Wissen

INFORMATIONSMATERIALIEN

Fach Mathematik

Einführung in die Kryptologie

FACH MATHEMATIK

Einführung in die Kryptologie

Inhaltsverzeichnis

EINLEITENDER ÜBERBLICK 1

Die Entwicklung der Kryptologie 1
Verschlüsselungsarten 3

DIE MONOALPHABETISCHE CHIFFRIERUNG 4

Das Verschlüsseln monoalphabetischer Texte 4
Ein praktisches Beispiel 5
Das Entschlüsseln monoalphabetischer Texte 5
Ein weiteres Beispiel 6
„Die tanzenden Männchen“ 7
Hilfsmittel zur Ver- und Entschlüsselung 9
Die Muster einer Sprache 9
Die Abhängigkeit der Häufigkeiten vom Klartext 9
Der Wortzwischenraum 9
Häufigkeiten von n-Grammen 9
Die Abhängigkeit von der Verfassersprache 10
Häufigkeitsgebirge 10
Der Koinzidenzindex (Kappa κ) einer Sprache 11
Die Verschleierung der Häufigkeiten 11
Zusammenfassung 12

DIE POLYALPHABETISCHE CHIFFRIERUNG 13

Die Idee der polyalphabetischen Chiffrierung 13
Die Theorie der Vigenère-Chiffrierung 13
Ein praktisches Beispiel 14
Die Theorie der Entschlüsselung 15
Der Kasiski - Test 15
Der Friedmann - Test 16
Die Bestimmung des Schlüsselwortes 17
Ein weiteres Beispiel 17
Zusammenfassung: 19

DIE PUBLIC-KEY-CHIFFRIERUNG 20

Das Prinzip des öffentlichen Schlüssels 20
Der RSA-Algorithmus 20
Die Erzeugung der Schlüssel 20
Die Anwendung der Schlüssel 21
Ein praktisches Beispiel 21
Die digitale Unterschrift mit RSA 22
Die Sicherheit von RSA 22

ANHANG 23

Quellenangabe 23
Lösung des Tauschchiffres 23
Übungsaufgabe monoalphabetische Chiffrierung 24
Lösung zur Übungsaufgabe 25
Das Vigenère - Quadrat 26
Der „Erweiterte Euklidische Algorithmus“ 26

Kapitel

1

Einleitender Überblick

Kryptologie ist die Kunst und Wissenschaft, Methoden zur Verheimlichung von Nachrichten zu entwickeln. Häufig wird dabei noch zwischen Kryptographie - der Wissenschaft von der Entwicklung von Kryptosystemen - und Kryptoanalyse - der Kunst des Brechens dieser Systeme - unterschieden. Die Begriffe Kryptologie und Kryptographie sind aus dem griechischen Wörtern κρυπτοσ (geheim), λογοσ (das Wort, der Sinn) und γραφειν (schreiben) gebildet.

Die Kryptologie beschäftigt sich mit der Ver- und Entschlüsselung von Informationen. Dieses Thema mutet zwar recht Antik an, wird aber in unserer zeit wieder verstärkt benötigt, denkt man nur an die Sicherheit im Internet, an Chipkarten und Paßwörter.

Aus der Geschichte der Kryptologie

Vor ungefähr 2500 Jahren verwendete die Regierung von Sparta eine trickreiche Methode zur Übermittlung geheimer Nachrichten. Sender und Empfänger mußten beide eine sogenannte Skytale haben; das waren zwei Zylinder mit denau dem gleichen Radius. Der Sender wickelte ein schmales Band aus Pergament spiralförmig um seinen Zylinder und schrieb dann der Länge nach seine Nachricht auf das Band. War nun das Band abgewickelt, konnte die Nachricht nur von einer Person gelesen werden, die einen Zylinder genau desselben Umfangs hatte.

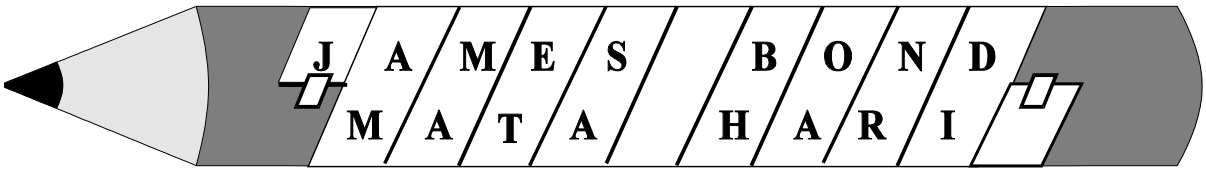


Abbildung 1: Einfache Nachbildung einer Skytale

JULIUS CAESAR verwendete darüber hinaus eine spezielle Methode der monoalphabetischen Chiffrierung - den Verschiebechiffre, d.h. er verschob die Buchstaben seine Klartextes um 3 Stellen bezüglich des Alphabetes nach links, so daß aus einem A ein d wurde:

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

Auch die **Staganographie**, eine Art von gedeckten Geheimschriften war bekannt. Diese Geheimschriften konnten entweder als unverfängliche, offen verständliche Nachricht oder in (winzigen) sichtbaren graphischen Details einer Schrift oder Zeichnung erscheinen. Letzteres bezeichnete man auch als Semagramm.

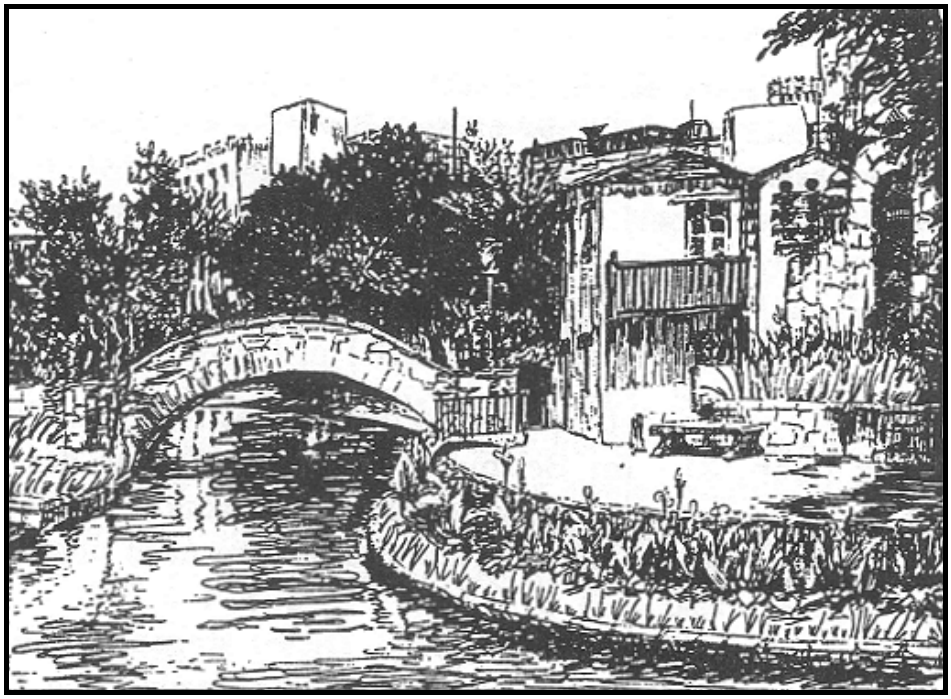


Abbildung 2: Semagramm. Die Nachricht steht im Morsecode, der aus kurzen und langen Grashalmen links von der Brücke entlang des Flusses und auf der kleinen Mauer gebildet wird. (aus Bauer)

Als eigentlicher Begründer der Kryptologie gilt L. B. ALBERTI, der 1466 erstmals den polyalphabetischen Schlüssel beschrieb. Parallel zur Weiterentwicklung der Kryptologie gab es auch Fortschritte in der Berechnung von Schlüsseln. Um 1400 gelang es den Arabern, Substitutionen zu brechen. G. B. DELLA PORTA löste erstmals einen polyalphabetischen Schlüssel. Wichtige Beiträge zur Kryptologie lieferten im 19. Jh. u.a. C. WHEATSTONE, F. BEAUFORT und FRIEDRICH W. KASISKI.

Die polyalphabetische Chiffrierung hat jahrtausendelang seine Bedeutung erhalten und wurde von den Deutschen noch im Zweiten Weltkrieg benutzt. Allerdings wurden dabei nicht Papierstreifen gegeneinander verschoben, sondern eine Maschine verwendet, in der sich Walzen mit eingravierten Buchstaben drehten. Durch Veränderung von Schaltungen, das heißt durch Neustecken von elektrischen Kontakten, konnte man den Code, also die Größe der einzelnen Verschiebungen, bei der Wahl eines jeden Buchstabens automatisch verändern. Der Empfänger besaß eine ähnliche Maschine, „ENIGMA“ genannt. (ENIGMA [griechisch.] - Rätsel, Geheimnis) Da er den jeweils verwendeten Code kannte, konnte er ihn in diese Maschine eingeben und erhielt dann durch Tippen des verschlüsselten Textes unmittelbar die entschlüsselte Botschaft. Auf deutscher Seite war man überzeugt, daß die kriegswichtigen Nachrichten vom Gegner nicht entziffert werden könnten. Man hatte sich aber getäuscht. Zwei Umstände machten es den Polen und Engländern möglich, den Code zu knacken:

1. Weil die Verschlüsselung jeweils nur durch Herstellung von verschiedenen Schaltungen erfolgte, gab es eine endliche Zahl von verschiedenen Codes. Es wurden somit nicht immer wieder neue Codes verwendet, sondern nach einiger Zeit alte nochmals eingesetzt. Durch diese Wiederholung wurden Ansatzpunkte geschaffen, die Verschlüsselung zu entziffern.

2. Die Engländer verfügten über die ersten leistungsfähigen elektronischen Rechenmaschinen. Dadurch konnten sie in Sekundenbruchteilen zahlreiche Zuordnungsmöglichkeiten ausprobiert, bis sie durch Zufall auf die richtige stießen.

Den Deutschen blieb verborgen, daß die Engländer ihre Nachrichten verstehen konnten, was nicht unwesentlich zur Entscheidung des Krieges beigetragen haben soll. Zu der Kryptoanalytikergruppe der Engländer gehörte unter anderen auch A. TURING, der nicht unwesentlichen Einfluß auf die Weiterentwicklung der noch jungen Computertechnik hatte. Wesentliche Veränderung für die Kryptologie brachte das Aufkommen des Computers mit sich. Unter dem Aspekt des Datenschutzes hat das Interesse an der Kryptologie ganz erheblich zugenommen. Andererseits bietet der Computer durch die Möglichkeit, große Datenmengen schnell analysieren zu können, auch neue Ansätze zum Brechen von Schlüsseln.

Verschlüsselungsarten

Eine sehr einfache Verschlüsselung erhalten wir, indem wir jedem Buchstaben ein festes Symbol zuordnen. Diese Verfahren heißen **monoalphabetisch**. Sie sind in der Regel, steht genügend Material zur Verfügung, leicht durch Häufigkeitsbetrachtungen zu berechnen. Wesentlich schwieriger ist es, **polyalphabetische** Geheimtexte, das sind solche, bei denen einem Buchstaben mehrere Symbole entsprechen können, zu brechen, weil hier statistische Erwägungen nicht ohne weiteres angewendet werden können. Die klassischen Verfahren haben den Nachteil, daß sich Sender und Empfänger über den zu verwendenden Schlüssel verständigen müssen, was eine zusätzliche Unsicherheit bedeutet. Dies entfällt bei den **Public-key-Systemen**, die seit 1976 entwickelt werden. Das bekannteste unter ihnen, das **RSA-Verfahren** (nach R. RIVEST, A. SHAMIR und L. ADLEMAN), verwendet die Primfaktorenzerlegung natürlicher Zahlen. Es ist nur so lange sicher, wie es keine wesentlich schnelleren Algorithmen zur Primfaktorenzerlegung gibt als die heute bekannten. Daneben setzt es die Kenntnis genügend vieler großer Primzahlen voraus.

Kapitel

2

Die monoalphabetische Chiffrierung

Monoalphabetische Chiffrierung besteht darin, das Klartextalphabet zu permutieren, d. h. die Buchstabenanordnung vertauscht wird. Unter der Annahme, daß das verwendete Alphabet 26 Buchstaben besitzt (deutsch - mit ä = ae, ö = oe, ü = ue und ß = ss), erhalten wir also
$26! = 403291461126605635584000000 \approx 4 \cdot 10^{26}$ Möglichkeiten der Anordnung der Buchstaben.

Das Verschlüsseln monoalphabetischer Texte

Zu den einfachsten Chiffren gehört die **Verschiebechiffre**, die schon von CAESAR verwendet wurde. Hierbei werden nur die Buchstaben in ihrer Reihenfolge verschoben. Einen solchen Geheimtext können wir einfach brechen, da für ein beliebiges Wort nur alle möglichen 26 Verschiebungen betrachtet werden müssen, um ein sinnvolles zu finden. Betrachten wir zum Beispiel `RBC`, so ergibt nur das Wort `ist` einen Sinn. Versuche nun den folgenden berühmten Satz zu dechiffrieren:

`LFK NDP VDK XQG VLH JWH!`[1]

Eine weitere Möglichkeit bietet die **Tauschchiffre**. Hierbei wird nicht einfach das gesamte Alphabet verschoben, sondern die Buchstaben untereinander getauscht. Mathematisch ausgedrückt heißt das: jedem Buchstabe des Klartextalphabetes wird gemäß der Reihenfolge die entsprechende natürliche Zahl zugeordnet. Multiplizieren wir den Wert eines jeden Klartextbuchstaben mit einer frei wählbaren Zahl, erhalten wir ein neues (i. allg. nicht eindeutiges) Geheimtextalphabet. Soll diese Abbildung eindeutig sein, müssen wir beachten, daß die Geheimzahl und die Anzahl der Klartextbuchstaben zueinander teilerfremd sind. Für ein Alphabet mit 26 Buchstaben sind also nur die Faktoren: 1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23 und 25 möglich. Wählen wir zum Beispiel 3 als Faktor, so entsteht folgendes Alphabet:

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

Natürlich können die beiden Verfahren auch miteinander kombiniert werden.

[1] Lösungen zu den Aufgaben sind zur **Kontrolle** im Anhang aufgeführt

Häufig wird die Methode des Schlüsselwortes verwendet, d.h. Sender und Empfänger vereinbaren ein **Schlüsselwort** und einen **Schlüsselbuchstaben**. Dies kann z.B. das fünfte Wort in der Bibel und der zweite Buchstabe des dritten Wortes sein. Somit kann die Chiffrierung jeden Tag mit anderen Voraussetzungen begonnen werden. Zur Vereinfachung wird folgendes angenommen:

Schlüsselwort: `GEHEIMSCHRIFT` Schlüsselbuchstabe: `E`

Zur Chiffrierung werden nun die im Schlüsselwort mehrfach auftretenden Buchstaben bei Wiederholung gestrichen, wir erhalten also `GEHIMSCRFT`. Dann wird der Rest des Schlüsselwortes unter das Klartextalphabet geschrieben, beginnend beim Schlüsselbuchstaben. Es folgt das Auffüllen der restlichen Alphabetbuchstaben.

Klartextalphabet: 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

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

Ein praktisches Beispiel

Wir wollen unserem Verleger die aktuelle Version des neuen Romans schicken. Da der benutzte Weg sehr unsicher ist, soll das Stück verschlüsselt werden. Als Schlüsselwort wurde „James Bond“ vereinbart, der Schlüsselbuchstabe soll das „Q“ sein. Wir entfernen zunächst das Leerzeichen aus dem Schlüsselwort und schreiben das Schlüsselwort beginnend beim Buchstaben „Q“ auf. Anschließend ergänzen wir die fehlenden Geheimtextbuchstaben, so daß keiner doppelt vorkommt:

Klartextalphabet: 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

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

Nun wandeln wir schrittweise den Klartext in den Geheimtext um, indem wir für den jeweiligen Klartextbuchstaben den darunterstehenden Geheimtextbuchstaben verwenden.

Der abgeschlossene Roman
IKA FGPKMHQVYMMKXK AYWFX

Los, hierher ihr beiden!! Wollt ihr wohl hoeren?! Hierher sag'
VYM, QRKAQKA RQA GKRIKX!! OYVVE RQA OYQV QYKAKX?! QRKAQKA MFP'

ich! Ja, so ist es brav! So - und jetzt macht ihr schoen Platz!
RHQ! TF, MY RME KM GAFB! MY - SXI TKECE WFHQE RQA MHQYKX ZVFEC!

Na bitte! Und jetzt bei Fuss! Na los! Bei Fuss hab' ich gesagt!
XF GREEK! SXI TKECE GKR LSMM! XF VYM! GKR LSMM QFG' RHQ PKMFPT!

Fuu...« »Mein Gott, Riebesehl«, stoehnt Gatti, »kannst du nicht
LSS...« »WKRX PYTT, ARKGKMKQV«, MEYKQXR PFEER, »UFXXME IS XRHQE

einmal deine Socken anziehen wie jeder andere auch??«
KRXWFV IKRXK MYHUKX FXCRKQKX ORK TKIKA FXIKAK FSHJ??«

[aus "STERN"; Hamburg; Heft 29/94 S. 74]

Das Entschlüsseln monoalphabetischer Texte

Um einen solchen Geheimtext zu entschlüsseln, müssen zwei Bedingungen erfüllt sein. Zum einen muß der Klartext in einer natürlichen Sprache verfaßt worden sein, und zum zweiten ein längeres Stück des Geheimtextes vorliegen. Die Analyse des Textes beruht auf der Häufigkeitsverteilung von Buchstaben

und Bigrammen (das sind zwei Buchstaben) in der Sprache. Für Deutsch sieht die Verteilungen wie folgt aus:

Buchstabe	Häufigkeit	Buchstabe	Häufigkeit	Bigramm	Häufigkeit
a	6,47%	n	9,84%	en	3,88%
b	1,93%	o	2,98%	er	3,75%
c	2,68%	p	0,96%	ch	2,75%
d	4,83%	q	0,02%	te	2,26%
e	17,48%	r	7,54%	de	2,00%
f	1,65%	s	6,83%	nd	1,99%
g	3,06%	t	6,13%	ei	1,88%
h	4,23%	u	4,17%	ie	1,79%
i	7,73%	v	0,94%	in	1,67%
j	0,27%	w	1,48%	es	1,52%
k	1,46%	x	0,04%		
l	3,49%	y	0,08%		
m	2,58%	z	1,14%		

Die Vorgehensweise zum Entschlüsseln ist folgende:

Man zählt die Häufigkeiten der Buchstaben im Geheimtext und findet so `e` und `n` und die Menge `{i,r,s,a,t}`. Durch Auszählen der Bigramme kann man dann `r,i,t,s,a` isolieren und schließlich über `ch` noch `c` und `h`, da das Bigramm `hc` fast nie vorkommt. Die Buchstaben `e,n,i,s,r,a,t,c` und `h` machen bereits schon rund 65% des Textes aus. Der Rest ergibt sich durch Probieren.

Ein weiteres Beispiel

Gegeben ist folgender monoalphabetisch verfaßter Geheimtext:

```
FWJNKYICW CAFFL NGXJMHGTK IWLLG FGMTG KYIPMGHGJFNLLGJ
PMGZGJ FWR FMHJWGTG FGMTG CWLVGT IWXG MYI HGHGT VRALU
GMTHGLWNKYIL ZGTGT HMTH GK KAPMGKA TMYIL XGKATZGJK PWK
FWYIL ZGMT INTZ JWNYIL GJ MFFGJ TAYI KA OMGR
```

1. **Zählen der einzelnen Buchstaben**

```
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
7  0  3  0  0 11 30  8 11 10 11 12 14  6  1  4  0  3  0 15  1  2 11  3  8  5
```

   Der Buchstabe G tritt am häufigsten auf, deshalb vermuteten wir:

$$G = e$$

   Da `n` der zweithäufigste Buchstabe ist, sehen wir, das `n` entweder `T` oder `M` sein muß. Aus der Gleichverteilung der Buchstaben `s,i,r,a,n,t` folgt, das sie `T,M,L,F,I,K,J` oder `W` sind.

$$n \in \{T, M\}$$

$$\{s,i,r,a,t,n\} \subset \{T, M, L, F, I, K, J, W\}$$

2. **Zählen der Bigramme, die mit e beginnen, also `e? = G?`**

```
GX  GT  GM  GH  GJ  GZ  GL  GK
 1   6   4   1   6   1   1   3
```

   Aus der Häufigkeitsverteilung der Bigramme folgt

$\{en, er\} = \{GT, GJ\}$ und damit `n = T` und `r = J`.

Wir suchen nun nach `ei` und `ie`, da diese mit gleicher Häufigkeit vorkommen.

So finden wir `i = M`, damit muß aber `s = K` sein.

3. **Zählen der Bigramme, die mit `e` enden, also `?e = ?G`**

```
NG  HG  LG  FG  TG  MG  ZG  WG  VG  XG
1   5   2   3   4   4   4   1   1   2
```

Da ie = MG und ne = TG bereits feststehen, gilt

$$\{te, de\} \subset \{HG, LG, FG, ZG, XG\}$$

Durch Vergleich mit obigen Mengen erhalten wir:

$$t \in \{L, F\};\ a \in \{L, F, I, W\};\ d \in \{H, L, F, Z, X\}$$

4. **Zählen der am häufigsten auftretenden Bigramme**

```
YI  GJ  GT  HG
8   6   6   5
```

Da `IY` nicht im Text vorkommt, liegt der Schluß zu `ch = YI` nah.

5. **Aufschreiben der gefundenen Buchstaben.**

```
FWJNKYICW CAFFL NGXJMHGTK IWLLG FGMTG KYIPMGHGJFNLLGJ
maruschka kommt uebrigens hatte meine schwiegermutter

PMGZGJ FWR FMHJWGTG FGMTG CWLVGT IWXG MYI HGHGT VRALU
wieder mal migraene meine katzen habe ich gegen zloty

GMTHGLWNKYIL ZGTGT HMTH GK KAPMGKA TMYIL XGKATZGJK PWK
eingetauscht denen ging es sowieso nicht besonders was

FWYIL ZGMT INTZ JWNYIL GJ MFFGJ TAYI KA OMGR
macht dein hund raucht er immer noch so viel
```

Durch einfache Tests findet man schnell die Buchstaben für `t` und `a` sowie:

```
t=L, a=W, u=N, w=P, g=H, m=F, b=X, d=Z, o=A, k=Z, l=R,
v=O, z=V, y=U
```

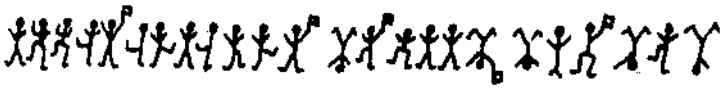
So heißt der nun geknackte Geheimtext:

> *Maruschka kommt - Übrigens hatte meine Schwiegermutter wieder mal Migräne - Meine Katzen habe ich gegen Zloty eingetauscht - denen ging es sowieso nicht besonders - Was macht Dein Hund? - Raucht er immer noch so viel?*

„Die tanzenden Männchen“

Die Möglichkeit, die Kryptoanalyse durch die Verwendung von Zeichen bzw. Hyroglyphen als Geheimtextalphabet zu erschweren, hat wenig Sinn, da die Eigenheiten der Sprache dadurch auf die Hyroglyphen übertragen werden. Dennoch wurde diese Art der monoalphabetische Chiffrierung auch gern von Krimi-Autoren verwendet, denn das Entschlüsseln des Textes ist verhältnismäßig einfach, so daß der Leser noch folgen kann. EDGAR ALLAN

POE verwendete diese Codierungsvariante in seinem Stück „Der Goldkäfer“, ARTHUR CONAN DOYLES Sherlock Holmes mußte sich in der Kurzgeschichte „Die tanzenden Männchen“ damit plagen. Hier ein kleiner Auszug aus dieser Geschichte:

> *Holmes hielt das Papier hoch, so daß die Sonne voll darauffiel. Es war eine aus einem Notizbuch herausgerissenen Seite. Die Zeichen waren mit Bleistift gemalt und sahen so aus:*
>
> 
>
> *Holmes betrachtete sie eine Zeitlang, faltete das Blatt vorsichtig zusammen und steckte es in die Brieftasche.*
> *»Das verspricht einen äußerst interessanten und ungewöhnlichen Fall« sagte er. ....*
> *Bei ihm war alles ruhig, außer daß eine lange Schriftzeile auf dem Sockel der Sonnenuhr erschienen war. Eine Kopie davon hatte er beigefügt, sie sah folgendermaßen aus:*
>
> 
>
> *Holmes beugte sich einige Minuten lang über den grotesken Fries und sprang dann plötzlich mit einem Ausruf der Überraschung und Bestürzung auf. ...*
> *»... Nachdem ich einmal erkannt hatte, daß die Symbole für Buchstaben stehen, und ich die Regeln anwandte, die für alle Arten von Geheimschriften gelten, war die Lösung nicht mehr schwierig. Die erste Nachricht, die man mir überlies, war so kurz, daß es unmöglich was, mir einiger Sicherheit mehr zu sagen, als daß*
>
> 
>
> *für E stand. Wie sie wissen, ist E der im Englischen gebräuchlichste Buchstabe, und er herrscht in einem solchen Maße vor, daß man erwarten kann, ihn selbst in einem kurzem Satz als den häufigsten zu finden. Von den fünfzehn Symbolen der ersten Botschaft kehrte eines viermal wieder, und so war es nur vernünftig, es als E anzunehmen. Nur ist es so, daß in einigen Fällen die Figuren Fähnchen tragen, in anderen nicht, aber es war an der Art, wie sich die Fähnchen verteilten, abzulesen, daß sie ein Wort vom anderen absetzten sollten. ... Nun, in dem einen Wort mit fünf Buchstaben habe ich bereits das E an zweiter und vierter Stelle. Es könnte ›sever‹ (trennen) oder ›lever‹ (Hebel) oder ›never‹ (niemals) bedeuten. Es steht außer Frage, daß letztere Bedeutung die wahrscheinlichstere wahr,...«*

Hilfsmittel zur Ver- und Entschlüsselung

Um Kryptoanalyse betreiben zu können, müssen wir uns mit den Gesetzmäßigkeiten der Sprache vertraut machen. Solche Normen hat jede Sprache und kann auch durch geschickte Chiffrierung nicht vollständig beseitigt werden.

Die Muster einer Sprache

Muster sind die Art und Weise, wie sich Buchstaben in einem Wort wiederholen. Sie werden über Ziffern ausgedrückt, wobei jeder neue Buchstabe auch eine neue Ziffer erhält, also z.B.:

OTTO ➔ 1221
NGRGUUV ➔ 1232445
PANAMAKANAL ➔ 12324252326

Solche Muster bleiben bei der monoalphabetischen Chiffrierung erhalten, d.h. enthält ein Geheimtext keine Muster, so ist er nicht durch monoalphabetischer Chiffrierung entstanden. Muster sind sehr hilfreich bei kurzen Texten.

Die Abhängigkeit der Häufigkeiten vom Klartext

Die Angaben über die Einzelbuchstaben schwanken und hängen außerdem vom Genre des Textes ab. So schreibt Beutelsbacher:

> *„Ein von Zitaten strotzender zoologischer Text über den Einfluß von Ozon auf die Zebras im Zentrum von Zaire wird eine andere Häufigkeitsverteilung ausweisen, als ein Traktat über die amourösen Adventüren des Balthasar Matzbach am Rande des Panamakanals.“*

Häufigkeiten sind um so schärfer, je länger der Text ist.

Der Wortzwischenraum

Sicherlich kommen wir auf die Idee, den Wortzwischenraum (Space) mit zu verschlüsseln. Dies führt dann zu einem Alphabet mit 27 Buchstaben. Da der Space im Deutschen nach dem e das häufigste „Zeichen“ und somit leicht zu enttarnen ist, läßt der professionelle Chiffrierer diesen Zwischenraum einfach weg. (Dies erschwert die Kryptoanalyse allerdings nur unwesentlich!)

Häufigkeiten von n-Grammen

N-Gramme sind Kolonnen von n Buchstaben. Die folgende Tabelle zeigt die Häufigkeiten für Bigramme im Deutschen und Englischem.

Bigramm engl.	Häufigkeit in %	Bigramm dt.	Häufigkeit in %
th	3,15	en	3,88
he	2,51	er	3,75
an	1,72	ch	2,75
in	1,69	te	2,26
er	1,54	de	2,00
re	1,48	nd	1,99
on	1,45	ei	1,88
es	1,45	ie	1,79
ti	1,28	in	1,67
at	1,24	es	1,52

Im Deutschen kommt `ch` sehr häufig vor, nahezu niemals `hc`. Desweiteren kommen `ei` und `ie` gleichhäufig vor.

Hier noch einige Trigramme:

deutsch	Häufigkeit in %	englisch	Häufigkeit in %
ein	1,22	the	3,53
ich	1,11	ing	1,11
nde	0,89	and	1,02
die	0,87	ion	0,75
und	0,87	tio	0,75
der	0,86	ent	0,73
che	0,75	ere	0,69

... und häufige Viergramme:

deutsch	icht, keit, heit, chon, chen, cher, urch, eich, ...

Aufschluß über den Ursprung des Textes kann auch die mittlere Wortlänge geben:

deutsch	5,9	italienisch	4,5
englisch	4,5	spanisch	4,4
französisch	4,4	russisch	6,3

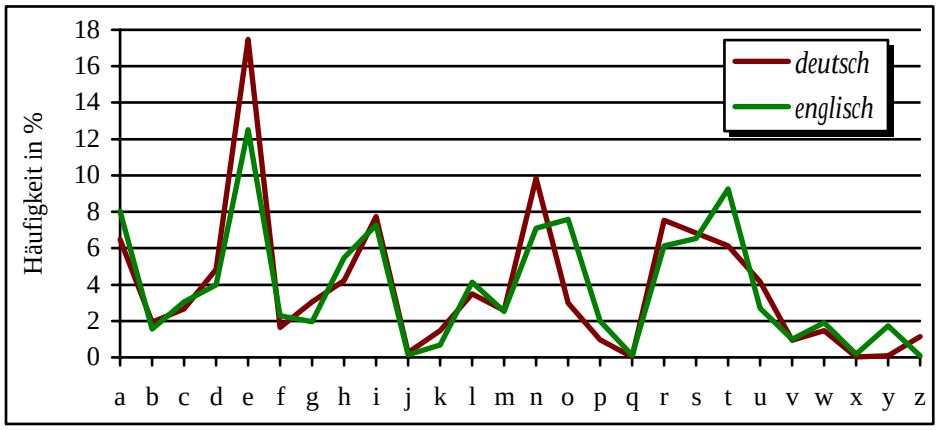
... aber auch die zehn häufigsten Wörter:

deutsch	die, der, und, den, am, in, zu, ist, daß, es
englisch	the, of, and, to, a, in, that, it, is, I
französisch	de, il, le, et, que, je, la, ne, on, les
italienisch	la, di, che, il, non, si, le, una, lo, in
spanisch	de, la, el, que, en, no, con, un, se, sa

Die Abhängigkeit von der Verfassersprache

Häufigkeitsgebirge

Jede Sprache hat ihre eigenen Häufigkeiten. Die Diagrammdarstellungen werden Häufigkeitsgebirge genannt.



Durch Häufigkeitsgebirge können wir bei genügend langer monoalphabetisch-chiffrierten Geheimtext feststellen, in welcher Sprache der Text verfaßt wurde.

Der Koinzidenzindex (Kappa κ) einer Sprache

Um das Kappa einer Sprache zu bestimmen, schreiben wir zwei gleichlange unterschiedliche Texte untereinander und zählt alle „Spalten“ mit gleichen Buchstaben. Anschließend teilen wir diese Zahl durch die Anzahl der Buchstaben in einer Zeile. Die mathematische Beschreibung von Kappa sieht so aus:

Für zwei Texte gleicher Länge

$T_x = x_1 x_2 \ldots x_n$ und $T_y = y_1 y_2 \ldots y_n$ definieren wir

$$\kappa(T_x, T_y) = \sum_{i=1}^{n} \frac{\delta(x_i, y_i)}{n} \quad \text{wobei} \quad \delta(x_i, y_i) = \begin{cases} 1 \ \textit{für } x = y \\ 0 \ \textit{für } x \neq y \end{cases} \text{ ist.}$$

Besser ist jedoch die Formel $\kappa = \frac{\sum_{i=1}^{26} n_i^2}{n(n-1)}$, wobei n die Gesamtzahl der Buchstaben ist und n_i die absolute Häufigkeit der einzelnen Buchstaben.

Es zeigt sich, daß jede Sprache ihr eigenes Kappa hat:

deutsch	7,62 %	spanisch	7,75 %
englisch	6,61 %	japanisch	8,19 %
französisch	7,78 %	russisch	5,29%
italienisch	7,38 %		

Kommt jeder Buchstabe (bei einem Alphabet aus 26 Buchstaben) mit der gleichen Wahrscheinlichkeit von $\frac{1}{26}$ vor, so ergibt sich für Kappa der Wert $\kappa = \frac{1}{26}$ = 3,85 %.

Durch Zerschneiden des Geheimtextes in zwei Hälften und der Ermittlung des Kappa der beiden so entstandenen Texte kann festgestellt werden, mit welcher Sprache der Klartext verfaßt wurde oder ob es eine Kunstsprache ist.

Die Verschleierung der Häufigkeiten

Die Invarianz ist der Schwachpunkt der monoalphabetischen Chiffrierung. Der professionelle Chiffrierer verschleiert deshalb die Häufigkeiten, indem er jedem Klartextbuchstaben mehrere Geheimtextzeichen (z.B. Ziffernpaare) zuordnet und zwar so viele, wie der Häufigkeit der Klartextbuchstaben entsprechen. Die einem Klartextbuchstaben zugeordneten Geheimtextzeichen heißen **Homophone**.

Beispiel:	e (17 Zeichen lt. Häufigkeit)	02 17 43 44 56 ...
	n (10 Zeichen)	07 18 19 86 ...
	i (8 Zeichen)	14 39 46 ...

Beim Chiffrieren wählt man die Homophone zufällig (z.B. durch Würfeln), so daß dann

ein	02 46 18 oder
	43 46 07 ist.

Die Kryptoanalyse ist nun schon wesentlich schwieriger, aber keinesfalls unmöglich!

Zusammenfassung

Um die Kryptoanalyse eines monoalphabetisch-chiffrierten Textes zu erschweren, bietet sich folgendes an:

- Kurz fassen!
- Kunstsprache mit gleich verteilten Buchstaben verwenden (Codebücher!)

  Im ersten Weltkrieg wurden hauptsächlich Codebücher verwendet. Jedoch war die Gefahr sehr groß, daß ein solches Buch in die Hände der Gegner fiel.

  Beispiel: auto = bfz oder panzer = terzt
- Verwendung von Homophonen
- Verwendung von Blendern, d.h. von sinnlosen Zeichen, die die Häufigkeiten von n-Grammen verschleiern.

  Beispiel: Klartextalphabet: a b c ... z
  Erweiterung: a b c ... z 0 1 ... 9

  Wenn im Klartext 16 mal ch auftritt, so fügt man an 16 beliebigen Stellen 0hc, 1hc, ... ein.
- Verwendung von Spreizern, d.h. einem Klartextbuchstaben wird nicht ein Geheimtextbuchstabe zugeordnet, sondern eine Buchstabenkolone, also

  a = B
  b = XC
  c = BAS (dann darf AS nicht als einzelnes Geheimtextzeichen auftreten)
  d = YI
  h = TO

  Entschlüsselung: BASTO kommt häufig vor, TOBAS dagegen nie!
- Vermeide Eigennamen, Standardfloskeln, „wahrscheinliche Wörter“, wie etwa Hauptquartier oder geheim.

> *„Ein ideal für die Chiffrierung vorbereiteter Klartext ist orthographisch falsch, sprachlich knapp und stilistisch grauenhaft.“ Bauer*

Kapitel

3

Die polyalphabetische Chiffrierung

Polyalphabetische Chiffrierung bedeutet, daß das Prinzip der monoalphabetischen Chiffrierung nach gewissen Regeln ständig verändert. Es wird also nicht der gesamte Klartext monoalphabetisch, sondern jede Buchstabengruppe mit einem anderen monoalphabetischen Schlüssel.

Die Idee der polyalphabetischen Chiffrierung

Eine polyalphabetische Chiffrierung kann z.B. für den Klartext 'abba' so erfolgen:

Klartextalphabet:	a b c d e ...
erstes Geheimtextalphabet:	**H** L W X D ...
zweites Geheimtextalphabet:	U **L** V W A ...
drittes Geheimtextalphabet:	N **A** R T I ...
viertes Geheimtextalphabet:	**D** Y Z L M ...

Damit wird aus 'abba' ganz einfach 'HLAD'. Als Konsequenz ergibt sich, daß die **Häufigkeiten und Muster verschwunden** sind! Allerdings müssen Empfänger und Sender die Regeln wissen, nachdem sich die Zuordnung Klartextalphabet zu Geheimtextalphabet ändert, i. allg. besitzen beide 26 Alphabete und müssen dazu das Anfangswort übermitteln. Die bekannteste Methode der polyalphabetischen Chiffrierung ist die **Vigenère-Chiffrierung.** Dazu benötigt man ein Vigenère-Quadrat, welches sich im Anhang befindet.

Die Theorie der Vigenère-Chiffrierung

Neben dem Vigenère-Quadrat benötigen wir noch ein Schlüsselwort. Um Klartext zu chiffrieren, schreibt man das Schlüsselwort, z.B.: VENUS periodisch über den Klartext:

V E N U S V E N U S V E N U S V
p o l y a l p h a b e t i s c h
K S Y S S G T U U T Z X V M U C

Jeder Klartextbuchstabe wird dann mit dem Geheimtextalphabet verschlüsselt, dessen erster Buchstabe im Vigenère-Quadrat über dem Klartextbuchstabe

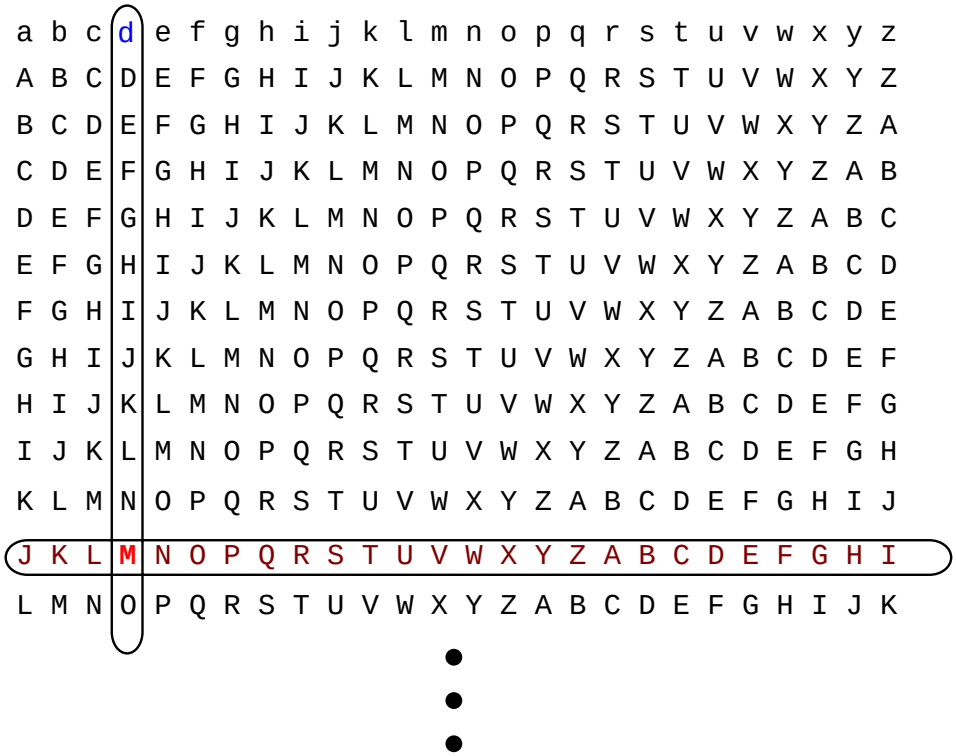
stehende Schlüsselwortbuchstabe ist. Also für den ersten Buchstaben des Beispiels von oben heißt das: Wenn man das erste p aus polyalphabetisch verschlüsselt will, muß man die Geheimtextalphabetzeile nehmen, die mit V beginnt. Dann sucht man sich im Klartextalphabet über den Vigenère-Quadrat das p, geht die Spalte nach unten und sucht den Geheimtextbuchstaben, der der „Schnittpunkt“ von Zeile und Spalte ist, also K. Die Dechiffrierung erfolgt analog.

Ein praktisches Beispiel

Dazu verwenden wir wiederum den Text aus Kapitel 2.2, den abgeschlossenen Roman, das Schlüsselwort sei JAMESBOND. Wir schreiben zunächst das Schlüsselwort über den Text.

JAMES BONDJ AMESB ONDJA ME
Derab gesch losse neRom an

Wir gehen mit den in die Zeile, die mit dem Schlüsselwortbuchstaben beginnt und suchen die Spalte, die mit dem Klartextbuchstaben beginnt. Am Schnittpunkt von Zeile und Spalte steht unser Geheimbuchstabe.



Und erhalten somit:

JAMES BONDJ AMESB ONDJA ME
Derab gesch losse neRom an
MEDET HSFFQ LAWKF BRUXM MR

Die Verschlüsselung der restlichen Textes ist eine mögliche Übungsaufgabe.

Die Theorie der Entschlüsselung

Die Kryptoanalyse von Vigenère-Chiffrierung ist leicht, **wenn das Schlüsselwort relativ kurz ist und der Geheimtext lang**. Das Knacken erfolgt in zwei Schritten:

1. Bestimmen der Länge des Schlüsseltextes (Kasiski- & Friedmann-Test)
2. Bestimmen des Schlüsselwortes selbst

Der Kasiski - Test

Der Test basiert auf folgender Idee:

Treten im Klartext gewisse Buchstabenfolgen häufig auf (Trigramme u.ä.), so werden sie stets gleich übersetzt, wenn ein Vielfaches des Schlüsselwortes dazwischen paßt.

Also:

```
Schlüsselwort:  V E N U S V E N U S V E N U S V E N U S
Klartext:             e i n       e i n   e i n
Geheimtext:           W D R       Z M A   W D R
```

Man sucht im Geheimtext sich wiederholende Zeichenfolgen und vermutet, das deren Abstand ein Vielfaches der Schlüsselwortlänge ist.

Beispiel:

```
UEQPC VCKAH VNRZU RNLAO KIRVG JTDVR VRICV IDLMY
IYSBC COJQS ZNYMB VDLOK FSLMW EFRZA VIQMF JTDIH
CIFPS EBXMF FTDMH ZGNMW KAXAU VUHJH NUULS VSJIP
JCKTI VSVMZ JENZS KAHZS UIHQV IBXMF FIPLC XEQXO
CAVBV RTWMB LNGNI VRLPF VTDMH ZGNMW KRXVR QEKVR
LKDBS EIPUC EAWJS BAPMB VSZCF UEGIT LEUOS JOUOH
UAVAG ZEZIS YRHVR ZHUMF RREMW KULKV KGHAH FEUBK
LRGMB JIHLI IFWMB ZHUMP LEUWG RBHZO LCKCW THWDS
ILDAG VNEMJ FRVQS VIQMU VSWMZ CTHII WGDJS XEOWS
JTKIH KEQ
```

Folge	Abstand	Primfaktorenzerlegung
KAH	128	2^7
JTD	50	$2 \cdot 5^2$
VIQM	265	$5 \cdot 5 \cdot 3$
TDMHZGNMWK	90	$2 \cdot 3 \cdot 3 \cdot 5$
MWK	75	$3 \cdot 5 \cdot 5$

Man bildet nun den größten gemeinsamen Teiler (ggT) der Abstandszahlen. In obigen Beispiel wäre dieser ggT = 1, d.h. die Schlüsselwortlänge eins. Dies scheidet aber aus, weil der Klartext dann monoalphabetisch verschlüsselt sein müßte. Nimmt man aber an, daß KAH nur zufällig mehrmals auftritt, dann ist der ggT der Abstände gleich 5. Das legt die Schlüsselwortlänge fünf nahe. Es könnte aber auch sein, daß sich VIQM und MWK zufällig wiederholen, dann wäre

der ggT(50, 90) = 10. **Also die Schlüsselwortlänge ist höchstwahrscheinlich fünf, könnte aber auch zehn sein!**

Der Friedmann - Test

Dieser Test wurde von Friedmann 1925 aufgestellt. Er dient dazu, die Größenordnung des Schlüsselwortes abzuschätzen.

Die Idee des Friedmanntestes ist nun folgende: Je länger das Schlüsselwort ist, desto regelmäßiger sind die Häufigkeiten verteilt, desto kleiner bzw. näher liegt κ_G an 3,85%.

Man nimmt nun an, die Schlüsselwortlänge sei x und schreibt den Geheimtext in x Spalten. Das sieht dann wie folgt aus:

S_1	S_2	...	S_x
B_{x+1}	B_{x+2}	...	B_x
B_{2x+1}	B_{2x+2}	...	B_{2x}
B_{3x+1}	B_{3x+2}	...	B_{3x}
:	:	:	:

Die Wahrscheinlichkeit, daß zwei beliebige Buchstaben in der gleichen Spalte stehen ist $\frac{1}{x}\cdot\frac{1}{x}+\frac{1}{x}\cdot\frac{1}{x}+\ldots+\frac{1}{x}\cdot\frac{1}{x}=x\cdot\frac{1}{x}\cdot\frac{1}{x}=\frac{1}{x}$.

Die Wahrscheinlichkeit, daß zwei beliebige Buchstaben in unterschiedlichen Spalten stehen, ist dann $1-\frac{1}{x}$.

Die Wahrscheinlichkeit, das zwei beliebige Buchstaben in einer Spalte gleich sind, ist κ_K (Klartextkappa). Die Buchstaben einer Spalte sind ja dann monoalphabetisch verschlüsselt!

Die Wahrscheinlichkeit, daß zwei Buchstaben aus verschiedenen Spalten gleich sind, ist 3,85% + ε, da über verschiedene Alphabete verschlüsselt wurde. Die Wahrscheinlichkeit, daß zwei beliebige Buchstaben des Geheimtextes gleich sind, ist also

$$\kappa_G=\frac{1}{x}\kappa_K+(1-\frac{1}{x})\cdot(3{,}85+\varepsilon)$$

$$\kappa_G=\frac{1}{x}(\kappa_K-3{,}85-\varepsilon)+3{,}85+\varepsilon$$

Da κ_K und κ_G leicht zu berechnen sind, kann die obige Formel nach x umgestellt werden, um die Schlüsselwortlänge zu bestimmen.

$$x=\frac{\kappa_K-3{,}85-\varepsilon}{\kappa_G-3{,}85-\varepsilon}$$

Für die deutsche Sprache ist das Klartextkappa 7,62 und kann in die Formel eingesetzt werden.

$$x=\frac{7{,}62-3{,}85-\varepsilon}{\kappa_G-3{,}85-\varepsilon}=\frac{3{,}77-\varepsilon}{\kappa_G-3{,}85-\varepsilon}\le\frac{3{,}77}{\kappa_G-3{,}85}$$

Im dem oben angeführten Beispiel ist κ_G = 4,39; d.h. $x \le 6{,}98$.

Kasiski- und Friedmann-Test legen also die Schlüsselwortlänge x = 5 nah.

Die Bestimmung des Schlüsselwortes

Da die Schlüsselwortlänge nun bekannt ist, kann das Schlüsselwort nun einfach gefunden werden. Sie die Länge des Schlüsselwortes gleich x. Schreibt man den Text wie folgt:

S_1	S_2	...	S_x
B_{x+1}	B_{x+2}	...	B_x
B_{2x+1}	B_{2x+2}	...	B_{2x}
B_{3x+1}	B_{3x+2}	...	B_{3x}
:	:	:	:

Jede Spalte wurde monoalphabetisch verschlüsselt, es genügt also das e zu finden, um herauszukriegen mit welchem Alphabet die entsprechende Spalte verschlüsselt wurde.

Spalte	häufigster Buchstabe	Schlüsselwortbuchstabe
1	V → e	R
2	E → e	A
3	H, D, U → e	D, Z, Q
4	N → e	I
5	S → e	O

Das Schlüsselwort lautet also RADIO. Nun kann der gesamte Geheimtext dechiffriert werden.

Ein weiteres Beispiel

Gegeben ist folgender Text:

```
KWCSS GXYUT ZBZMU CMRFY JZZNZ HMEBS WMEXA ZMZAW
IATBS ABUUK NMZHT ZAKCE HBVLY ZPVCE OMONT PKYML
VJVGW CZRFK ZQEYF FTRLL ZFKVM XPJNS WMEXS MAKYD
GMEES IVRVW MURHV VZWHA XPKPW MOVMK ZVUUK NLVLY
ZPVCE OMONV ZVBFS MBVRL ZQEXW PBZAT ZAKCE HMEGM
NAQOE WMZMH DMCCK OMJHA XPKGG ZOICU CLRMK DMVCF
ZURFY JZZNZ HCJXW MOVBW DUKYP OJLWZ NBRVW IQDED
VZKYP OMZHE VTVOF YMZHS ILVLW NURFK ZVKMH MQTBL
JPEYV VAJYK YIWOW MMZHW MMXYD BQSNV DMUYE ZUGZS
ZVXYJ BMEUM NIXNO VVEYJ ZCEXO VVEYJ NMENK KZZWZ
OMJCK OMENK XPVCV ZVUXS NARHB ZLVLK OMCFW YMJEJ
TXKIY MIDGK YMIMU CTLYK NMCYA ILVOL DOUYF FTRLL
ZFKVM XPJNS WMETM EMUYE BMYYA HBVRL WCTBK OISYF
AMJND ZOK
```

Wir führen zunächst den Kasiski-Test durch. Dazu suchen wir im Text nach gleichen Textfolgen:

KWCSS GXYUT ZBZMU CMRFY JZZNZ HMEBS WMEXA ZMZAW
IATBS ABUUK NMZHT ZAKCE HBVLY ZPVCE OMONT PKYML
VJVGW CZRFK ZQEYF FTRLL ZFKVM XPJNS WMEXS MAKYD
GMEES IVRVW MURHV VZWHA XPKPW MOVMK ZVUUK NLVLY
ZPVCE OMONV ZVBFS MBVRL ZQEXW PBZAT ZAKLE HMEGF
NAQOE WMZMH DMICK OMJHA XPKGG ZOICU CLRMK DVVCF
ZURFY JZZNZ HCJXW MOVBW DUKYP OJLWZ NBRVW SQDED
VZKYP OMZHE VTVOF YMZHS ILVLW NURFK ZVKMH MQTBL
JPEYV VAJYK YIWOW MMZHW MMXYD BQSNV DMUYE ZUGZS
ZVXYJ BMEUM NIXNO VVEYJ ZCEXO VVEYJ NMENK KZZWZ
OMJCK OMENK XPVCV ZVUXS NARHB ZLVLK OMIFW YMJEJ
TXKIY MIDGK YMIMU CTLYK NMIYA ILVHL DOUYF FTRLL
ZFKVM XPJNS WMENM ENUYE BMYYA HBVRL WCTBK OISYF
AMJND ZOK

Folge	Abstand	Primfaktorenzerlegung
SWMEX	80	$2^2 \cdot 5$
UUK	105	$3 \cdot 5 \cdot 7$
OMO	95	$5 \cdot 19$
YFFTRLLZFKVMXPJNSWME	380	$2^2 \cdot 5 \cdot 19$
ZVU	265	$5 \cdot 53$

Wir bilden nun den größten gemeinsamen Teiler der Abstandszahlen. Dieser ggT ist fünf, d.h. die Schlüsselwortlänge hat vermutlich die Länge fünf.

Wir führen jetzt den Friedman-Test aus. Dazu bestimmen wir die Anzahl der Buchstaben und ihre absolute Häufigkeit. In unserem Fall ist n = 528 und

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
16	17	20	11	27	15	8	15	12	17	30	21	51	20	22	13	6	13	14	13	18	39	22	16	29	43

Damit können wir die Formel $\kappa = \frac{\sum_{i=1}^{26} n_i^2}{n(n-1)}$ anwenden und erhalten

$$\kappa = \frac{13644}{528 \cdot 527} = 4{,}9\%\,.$$

Nun ist es möglich, die Schlüsselwortlänge zu bestimmen:

$x \approx \frac{3{,}77}{4{,}9 - 3{,}85} \approx 3{,}6$. Wir vermuten also eine Schlüsselwortlänge von drei.

Der Vergleich der beiden Methoden zeigt, daß kein eindeutiges Resultat entsteht. Wir probieren zunächst die Vermutung der Schlüsselwortlänge fünf, da der Kasiski-Test eindeutig funktionierte. Dazu schreiben wir den Text in Spalten zu je fünf Zeichen, also:

KWCSS
GXYUT
ZBZMU
CMRFY
JZZNZ ...

Wir suchen nun in jeder Spalte den häufigsten Buchstaben und vermuten das dieser für e steht.

Spalte	häufigster Buchstabe	Schlüsselwortbuchstabe
1	Z → e	V
2	M → e	I
3	V → e	R
4	Y → e	U
5	W, K → e	S, G

Wie unschwer zu sehen ist, ist das Schlüsselwort vermutlich VIRUS. Nun können wir den restlichen Text dechiffrieren. Das Resultat lautet:

> *Polyalphabetische Algorithmen haben die Eigenschaft, dass ein bestimmter Geheimtextbuchstabe mehr als einen Klartextbuchstaben darstellen kann. Aber man darf nicht vergessen, dass der Geheimtext den Klartext eindeutig bestimmen muss. Zum Beispiel ist es nicht moeglich, das Sie einem Algorithmus der Geheimtextbuchstaben im Klartext einmal e und ein anderes mal s entspricht, ohne dass es dafuer eine Regel gibt, die dem Empfaenger genau sagt, wann er e und wann er s entspricht. Es ist entscheidend, dass an jeder Stelle des Kryptogramms der Schluessel eindeutig den Klartextbuchstaben zu jedem Geheimtextbuchstaben festlegt.*

Zusammenfassung:

- Virgenére-Chiffrierung mit kurzem Schlüsselwort und langen Text sind einfach zu knacken.
- Also kurz fassen und lange Schlüsselwörter verwenden!
- Das Problem ist die Übermittlung des Schlüsselwortes. Dieses aber kann zu geeigneten Zeiten über sichere Kanäle übermittelt werden.

Kapitel

4

Die Public-Key-Chiffrierung

Die Gefahren der monoalphabetischen und polyalphabetischen Chiffrierung liegen in der Übermittlung des Schlüssels zum Dechiffrieren. Beim Public Key [engl. öffentlicher Schlüssel] besitzt der Sender ein solches geheimes Wort nicht, sondern verwendet ein allen Nutzern zu Verfügung stehendes Kodewort.

Das Prinzip des öffentlichen Schlüssels

Das Public-Key-Prinzip ist sehr einfach zu verstehen: Damit der Sender eine Nachricht verschlüsseln kann, benutzt er den öffentlichen Schlüssel des Empfängers. Dieser Schlüssel kann z. B. in einer Datenbank für jeden zugänglich gespeichert sein. Der Empfänger allein ist im Besitz eines privaten Schlüssels, mit dem er die codierte Information wieder in Klartext umwandeln kann. Anschaulich gesprochen heißt das, jeder kann einen Brief in einen Briefkasten einwerfen, aber nur der Besitzer des Briefkastenschlüssels wird die Post erhalten.

Die wohl bekannteste Public-Key-Kodierung geht auf RON RIVEST, ADI SHAMIR und LEONARD ADLEMAN zurück und wird **RSA-Algorithmus** bezeichnet.

Der RSA-Algorithmus

Der RSA-Algorithmus greift in vollen Zügen in die Mathematik hinein, speziell in den Bereich der Zahlentheorie. Dennoch ist das Prinzip auch für einen Laien mit Schulmathematik verständlich.

Die Erzeugung der Schlüssel

Bevor überhaupt ver- bzw. entschlüsselt werden kann, muß von einer Zentrale der öffentliche und private Schlüssel erzeugt werden. Der private Schlüssel besteht aus einer ganzen positiven Zahl d, der öffentliche Schlüssel aus den beiden ganzen positiven Zahlen e und n. Um diese Zahle festzulegen wählt die Zentrale für jeden Teilnehmer zwei große Primzahlen p und q und bildet deren Produkt $n = p \cdot q$. Damit hat sie den ersten Bestandteil des öffentlichen Schlüssels. Als e, den zweiten Teil des öffentlichen Schlüssels des Teilnehmers, legt sie eine Zahl fest, die mit der Zahl $(p - 1)(q - 1)$ keinen gemeinsamen Teiler hat. Der größte gemeinsame Teiler von e und $(p - 1)(q - 1)$ soll also 1 sein. Der

private Schlüssel des Teilnehmers ist dann die Zahl *d*, die folgende Bedingungen erfüllt:

- *d* ist kleiner oder gleich $(p - 1)(q - 1)$,
- das Produkt $e \cdot d$ liefert nach Division durch $(p - 1)(q - 1)$ den Rest 1, also $e \cdot d \bmod (p - 1)(q - 1) = 1$

Dem Teilnehmer werden nun die Zahl *d* geheim als privater Schlüssel mitgeteilt; die Zahlen *n* und *e* werden veröffentlicht und die Werte *p* und *q* sollten von der Zentrale vernichtet werden.

Die Anwendung der Schlüssel

Jeder kann nun eine Nachricht an der Teilnehmer senden. Dazu muß er den Klartext an Form einer Zahl m darstellen, die nicht größer als n ist. Um nun die Zahl *m* (z. B. ASCII-Codierung) zu verschlüsseln, rechnet der Sender den Divisionsrest aus, der bei Division von m^e und *n* entsteht, also

$$c := m^e \bmod n$$

Diese Zahl *c* ist der Geheimtext, der zum Klartext *m* gehört.

Der Empfänger, der die chiffrierte Botschaft *c* erhält, entschlüsselt sie mit der Rechnung

$$m := c^d \bmod n.$$

Daß der Empfänger hier wirklich wieder *m* zurückerhält, die Verschlüsselungvorschrift also wieder den Klartext liefert, beruht auf dem „Satz von EULER“, einem über 200 Jahre alten Ergebnis der Zahlentheorie.

Ein praktisches Beispiel

Die Schlüssel

Um den öffentlichen und den privaten Schlüssel festzulegen, wählt man zufällig zwei große Primzahlen *p* und *q*. In der Praxis müssen *p* und *q* mehr als 256 Bits lang sein; im folgenden Beispiel beschränken wir uns der Übersichtlichkeit halber auf 11stellige Primzahlen, nämlich

$p = 37419669101$ und $q = 11110693267$.

Das Produkt dieser Zahlen ist $n = 415758465533848642967$. Damit ist die erste Zahl des öffentlichen Schlüssels bekannt. *e*, die zweite Zahl dieses Schlüssels muß zu $(p - 1)(q - 1) = 415758465485318280600$ teilerfremd sein. Das geht beispielsweise mit $e := 2^{16} + 1 = 65537$; speziell diese Zahl hat obendrein eine einfache binäre Darstellung, was die Rechnung vereinfacht.

Es bleibt der private Schlüssel d zu bestimmen. Er ist die einzige Zahl zwischen 1 und $(p - 1)(q - 1)$ mit der Eigenschaft, daß $e \cdot d \bmod (p - 1)(q - 1) = 1$.

Um aus dieser Gleichung *d* zu bestimmen, verwendet man den „Erweiterten Euklidischen Algorithmus“[2]. Das Ergebnis ist $d = 16481384459631305873$.

Die Chiffrierung

Jetzt ist man in der Lage, Nachrichten zu verschlüsseln. Dazu sind die Nachrichten zunächst in eine Zahlenfolge zu übersetzen. Es bietet sich die ASCII-Kodierung an; zur Übersichtlichkeit wählen wir aber folgendes Verfahren: a bis z werden durch die Zahle 01 bis 26 kodiert, das Leerzeichen mit 00. Die Nachricht `Kryptologie macht Spass` entspricht damit der Zahlenfolge:

11182516201512150709050013010308200019160119 19

[2] siehe dazu im Anhang nach

Diese Zahl läßt sich nicht als Ganzes verschlüsseln, weil sie größer ist als *n*, die erste Zahl des öffentlichen Schlüssels. Also teilt man die Zahlenfolge der Nachricht in Blöcke auf, die jeweils einzeln kodiert werden. Der erste dieser Blöcke ist

$$m := 11182516201512150709.$$

m wird nun verschlüsselt zu

$$c := 11182516201512150709^{65537} \bmod 415758465533848642967$$
$$= 71043117991897565951.$$

Entsprechend geht man mit den weiteren Blöcken um.

Dechiffrierung

Der Empfänger, der diese Zahl erhält, wendet darauf seinen privaten Schlüssel *d* an:

$$m := 71043117991897565951^{1648138445963130587 3} \bmod 415758465533848642967$$
$$= 11182516201512150709$$

... und erhält so den ersten Teil der Nachricht zurück.

Die digitale Unterschrift mit RSA

Mit Hilfe des RSA läßt sich auch eine digitale Unterschrift verwirklichen: Ein Teilnehmer „unterschreibt“ eine Nachricht m, indem er sie mit seinem geheimen Schlüssel *d* kodiert: Er berechnet

$$s := m^d \bmod n$$

und er veröffentlicht diese „Signatur“ *s* und die Nachricht *m* selbst. Diese unterschriebene Nachricht kann jetzt von jedem überprüft werden: Liefert die Rechnung

$$s^e \bmod n$$

wieder die bekannte Nachricht *m* zurück, dann ist die Unterschrift gültig und die Nachricht nicht verändert worden.

Die Sicherheit von RSA

Wie sicher ist RSA? Dieser Algorithmus basiert darauf, daß es schwer ist, große Zahlen in ihre Primfaktoren zu zerlegen. Für den RSA-Algorithmus werden Zahlen mit mindestens 512 Bits (155 Dezimalstellen) empfohlen. Zum Vergleich: Der Weltrekord der Faktorisierung beliebiger Zahlen liegt derzeit bei 116 Dezimalstellen.

Weil Zahlen mit 20 Dezimalstellen heute von jedem Computeralgebra-System wie Derive, Maple oder Mathematica auf PC-Basis in Sekundenschnelle faktorisiert werden können, ist von Produkte, die RSA mit 64-Bit-Zahlen.

Anhang

Quellenangabe

- Beutelsbacher, Albrecht. Kryptologie. 3. Aufl. Braunschweig/Wiesbaden: Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, 1992.
- Bauer, Friedrich L.: Kryptologie: Methoden und Maximen. 2. Aufl. Berlin, Heidelberg, New York: Springer-Verlag, 1994
- Beutelsbacher, Albrecht. „Hinter Schloß und Riegel?“. mc Mai 1994: 88ff.
- Brockhaus-Enzyklopädie. 19. Aufl. Mannheim 1990, Stichwort: Kryptologie
- Zachmann. Die große Bertelsmann Lexikothek: Naturwissenschaft und Technik. Bd. 2. Gütersloh 1991. S. 76ff.
- Poe, Edgar Allan. Die Abenteuer eines gewissen Hans Pfaall. Unheimliche und phantastische Erzählungen. Der Goldkäfer: Verlag Neues Leben Berlin, 1977.
- Doyle, Arthur Conan. Die Wiederkehr des Sherlock Holmes. Die tanzenden Männchen: Gustav Kiepenheuer Verlag Leipzig, Weimar, 1984.
- Vorlesungsmitschrift der Vorlesung Kryptologie Prof. Böttcher, TU Chemnitz-Zwickau
- Vorlesungsskript der Vorlesung Kryptologie Prof. Horster, TU Chemnitz-Zwickau

Lösung des Tauschchiffres

Der berühmte Spruch wurde mittels Caesar-Chiffrierung codiert und lautet:

```
Ich kam, sah und siegte!
```

Übungsaufgabe[3] monoalphabetische Chiffrierung

Gegeben ist folgender monoalphabetischer Text. Versuche diese Nachricht zu knacken!

IU MGF IWH UVHHWKIF DVFKIH. RWI SGOHIWRITAUI AGYYI IU UWTA GOC IWHID UYIWH KISOIYXWTA KIDGTAY. EXVIYSXWTA NGD RIF COTAU GOU RID MGXR KIFGHHY. IF UYOIFSI GOC RWI ITAUI SO. RG NGD RWI NFIOSVYYIF ZVD KGOD OHR CWIX GOC RIH COTAU UWI UGA WAH GH OHR VICCHIYI WAFIH DOHR. GXU RIF COTAU RWI UEWYSIH KWCYSGIAHI UGA, FGHHYI IF UV UTAHIXX IF NVHHYI RGZVH. OHR MIHH IF HWTAY KIUYVFKIH UV FIHHY IF HVTA AIOYI.

[3] aus dem Vorlesungsmaterial (von Eva Böttcher)

Lösung zur Übungsaufgabe

Hier nun die Lösung - aber nicht schummeln!

Es war ein sonniger Morgen. Die Zauneidechse hatte es sich auf einem Stein gemuetlich gemacht. Ploetzlich kam der Fuchs aus dem Wald gerannt. Er stuerzte auf die Echse zu. Da kam die Kreuzotter vom Baum und fiel auf den Fuchs. Sie sah ihn an und oeffnete ihren Mund. Als der Fuchs die spitzen Giftzähne sah, rannte er so schnell er konnte davon. Und wenn er nicht gestorben, so rennt er nich heute.

Das Vigenère - Quadrat

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

Der „Erweiterte Euklidische Algorithmus“

Um den privaten Schlüssel d beim RSA-Verfahren zu bestimmen, ist die Gleichung

$$e \cdot d \bmod (p - 1)(q - 1) = 1$$

bei gegebenem e, p und q zu lösen. Wir führen das an einem Beispiel vor, nämlich der Gleichung

$$41d \bmod 192 = 1.$$

Jetzt rechnet man:

$$192 = 4 \cdot 41 + 28$$
$$41 = 1 \cdot 28 + 13$$
$$28 = 2 \cdot 13 + 2$$
$$13 = 6 \cdot 2 + 1.$$

Nun löst man diese Gleichungskette von unten nach oben auf:

$1 = 13 - 6 \cdot 2$	| Umformung der vierten Zeile
$= 13 - 6 \cdot (28 - 2 \cdot 13)$	| die 2 aus der dritten Zeile einsetzen
$= 13 \cdot 13 - 6 \cdot 28$	| 13 zusammenfassen
$= 13 \cdot (41 - 1 \cdot 28) - 6 \cdot 28$	| 28 zusammenfassen
$= 13 \cdot 41 - 19 \cdot 28$	| die 28 aus der ersten Zeile einsetzen
$= 13 \cdot 41 - 19 \cdot (192 - 4 \cdot 41)$	| 41 zusammenfassen
$= 89 \cdot 41 - 19 \cdot 192$	

Daraus liest man ab:

$89 \cdot 41 = 19 \cdot 192 + 1$. Damit weiß man:

$41 \cdot 89 \bmod 192 = 1$, und das gesuchte d ist also 89.