEU NJEMP XRUNJ SNIKR NILBN RJNEC PKKCM ECILQ
NS ND ERNAU CHDER WEISH EITLE HRENM USSMA NMIT

NJOEP NONER FNJNE UMKKU INKCI LQNJK LMEUO NKXRM
ER NU E ENH  EREN DASSD IESMI T ERS TAND  ESCHA

RSMJR NJJBN RJNJB MNCGN BUMCM VPEUC FJILY UINKN
HWARH ERRLE HRERL AEM E LDAMA  UNDM  RIT  DIESE

ANIUN ECFXR LNEIR EUMJP CEIXR LBNIU NEUNE ESNJA
BEIDE NM CH TENIH NDARU MNICH TLEID ENDEN NWERB

FNKNK LJNIX RNCMX RLOIA LEIXR LMPDU NEBNR JNJMX
 ESES TREIC HEMAC HT IB TNICH TAU D ENLEH RERAC

RL
HT
```


Niemand, der schon je ein Kreuzworträtsel gelöst hat, sollte Schwierigkeiten haben mit den wenigen Lücken, die jetzt noch übrig sind; der Klartext springt nun förmlich ins Auge: Natürlich kann die Lücke im ersten Block der ersten Zeile nur für ein O stehen, und die Lücke in der zweiten Zeile kann nur ein G sein. Selbst in der vierten Zeile, wo es in einem Wort noch drei Lücken gibt, fällt es nicht schwer, diese zu schließen. Mit Wortgrenzen und Satzzeichen geschrieben ist der Klartext

Also lautet ein Beschluss:
Dass der Mensch was lernen muss. –
Nicht allein das A-B-C
Bringt den Menschen in die Hoeh';
Nicht allein im Schreiben, Lesen
uebt sich ein vernuenftig Wesen;
Nicht allein in Rechnungssachen
Soll der Mensch sich Muehe machen;
Sondern auch der Weisheit Lehren
Muss man mit Vergnuegen hoeren.

Dass dies mit Verstand geschah,
War Herr Lehrer Laempel da. –
– Max und Moritz, diese beiden,
Mochten ihn darum nicht leiden;
Denn wer boese Streiche macht,
Gibt nicht auf den Lehrer acht.

(WILHELM BUSCH verwendet natürlich sowohl Umlaute als auch das „ß“; zum besseren Vergleich mit dem Kryptogramm habe ich sie aber so gelassen, wie ich sie vor der Verschlüsselung umschrieben habe.)

Damit haben wir also mit relativ geringem Aufwand ein Kryptogramm entschlüsselt, bei dem es auf den ersten Blick so aussah, als sei das nur möglich durch Ausprobieren von über 403 Quadrillion Möglichkeiten.

Wir haben zur Entschlüsselung zwar einen Computer als Hilfsmittel benutzt, aber auch wenn uns der das Leben etwas bequemer gemacht hat, war er nicht wirklich erforderlich: Die Buchstaben im Kryptogramm hätten wir mit nur unwesentlich größerem Aufwand auch von Hand durchzählen können, und die Kontaktdiagramme wurden in der Zeit,

als man noch alles von Hand machen mußte, gezeichnet, indem man einfach Querstriche über oder unter einen Buchstaben zeichnete, wenn dieser vor oder nach dem gerade untersuchten auftrat.

Nachdem die acht bis zehn häufigsten Buchstaben identifiziert sind, kann der Computer ohnehin nicht mehr weiterhelfen; jetzt ist eher Erfahrung mit der Sprache gefragt, am besten solche, die durch das Lösen vieler Kreuzworträtsel erworben wurde.

Als Kuriosität am Rande sei erwähnt, daß die Engländer im zweiten Weltkrieg das Personal für *Bletchley Park*, ihre Dechiffrierzentrale, tatsächlich zum Teil dadurch rekrutierten, daß sie Kreuzworträtselwettbewerbe in Zeitungen plazierten und die Sieger als potentielle neue Mitarbeiter ins Visier nahmen.

Die Häufigkeitsdiagramme für die seltenen Buchstaben nützen natürlich fast nie sonderlich viel, da hier die Schwankungen zwischen verschiedenen Texten besonders groß sind. Das beste Beispiel dafür ist das obige Diagramm zum Buchstaben Y: Es kommt im wesentlichen dadurch zustande, daß es in JEAN PAULS Roman einen häufig erwähnten Badearzt namens *Strykius* gibt.

Ein praktisches Problem bei der Schlüsselvereinbarung der hier betrachteten Substitutionschiffren bestand darin, daß eine Permutation der 26 Buchstaben im allgemeinen nur schwer zu übermitteln und auch zu merken ist, was unter anderem auch KERCKHOFFS dritte Forderung verletzt. Die klassische Lösung ging einfach aus von einem Wort oder einer Folge von Wörtern und schrieb diese von links nach rechts unter die Buchstaben A, B, C, ... Dabei mußte natürlich jeder Buchstabe, der schon dasteht, gestrichen werden. Falls am Ende noch nicht alle 26 Positionen gefüllt waren, nahm man dazu die noch verbleibenden Buchstaben in alphabetischer Reihenfolge.

Geht man aus von „Max und Moritz“, führt dies auf die Verschlüsselungstabelle

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

Auf diese Weise wurde das gerade behandelte Beispielkryptogramm verschlüsselt.

§ 2: Polyalphabetische Substitutionen

Da eine bloße Permutation des „Alphabets“ der Buchstaben oder Bytes offensichtlich keine nennenswerte Sicherheit bietet, kann man als nächstes versuchen, mehrere solcher Permutationen anzuwenden. Natürlich hat es keinen Sinn, sie alle jeweils auf denselben Buchstaben anzuwenden – das Produkt mehrerer Permutationen ist schließlich wieder nur eine einfache Permutation. Stattdessen unterteilt man die Nachricht in Blöcke einer festen (geheimgehaltenen) Länge n, wählt n Permutationen $\pi_i \in \mathfrak{S}$ und verschlüsselt jeweils das i-te Zeichen eines Blocks mit der Permutation π_i.

Am bekanntesten ist die sogenannte VIGENÈRE-Chiffren, benannt nach BLAISSE DE VIGENÈRE, der sie 1586 in seinem Buch *Traictés des chiffres ou secrètes manières d'escrire* beschrieb; sie wurden allerdings bereits im 1518 erschienenen Buch *Polygraphia* des Abbé JEAN TRITHÈME erwähnt und gehen laut VIGENÈRE zurück auf die Hebräer. Hier sind alle π_i CAESAR-Chiffren; der Einsatz allgemeiner Permutationen π_i dürfte jedoch wenn überhaupt nur unwesentlich jünger sein.

L. SACCO, der ehemalige Chef des Chiffrierdienstes der italienischen Armee, schreibt in seinem *Manuel de cryptographie* (Paris, 1951):

> *Vigenère réussit à produire un chiffre plus faible et moins commode que les précédents, qui n'en connut pas moins une fortune imméritée, particulièrement dans les cent dernières années, où il fut adopté par de nombreuses armées, même après la publication d'un moyen de décryptement (1863), indice évident de décadence dans la domaine de la cryptographie.*

Die VIGÈNERE-Chiffre oder ähnliche Verfahren wurden bis Ende des neunzehnten Jahrhunderts unter anderem von der österreichischen, der französischen und der italienischen Armee benutzt; von letzterer sogar bis 1917. In verschiedenen Computerprogrammen ist ähnliches heute noch zu finden.

Um zu verstehen, warum SACCO trotzdem so abfällig darüber spricht, müssen wir ihre Sicherheit etwas genauer betrachten und insbesondere sehen, wann sie mit welchem Aufwand geknackt werden kann.

Hauptproblem des Kryptanalytikers ist offensichtlich die Bestimmung der Blocklänge n, denn sobald diese bekannt ist, weiß man, welche Buchstaben mit demselben Alphabet verschlüsselt wurden und kann (leicht modifiziert) die kryptanalytischen Verfahren aus dem vorigen Paragraphen anwenden.

Mit dieser Ermittlung der Periode hatten die Kryptanalytiker im letzten Jahrhundert lange ihre Probleme; den ersten Ansatz fand der preußische Offizier FRIEDRICH W. KASIKI (1805–1881) und veröffentlichte ihn 1863 in einem damals kaum beachteten nur 95 Seiten dicken Buch mit dem Titel *Die Geheimschriften und die Dechiffrierkunst.* Seine Idee war die folgende: Gewisse Wörter und Buchstabenkombinationen wie etwa die bestimmten Artikel sind in fast allen Texten sehr häufig. Wenn nun ein langer Text mit einem polyalphabetischen Verfahren kurzer Periode verschlüsselt wird, ist es sehr wahrscheinlich, daß solche Buchstabenfolgen mehrfach auf die gleiche Weise verschlüsselt werden. Man suche daher im Kryptogramm nach zweimal vorkommenden Buchstabenfolgen (Heute nennt man so etwas ein KASIKI-Paar) und berechne deren Abstand. Die Periode sollte ein Teiler von relativ vielen dieser Abstände sein, wobei Abstände, die zu langen Buchstabenfolgen gehören, natürlich höher zu gewichten sind als solche, die etwa nur zu Buchstabenpaaren gehören: Bei letzteren ist die Wahrscheinlichkeit, daß es sich um eine zufällige Koinzidenz handelt, erheblich größer.

Die Suche nach KASIKI-Paaren ist recht aufwendig, und sie führen im allgemeinen nur dann zu einer Lösung, wenn das Kryptogramm erheblich länger ist als der verwendete Schlüssel. Eine Alternative fand um 1920 der wohl bedeutendste der klassischen Kryptologen, WILLIAM FRIEDMAN (1891–1969). Er wurde als WOLFE FRIEDMAN in Rußland geboren, aber als seine Eltern 1892 nach Amerika emigrierten, änderten sie seinen Vornamen. Er studierte zunächst Landwirtschaft, spezialisierte sich später auf Genetik und bekam 1915 eine Stelle als Genetiker bei dem Textilkaufmann GEORGE FABYAN, auf dessen Gut Riverbank in Geneva, Illinois. Dieser unterhielt dort Laboratorien für Akustik, Chemie, Genetik und Kryptologie – letztere mit dem Ziel zu beweisen, daß BACON der wahre Autor der SHAKESPEAREschen Schriften sei. Dadurch mußte sich FRIEDMAN zwangsläufig auch für Kryptologie interessieren

und war damit so erfolgreich, daß er bald Leiter der Laboratorien sowohl für Genetik als auch für Kryptologie war.

Mit dem Kriegseintritt der Vereinigten Staaten im April 1917 mußte sich auch die amerikanische Armee für Kryptographie interessieren, und da es außer Riverbank kein amerikanisches Zentrum für Kryptologie gab, wurden nicht nur aufgefangene Kryptogramme dorthin geschickt, sondern auch Armee-Offiziere, die bei FRIEDMAN in Kryptanalyse ausgebildet werden sollten. Diese Kurse wurden nach Kriegsende fortgesetzt, und 1921 verließ FRIEDMAN Riverbank, um anschließend in verschiedenen militärischen Funktionen sowohl Codes zu entwerfen als auch Codes zu knacken. Von ihm stammt das Wort *Kryptanalyse,* das das unbefugte Dechiffrieren vom legitimen unterscheidet.

Seine 1925 perfektionierte Idee zur Bestimmung der Periode ist folgende: Man betrachte für eine natürliche Zahl n die Wahrscheinlichkeit dafür, daß ein Buchstabe mit seinem n-ten Nachfolger übereinstimmt.

Falls n ein Vielfaches der Periode ist, wurden die beiden Buchstaben mit derselben Permutation verschlüsselt; da Summen nicht von der Reihenfolge der Summanden abhängen, ist die Wahrscheinlichkeit also

$$\sum_{i=1}^{26} p_i^2 \,,$$

wobei p_i die Häufigkeit des i-ten Buchstabens im Klartext ist.

Falls n dagegen *kein* Vielfaches der Periode ist, stammen ein Buchstabe und sein n-ter Nachfolger bei hinreichend langer Periode fast immer aus verschiedenen Permutationen, man kann sie also in allererster Näherung als voneinander unabhängig ansehen. Dann ist die Wahrscheinlichkeit einer Koinzidenz ungefähr gleich

$$\sum_{i=1}^{26} \left(\frac{1}{26}\right)^2 = \frac{1}{26} \,.$$

Betrachtet man die deutsche Sprache auf Grundlage von *Dr. Katzenbergers Badereise,* so ist

$$\sum_{i=1}^{26} p_i^2 \approx 0{,}0789 \quad \text{verglichen mit} \quad \frac{1}{26} \approx 0{,}0385 \,,$$

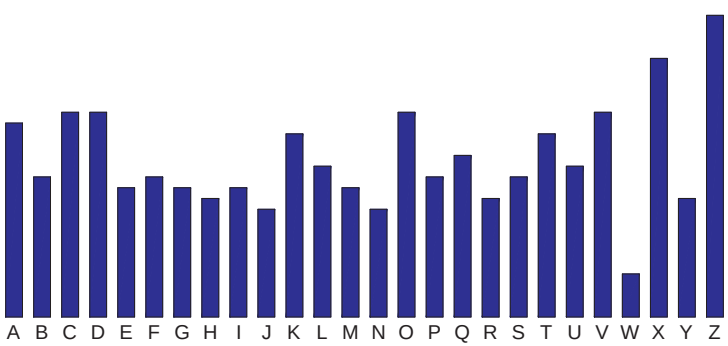
die beiden Werte unterscheiden sich also deutlich. In anderen Sprachen erhält man zwar andere Werte, aber der Unterschied ist auch dort unübersehbar.

In einem relativ kurzen Kryptogramm wird man selbstverständlich andere Werte berechnen, aber trotzdem sollte es im allgemeinen für gleiche Alphabete mehr Koinzidenzen geben als für verschiedene.

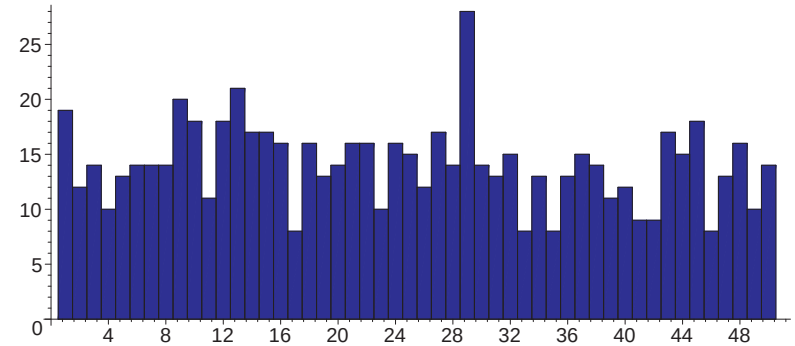
Betrachten wir als erstes Beispiel das Kryptogramm

```
PDOAA KKMQB LYORJ FTLXM OQGYU XTKCQ LXLVB ATCBU MJEDM SQZJJ
OZPHT AEZZD FFRSK XTYZV MVVZS QTZUO CTGDX ENOGX XGCOI GXHBN
OZXCC COJXJ PBQTV XTDOF ZRZND FHADX LZCQC NPBSL HTDVM ESKSP
YFCDB QHCEV ZDVIA DRKTR PGJDQ RFUUW AKQXP UZQVD AMXLF XGFGC
BTCFT KYRES GSALT VGZWT VSQOP UCALM ZFGMA VNDOZ ZNJOA HVDDQ
CMONK CPPBU ZZZYK PYRAD LZLYV GXNUB IURKX OEBBZ DNAWE UOVKH
TPISF DLIIQ LOXHO PAIZZ CAWEV XYSXU IZVUQ MAICE SKYXL AIZEH
KVNCR KUXYH IFKMK BJVHO TRSXX ZIEZJ
```

Auszählen der Buchstabenhäufigkeiten zeigt, daß die Buchstabenhäufigkeiten sehr viel gleichmäßiger verteilt sind als in deutschem Text; dies deutet darauf hin, daß keine *monoalphabetische*, sondern eine *polyalphabetische* Substitution angewandt wurde, d.h. die verschiedenen Buchstaben des Kryptogramms wurden nicht alle mittels ein und derselben Permutation aus den Klartextbuchstaben berechnet.



Buchstabenhäufigkeiten des Beispielkryptogramms

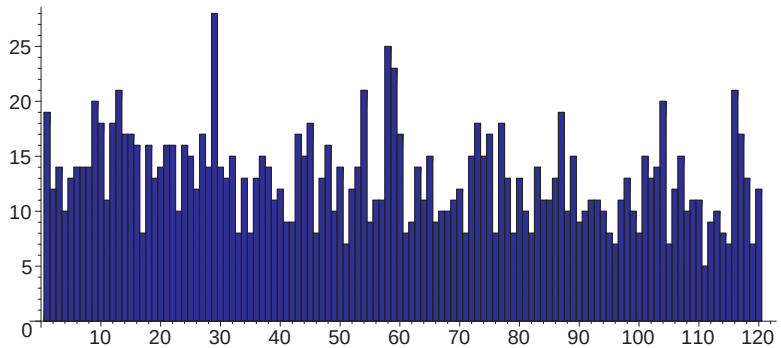


Koinzidenzen im Beispielkryptogramm für Abstände bis 50

Für das obige Kryptogramm zeigt Abbildung vier die Verteilung der Koinzidenzen bei n Buchstaben Abstand. Das erste Maximum bei $n = 1$ ist natürlich nicht ernst zu nehmen; sonst hätten wir eine monoalphabetische Substitution, deren Häufigkeitsverteilung deutlich anders aussieht als die hier beobachtete. Auch die Spitzen bei $n = 9$ und $n = 13$ sind wohl eher zufällig, denn bei $n = 18$ und $n = 26$ sind die Anzahlen eher klein. Unübersehbar ist das absolute Maximum bei $n = 29$; um zu sehen, ob dies das „richtige“ Maximum ist, müssen wir allerdings etwas mehr Werte berechnen. Die nächste Abbildung zeigt die entsprechenden Abstände bis einschließlich 120, und man sieht doch recht deutliche Ausschläge bei $n = 58$, 67 und 116. Wir wollen daher als erstes versuchen, das Kryptogramm unter der Annahme zu dechiffrieren, daß $n = 29$ ist.

Wir arbeiten also in diesem Ansatz mit der Hypothese, daß zwei Buchstaben genau dann mit derselben Permutation verschlüsselt sind, wenn ihr Abstand ein Vielfaches von 29 ist.

Deshalb teilen wir das Kryptogramm auf in 29 Buchstabenfolgen, die jeweils mit derselben Permutation verschlüsselt sein sollten. Ein Kryptanalytiker würde nun die 29 Häufigkeitsverteilungen dazu betrachten; da sie meisten (ganz grob) ungefähr so aussehen wie die von CAESAR-Substitutionen, würde er VIGENÈRE als wahrscheinlichste Möglichkeit ansehen.



Koinzidenzen im Beispielkryptogramm für Abstände bis 120

Die Entzifferung von gleich 29 CAESAR-Chiffren ist allerdings alles andere als kurzweilig; idealerweise sollten wir also auch das noch automatisieren.

Zum Glück läßt sich FRIEDMANs Idee auch darauf anwenden: p_i für $1 \leq i \leq 26$ sei die Wahrscheinlichkeiten dafür, daß ein Buchstabe in deutschem Klartext gleich dem i-ten Buchstaben des Alphabets ist, q_i entsprechend die, daß dies ein Buchstabe des Kryptogramms ist. Mit $\oplus$ bezeichnen wir eine Addition, bei der das Ergebnis durch Reduktion modulo 26 in das Intervall von 1 bis 26 gezwungen wird.

Falls r die Anzahl ist, um die wir zur Entschlüsselung verschieben müssen, ist $p_i \approx q_{i+r}$, also

$$\sum_{i=1}^{26} p_i q_{i\oplus r} \approx \sum_{i=1}^{r} p_i^2 \,.$$

Für jeden anderen Wert von r ist die Korrelation zwischen p_i und $q_{i\oplus r}$ schwächer, die Summe sollte also kleiner sein. Der Effekt ist sicherlich nicht so ausgeprägt, wie bei FRIEDMANs Test, und zumindest bei kurzen Kryptogrammen kann das Maximum auch gelegentlich bei einem falschen r liegen, aber das richtige r sollte im allgemeinen relativ weit oben liegen.

Wendet wir dies hier an, erhalten wir folgende Kandidaten für Schlüsselbuchstaben, wobei links jeweils der mit der größten $\sum p_i q_i$ steht, danach

die vier mit den nächstkleineren:

```
N 0,085 E 0,057 Z 0,055 O 0,050 A 0,048
A 0,083 E 0,062 N 0,047 Z 0,046 K 0,045
T 0,080 D 0,068 C 0,064 G 0,059 P 0,057
H 0,077 I 0,054 L 0,051 Y 0,049 G 0,049
M 0,077 Z 0,067 Q 0,055 D 0,051 I 0,051
B 0,055 I 0,054 F 0,053 J 0,053 P 0,048
P 0,074 T 0,071 C 0,059 M 0,058 D 0,056
T 0,069 G 0,055 E 0,052 R 0,050 I 0,048
A 0,081 N 0,070 J 0,061 L 0,050 K 0,047
G 0,092 C 0,075 F 0,064 P 0,056 Q 0,048
S 0,079 W 0,066 J 0,060 F 0,055 N 0,054
S 0,061 F 0,060 R 0,053 W 0,052 J 0,047
E 0,080 R 0,056 D 0,054 I 0,052 V 0,049
I 0,070 M 0,062 W 0,056 L 0,052 P 0,051
I 0,089 E 0,059 H 0,053 R 0,053 W 0,051
A 0,080 E 0,079 N 0,074 R 0,055 G 0,052
A 0,067 K 0,061 B 0,054 O 0,050 Z 0,050
R 0,076 N 0,072 W 0,056 B 0,054 I 0,049
K 0,072 B 0,060 E 0,059 V 0,056 U 0,051
R 0,075 E 0,055 N 0,054 Q 0,050 D 0,049
Y 0,069 S 0,059 P 0,058 J 0,054 W 0,054
P 0,099 C 0,061 L 0,053 M 0,051 Y 0,049
T 0,092 X 0,071 G 0,061 K 0,050 H 0,045
O 0,080 K 0,068 B 0,066 F 0,063 S 0,047
L 0,059 P 0,056 W 0,055 M 0,051 X 0,050
O 0,083 K 0,068 B 0,058 G 0,056 N 0,052
G 0,080 C 0,068 K 0,062 D 0,055 R 0,050
I 0,083 E 0,059 Z 0,053 V 0,053 R 0,051
E 0,107 N 0,059 A 0,056 R 0,054 I 0,048
```

Die Buchstabenfolge in der ersten Spalte legt nahe, daß hier kryptographisch unsorgfältig gearbeitet wurde: Der Schlüssel wurde wohl nicht zufällig gewählt, sondern als sinnvoller Teil der deutschen Sprache, wobei wir aber nicht alle Buchstaben auf Anhieb richtig erraten haben. Wir können nun entweder versuchen, den Schlüssel anhand der Buch-

staben aus den folgenden Spalten zu erraten (was hier wohl selbst ohne diese nicht sonderlich schwerfällt), oder aber wir entschlüsseln einfach mit dem wahrscheinlich falschen Schlüssel und korrigieren anhand des entschlüsselten Textes. Wenn wir diesen in Zeilen der Länge 29 aufschreiben, stehen jeweils die Buchstaben, die zum gleichen Alphabet gehören, untereinander, so daß es für jeden Schlüsselbuchstaben mehrere Überprüfungsmöglichkeiten gibt.

Der Entschlüsselung sungsversuch führt auf folgendes Ergebnis: (In der ersten Zeile steht der Schlüssel.)

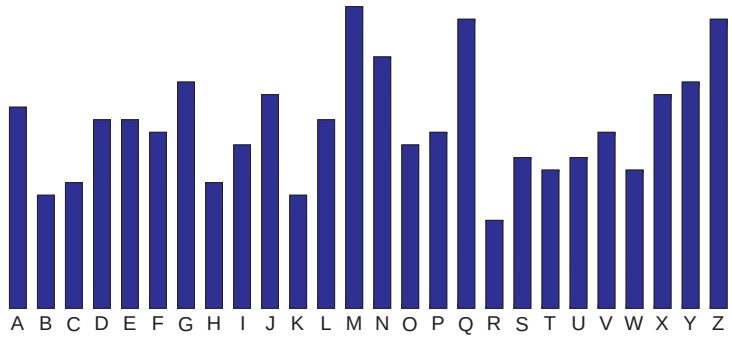
```
N A T H M B P T A G S S E I I A A R K R Y P T O L O G I E

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

Damit sollte wohl jeder Leser den Klartext rekonstruieren können. Bei einer Schlüssellänge von 29 und einer Textlänge von 480, bei etwa 16,5 Zeichen pro Alphabet also, ist das Vigenère-Verfahren somit schon völlig unsicher. Bei noch mehr Buchstaben pro Alphabet wird das Knacken noch einfacher: Betrachten wir als Beispiel denselben Klartext wie oben, chiffriert mit einem kürzeren Schlüssel:

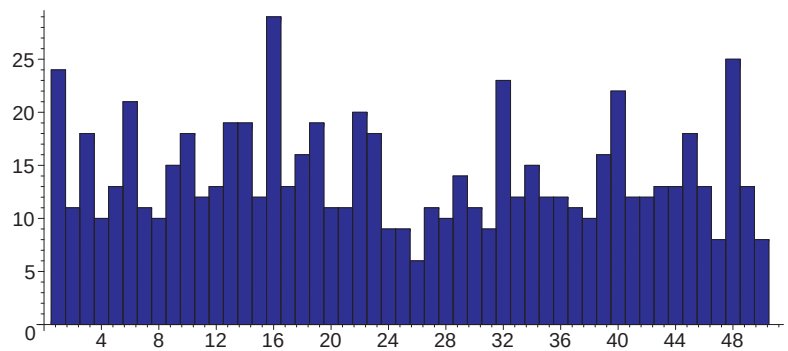
```
PZWDV AKLNZ ZDMDE MGYNZ VNGSC DVFAD YTFDP PVKOS BFMYT SUDND
JOMAO MJVIQ BMQUQ MZAAV XNAEO UXQFX IDXNM UYHDR AFEHO AQVZN
JPRIQ EBUGC QGRVJ XPLXS IROTX LMMUF ZILPT KKIHM MHWXD NYIIJ
NESVD VTGDX ZZSOA JNJEU ZZLHQ LUXMA CLXJE EZPXQ NXUYJ IIBYW
ETCFN MSXZH FOPLS FPZFG GCUGR JWHIA OPQEY PTLUM MPHCN BKWAZ
```



```
IQGCQ KNZNY MUGGO TCXND ELQYN KTVSR WKCQF ZFBWZ WJLLX IEGGA
FHZYA MRVBP QJNNV QAQQG PYJMM YYYAE WQBCQ GEOZY QLTOW YMQLH
ZWMGQ ZDLXF JJOME SGGSZ SBMTK NJJVY
```



Häufigkeitsverteilung der Buchstaben im Kryptogramm



Koinzidenzdiagramm dazu

Für dieses Kryptogramm sehen wir wieder anhand der relativ gleichmäßigen Verteilung der Buchstabenhäufigkeiten, daß es sich um eine polyalphabetische Verschlüsselung handeln muß.

Das Diagramm der Koinzidenzen bei n Buchstaben Abstand hat bei $n = 6$ eine erste Spitze; diese kann jedoch ignoriert werden, da es bei

n = 12 und n = 24 keine Maxima gibt. Die Maxima bei n = 16, 32 und 48 dagegen lassen eine Periodenlänge von 16 als wahrscheinlich erscheinen. Es gibt zwar auch eine ziemlich hohe Spitze bei n = 40, was vielleicht doch auf eine Periode acht hindeutet, da allerdings weder n = 8 noch n = 24 zu sonderlich großen Häufigkeiten führen, spricht das Diagramm doch eher für n = 16. Der pq-Test schlägt folgende Schlüsselbuchstaben vor:

```
N 0,066 M 0,052 Q 0,050 R 0,048 J 0,046
E 0,079 A 0,056 F 0,055 I 0,050 N 0,050
U 0,079 H 0,058 E 0,057 Q 0,056 D 0,051
E 0,064 D 0,055 P 0,049 F 0,048 T 0,048
R 0,083 V 0,059 E 0,050 I 0,046 Q 0,045
S 0,087 B 0,051 O 0,049 R 0,048 D 0,048
T 0,077 P 0,070 G 0,068 F 0,047 K 0,045
U 0,067 H 0,052 L 0,052 Y 0,052 V 0,049
D 0,075 Z 0,060 Q 0,056 U 0,049 H 0,046
I 0,072 Z 0,065 M 0,055 E 0,047 N 0,046
E 0,069 R 0,058 I 0,049 P 0,048 F 0,047
N 0,084 R 0,053 J 0,049 W 0,048 A 0,047
G 0,076 K 0,050 Z 0,049 X 0,047 Y 0,046
A 0,078 W 0,073 B 0,063 F 0,051 N 0,050
N 0,090 J 0,053 W 0,052 A 0,047 R 0,042
G 0,091 X 0,057 T 0,052 K 0,048 F 0,047
```

Hier steht gleich in der ersten Spalte der sehr wahrscheinlich aussehende Schlüssel `NEUERSTUDIENGANG`, der auch in der Tat zu verständlichem Klartext führt (entnommen aus einer Informationsbroschüre der Fakultät für Mathematik und Informatik für Schüler zu der Zeit, als der gerade noch existierende Diplomstudiengang noch neu war):

> *Der integrierte Studiengang Mathematik und Informatik bietet Ihnen eine berufsorientierte wissenschaftliche Ausbildung in den Fächern Mathematik und Informatik. Erst nach zwei Jahren entscheiden Sie sich für eine der beiden Ausrichtungen „Mathematik“ und „Informatik“ und damit auch für einen der beiden Abschlüsse „Diplom-Mathematiker“ oder „Diplom-Informatiker“. Während der ersten beiden Studienjahre sind alle Veranstaltungen gemeinsam.*

Informationstheoretische Betrachtungen zeigen, daß schon bei weniger als zwei Buchstaben pro Alphabet der Klartext und der Schlüssel mit hoher Wahrscheinlichkeit eindeutig durch den Chiffretext bestimmt sind. Zu einer auf der Informationstheorie beruhenden Kryptanalyse benötigt man allerdings praktisch *unbeschränkte* Ressourcen, denn die Größen, mit denen hier gerechnet wird, sind definiert als Summen über den Raum *aller* möglicher Schlüssel; im Falle des obigen Schlüssels der Länge 29 wäre das eine Summe über $26^{29} \approx 10^{42}$ Summanden. Kryptanalytiker suchen daher nach weniger aufwendigen Verfahren – und finden Sie auch in vielen Fällen.

§3: Der one time pad

Der *one time pad* („Einmalblock“) ist ein Spezialfall des VIGENÈRE-Verfahrens; daß er trotzdem einen eigenen Paragraphen bekommt, liegt an einem entscheidenden Punkt: Während die typische VIGENÈRE-Chiffre, wie wir im vorigen Paragraphen gesehen haben, keinerlei Sicherheit garantiert, ist der *one time pad* eines von nur zwei in dieser Vorlesung behandelten Verfahren, die beweisbare Sicherheit bieten. (Das zweite ist die Quantenkryptographie, die allerdings tatsächlich einfach ein Protokoll ist, um über räumliche Distanzen hinweg Schlüssel für den *one time pad* zu vereinbaren.)

Der *one time pad* hat seinen Namen von der Art und Weise, wie er früher benutzt wurde: Gedruckt wurden zwei identische Exemplare eines Blocks, der auf jeder Seite eine zufällige Folge von Buchstaben enthält; jeder der beiden Kommunikationspartner bekommt ein Exemplar.

Zur Verschlüsselung einer Nachricht nimmt der Absender das oberste Blatt, verschlüsselt den ersten Buchstaben der Nachricht à la CAESAR mit dessen erstem Buchstaben, den zweiten mit dem zweiten *usw.* Wenn alle Buchstaben auf dem Blatt aufgebraucht sind oder die Nachricht vollständig verschlüsselt ist, wird das Blatt vernichtet und es geht weiter mit dem nächsten Blatt. Der Empfänger kann die Nachricht mit Hilfe seines identischen Blocks problemlos entschlüsseln.

Falls KCHQR OFVFN FVSLA XRQBV E die ersten Buchstaben auf der aktuellen Seite sind, wird also die Nachricht „Angriff im Morgengrauen“

verschlüsselt gemäß

```
  ANGRI FFIMM ORGEN GRAUE N
+ KCHQR OFVFN FVSLA XRQBV E
= LQOIA ULESA UNZQO EJRWA S
```

Was kann ein Gegner mit diesem Chiffretext anfangen?

Wenn er das Verfahren kennt, weiß er, daß für die Verschlüsselung dieser Nachricht aus 21 Buchstaben auch 21 Schlüsselbuchstaben verwendet wurde; dazu gibt es $26^{21} \approx 5{,}181318713 \cdot 10^{29}$ Möglichkeiten, also über Tausend mal so viele wie bei der allgemeinen monoalphabetischen Substitution. Wie wir dort gesehen haben, können wir allerdings auf solche Zahlen nichts geben. Die Sicherheit des *one time pad* beruht auf einer ganz anderen Überlegung:

Angenommen, ein Gegner kommt irgendwie auf den richtigen Schlüssel KCHQR OFVFN FVSLA XRQBV E und entschlüsselt die Nachricht damit:

```
  LQOIA ULESA UNZQO EJRWA S
- KCHQR OFVFN FVSLA XRQBV E
= ANGRI FFIMM ORGEN GRAUE N
```

Damit kennt er die Nachricht. Aber weiß er das?

Da die Schlüssel zufällig gewählt wurden, ist KCHQR OFVFN FVSLA XRQBV E genauso wahrscheinlich wie etwa JYFUT PJSXN PZTVJ MWWCG J, und damit würde via

```
  LQOIA ULESA UNZQO EJRWA S
- JYFUT PJSXN PZTVJ MWWCG J
= BRING EBLUM ENFUE RMUTT I
```

mit einem völlig anderen Ergebnis entschlüsselt. Offensichtlich gibt es für jede Buchstabenfolge der Länge 21 genau einen Schlüssel, der genau diese Folge als „Entschlüsselung“ liefert, und da alle Schlüssel dieselbe Wahrscheinlichkeit haben, gilt dasselbe auch für alle Nachrichten der Länge 21.

Natürlich sind aus Sicht des Gegners, der im allgemeinen durchaus Informationen über das Umfeld der Kommunikation hat (warum sonst sollte

er schließlich abhören?) nicht alle diese Nachrichten gleich wahrscheinlich, aber durch das Auffangen des Chiffretexts bekommt er keinerlei neue Informationen, die seine Einschätzung der relativen Wahrscheinlichkeit der verschiedenen Möglichkeiten verändern könnten. Dieses Phänomen bezeichnet man als absolute informationstheoretische Sicherheit.

Er lernt allerdings zwei Dinge: Erstens, daß der Absender eine Nachricht an den Empfänger schickte und zweitens, wie lange diese Nachricht war. Auch das läßt sich verhindern, indem der Absender regelmäßig zu festgesetzten Zeiten eine Nachricht an den Empfänger schickt; falls es nichts zu sagen gibt, schickt er einfach irgendeinen Text.

Der *one time pad* wurde wohl erstmalig vom amerikanischen General VERNAM im ersten Weltkrieg benutzt; man redet daher gelegentlich auch von der VERNAM-Chiffre. Im zweiten Weltkrieg kommunizierte London auf diese Weise mit der französischen *Résistance*, und später sicherten FIDEL CASTRO und CHE GUEVARA ihre Kommunikation auf diese Weise.

Als *high tech*-Variante davon entstand im Kalten Krieg das *Rote Telephon* zwischen dem Weißen Haus und dem Kreml: Damals hielten viele (wohl zu Recht) die Gefahr eines Atomkriegs aus Versehen für erheblich größer als die eines absichtlichen. Um ersteren etwas weniger wahrscheinlich zu machen, einigten sich die beiden Großmächte im Juni 1963 in Genf darauf, das sogenannte *Rote Telephon* einzurichten; es funktioniert seit dem 30. August 1963.

Natürlich handelt es sich dabei nicht wirklich um ein Telephon, denn zu keinem Zeitpunkt des kalten Krieges reichten die Sprachkenntnisse eines amerikanischen Präsidenten oder eines Generalsekretärs der KPdSU auch nur für ein direktes Gespräch über das Wetter. Tatsächlich geht es um eine Fernschreibverbindung mit je vier (bei Siemens Mannheim gebauten) Fernschreibern an beiden Enden: jeweils zwei mit lateinischem und zwei mit kyrillischem Alphabet. Bislang verbrachten sie ihre meiste Zeit damit, stündliche Testnachrichten zu drucken wie amerikanische Baseball-Ergebnisse oder TURGENJEWs *Aufzeichnungen einer Jägers*.

Aus Sicherheitsgründen wurden zwei Leitungen eingerichtet, eine entlang der Route Washington-London-Kopenhagen-Stockholm-Helsinki-Moskau, die andere via Tanger. Natürlich war es unmöglich, diese Leitungen auf ihrer ganzen Länge zu überwachen, so daß niemand ausschließen konnte, daß irgendwo zwischen Moskau und Washington eine vertrauliche Kommunikation abgehört oder – mit potentiell sehr viel katastrophaleren Folgen – eine gefälschte Nachricht eingespielt wurde. Aus diesem Grund mußten alle Nachrichten verschlüsselt werden

Dazu diente folgende einfache Variation des *one time pads:* Von Zeit zu Zeit tauschten die beiden Seiten per Kurier Magnetbänder mit zufallserzeugten Bitfolgen aus. Jedesmal, wenn eine Nachricht übermittelt werden sollte, übersetzte der Fernschreiber diese in eine Bitfolge, d.h. in einen Vektor $\vec{v}$ aus einem Vektorraum $\mathbb{F}_2^N$. Aus den ersten N bislang noch nicht benutzten Bits auf dem Magnetband wurde dazu ein weiterer Vektor $\vec{w} \in \mathbb{F}_2^N$ gebildet, und tatsächlich übertragen wurde die Summe $\vec{s} = \vec{v} + \vec{w}$.

Am anderen Ende der Leitung, wo eine Kopie des Magnetbands vorlag, war $\vec{w}$ bekannt, so daß die Nachricht

$$\vec{v} = \vec{v} + \vec{0} = \vec{v} + (\vec{w} + \vec{w}) = (\vec{v} + \vec{w}) + \vec{w} = \vec{s} + \vec{w}$$

leicht rekonstruiert werden konnte.

Ein Lauscher ohne Magnetband konnte nur die Länge N der Nachricht ermitteln, was bei seitenlangen in Diplomatensprache formulierten Texten so gut wie keine konkrete Information liefert – ganz abgesehen davon, daß man nie ausschließen kann, daß es sich vielleicht einfach um eine zusätzliche Testnachricht zu Wartungszwecken handelt.

Wichtig ist auch, daß jemand, der einfach irgendeinen Vektor $\vec{s}$ in die Leitung einspielt, so gut wie keine Chance hat, daß nach Addition von $\vec{w}$ daraus verständlicher Text wird; die Manipulation wird also mit an Sicherheit grenzender Wahrscheinlichkeit entdeckt.

In alltäglicheren Anwendungen ist der mit dem *one time pad* verbundene Aufwand meist zu hoch; man muß notgedrungen mit Schlüsseln arbeiten, die deutlich kürzer sind als die (Summe der) Nachrichten, die damit verschlüsselt werden.

Informationstheoretische Berechnungen legen nahe, daß VIGENÈRE bereits unsicher wird, wenn die Nachricht nur um 30% länger ist als der Schlüssel. der Klartext nur etwa 30% länger ist als der Schlüssel. Wenn man bedenkt, daß es Programme gibt, die lange Dokumente oder gar eine gesamte Festplatte verschlüsseln in Abhängigkeit von einem nur wenige Zeichen langen Schlüssel mittels eines byte- statt buchstabenbasierten VIGENÈRE-Systems, wundert es nicht, daß so viele Programme auf dem Markt sind, die „vergessene“ Passwörter rekonstruieren.

Wie der bislang größte (bekannt gewordene) Unfall der sowjetischen Kryptographie zeigt, kann selbst der *one time pad* bei falscher Anwendung unsicher werden: Aus irgendeinem Grund wurden Anfang 1942 eine Zeit lang von jedem Einmalblock nicht zwei, sondern vier Exemplare produziert. Die „überflüssigen“ Exemplare wurden nicht zerstört, sondern wanderten ins Vorratslager und wurden später, als ihre Herkunft längst vergessen war, benutzt. Dies gab den amerikanischen und britischen Geheimdiensten die Möglichkeit, einen Teil der geheimsten sowjetische Kommunikation zu entschlüsseln, obwohl zusätzlich zum *one time pad* vorher noch eine Verschlüsselung nach einem Codebuch eingesetzt wurde. Die rekonstruierten Dokumente kann inzwischen auch die interessierte Öffentlichkeit nachlesen: Man suche unter `www.nsa.gov` nach *venona*.

§4: Transpositionschiffren

Permutationen lassen sich nicht nur anwenden, um ein Alphabet zu permutieren; sie können auch verwendet werden, um die Reihenfolge der Buchstaben des Klartexts so zu verändern, daß dieser nicht mehr erkennbar ist.

Konkret wählt man eine (geheime) Blocklänge n und eine Permutation $\pi \in \mathfrak{S}_n$. Der Klartext wird aufgeteilt in Blöcke der Länge n, wobei der letzte eventuell noch mit zufällig gewählten Buchstaben aufgefüllt werden muß; sodann wird der Block $a_1 a_2 \dots a_n$ ersetzt durch $a_{\pi(1)} a_{\pi(2)} \dots a_{\pi(n)}$.

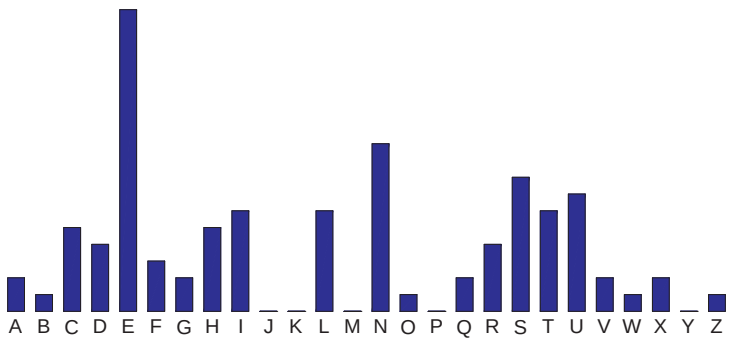
Da nur die Reihenfolge der Buchstaben verändert wurde, ist die Verteilung der Buchstabenhäufigkeiten im Kryptogramm exakt dieselbe

wie im Klartext; man erhält also ein Histogramm, daß genauso aussieht, wie man es von deutschem Klartext erwartet. Auf diese Weise erkennt ein Kryptanalytiker Transpositionschiffren.

Betrachten wir als Beispiel das Kryptogramm

```
CSHSI EUNUE EERRQ VNGDB EUINW HINES ZCUTE NELQL
UHNXI DTCRX EFTTF ESLSD TOALI TEESE DNLSU CEEHS
NRGLE VUEFE AINCN H
```

Wie das Diagramm zeigt, entsprechen die Buchstabenhäufigkeiten darin in der Tat der von deutschem Klartext:



Es hat eine Länge von 96 Zeichen; die Blocklänge n sollte also ein Teiler von 96 sein. (Ein geschickter Kryptograph würde freilich am Ende noch weitere Buchstaben anfügen, so daß das Kryptogramm keinen Hinweis auf n liefert.)

Versuchen wir etwa, das Kryptogramm zu entschlüsseln unter der Annahme, die Schlüssellänge sei gleich sechs. Dazu teilen wir es auf in Blöcke der Länge sechs:

1 CSHSIE	5 WHINES	9 RXEFTT	13 LSUCEE
2 UNUEEE	6 ZCUTEN	10 FESLSD	14 HSNRGL
3 RRQVNG	7 ELQLUH	11 TOALIT	15 EVUEFE
4 DBEUIN	8 NXIDTC	12 EESEDN	16 AINCNH

Da im ersten Block sowohl ein C als auch ein H vorkommen, im zweiter aber kein H steht, ist zu erwarten, daß C und H im Klartext ein CH sind;

in Klartext sollte also die erste Spalte der obigen Blockdarstellung links vor der dritten stehen.

Im dritten Block steht an dritter Stelle ein Q, aber es gibt kein U. Das nächste U steht eine Zeile tiefer an vierter Stelle. Falls Q und U im Klartext ein QU bilden, muß also die dritte Spalte im Klartext die letzte sein und die vierte die erste.

Im sechsten Block haben wir ein C an zweiter Stelle, aber das nächste H steht eine Zeile tiefer an letzter Stelle; dies würde dafür sprechen, daß die zweite Spalte im Klartext an sechster Stelle steht und die sechste an erster.

Im siebten Block steht in der dritten Spalte ein Q und des einzige dazu passende U steht in Spalte fünf; denn im im achten Block gibt es keines. Demnach sollte die dritte Spalte im Klartext vor der fünften stehen.

In Block acht steht ein C, aber weder im achten noch im neunten Block gibt es ein H oder K.

In Spalte vier des dreizehnten Blocks steht ein C; dazu paßt eigentlich nur das H in der ersten Spalte des nächsten Blocks, d.h. die vierte Spalte sollte letzte sein und die erste auch im Klartext die erste.

Der letzte Block schließlich legt nahe, daß die vierte Spalte vor der sechsten stehen sollte.

Natürlich muß nicht jede dieser Folgerungen richtig sein; es kommt vor, daß nach einem C in deutschem Klartext weder ein H noch ein K steht, und es kommt auch vor, daß nach einem Q kein U steht. Hier haben aber so viele Widersprüche zwischen den getroffenen Folgerungen, daß so etwas der Regelfall sein müßte – dies spricht eher dagegen, daß die Verschlüsselung mit Blocklänge sechs erfolgte.

Orden wir den Chiffretext in Blöcke der Länge acht, bietet sich folgendes Bild:

1 CSHSIEUN	4 WHINESZC	7 RXEFTTFE	10 LSUCEEHS
2 UEEERRQV	5 UTENELQL	8 SLSDTOAL	11 NRGLEVUE
3 NGDBEUIN	6 UHNXIDTC	9 ITEESEDN	12 FEAINCNH

Hier legt Block eins nahe, daß die erste Spalte unmittelbar vor der dritten stehen sollte. Block zwei kann ein QU realisieren, wenn entweder Spalte sieben im Klartext vor der ersten steht, oder aber sie steht ganz hinten und Spalte sechs ganz vorne. Block vier legt nahe, daß Spalte acht vor Spalte zwei stehen sollte, bei Block fünf haben wir dieselbe Situation wie bei Block zwei und bei Block sechs wie bei Block vier. Block zehn plaziert Spalte vier vor Spalte sieben und Block zwölf sechs vor acht.

Vermuten wir bei Block zwei die erste Möglichkeit, so stehen im Klartext einerseits die Spalten 4, 7, 1, 3 als Viererblock nebeneinander, andererseits Spalten 6, 8, 2 als Dreierblock. Entweder am Anfang oder am Ende oder dazwischen steht die fünfte Spalte.

Die möglichen Reihenfolgen sind also 47136825, 47135682, 54713682, 68247135, 68254713 und 56824713. Betrachtet man den Effekt dieser Permutationen auf die erste Zeile, ist klar, daß nur die erste Möglichkeit in Frage kommt; der Klartext ist also

SUCHEN SIE QUERVERBINDUNGEN ZWISCHEN
QUELLTEXT UND CHIFFRETEXT DAS SOLLTE
DIE ENTSCHLUESSELUNG VEREINFACHEN

Bei diesem Kryptogramm erforderte die Kryptanalyse wegen der relativ vielen Paare CH und QU wenig Arbeit; bei schwierigeren Texte hat er nicht unbedingt viele sichere Regeln, jedoch gibt es durchaus noch andere ziemlich sichere Kombinationen: So steht zum Beispiel, wie die Kontaktdiagramme zeigen, vor einem D sehr häufig ein N, und nach einem B oder G kommt zwar nicht immer, aber doch oft ein E, nach V steht meist O oder E. Bei kurzen Kryptogrammen mit wenigen typischen Buchstabenpaaren kann er für jedes Paar von Spalten die Wahrscheinlichkeiten dafür ausrechnen, daß die eine vor der anderen steht; auch das macht die Situation viel klarer. Oft springen dem Betrachter bei hinreichend langen Blöcken (die man aus Sicherheitsgründen braucht) auch Wörter ins Auge, die weiterhelfen können. Hier ist das Sprachgefühl des Kryptanalytikers oft wichtiger als jede mathematische Analyse.

Transpositionschiffren spielten bis zum ersten Weltkrieg eine große Rolle in der deutschen Militärkryptographie, allerdings wurden sie nie allein angewendet, sondern nur als Teil eines zweistufigen Verfahrens.

Der Grund dürfte klar sein: Gegen einem Angriff mit bekanntem Klartext einer Länge, die über der Blocklänge liegt, bieten Transpositionschiffren keinerlei Schutz. Heute sind sie (zu Recht) weitgehend vergessen.

§5: Rotormaschinen

Alle bislang behandelten Kryptoverfahren sind hinreichend einfach, daß Nachrichten einfach mit Bleistift und Papier ver- und entschlüsselt werden können. Mit Ausnahme des logistisch sehr aufwendigen *one time pad* lassen sie sich allerdings auch alle relativ einfach knacken.

Viele klassische Kryptographen versuchten, die Sicherheit des *one time pad* mit der Einfachheit der VIGENÈRE-Chiffre zu kombinieren: Letztere könnte man auffassen als einen *one time pad*, bei dem die immer gleiche Buchstabenfolge endlos wiederholt wird, und das ist – wie wir gesehen haben – extrem unsicher.

Alternativ könnte man versuchen, aus einer relativ kurzen Schlüsselinformation eine komplizierte Buchstabenfolge zu erzeugen, die sich idealerweise verhält wie eine Zufallsfolge. Schließlich haben selbst viele Taschenrechner heute eine Taste, die Zufallszahlen erzeugt, und auch die meisten Programmiersprachen bieten „Zufallsgeneratoren“ an, manchmal auch unter dem korrekteren Namen „Pseudozufallsgenerator“.

Leider liest man immer wieder Untersuchungen, die vielen dieser Programme miserable Ergebnisse attestieren, aber es gibt durchaus Algorithmen, die Folgen liefern, die sich beispielsweise für die Zwecke einer Simulation nicht wesentlich von echt zufällig gewählten Zahlen unterscheiden – wie immer auch man letztere bekommen kann. (Wir werden uns am Ende der Vorlesung auch damit beschäftigen.)

Aber, und hier kommt das zweite *leider*, ein Algorithmus der gute Ergebnisse für Simulationszwecke liefert, liefert nicht notwendigerweise auch gute Ergebnisse für kryptographische Zwecke: Diese beiden Anwendungen sind praktisch disjunkt, denn die Algorithmen, die derzeit als kryptographisch sicher gelten, sind für Simulationszwecke viel zu aufwendig, und die bewährten Algorithmen, die für gute Simulationen eingesetzt werden, sind kryptographisch völlig unsicher.

Um 1920, als unabhängig voneinander in mehreren Länder erste Verschlüsselungsmaschinen auftauchten, wußte man von diesen Algorithmen des Computerzeitalters natürlich noch nichts; man versuchte einfach, aus relativ kurzer und gut übermittelbarer Schlüsselinformation einen möglichst komplexen Schlüsselstrom zu erzeugen.

Wie üblich in der Kryptographie führten nahezu identische Ansätze zu dramatisch verschiedenen Ergebnissen:

Die Ansätze waren für Laien praktisch ununterscheidbar: Praktisch alle Maschinen bestanden aus sogenannten *Rotoren.* Dabei handelte es sich um Räder mit (im mitteleuropäischen Bereich) 26 äquidistanten elektrischen Kontakten auf der Oberfläche. Durch das Innere liefen Drahtverbindungen zwischen diesen Kontakten, die jeweils zwei miteinander verbanden; ein Strom der an einem dieser Kontakte angelegt wurde, verließ den Rotor also an einer festen anderen Stelle. Dadurch wurde eine Permutation aus $\mathfrak{S}_{26}$ realisiert, die sich als Produkt von 13 elementfremden Transpositionen schreiben ließ.

Die Tastatureingabe wurde über mehrere solcher Rotoren geleitet, so daß effektiv ein Produkt von drei bis fünf solcher Permutationen realisiert wurde. Für sich allein bietet das natürlich keinerlei Sicherheit, denn das Produkt von egal wievielen Permutationen ist schließlich wieder nur eine einfache Permutation. Deshalb drehte sich einer der Rotoren nach jedem Buchstaben um eins weiter um so eine andere Permutation zu realisieren.

Aus Sicht eines Kryptanalytikers ist auch eine solche reguläre Bewegung kein großes Hindernis; als zusätzliche Sicherheit gehört daher zu einer Rotormaschine, daß sich auch weiteren Rotoren gelegentlich aber nicht immer weiterdrehen. Bei den ersten Rotormaschinen wie etwa der von HEBERN geschah dies in einer völlig regelmäßigen Weise: Der fünfte Rotor bewegte sich nach jedem Buchstaben um eins weiter, der erste nach je 26 und der dritte nach je 676 Buchstaben. Der zweite und der vierte Rotor änderten ihre Position nicht. Bei anderen Maschinen wie der Enigma wurde die Bewegung der Rotoren durch unregelmäßig auf den einzelnen Rotoren angeordneten Haken realisiert und war damit auch von der Rotorkonfiguration abhängig.

Zur Entschlüsselung wird der Chiffretext einfach in eine identisch aufgebaute Maschine mit gleicher Anfangsstellung aber umgekehrter Rotorreihenfolge gegeben. (Das Bild hier zeigt eine Enigma-Maschine mit drei Rotoren.)

Bei der Enigma, im Gegensatz zu anderen Rotormaschinen, war der letzte Rotor fest und leitet den Strom, nach einer weiteren Permutation, zurück durch die beweglichen Rotoren. Dadurch kann die Entschlüsselung mit derselben Rotorenkonfiguration erfolgen.

Im WWW sind verschiedene Simulatoren sowohl der vielen deutschen Enigma-Modelle als auch anderer Rotor-Maschinen zu finden; eine Enigma der deutschen Abwehr mit vier Rotoren findet man beispielsweise unter `http://home.caiway.nl/~antonh/enigma_ga.html` . Die mit solchen Maschinen erzielbaren Perioden liegen im Bereich über 100 000; wie jeder Leser inzwischen hoffentlich verstanden hat, ist das aber freilich keine Garantie für Sicherheit.

In der Tat wurde eine der ersten Rotormaschinen, der Kryptograph des deutschen Ingenieurs ALEXANDER VON KRYHA, die in Europa wie auch Amerika recht erfolgreich an Geschäftsleute verkauft wurde, im Januar 1933 auch der amerikanischen Armee zum Kauf angeboten. Diese ließ sich zu Testzwecken ein damit verschlüsseltes Kryptogramm aus 1135 Buchstaben (200 Wörtern) geben, und leitete dieses an ihren Chefkryptanalytiker ROBERT FRIEDMAN weiter. Diesem gelang es, zusammen mit drei Mitarbeitern, das Kryptogramm (ohne Maschinen) in zwei Stunden und 41 Minuten zu entschlüsseln. Die Öffentlichkeit erfuhr davon natürlich nichts; auch um 1950 verschlüsselte das Auswärtige Amt in Bonn noch seine diplomatische Korrespondenz mit solchen Maschinen.

Die Enigma-Maschine war deutlich sicherer, wurde aber trotzdem bereits während des zweiten Weltkriegs fast routinemäßig geknackt. Ein Grund dafür war, daß während der Gültigkeitsperiode eines Schlüssels (d.h. während eines Tages) alle Nachrichten mit derselben Anfangsstellung verschlüsselt werden müssen oder aber die Anfangsstellung auf andere Weise übermittelt werden muß.

Die deutsche Wehrmacht umging dieses Problem in einigen Netzen

dadurch, daß mit dem jeweiligen Tagesschlüssel nur ein speziell für die folgende Nachricht gültiger zufällig gewählter Schlüssel übermittelt wurde; danach wurde die Maschine gemäß diesem Schlüssel in eine

neue Anfangsstellung gebracht und die Nachricht verschlüsselt.

Grundsätzlich ist das eine recht sichere Option; falls allerdings durch einen Übermittlungsfehler bei den ersten drei Buchstaben diese nicht korrekt beim Empfänger ankommen, kann dieser die Nachricht nicht entschlüsseln und muß um Wiederholung bitten.

Dies war den Militärs zu umständlich; deshalb wurden die drei Buchstaben der Anfangsstellung nicht einmal, sondern zweimal übertragen. Sobald die Kryptanalytiker in der britischen Dechiffrierzentrale in *Bletchley Park* dies merkten, hatten sie eine entscheidende Zusatzinformation: Die ersten sechs Buchstaben einer jeden Übertragung gehören zu einem Klartext, der aus zweimal derselben Dreiergruppe von Buchstaben besteht. Der Tagesschlüssel muß daher die Eigenschaft haben, daß alle diese aufgefangenen Sechsergruppen bei Entschlüsselung damit zu zwei aufeinanderfolgenden identischen Dreiergruppe werden, was bei zufällig gewählten Schlüsseln natürlich fast nie der Fall ist.

Um zu sehen, für welche(n) Schlüssel das funktioniert, muß man im Prinzip alle ausprobieren, was bei $26^3 = 17\,576$ möglichen Anfangsstellungen und anfänglich $3! = 6$, später $5 \cdot 4 \cdot 3 = 60$ möglichen Rotorkombinationen für die damalige Zeit ein beträchtlicher Aufwand war. Hierzu wurden in *Bletchley Park* spezielle Geräte konstruiert, die sogenannten *Bomben*, die mit hoher Geschwindigkeit viele Rotoren parallel arbeiten ließen.

Viele Angriffe waren auch Angriffe mit bekanntem Klartext, sogenannten *cribs*, die ebenfalls mit Hilfe der Bomben auf mögliche Schlüssel untersucht werden konnten. Meist gelang es, bis etwa elf Uhr morgens den Tagesschlüssel zu ermitteln und ab dann alle Kommunikation zu entschlüsseln.

Die Verdrahtung der einzelnen Rotoren war schon vor dem Krieg von polnischen Kryptanalytikern geknackt worden, nachdem der Deutsche HANS-THILO SCHMIDT (1888-1943) die Gebrauchsanweisung für die Enigma-Maschine und die Liste der Schlüssel für September und Oktober 1932 an einen französischen Kryptographen namens BERTRAND verkauft hatte. Dieser konnte die Maschine zwar nicht knacken, gab die Informationen aber nach England und nach Polen weiter; einer Gruppe

polnischer Kryptographen unter Leitung von MARIAN ADAM REJEWSKI (1905–1980) gelang dann die Rekonstruktion der drei damals gebräuchlichen Rotoren und somit der Maschine.

Diese Rotoren blieben während des gesamten zweiten Weltkriegs im Einsatz; kein Rotor wurde je aus dem Verkehr gezogen. Zwar wurden in manchen Netzwerken wie etwa bei den U-Booten ein oder zwei neue Rotoren eingeführt, aber da diese zusammen mit den alten benutzt wurden, hatten die Kryptanalytiker dann nur noch das Problem, die Verdrahtung *eines* unbekannten Rotors zu ermitteln, was angesichts des großen Nachrichtenvolumens stets gelang.

Nach dem Krieg verkaufte die britische Regierung übrigens die erbeuteten Enigma-Maschinen an Regierungen ihrer Ex-Kolonien ohne denen über die erfolgreiche Entschlüsselung zu berichten; damit war sie erstens immer gut über die Vorkommnisse in diesen Ländern informiert und verdiente zweitens noch daran.

Seit etwa 1970 werden Rotormaschinen nicht mehr eingesetzt – außer in UNIX. Das dortige `crypt(1)`-Kommando simuliert eine Rotormaschine mit *einem* Rotor, der 256 Ein- und Ausgänge hat. Sehr sicher ist das nichts; selbst in der `man`-Seite dazu heißt es *Methods of attack on such machines are widely known; thus crypt provides minimal security.* Für Paßwörter verwendet UNIX daher auch nicht dieses Kommando, sondern das auf einem „gesalzenen“ DES beruhende deutlich bessere `crypt(3)`-Unterprogramm. Es gibt inzwischen ein universelleres Kommando `mcrypt`, das sich zwar mit Option `--enigma` genauso verhält wie `crypt(1)`, das aber auch alternative Algorithmen anbietet, die nach heutigen Standards sicher sind, z.B. AES und triple-DES. Es ist unter der GNU public licence erhältlich unter `mcrypt.sourceforge.net`.

§6: Literaturhinweise

Das (dickleibige) Standardwerk zur alten Kryptographie mit Schwerpunkt auf der geschichtlichen Darstellung ist das bereits im letzten Kapitel erwähnte Buch

DAVID KAHN: *The Codebreakers – the comprehensive history of secret communication from ancient time to the internet,* Scribner, New York, [2]1996

Für zahlreiche ältere Kryptologen war die erste Auflage von 1967 der Einstieg in ihr Arbeitsgebiet.

Deutlich kürzer, billiger und verfahrensorientierter ist

HELEN FOUCHÉ GAINES: *Cryptanalysis – a study of ciphers and their solution,* Dover, New York, 1956 *(Originalausgabe 1939)*

L. SACCO: *Manuel de cryptographie,* Payot, Paris, 1951

beschreibt die klassischen Verfahren und ihre Kryptanalyse aus seiner Sicht als Chef des Chiffrierdienstes der italienischen Armee. Eine Reihe entsprechender Bücher gibt es auch von WILLIAM FRIEDMAN, jedoch sind diese in Deutschland wenn überhaupt nur schwer zu finden.

Rotormaschinen und ihrer Kryptanalyse ist das Buch

CIPHER A. DEAVOURS, LOUIS KRUH: *Machine cryptography and modern cryptanalysis,* Artech House, Dedham MA, 1985

gewidmet. „Modern“ bezeichnet hier den Zeitraum von etwa 1920–1970. Weitere Informationen zur Kryptanalyse von Rotormaschinen im zweiten Weltkrieg findet man im Buch von KAHN sowie bei

BENGT BECKMAN: *Codebreakers – Arne Beurling and the Swedish crypto program during Word War II,* American Mathematical Society, Providence, R.I., 2002

Moderne Lehrbücher mit Kapiteln über klassische Kryptographie sowie über statistische und informationstheoretische Ansätze zur Kryptanalyse sind

JAN C.A. VAN DER LUBBE: *Basic Methods of cryptography,* Cambridge University Press, 1998

und

ALAN G. KONHEIM: *Computer Security and Cryptography,* Wiley, 2007

sowie dessen Vorgänger

ALAN G. KONHEIM: *Cryptography – A Primer,* Wiley, 1981