# Kapitel 1
# Klassische Verfahren der Kryptologie

Das wichtigste vorab: Abgesehen von einer einzigen Ausnahme ist keines der hier vorgestellten Verfahren auch nur ansatzweise sicher. Die meisten lassen sich ohne große Kenntnisse und ohne großen Aufwand von Hand knacken; einem Experten, der mit Computer arbeitet, bereitet (abgesehen von der einzigen Ausnahme) keines nennenswerte Schwierigkeiten.

Daraus sollte man folgern, daß diese Verfahren heutzutage keine Rolle mehr spielen, aber dem ist leider nicht so: Viele Verschlüsselungsalgorithmen für Festplatten oder in Textverarbeitungsprogrammen arbeiten noch heute mit Algorithmen wie dem VIGENÉRE-Verfahren, das wir gleich als völlig unsicher erkennen werden. Der einzige „Fortschritt" besteht darin, daß nun nicht mehr die 26 Buchstaben des Alphabets die Bausteine einer Nachricht bilden, sondern die 256 möglichen Bytes. Das ist nicht nur für den Anwender eine große Bequemlichkeit, sondern auch für den Kryptanalytiker, denn wie wir im nächsten Kapitel sehen werden, erhöht es die Redundanz der Nachricht deutlich und bietet damit bessere Ansatzpunkte zur unbefugten Entschlüsselung.

Die alten Kryptologen hatten das verstanden: Sie verzichteten bewußt auf die Unterscheidung zwischen Groß- und Kleinbuchstaben und ließen auch alle Satz- und Leerzeichen weg. Für den legitimen Empfänger ist das nur eine kleine Unbequemlichkeit; für den illegitimen Kryptanalytiker bedeutet es zumindest bei kurzen Nachrichten eine deutliche Erschwernis.

Sinn dieses Kapitels ist nicht in erster Linie die Vorstellung alter Verfahren; es soll in erster Linie eine Vorstellung davon vermitteln, mit welch



geringem Aufwand man gelegentlich Verschlüsselungsverfahren von (auf den ersten Blick) gewaltiger Komplexität knacken kann und auf die informationstheoretischen Ansätze des nächsten Kapitels vorbereiten.

## §1: Die Anfänge

Wo nur eine verschwindend kleine Minderheit lesen kann, ist *jede* Schrift eine Geheimschrift; viele der frühen Hochkulturen wie etwa die chinesische sahen daher keinen Grund für die Entwicklung von Verschlüsselungstechniken.

Auch wenn der Sender und Empfänger eine gemeinsame Sprache haben, die ihrer Umgebung unbekannt ist, gibt es keine Notwendigkeit für eine Verschlüsselung. Dies war selbst im zwanzigsten Jahrhundet noch durchaus praktikabel: Beispielsweise beschäftigte die amerikanische Armee im zweiten Weltkrieg Navajos als *code talkers.* Da indianische Sprachen eine sehr komplexe Grammatik und eine schwierige Phonetik haben, sind sie nur schwer erlernbar und noch schwieriger nachzuahmen: Damals waren nur 28 Nicht-Navajos bekannt, die der Sprache mächtig waren. Aus diesem Grund hatte die Kommunikation in Navajo eine größere Sicherheit als die immer wieder geknackte maschinelle Kryptographie.

Ähnlich zeigte sich nach dem zweiten Weltkrieg, beim UN-Einsatz im Kongo, daß das irische Kontingent die effektivste Kryptographie hatte: Die Soldaten unterhielten sich einfach auf gälisch.

Ein historisches Beispiel, wie gelegentlich selbst weitbekannte Sprachen als Kryptosystem dienen können, bietet der Burenkrieg: Dort nutzten die Engländer aus, daß nur wenige Buren nennenswerte Bildung hatten und verwendeten Latein als Kryptographie. Die Einnahme der Stadt *Sindh* etwa wurde gemeldet mit der Nachricht *pecavi,* auf englisch *I have sinned.*

Anderswo schützte man Nachrichten dadurch, daß selbst das *Vorhandensein* einer Nachricht verschleiert wurde; von einem berühmten Beispiel etwa berichtet HERODOT in seinen *Historien:*

Ein gewisser HISTIAEUS, der am persischen Hof lebte, wollte seinen Schwiegersohn, den Tyrannen ARISTAGORAS von Miletos, zu einem Aufstand gegen die Perser bewegen. Dazu rasierte er einem zuverlässigen Sklaven die Haare ab, tätowierte seine Botschaft in dessen Kopfhaut und schickte ihn, nachdem die Haare nachgewachsen waren, zu ARISTAGORAS mit dem Auftrag, sich von diesem die Haare schneiden zu lassen.

Heute bezeichnet man dieses Verschleiern der bloßen Existenz einer Nachricht als *Steganographie;* wo sie erfolgreich ist, besteht keine Notwendigkeit mehr, sich um *Kryptographie* zu kümmern.

In vielen Situationen ist Steganographie undurchführbar, da entweder der Gegner weiß, daß Nachrichten ausgetauscht werden, oder aber weil einfach der Umfang der Kommunikation zu groß ist. Typisches Beispiel ist die militärische Kommunikation, für die erstmals die Spartaner um 487 v.Chr. ein kryptographisches Verfahren benutzten, den sogenannten *Skytale:* Ein Lederband wurde um ein Stück Holz gewickelt, die Nachricht in senkrechten Spalten auf das Leder geschrieben und dieses anschließend von einem Soldaten als Gürtel getragen. Der Empfänger mußte ein identisches Stück Holz haben, konnte das Band darum wickeln und sodann die Nachricht lesen.

Auch in Indien war Kryptographie zu politischen und militärischen Zwecken schon früh bekannt; kurz vor 300 v.Chr. schrieb KAUTILYA sein Buch *Artha-sastra,* in dem er (für Diplomaten) nicht nur die Verwendung von Codes erklärt, sondern auch, wie man diese knacken kann – spätestens seit damals gibt es also zusätzlich zur Kryptographie auch die *Kryptanalyse.* Rund 700 Jahre später bezeichnete VATSAYANA in der *Kama sutra* Kryptographie in Schrift und Wort als zwei der 64 Fähigkeiten, über die jede Frau verfügen sollte.

Der älteste heute noch verwendete Code geht auf CAESAR zurück und ist ein deutlicher Rückschritt hinter die Spartaner und Inder: Genau wie die Römer selbst war auch ihre Kryptographie sehr primitiv. Sie bestand im Falle CAESARS einfach darin, das Alphabet zyklisch um drei Stellen nach rechts zu verschieben, d.h.

$$A \to D, \quad B \to E, \quad ,\ldots, \quad X \to A, \quad Y \to B, \quad Z \to C\,.$$



Heute bezeichnet man jeden Code, der das Alphabet zyklisch um eine feste Anzahl $n$ von Stellen verschiebt, als CAESAR-*Chiffre;* im Internet bedeutsam ist der Fall $n = 13$, die sogenannte `ROT-13`-Chiffre des Usenet. Sie dient natürlich nicht dazu, Informationen geheimzuhalten: Ihr Gebrauch ist lediglich ein Hinweis darauf, daß der verschlüsselte Text nicht nach jedermanns Geschmack ist und jemand, der ihn trotzdem entschlüsselt, sich nicht über den Inhalt beklagen sollte.

## §2: Arten kryptanalytischer Angriffe

In seiner Originalform hat CAESARs Verfahren einen wesentlichen Nachteil: Da alles eindeutig festgelegt ist, kann jeder, der irgendwann die Methode erlernte, auch künftig alle Nachrichten lesen – angesichts der zeitlich sehr variablen Bündnisse im römischen Reich sicherlich nicht ganz unproblematisch.

Auch ein Gegner, der es irgendwie geschafft hat, *eine* Nachricht zu entschlüsseln, hat künftig die Möglichkeit *jede* Nachricht zu entschlüsseln. Im Rom des ersten Jahrhunderts, wo wohl niemand die *Artha-sastra* kannte, war das vielleicht kein sehr großes Problem; für die heutigen Massenanwendungen der Kryptographie im Internet und im Bankenbereich ist eine solche Chiffre aber wertlos: Selbst die Militärs gehen schon seit mindestens einem halben Jahrhundert davon aus, daß das *grundsätzliche* Verschlüsselungsverfahren über kurz oder lang auch dem Gegner bekannt sein wird – auch wenn sie selbstverständlich weiterhin alles tun, um dies zu verhindern. Die Spartaner hatten hier vorgesorgt: Auch wer ihr Verfahren kannte, hatte ohne das jeweils passende Stück Holz Schwierigkeiten, die Nachricht zu lesen.

Abgesehen von der Original CAESAR-Chiffre hat daher praktisch jedes kryptographische Verfahren zwei Komponenten: Einmal die *Verschlüsselungsmethode,* bei der man realistischerweise davon ausgehen muß, daß man sie nicht unbegrenzt geheimhalten kann, und dann noch einen *Schlüssel,* der bei der Anwendung dieser Methode benutzt wird und der häufig gewechselt werden muß. Als *sicher* darf man ein Kryptoverfahren nur dann betrachten, wenn ein Gegner trotz genauer Kenntnis des Verfahrens nicht in der Lage ist, einen Text ohne Kenntnis des Schlüssels zu dechiffrieren.

Der letzte Satz klingt vernünftig, ist aber unvollständig, da er eine wesentliche Frage ausklammert: *Was darf der Gegner alles wissen, bevor die Sicherheit des Verfahrens* (der Schlüssel) *gefährdet ist?*

Die offensichtliche Antwort scheint zu sein, daß er den chiffrierten Text und das Verfahren kennt. Das ist aber leider viel zu optimistisch: Schon 725 beschrieb ABU YAHMADI, der Autor des ersten arabischen Wörterbuchs, wie er ein altes byzantinisches Kryptogramm löste ausgehend von der Annahme, daß es mit der Formel „Im Namen Gottes“ beginne – für einen Text aus der fraglichen Zeit fast eine Selbstverständlichkeit.

Heute beginnen wir Texte nicht mehr mit der Formel „Im Namen Gottes“, aber wir verwenden Word-Dateien, zip-Dateien *usw.*, die mit *magic bytes* und anderer Vorspann-Information beginnen; einige von uns versenden e-mails, bei denen die ASCII-Kunst oder die Visitenkarte am Ende länger ist als die Nachricht selbst; Militärs beginnen Nachrichten typischerweise mit Worten wie „Wehrmachtskommando XY an ...“; Banken haben sich für elektronische Überweisungen auf ziemlich rigide Formate geeinigt; jeder Besitzer einer EC-Karte kennt seine Geheimzahl und die sonstige Kontoinformation auf der Karte und kann, so er einen Magnetkartenleser hat, auch die verschlüsselte Version auf der Karte lesen: Falls er mit dieser Information (eventuell auch von vielen EC-Karten statt nur von einer) die zehnstellige Schlüsselzahl des Systems ermitteln könnte, müßten sofort alle Geldautomaten abgeschaltet werden.

Ein Gegner kennt daher in vielen Fällen nicht nur den Chiffretext, sondern er kennt auch zumindest einen Teil des zugehörigen Klartexts. Ein gutes Kryptoverfahren muß sicherstellen, daß er trotzdem nicht in der Lage ist, den Schlüssel zu identifizieren und damit auch den ihm unbekannten Rest zu entschlüsseln.

In manchen Situationen ist der Gegner sogar in einer noch besseren Position: Er kann sich zu von ihm frei gewähltem Klartext den Chiffretext verschaffen. Eine solche Situationen könnte beispielsweise auftreten, wenn ein Diplomat ein Schreiben übergibt, bei dem er ziemlich sicher



sein kann, daß es anschließend verschlüsselt weitergegeben wird. Kurz vor Ausbruch des zweiten Weltkriegs übergab sogar einmal ein japanischer Diplomat ein Schreiben an das amerikanische Außenministerium mit der Bitte, es nach Japan zu übermitteln, da er den Beamten der eigenen Chiffrierabteilung mißtraue – ob dahinter ein Versuch einer solchen Attacke lag, ist bis heute unbekannt. (Die Amerikaner übermittelten das Dokument.)

Heute bilden auch Paßwortdateien wenig gesicherter Computersysteme einen Ansatz für solche Attacken: Jeder Benutzer kann sein Paßwort beliebig oft ändern und auf jeden beliebigen Wert setzen; das Ergebnis wird verschlüsselt in einer Paßwortdatei gespeichert, die auf manchen Computern für jeden Benutzer lesbar ist. Natürlich muß das Verschlüsselungsverfahren sicherstellen, daß trotzdem niemand in der Lage ist, die Paßwörter *anderer* Benutzer zu entschlüsseln.

Auch Smartcards bieten Ansätze für Angriffe mit frei wählbarem Klartext: Sichere Kryptosysteme arbeiten mit Schlüssellängen, die weit jenseits dessen liegen, was sich ein Mensch merken kann; sie können daher nur mit Smartcards oder etwas ähnlichem funktionieren. Diese sind natürlich *(?)* zusätzlich durch eine Geheimzahl oder ein Paßwort gesichert, aber viele unbedarfte Benutzer stellen praktisch sicher, daß ihre Karte nie ohne Paßwort gestohlen werden kann. Damit ist der Dieb in der Lage, die Karte zum Lesen verschlüsselter Nachrichten, zur elektronischen Unterschrift mit dem Namen des Bestohlenen und in Zugangskontrollsystemen verwenden – bis der Diebstahl entdeckt wird.

Eine vielseitig einsetzbare Karte wird sicherlich schnell vermißt werden; danach wird der in die Karte kodierte Schlüssel für ungültig erklärt und somit nutzlos. Ein geschickter Dieb muß also die Karte wieder zurückgeben, *bevor* der Verlust bemerkt wird. Die Schlüsselinformation kann er nicht kopieren: Die steht in einem auslesesicheren Register. Er kann aber beliebig viel Eingabetext seiner Wahl von der Karte bearbeiten lassen, und ein gutes Kryptosystem muß sicherstellen. daß er trotzdem nichts über den Schlüssel erfährt und ohne die Karte später nicht mehr in der Lage ist, verschlüsselte Nachrichten an den Besitzer zu dechiffrieren, in dessen Namen zu unterschreiben oder sich als diesen auszugeben.

## §3: Kryptanalyse der Caesarchiffre

Es ist klar, daß auch eine CAESAR-Substitution mit unbekannter Verschiebung $n$ keinen Schutz bietet gegen einen Angreifer, der Klartext und zugehörigen Chiffretext kennt: Sobald er nur für einen einzigen Buchstaben sowohl diesen als auch seine Verschlüsselung kennt, kann er $n$ berechnen.

Aber auch ein Angreifer, der allein mit Chiffretext arbeiten muß, hat keine großen Schwierigkeiten: Er kann beispielsweise einfach alle 26 Möglichkeiten $n = 0, \ldots, 25$ durchprobieren; sobald sinnvoller Klartext entsteht, hat er den Schlüssel gefunden. Informationstheoretische Überlegungen zeigen, daß im Allgemeinen schon nach zwei Buchstaben klar ist, ob es sich bei einer Probeentschlüsselung um deutschen Klartext handeln kann oder nicht.

Es geht aber noch einfacher: Bekanntlich kommen die 26 Buchstaben des Alphabets in deutschem Text nicht mit gleicher relativer Häufigkeit vor. Abbildung eins zeigt die Häufigkeitsverteilung der 260 238 Buchstaben aus dem Roman *Dr. Katzenbergers Badereise* von JEAN PAUL. (Wer nachzählen möchte, findet diesen und viele andere Texte im Internet auf den Seiten des Projekts Gutenberg unter `http://gutenberg.aol.de/` und bei mehreren mirror sites.)

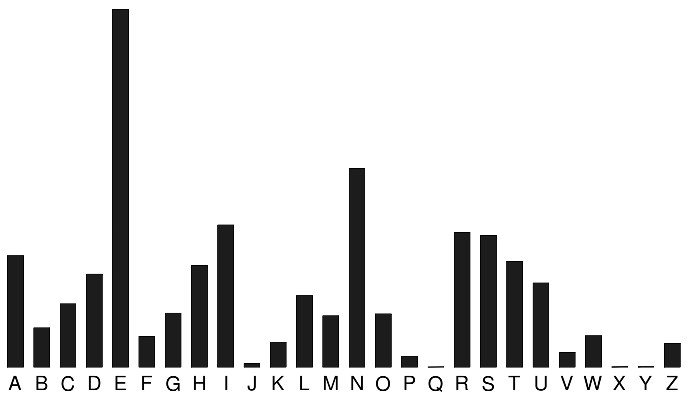



Abb. 1: Buchstabenhäufigkeiten in deutschem Text



In Prozentzahlen ausgedrückt sieht das Ergebnis so aus:

| A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|
| 5,75 | 2,03 | 3,27 | 4,81 | 18,49 | 1,58 | 2,79 | 5,25 | 7,35 |

| J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|
| 0,19 | 1,29 | 3,69 | 2,65 | 10,27 | 2,75 | 0,57 | 0,01 | 6,95 |

| S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|
| 6,81 | 5,46 | 4,35 | 0,76 | 1,62 | 0,02 | 0,05 | 1,23 |

Wie zu erwarten war, ist also E mit großem Abstand der häufigste Buchstabe, während Q, X und Y sehr selten vorkommen.

Wenn ich einen anderen Text als *Dr. Katzenbergers Badereise* ausgezählt hätte, wären die Ergebnisse natürlich anders gewesen: Das Spiegel-Interview mit dem Nobelpreisträger HORST STÖRMER etwa hat 5455 Buchstaben, die sich prozentual folgendermaßen verteilen:

| A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|
| 6,00 | 2,07 | 2,84 | 4,16 | 17,23 | 1,50 | 2,97 | 4,68 | 8,18 |

| J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|
| 0,26 | 1,20 | 3,81 | 3,27 | 10,45 | 3,04 | 1,05 | 0,07 | 6,93 |

| S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|
| 6,84 | 5,56 | 4,42 | 0,73 | 1,43 | 0,02 | 0,33 | 1,03 |

Zumindest in absoluten Prozentzahlen sind die Unterschiede praktisch vernachlässigbar; die *relativen* Unterschiede sind bei den häufigen Buchstaben ebenfalls unbedeutend, bei den seltenen allerdings teilweise recht dramatisch.

Bedenkt man aber, daß 0,02% X bei einem Text aus 5455 Buchstaben gleichbedeutend ist mit nur einem X und 0,07% J mit vier J, so sieht man, daß auch diese Abweichungen minimale Zufallsschwankungen sind.

Für sehr kurze Texte spielen solche Zufallsschwankungen eine große Rolle: Falls wir etwa das Kryptogramm

```
DQJUL IILPP RUJHQ JUDXH Q
```

entschlüsseln wollen, zeigt uns Abbildung zwei, daß es dort gleich drei häufigste Buchstaben gibt: J, Q und U kommen je dreimal vor. Entschlüsselt man unter der Annahme, daß J dem E entspricht, ist $n = 5$ und das Entschlüsselungsergebnis `YLEPG DDGKK MPECL EPYSC L` enthält zwar drei E, hat aber sonst nur wenig mit deutschem Text zu tun. Auch die Entschlüsselungsversuche

`REXIZ WWZDD FIXVE XIRLV E`

und

`NATEV SSVZZ BETRA TENHR A`,

die von $E \to Q$ (d.h. $n = 12$) und $E \to U$ (d.h. $n = 16$) ausgehen, liefern kein vernünftiges Ergebnis.

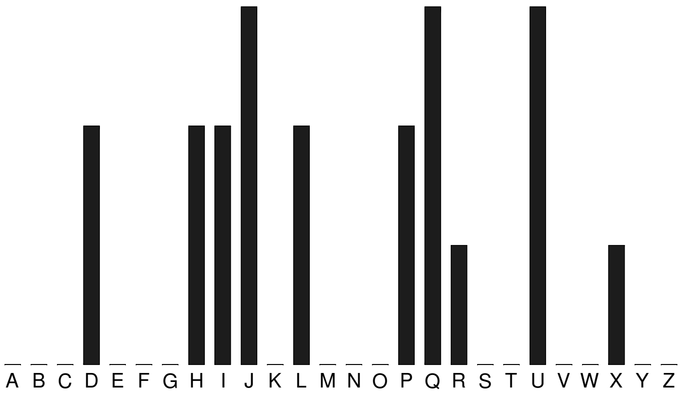



Abb. 2: Buchstabenhäufigkeiten eines kurzen Kryptogramms

Ein klassischer Kryptanalytiker wäre anders vorgegangen: Er wäre sich beim Blick auf Abbildung zwei praktisch sicher gewesen, daß die drei Leerstellen am Ende und die beiden am Anfang wohl nur den fünf seltenen Buchstaben von V bis Z entsprechen können, so daß Abbildung eins um drei Stellen verschoben wurde, d.h. A wurde als D verschlüsselt *usw.,* wie es schon CAESAR praktizierte. Dies führt zum in fast jedem Lehrbuch der klassischen Kryptographie zu findenden Klartext `ANGRIFF IM MORGENGRAUEN`.



Ein moderner Kryptograph möchte sich mit trivialen Kryptogrammen wie dem gerade betrachteten nicht mehr selbst beschäftigen; für solche Dreckarbeit hat er schließlich seine Computer. Denen springt keine Lücke *ins Auge,* und sie haben auch Schwierigkeiten, Lücken *exakt* von Nichtlücken zu unterscheiden. Diese Schwierigkeiten hatte der klassische Kryptanalytiker zwar auch, aber er konnte sich da im Gegensatz zu einem Computer meist auf sein Gefühl verlassen – auch wenn ihn das immer wieder einmal gewaltig in die Irre schickte.

Computer dagegen lieben Zahlen, und auch dazu können ihnen die klassischen Kryptanalytiker verhelfen: Sei $p_i$ die Häufigkeit des $i$-ten Buchstabens des Alphabets in deutschen Texten, geschätzt etwa durch die Prozentzahlen in obiger Tabelle, und sei $q_i$ die Häufigkeit dieses Buchstabens im zu untersuchenden Kryptogramm. Im obigen Beispiel wäre also

$$q_{10} = q_{17} = q_{21} = \frac{3}{21} = \frac{1}{7}\,,$$

da J, Q und U jeweils dreimal vorkommen;

$$q_4 = q_8 = q_9 = q_{12} = q_{16} = \frac{2}{21}\,,$$

da D, H, I, L und P je zweimal vorkommen,

$$q_{17} = q_{24} = \frac{1}{21}$$

für die beiden einfach vorhandenen Buchstaben R und X, und alle übrigen $q_i$ verschwinden.

Dazu kann man die Zahlen

$$r_n = \sum_{i=1}^{26} p_i q_{i\oplus n}$$

betrachten, wobei $i \oplus n$ die auf den Bereich von 1 bis 26 normierte Summe von $i$ und $k$ modulo 26 ist, d.h.

$$i \oplus n = 1 + \big((i + n - 1) \bmod 26\big)\,.$$

Um die Bedeutung der $r_n$ zu verstehen, betrachten wir zunächst

$$r_0 = \sum_{i=1}^{26} p_i q_i\,.$$

Da $(p_i - q_i)^2 = p_i^2 + q_i^2 - 2p_iq_i$ nicht negativ sein kann, ist

$$p_iq_i \le \frac{1}{2}(p_i^2 + q_i^2) \quad \text{und} \quad r_0 = \sum_{i=1}^{26} p_iq_i \le \frac{1}{2}\sum_{i=1}^{26}(p_i^2 + q_i^2)\,.$$

Die beiden Seiten dieser Ungleichung unterscheiden sich dabei umso mehr, je größer $\sum(p_i - q_i)^2$ ist, je mehr sich also die beiden Häufigkeitsverteilungen unterscheiden. Die Ungleichung wird genau dann zur Gleichung, wenn $p_i = q_i$ ist für alle $i$, wenn also die Buchstaben im Kryptogramm dieselbe Häufigkeitsverteilung haben wie die Buchstaben der deutschen Sprache.

Für die übrigen $r_n$ läßt sich genauso argumentieren: Da die Summe der $p_i^2$ nicht von der Reihenfolge der Summanden abhängt, ist auch

$$r_n = \sum_{i=1}^{26} p_iq_{i\oplus n} \le \frac{1}{2}\sum_{i=1}^{26}(p_i^2 + q_{i\oplus n}^2) = \frac{1}{2}\sum_{i=1}^{26}(p_i^2 + q_i^2)$$

mit Gleichheit genau dann, wenn $p_i = q_{i\oplus n}$ ist für alle $i$.

Wir erwarten also einen großen Wert von $r_n$, wenn die um $n$ verschobene Häufigkeitsverteilung des Kryptogramms gut mit der Häufigkeitsverteilung der deutschen Sprache übereinstimmt oder, anders ausgedrückt, wenn das à la CAESAR um $n$ Buchstaben nach vorne verschobene Kryptogramm eine Häufigkeitsverteilung hat, die gut mit der deutscher Texte übereinstimmt.

Im vorliegenden Fall sind die größten Werte

$$r_{16} = 0{,}0707, \quad r_3 = 0{,}0661, \quad r_{12} = 0{,}0541 \quad \text{und} \quad r_7 = 0{,}0538\,.$$

Der erste Versuch mit $n = 16$ führt, wie wir bereits wissen, zu keinem vernünftigen Ergebnis; der zweite mit $n = 3$ entschlüsselt aber das Kryptogramm.

## §4: Vigenère-Chiffren

CAESAR-Substitutionen sind so einfach zu entschlüsseln, weil es erstens nur 26 Verschlüsselungsmöglichkeiten gibt und zweitens die Häufigkeitsverteilung der Buchstaben viele Anhaltspunkte gibt. Zur Verbes-



serung des Verfahrens muß also erstens die Anzahl der Verschlüsselungsmöglichkeiten deutlich erhöht werden, und zweitens sollte man versuchen, die Häufigkeitsverteilung der Buchstaben im Chiffretext möglichst homogen zu machen.

Beide Ziele erreicht die VIGENÈRE-Chiffre, indem sie mehrere CAESAR-Substitutionen miteinander mischt: Schlüssel ist eine (nicht zu kurze) Buchstabenfolge. Diese schreibt man so oft hintereinander, bis eine Buchstabenfolge entsteht, die mindestens genauso lang ist wie der zu verschlüsselnde Text, und man verschlüsselt den $i$-ten Klartextbuchstaben dadurch, daß man ihn zum $i$-ten Buchstaben der aus dem Schlüssel entstandenen Buchstabenfolge „addiert". Bei dieser „Addition" wird A mit 1, B mit 2 identifiziert *usw.*, und ein Additionsergebnis größer 26 wird durch Subtraktion von 26 wieder in den richtigen Bereich gebracht. Beispielsweise ist also

$$A + A = B, \quad B + B = D, \quad A + Y = Z \quad \text{und} \quad B + Y = A\,.$$

Verschlüsselung von `Angriff im Morgengrauen` mit dem Schlüssel `Krypto` führt somit zu

```
  ANGRI FFIMM ORGEN GRAUE N
+ KRYPT OKRYP TOKRY PTOKR Y
= LFFHC UQALC IGRWM WLPFW M
```

VIGENÈRE-Chiffren sind 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.

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.

Für einen Schlüssel mit nur einem Buchstaben haben wir einfach eine CAESAR-Substitution, die praktisch keinerlei Schutz bietet – es sei denn, die Nachricht sei nur einen Buchstaben lang. In diesem Fall bietet die CAESAR-Chiffre perfekte Sicherheit gegen eine Attacke nur mit Chiffretext, denn genau wie der Schlüsselbuchstabe kann auch der Klartextbuchtabe beliebig sein.

Genauso bietet die VIGENÈRE-Chiffre absolute Sicherheit gegen einen Angriff nur mit Chiffretext, wenn der Schlüssel mindestens genauso lang ist wie die Nachricht und zufällig gewählt wird. Verschlüsseln wir die wohlbekannte Nachricht mit einem zufällig gewählten Schlüssel der Länge 21, so erhalten wir etwa

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

Nun könnte möglicherweise ein Gegner irgendwie den Schlüssel erraten und damit auf den Klartext stoßen. Das nützt ihm aber nur wenig, denn es gibt noch viele andere Klartexte, die zu diesem Kryptogramm passen: Wenn er andere Schlüssel ausprobiert, erhält er beispielsweise auch die Entschlüsselung

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

und entsprechend läßt sich auch jeder andere Klartext mit 21 Buchstaben produzieren; er kann aus dem Kryptogramm also keine weitere



Information ziehen, als daß der Klartext 21 Buchstaben lang war. Wenn in einem Kryptosystem der Schlüssel mindestens genauso komplex ist wie der Chiffretext, läßt sich durch geschickte Wahl von Schlüssel und Verschlüsselungsverfahren fast jede Botschaft in die Chiffre hineinlesen – was von Personen, die Nachrichten von Außerirdischen bekommen und/oder Verschwörungstheorien beweisen, bewußt oder unbewußt durchaus auch ausgenutzt wird.

Für die Sicherheit eines VIGENÈRE-Verfahrens mit Schlüsserllänge = Nachrichtenlänge ist ganz wesentlich, daß der Schlüssel *zufällig* gewählt wird; er darf auf keinen Fall beispielsweise ein Buch sein, denn die „Summe“ (im VIGENÈRE-Sinne) zweier deutscher Klartexte hat noch so viele statistische Gesetzmäßigkeiten, daß die beiden Summanden relativ einfach ermittelt werden können.

Das gerade vorgestellte perfekt sichere Verfahren bezeichnet man in der Kryptographie als *one time pad,* da es früher meist mit Schreibblöcken realisiert wurde: Man erzeugte durch einen Zufallsprozeß Buchstabenfolgen und notierte diese auf zwei Schreibblöcken in gleicher Weise, bis die Blöcke voll waren. Diese beiden Blöcke gingen dann an die beiden Kommunikationspartner, die für ihre $i$-te Nachricht die Buchstabenfolge auf Blatt $i$ des Blocks als Schlüssel benutzten; danach wurde dieses Blatt vernichtet. Als erstes benutzte diese Technik wohl der amerikanische General VERNAM im ersten Weltkrieg; im zweiten Weltkrieg kommunizierte London auf diese Weise mit der französischen *Résistance,* und später sicherten auch 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. Dabei handelte es natürlich nicht wirklich um kein Telephon, über das sich etwa Carter und Breschnjew in einen russisch-englischen Kauderwelsch unterhalten konnten, sondern um eine Telexverbindung mit je einem russischen und einem englischen Fernschreiber an den beiden Endpunkten. Zur Sicherung dieser Verbindung dienten Magnetbänder mit Zufallsbits, die über Kuriere ausgetauscht und genau nach Art des *one time pad* benutzt wurden.

In alltäglicheren Anwendungen ist ein solcher Aufwand meist zu hoch; man muß notgedrungen mit Schlüsseln arbeiten, die deutlich kürzer sind als die (Summe der) Nachrichten, die damit verschlüsselt werden müssen. Wir wollen sehen, wie sich das VIGENÈRE-System dann verhält.

Betrachten wir ein Beispiel-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
```

Abbildung drei 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.

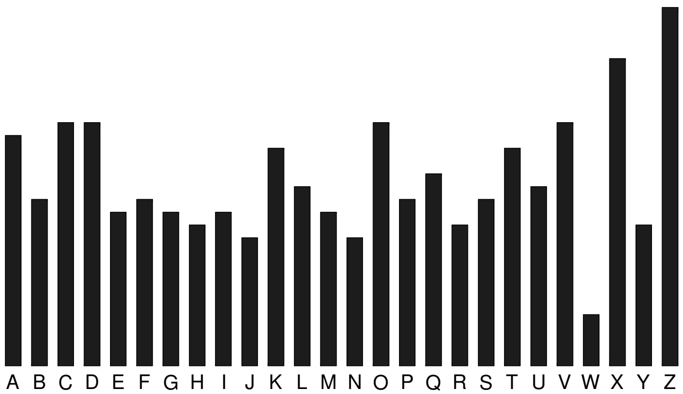



Abb. 3: Buchstabenhäufigkeiten des Beispielkryptogramms



Es gibt polyalphabetische Substitutionen wie etwa den *one time pad,* die ohne Kenntnis von Verfahren und Schlüssel nicht entschlüsselbar sind. Erfolg wird der Kryptanalyst immer dann haben, wenn er die Buchstaben identifizieren kann, die mit derselben Permutation verschlüsselt wurden, und wenn er für (fast) jede Permutation genügend viele Buchstaben hat, um statistische Methoden anzuwenden.

Am einfachsten ist dies bei *periodischen* Verfahren, d.h. wenn die anzuwendende Permutation so wie beim VIGENÈRE-Verfahren periodisch wechselt. In diesem Fall muß er nur die Periode feststellen, um die mit derselben Permutation verschlüsselten Buchstaben zusammenzufassen.

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 als 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 dort 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 in typischem 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 zufallsverteilt 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

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

die beiden Werte unterscheiden sich also deutlich.

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.

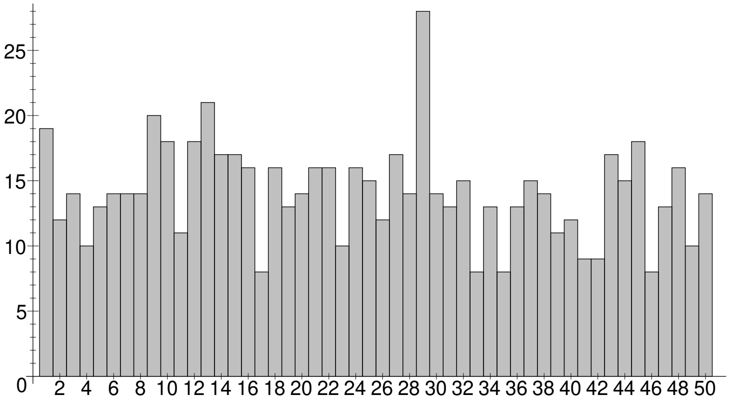



Abb. 4: Koinzidenzen im Beispielkryptogramm für Abstände bis 50

Für das obige Kryptogramm zeigt Abbildung vier die Verteilung der Koinzidenzen bei $n$ Buchstaben Abstand. Das erste Maximum bei $n = 1$ ist natürlich nicht ernstzunehmen; sonst hätten wir eine monoalphabetische Substitution, deren Häufigkeitsverteilung deutlich anders aussieht als Abbildung drei. 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. Abbildung fünf zeigt die entsprechenden Abstände bis einschließlich 120, und man sieht doch recht deutliche Ausschläge

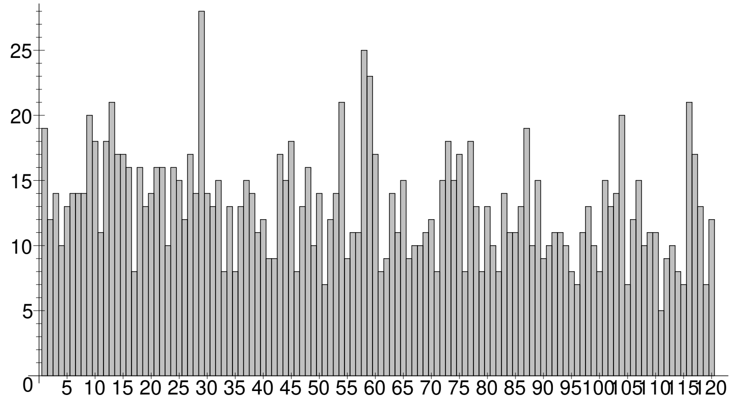



Abb. 5: Koinzidenzen im Beispielkryptogramm für Abstände bis 120

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 ungefähr so aussehen wie die von CAESAR-Substitutionen, würde er VIGENÈRE als wahrscheinlichste Möglichkeit ansehen.

In erster Näherung kann man CAESAR-Chiffren über den $pq$-Test entschlüsseln; wendet man dies hier an, erhält man 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 Buchstaben 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 Entschlüsselungsungsversuch 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
```

Für dieses Kryptogramm zeigt Abbildung sechs die Verteilung der Koinzidenzen bei $n$ Buchstaben Abstand. Die Spitze bei $n = 6$ kann ignoriert werden, da es bei $n = 12$ und $n = 24$ keine Maxima gibt; die Maxima bei $n = 16$, 32 und 48 lassen eine Periodenlänge von 16 als wahrscheinlich



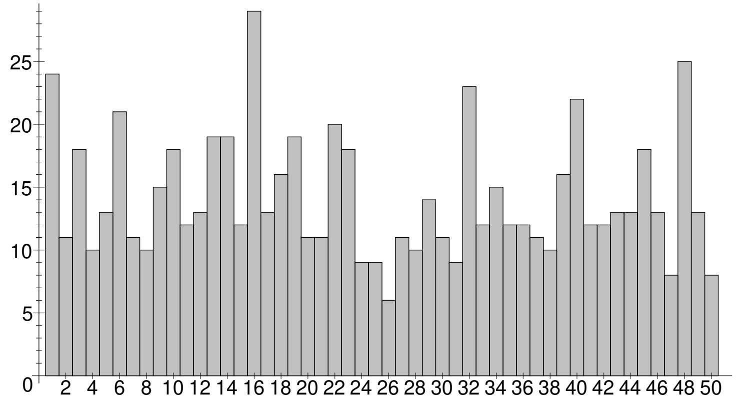



Abb. 6: Koinzidenzen bei kürzerer Schlüssellänge

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):

> *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.

Bei VIGENÈRE-Chiffren muß man realistischerweise davon ausgehen, daß sie bereits anfangen unsicher zu werden, wenn der Klartext nur etwa 30% länger ist als der Schlüssel; wirklich sicher sind sie also praktisch nur in Form des *one time pad.* Wenn man nun 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.



## §5: Substitutionschiffren

Die Entschlüsselung von CAESAR- und VIGENÈRE-Chiffren wird vor allem dadurch erleichtert, daß es pro Alphabet nur 26 mögliche Schlüssel gibt. Selbstverständlich gibt es keinen Grund, warum man sich auf zyklische Permutationen der Buchstaben des Alphabets beschränken muß; betrachtet man stattdessen beliebige Permutationen, gibt es

$$26! = 403291461126605635584000000$$

Möglichkeiten, also 403 Quadrillionen 291 Trilliarden 461 Trillionen 126 Billiarden 605 Billionen 635 Milliarden und 584 Millionen. Wie schon GIROLOMAO CARDANO (1501–1576, berühmt unter anderem durch die Lösungsformel für die kubische Gleichung, an der er völlig unschuldig ist) bemerkte, ist diese Zahl so groß, daß selbst viele Bücher nicht ausreichten, um alle diese Permutationen zu fassen.

Heutzutage benutzen wir Computer zur Entschlüsselung, und die können deutlich mehr Information verarbeiten als selbst „viele" Bücher zu fassen vermögen. Ein Computer mit einer Taktfrequenz von 1 GHz führt pro Sekunde eine Milliarde Zykeln aus, pro Stunde also 3,6 Billionen, pro Tag 86,4 Billionen, pro Jahr 31,5 Billiarden. Das Alter unseres Universums wird auf etwa zehn Milliarden Jahre geschätzt; falls ein solcher Computer diese ganze Zeit hindurch gerechnet hätte, hätte er also mittlerweile etwa 300 Quadrillionen Zykeln durchlaufen. Selbst unter der (völlig unrealistischen) Annahme, daß der Computer pro Zykluszeit einen Schlüssel durchprobieren kann, hätte er bis heute noch nicht alle Möglichkeiten erschöpft.

Da Kryptanalytiker nur selten alle möglichen Schlüssel durchprobieren, ist diese Rechnung natürlich ohne jegliche Relevanz für die Sicherheit der Chiffre; in der Tat lassen sich schon relativ kurze Kryptogramme mit Chiffretext allein entschlüsseln.

Dazu reichen nun allerdings nicht mehr die Häufigkeitsverteilungen der einzelnen Buchstaben, denn dazu gibt es doch zu viele Buchstaben mit sehr ähnlicher Häufigkeit. Abbildung sieben zeigt die bereits aus Abbildung eins bekannte Häufigkeitsverteilung bei Anordnung der

Buchstaben nach abnehmender Häufigkeit; wie man sieht, sind die Unterschiede zwischen benachbarten Buchstaben oft so minimal, daß eine sichere Identifikation allein aufgrund dieser Unterschiede praktisch ausgeschlossen ist.

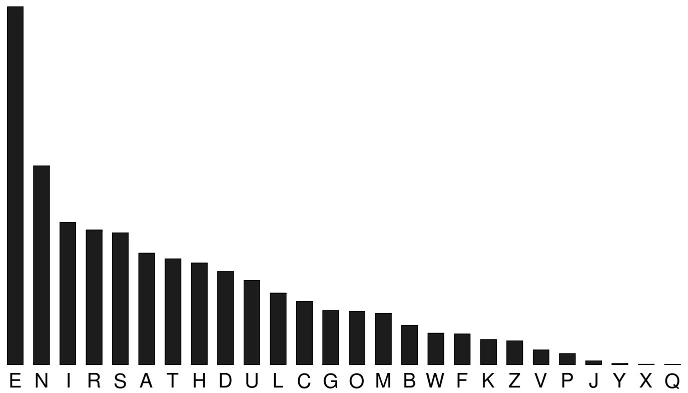



Abb. 7: Nochmals die Buchstabenhäufigkeiten in deutschem Text

Ein Buchstabe ist aber relativ gut charakterisiert durch seine Nachbarschaft: Auf ein „C“ folgt praktisch immer ein „H“, ein „D“ vor einem „E“ ist sehr viel häufiger als eines danach *usw.* Abbildung acht zeigt die in der Tat stark schwankenden Häufigkeiten von Buchstabenpaaren; die Achsen sind dabei der Einfachheit halber mit den Rangummern der Buchstabens im Alphabet beschriftet.

Das mag nicht besonders nützlich erscheinen, da schließlich die Nachbarn eines Buchstaben genausowenig bekannt sind wie der Buchstabe selbst. Bekannt ist aber die Rangordnung der Buchstaben in der Häufigkeitsverteilung des Kryptogramms, und diese sollte sich nur wenig von der der zugehörigen Klartextbuchstaben in einem Referenztext unterscheiden. Trägt man daher für jeden Buchstaben nach oben Balken auf, die angeben, wie oft der häufigste, zweithäufigste usw. Buchstabe *hinter* ihm steht und nach unten entsprechende Balken für die Buchstaben vor ihm, erhält man ein Doppelhistogramm, das sich zumindest für einiger-



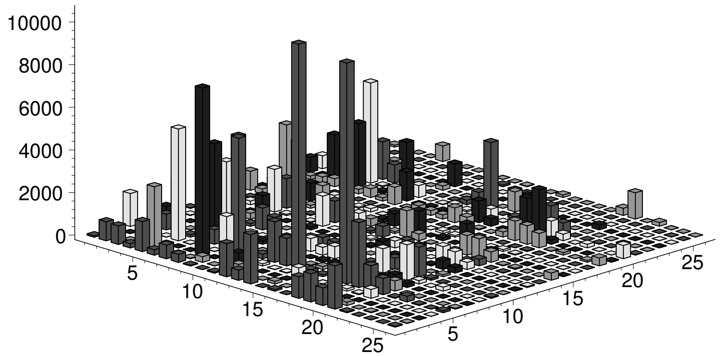



Abb. 8: Häufigkeiten von Buchstabenpaaren in deutschem Text

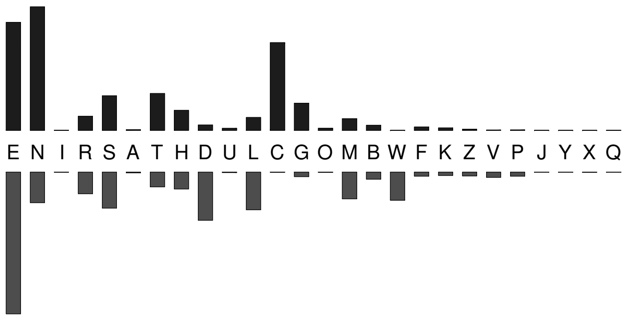
maßen häufige Buchstaben nicht allzusehr unterscheiden sollte von dem des zugehörigen Klartextbuchstabens in einem Referenztext.

Für die praktische Anwendung dieser Idee brauchen wir zunächst die entsprechenden Doppelhistogramme für deutschen Klartext. Im folgenden sind für die 26 Buchstaben des Alphabets, nach Häufigkeit geordnet, die entsprechenden Kontaktdiagramme für *Dr. Katzenbergers Badereise* aufgezeichnet:

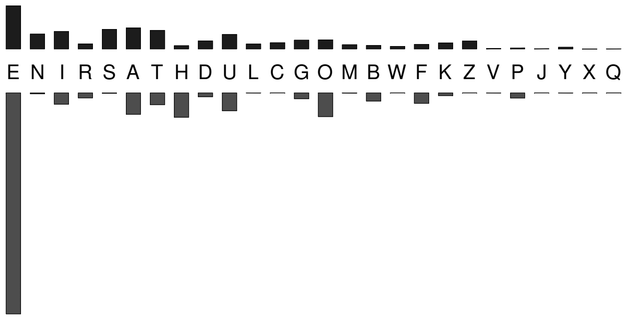
Kontaktdiagramm des Buchstabens E

Kontaktdiagramm des Buchstabens N

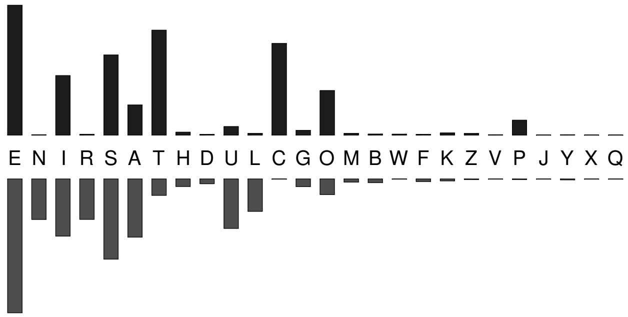

Kontaktdiagramm des Buchstabens I




Kontaktdiagramm des Buchstabens R



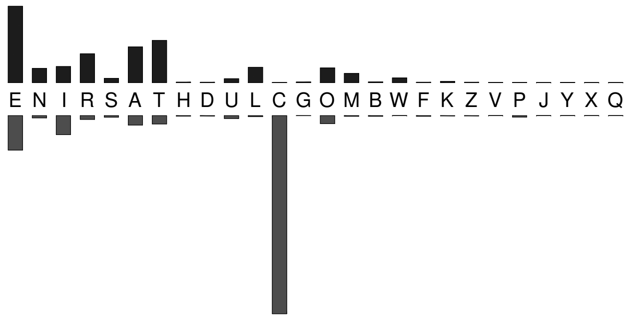

Kontaktdiagramm des Buchstabens S



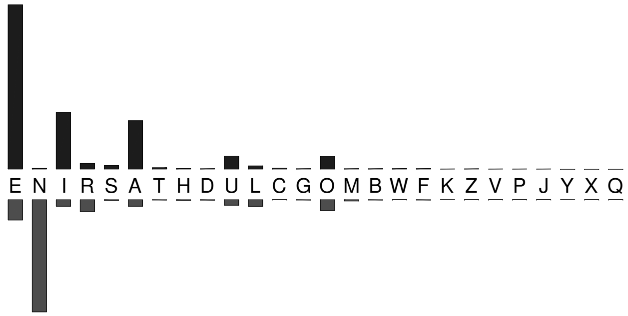

Kontaktdiagramm des Buchstabens A


Kontaktdiagramm des Buchstabens T

Kontaktdiagramm des Buchstabens H




Kontaktdiagramm des Buchstabens D

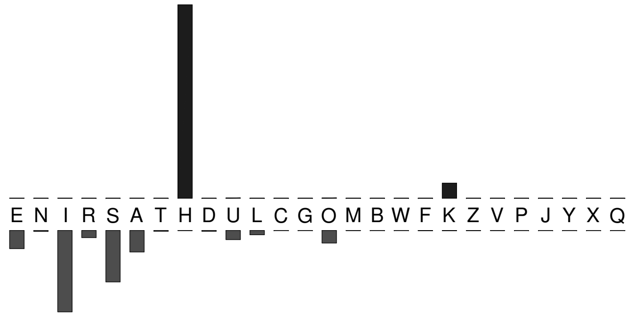



Kontaktdiagramm des Buchstabens U

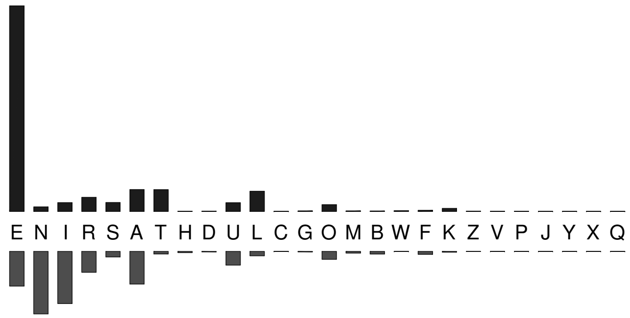



Kontaktdiagramm des Buchstabens L


Kontaktdiagramm des Buchstabens C

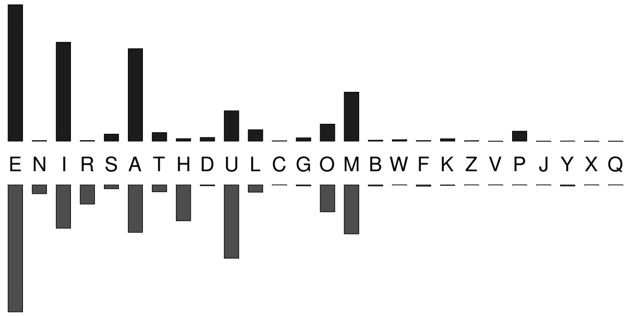



Kontaktdiagramm des Buchstabens G



Kontaktdiagramm des Buchstabens O

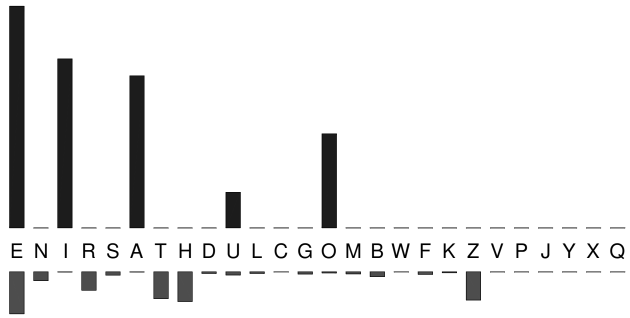

Kontaktdiagramm des Buchstabens M



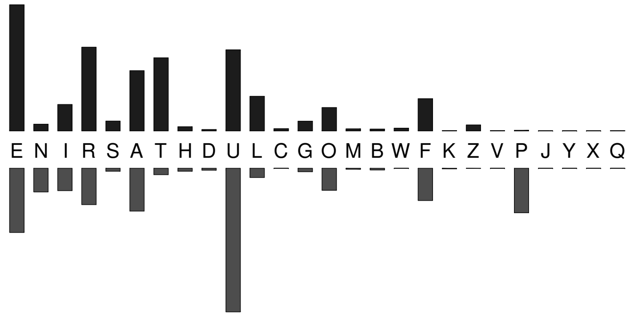

Kontaktdiagramm des Buchstabens B



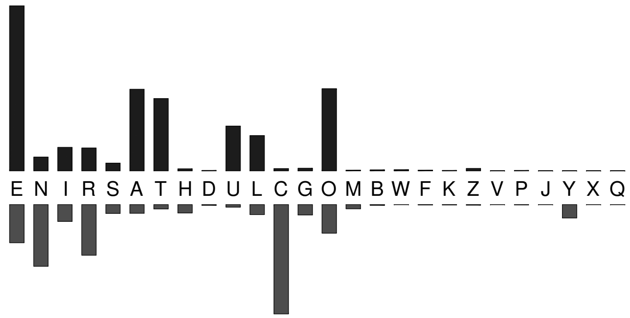

Kontaktdiagramm des Buchstabens W



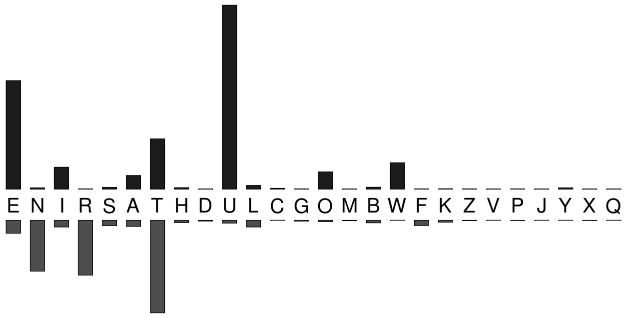

Kontaktdiagramm des Buchstabens F



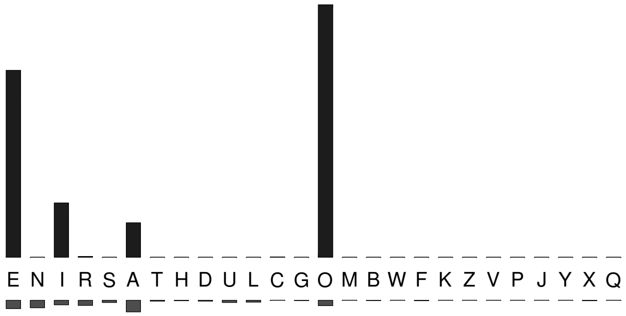

Kontaktdiagramm des Buchstabens K



Kontaktdiagramm des Buchstabens Z



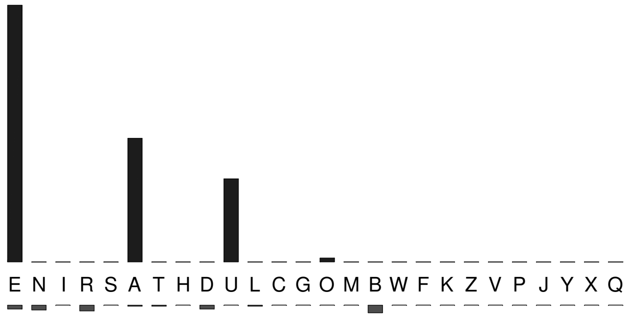

Kontaktdiagramm des Buchstabens V



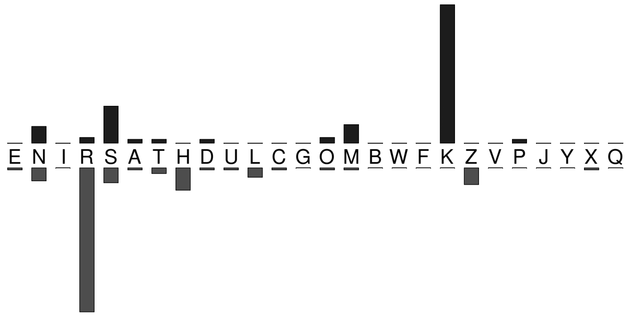

Kontaktdiagramm des Buchstabens P



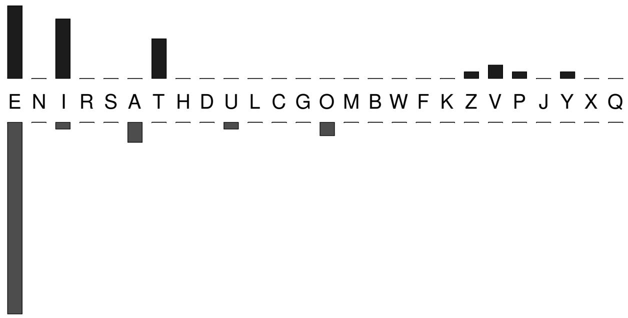

Kontaktdiagramm des Buchstabens J




Kontaktdiagramm des Buchstabens Y




Kontaktdiagramm des Buchstabens X



Kontaktdiagramm des Buchstabens Q


Wie man sieht, verhalten sich die Buchstaben in der Tat sehr verschieden: „E“ tritt besonders gerne mit „N“ und “R“ auf, wobei beide eher hinter als vor dem „E“ stehen; „R“ hat in der überwiegenden Anzahl der Fälle ein „E“ vor sich, während die Buchstaben dahinter sehr viel gleichmäßiger verteilt sind; „S“ tritt vor und nach fast allen der häufigen Buchstaben auf, und auch „A“ hat unter diesen nur wenig ausgeprägte Präferenzen; „C“ und „H“ charakterisieren sich gegenseitig wegen der häufigen „CH“ und „SCH“ und so weiter.

Aufgrund seiner Erfahrung mit der deutschen Sprache wird wohl jeder leicht Erklärungen für die genannten Phänomene finden; etwas seltsam erscheint aber, daß beispielsweise „Y“ vorzugsweise auf „R“ folgt und ein „K“ hinter sich stehen hat. Dies ist in der Tat im Bereich der Zufallsschwankungen und nur mit Kenntnis des ausgezählten Textes verständlich: Der Zweck von Dr. Katzenbergers Badereise bestand darin, den Rezensenten seines Buchs zu verprügeln, einen Badearzt namens Strick, der sich St*ryk*ius nennt.

Für seltene Buchstaben sind obige Diagramme also mit Vorsicht zu genießen, wenn auch natürlich der allerseltenste Buchstabe „Q“ in der Tat durch das nachfolgende „U“ charakterisiert ist.

Für die häufigeren Buchstaben aber sind die Diagramme eine gute Charakterisierung; man muß nur bedenken, daß es bei einem kurzen Kryptogramm natürlich Schwankungen nicht nur in der relativen Höhe



der Balken gibt, sondern auch bei der Rangfolge: Wie wir bereits bei den CAESAR-Chiffren gesehen haben, ist „E“ nicht immer der häufigste Buchstabe.

Betrachten wir als Beispiel das Kryptogramm

```
KAPUY WOYEU NAWON QANSW HEUIE ONGQE BBHWQ OWOON
EEUIW HEGQH LOVFM EOFWY NGQUE BGALG QNFME OFWYV
EKAOI WHWOQ EUHMA TCEOT WNHLO VNAYE CQEUL HAHIE
OYWLH WQIHC EODLW UWHEO HWEIW HIACV NMWCT LOYUE
CYEGH PUSCA GEUIE VWAUE HECEW EIHEY NEQEO HAOKA
PUOLW YLANN MAOYE OYIEG QHKAO NEEUF LNEEU IWHWQ
IQABN HEEHL OVHCA NNKAI QWUBN ECFHL WNFLI ECFHN
HWBHL WYLAN NHCWO THIWU GQLOV NSCWG QHVEF LYCWO
NHLOV UEGQH IEOGQ IEUVE OOLOV MEOOY WOYLW YLANN
LWNFL CNGQU EBNHE HHKAO IIEOO MAHAV TCEOT MECLO
VELBI LUHWI AUAGQ SBWBB
```

Die Häufigkeitsverteilung der Buchstaben ist in Abbildung neun zu sehen; sie ist offensichtlich nicht einfach eine verschobene Häufigkeitsverteilung für Klartext, wie wir sie bei CAESAR-Substitutionen hatten.

Abbildung zehn zeigt aber, daß zumindest die sortierte Häufigkeitsverteilung doch recht gut der in Abbildung sieben dargestellten entspricht; man kann also davon ausgehen, daß es sich bei Kryptogramm um eine monoalphabetische Substitution handelt.

Eine Probeentschlüsselung einfach aufgrund des Vergleichs von Abbildung zehn mit Abbildung sieben führt auf folgenden Text:

```
ZAPTU RNUET SARNS DASVR IETHE NSGDE MMIRD NRNNS
EETHR IEGDI CNOWF ENWRU SGDTE MGACG DSWFE NWRUO
EZANH RIRND ETIFA KLENK RSICN OSAUE LDETB IAIHE
NURBI RDHIL ENJCR TRIEN IREHR IHALO SFRLK CNUTE
LUEGI PTVLA GETHE ORATE IELER EHIEU SEDEN IANZA
PTNBR UBASS FANUE NUHEG DIZAN SEETW CSEET HRIRD
HDAMS IEEIC NOILA SSZAH DRTMS ELWIB RSWCH ELWIS
IRMIB RUBAS SILRN KIHRT GDCNO SVLRG DIOEW CULRN
```

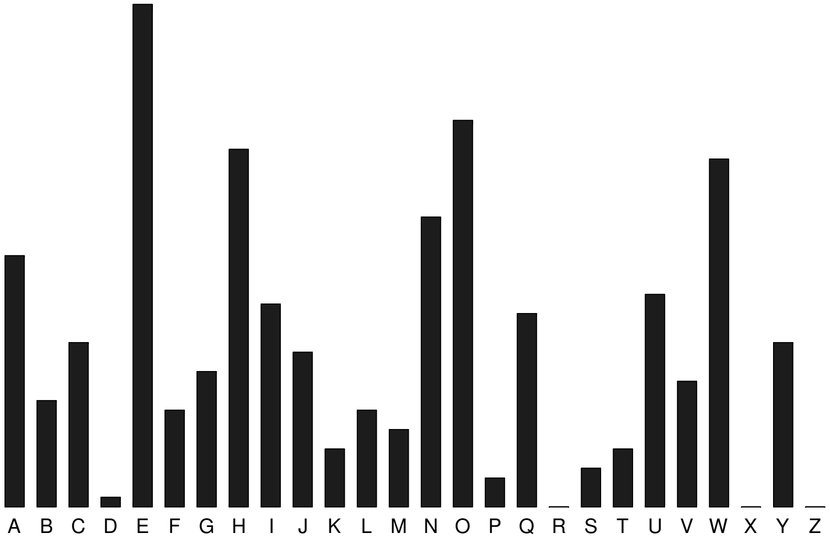



Abb. 9: Buchstabenhäufigkeiten im Kryptogramm

```
SICNO TEGDI HENGD HETOE NNCNO FENNU RNUBR UBASS
BRSWC LSGDT EMSIE IIZAN HHENN FAIAO KLENK FELCN
OECMH CTIRH ATAGD VMRMM
```

Das ist ganz offensichtlich kein deutscher Klartext, und auch mit viel Sprachgefühl fällt es schwer, hier Kandidaten für deutsche Wörter zu sehen. Wir müssen also, zumindest für die häufigeren Buchstaben, deren Vorgänger und Nachfolger betrachten.

Der häufigste Buchstabe ist, wie wir das auch von deutschem Klartext gewohnt sind, das „E“; betrachten wir sein Kontaktdiagramm:



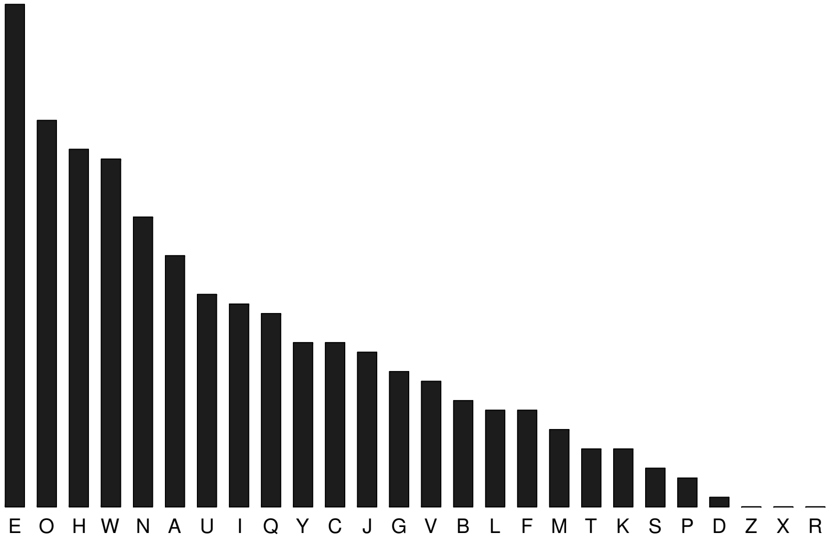
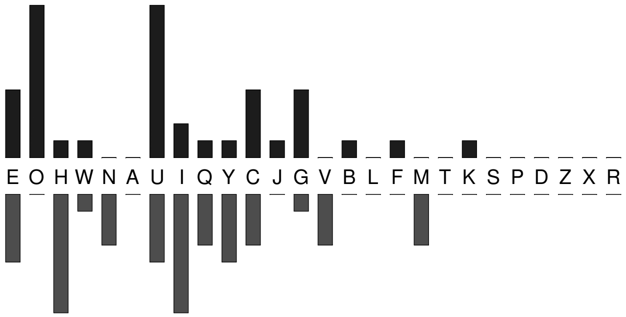



Abb. 10: Sortierte Buchstabenhäufigkeiten im Kryptogramm

Kontaktdiagramm des Buchstabens E




Schon ein flüchtiger Blick zeigt, daß sich dieses Diagramm in wesentlichen Punkten von dem des Klartext- „E“ unterscheidet, denn dieses kommt nur selten verdoppelt vor, und hat in erster Linie sehr häufige Buchstaben hinter sich stehen.

Auch um ein „N“ kann es sich nicht handeln, denn dazu gibt es zu viele Balken nach unten. „I“ ist unmöglich, da Doppel-I praktisch nicht vorkommt, „R“ mit seinem sehr charakteristischen Kontaktdiagramm ist es auch nicht, ebensowenig „S“, das sowohl vor als auch nach sich eher etwas häufigere Buchstaben bevorzugt.

Die Übereinstimmung mit „A“ ist ebenfalls alles andere als optimal, aber immerhin kommt „A“ etwas häufiger als die anderen Vokale verdoppelt vor und hat sehr häufig den zweithäufigsten Buchstaben sowie Buchstaben mittlerer Häufigkeit nach sich bei wenig ausgeprägten Präferenzen für die Vorgänger.

Die Diagramme der nächsten einigermaßen häufigen Buchstaben in deutschem Klartext passen wieder deutlich schlechter. Da der häufigste Buchstabe des Kryptogramms wohl schon unter den zehn häufigsten Buchstaben der deutschen Sprache zu finden sein sollte und mit relativ hoher Wahrscheinlichkeit ein Vokal sein dürfte, spricht alles in allem gesehen doch vieles für „A“. Zumindest falls der Chiffrebuchstabe „O“ für „N“ stehen sollte, paßt das Diagramm einigermaßen zum „A“.

Die (vorläufig) endgültige Entscheidung sollten wir also erst fällen, wenn wir auch das Kontaktdiagramm des zweithäufigsten Chiffrebuchstabens „O“ betrachtet haben.

Kontaktdiagramm des Buchstabens O


Dieses Diagramm entspricht in der Tat recht gut dem des Klartextbuchstabens „N“; es spricht also vieles für die Entschlüsselung

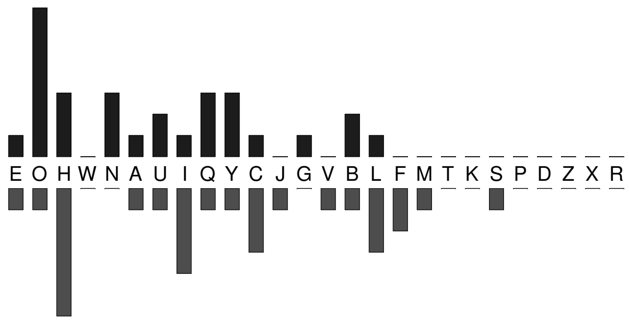
$$E \to A, \quad O \to N\,.$$



Kontaktdiagramm des Buchstabens H

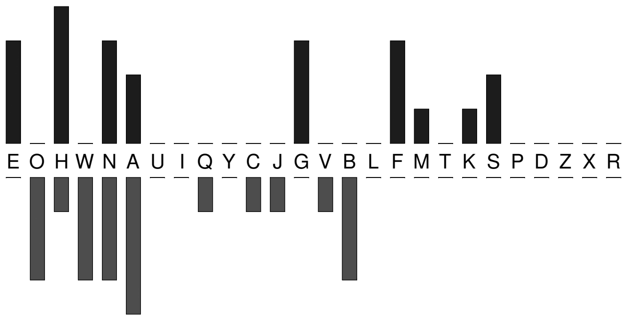

Der dritthäufigste Kryptogrammbuchstabe „H“ tritt sehr häufig verdoppelt auf, kann also kein „E“, „I“ oder „R“ sein. „S“ und „T“ sind Möglichkeiten, wobei „S“ auf den ersten Blick optisch günstiger aussieht. Bedenkt man aber, daß der erste Balken bei den einen Diagrammen für „E“ und bei den anderen für „A“ stehen, spricht doch vieles eher für „T“: „S“ steht häufiger nach „A“ als davor, bei „T“ ist es eher umgekehrt; auch die relativ scharf ausgeprägten oberen Balken beim „S“ passen nicht richtig. Gegen „T“ spricht vor allem der auffällige „E“-Balken, aber da wir bereits beim dritthäufigsten Buchstaben des Kryptogramms sind und noch immer kein „E“ gefunden haben, ist dieses hier wohl doch etwas unterrepräsentiert.

Kontaktdiagramm des Buchstabens W



Das „W“ im Kryptogramm hat gerne ein Klartext-„N“ nach sich und ein Klartext-„T“ sowohl vor als auch nach sich, tritt praktisch nie verdoppelt auf und hat einen Buchstaben mittlerer Häufigkeit, der oft danach steht; all dies spricht sehr für ein „I“.

Kontaktdiagramm des Buchstabens N




Der Klartextbuchstabe zu „N“ wird oft von „A“, „T“ und sich selbst gefolgt und steht oft nach „N“; damit dürfte fast klar sein, daß es sich um ein „S“ handeln muß.

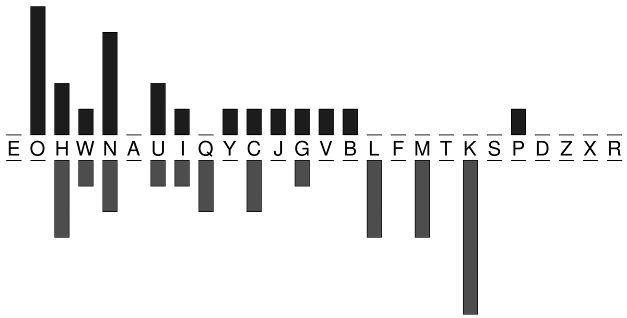
Damit haben wir mit

$$E \to A, \quad O \to N, \quad H \to T \quad \text{und} \quad N \to S$$

die fünf häufigsten Buchstaben des Kryptogramms entschlüsselt und können zumindest für die mit den jeweils korrekten Balken der Klartextdiagramme vergleichen.

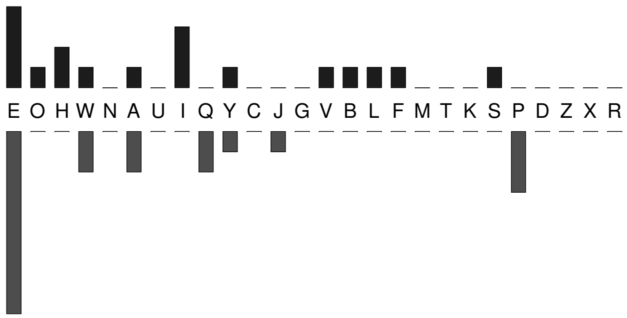


Kontaktdiagramm des Buchstabens A




Das Diagramm zum Chiffretext-„A“ gibt uns gleich zwei Identifizierungen: Ein Buchstabe, der weder mit sich selbst noch mit dem Klartext-„A“ zusammen auftritt, dafür aber häufig ein „N“ oder „S“ nach sich führt, „T“ zu beiden Seiten und einen eher seltenen Buchstaben oft vor sich, kann wohl nur „O“ sein, und das häufig davorstehende „K“ ist somit ein „V“.

Kontaktdiagramm des Buchstabens U




„U“ wird entschlüsselt durch einen Buchstaben, der die bereits bekannten Vokale gerne sowohl vor als auch hinter sich hat, sich selbst allerdings praktisch nie, die Konsonanten „N“, „S“ und „T“ stehen häufiger dahinter, aber praktisch nie davor, und eine ganze Reihe eher seltener Buchstaben steht immer wieder einmal dahinter – hier fällt eine klare Identifikation schwer.

Wir können aber schauen, ob wir mit den bereits gefundenen acht Klartextbuchstaben schon Wörter im Kryptogramm erkennen können:

```
KAPUY WOYEU NAWON QANSW HEUIE ONGQE BBHWQ OWOON
VO    IN A  SOINS  OS I TA  A NS  A   TI  NINNS
EEUIW HEGQH JOVFM EOFWY NGQUE BGAJG QNFME OFWYV
AA  I TA  T  N    AN I  S   A   O    S  A N I
EKAOI WHWOQ EUHMA TCEOT WNHJO VNAYE CQEUL HAHIE
AVON  ITIN  A T O   AN  IST N  SO A   A   TOT A
OYWLH WQIHC EODJW UWHEO HWEIW HIACV NMWCT JOYUE
N I T I  T  AN  I  ITAN TIA I T O   S I    N  A
CYEGH PUSCA GEUIE VWAUE HECEW EIHEY NEQEO HAOKA
A T     O  A  A  IO A TA AI A TA  SA AN TONVO
PUOLW YLANN MAOYE OYIEG QHKAO NEEUF JNEEU IWHWQ
N I   OSS  ON A N  A   TVON SAA    SAA   ITI
IQABN HEEHJ OVHCA NNKAI QWUBN ECFHL WNFJI ECFHN
O S TAAT  N T O SSVO   I  S A  T  IS    A  TS
HWBHL WYLAN NHCWO THIWU GQJOV NSCWG QHVEF JYCWO
TI T  I  OS ST IN  T I     N  S  I   T A     IN
NHJOV UEGQH IEOGQ IEUVE OOJOV MEOOY WOYLW YLANN
ST N   A  T  AN    A  A NN N   ANN  IN  I   OSS
LWNFJ CNGQU EBNHE HHKAO IIEOO MAHAV TCEOT MECJO
IS    S    A STA TTVON   ANN  OTO    AN   A  N
VEJBI JUHWI AUAGQ SBWBB
A      TI  O O     I
```

Gleich in der ersten Zeile springt „a-so ins“ ins Auge, also steht wohl „U“ für „L“. Damit wir das nächste Wort zu „-os-ital“, d.h. „Q“ steht für „H“ und „S“ für „P“.

Nachdem wir zusätzlich zum schon länger bekannten „S“ auch das „H“ kennen, können wir nach „SCH“ suchen; ein Kandidat dafür ist die fünfte Gruppe der zweiten Zeile, was „G“ als Verschlüsselung von „C“ ergibt.

Damit wird der Text schon stellenweise verständlich: Das zweite Wort muß wohl „ging“ sein, vor dem „avon“ zu Beginn der dritten Zeile kann



wohl nur ein „D“ stehen und so weiter. Der Rest ist ein reines Puzzlespiel und führt zum Klartext

> *Voyl ging also ins Hospital. Man schafft ihn in'n Saal mit achtundzwanzig Schlafcouchs, …zwanzig davon mit Inhalt, wo krank ist und sogar halbtot. Man gibt ihm Tranquilitantia mit Mordswirkung (Largactyl, Procalmadiol, Atarax). Am Tag sah Anton Voyl 'n Big Boss, wo'n Gang macht von Saal zu Saal; mit ihm Hofstaat und Tross vom Hilfsarzt bis zum Arztstift; Big Boss trinkt Milch und spricht dazu, grinst und lacht manchmal … Dann und wann ging Big Boss bis zur Schlafstatt von'm Mann, wo todkrank war und aufm ultimo Loch pfiff …*

Der Text stammt aus der deutschen Übersetzung des Romans *La disparaition* von GEORGE PEREC *(Edition Denoël, Paris,* 1969*).* Dieser französische Autor zahlreicher Romane, hatte nach einem feuchtfröhlichen Abend gewettet, daß er einen Roman schreiben könne, ohne den Buchstaben „E“ zu verwenden; er gewann. Die deutsche Übersetzung von EUGEN HELMLÉ mit dem Titel *Anton Voyls Fortgang (Zweitausendeins,* $^{3}$1998*)* hält sich auch an die Vorgabe, verwendet aber Umlaute – daher die Auslassungspunkte im obigen Text.

Wie man sieht, lassen sich die einzelnen Buchstaben also selbst dann noch recht gut anhand ihrer Umgebung identifizieren, wenn ihre Rangfolge nicht mehr ganz der für deutsche Texte üblichen entspricht.

Eine hier nicht behandelte Möglichkeit, sich Buchstaben zu verschaffen, liegt auch im Erraten von Wörtern aus der Botschaft. Da man beim Abhören meist weiß, wonach man sucht, gibt es oft Wörter, die mit hoher Wahrscheinlichkeit im Text vorkommen sollten. Falls ein solches Wort einen oder gar mehrere Buchstaben mehrfach enthält, kann man im Chiffretext nach demselben Muster suchen. Ein ideales Wort wäre beispielsweise *Substitution:* Es muß von jeder monoalphabetischen Substitution in eine Folge der Art

$$XY pX ZW ZY ZW qr$$

transformiert werden, und eine Folge dieser Art kann praktisch nur für das Wort *Substitution* stehen. Damit sind sieben Buchstaben identifiziert.

## §6: Polyalphabetische Substitutionen

Als nächste Stufe der Komplexität bietet sich natürlich an, aufeinanderfolgende Buchstaben mit verschiedenen Permutationen zu verschlüsseln, d.h. also eine Art von VIGENÈRE-Verfahren mit allgemeinen monoalphabetischen Substitutionen anstelle der doch sehr speziellen CAESAR-Substitutionen.

Leider steigt dadurch, sofern die Anzahl der Permutationen nicht in die Größenordnung der Nachrichtenlänge kommt, die Sicherheit nur unwesentlich an, denn die Verfahren aus dem vorigen Paragraphen funktionieren auch für diesen Fall: Man stellt zunächst, genau wie bei den VIGENÈRE-Chiffren, die Periode fest; danach weiß man, welche Buchstaben mit derselben Permutation verschlüsselt wurden. Jede dieser Buchstabenmengen wird nach Häufigkeit geordnet, und man kann wieder Kontaktdiagramme aufstellen. Dabei muß nur beachtet werden, daß die rechten und linken Nachbarn nicht aus derselben Gruppe kommen, sondern aus je einer anderen festen Gruppe, so daß Balken nach oben und nach unten nicht mehr für denselben Buchstaben stehen müssen. Abgesehen von dieser kleinen Erschwernis ändert sich nichts, und ein erfahrener Kryptanalytiker kann ab etwa 30–40 Buchstaben pro Alphabet praktisch sicher sein, daß er die Nachricht entschlüsseln kann.

Polyalphabetische Chiffren mit sehr großen Periodenlängen erzeugten die sogenannten *Rotormaschinen,* die in den Zwanzigerjahren aufkamen und die Kryptographie des zweiten Weltkriegs dominierten. Ein Rotor ist eine Scheibe, die auf jeder Seite in gleichmäßigen Abständen 26 Kontakte hat. Jeder Kontakt auf der Vorderseite ist mit einem der Kontakte auf der Rückseite verbunden, wobei nur darauf zu achten ist, daß diese Verbindungen spiegelsymmetrisch angebracht werden, so daß die so realisierte Permutation ein Produkt von dreizehn Transpositionen ist.

Zum Einsatz der Maschine wählt man, je nach Maschine, drei bis fünf von mehreren verfügbaren Rotoren aus und setzt diese in einer festzulegenden Reihenfolge und Anfangsstellung in die Maschine ein. Der erste Buchstabe wird dann von einem festen Kontakt auf der einen Seite der Maschine durch die Rotoren in dieser Stellung an einen der 26 festen



Kontakte auf der anderen Seite der Maschine übertragen; das Chiffrierergebnis ist der diesem Kontakt zugeordnete Buchstabe. Danach bewegen sich die Rotoren automatisch in einer nicht ganz regelmäßigen Weise weiter, so daß für den nächsten Buchstaben eine andere Permutation realisiert wird und so weiter. Zur Entschlüsselung wird der Chiffretext einfach in eine identisch aufgebaute Maschine mit gleicher Anfangsstellung aber umgekehrter Rotorreihenfolge gegeben.

(Unter http://library.thinkquest.org/28005/flashed/timemachine/courseofhistory/enigmaapp.shtml kann man ausprobieren, wie das bei er deutschen Enigma-Maschine funktionierte.)

Die mit solchen Maschinen erzielbaren Perioden liegen im Bereich über 100 000; trotzdem wurden sie schon im zweiten Weltkrieg häufig geknackt.

Ein Grund dafür ist, daß während der Gültigkeitsperiode eines Schlüssels alle Nachrichten mit derselben Anfangsstellung verschlüsselt werden müssen (oder aber die Anfangsstellung auf eine kryptographisch unsichere Weise übermittelt werden muß, was sich gerade bei der deutschen Enigma als noch schlimmer herausstellte) so daß damit zumindest für die ersten Buchstaben relativ viele Chiffrebuchstaben zusammenkommen. Das weitere Problem, daraus die Anfangsstellung der Maschine zu ermitteln, lösten die Engländer im zweiten Weltkrieg mit speziell gebauten Maschinen, den sogenannten „Bomben“, die mechanisch alle möglichen Anfangsstellungen durchprobierten. Die Verdrahtung der einzelnen Rotoren war schon vor dem Krieg von den Polen geknackt worden.

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 nicht, und auch UNIX selbst verwendet beispielsweise für Paßwörter das auf einem „gesalzenen“ DES beruhende deutlich bessere `crypt(3)`-Unterprogramm.

## § 7: Literaturhinweise

Das (dickleibige) Standardwerk zur alten Kryptographie mit Schwerpunkt auf der geschichtlichen Darstellung ist

David Kahn: *The Codebreakers – the comprehensive history of secret communication from ancient time to the internet,* Scribner, New York, $^2$1996

Deutlich kürzer und billiger, mit Schwerpunkt auf der Darstellung der Verfahren, ist

Helen Fouché Gaines: *Cryptanalysis – a study of ciphers and their solution,* Dover, New York, 1956 *(Nachdruck einer Originalausgabe von 1939)*

Rotormaschinen und ihrer Kryptanalyse gewidmet ist das Buch

Cipher A. Deavours, Louis Kruh: *Machine cryptography and modern cryptanalysis,* Artech House, Dedham MA, 1985

Das Wort „modern“ bezeichnet hier den Zeitraum von etwa 1920–1970.

Ein wirklich modernes Lehrbuch mit Kapitel über klassische Kryptographie sowie über statistische und informationstheoretische Ansätze zur Kryptanalyse ist

Jan C.A. van der Lubbe: *Basic Methods of cryptography,* Cambridge University Press, 1998

Eine sehr ausführliche Darstellung der statistischen und informationstheoretischen Ansätze findet man in

Alan G. Konheim: *Cryptography: a primer,* Wiley, New York, 1981