Karlsruhe Institute of Technology

Schnupperkurs Mathematik 2019

Verschlüsselung - Von Caesar bis zum TXDQWHQFRPSXWHU

Daniel Weiß

Sommersemester 2019

Inhaltsverzeichnis

Kapitel 1

Einführung

Die **Kryptologie** (altgriechisch $\kappa\rho\upsilon\pi\tau\acute{o}\varsigma$ (kryptós) verborgen, geheim; $\lambda\acute{o}\gamma o\varsigma$ (lógos) Wort, Lehre) ist eine Wissenschaft, die sich mit der Verschlüsselung und Entschlüsselung von Informationen bzw. noch etwas allgemeiner mit Informationssicherheit beschäftigt. Sie bedient sich dabei Methoden und Techniken der Mathematik, Informatik, Linguistik und Quantentheorie.

Die Kryptologie umfasst unter anderem die Kryptographie und die Kryptoanalyse:

> **Kryptographie:** Beschäftigt sich mit der Entwicklung von Verschlüsselungsverfahren. Der Begriff Kryptographie wird heutzutage auch als Synonym für Kryptologie verwendet.
>
> **Kryptoanalyse:** Beschäftigt sich mit der Entwicklung von Methoden, um Informationen aus verschlüsselten Texten (ohne Kenntnis des Schlüssels) zu erhalten.

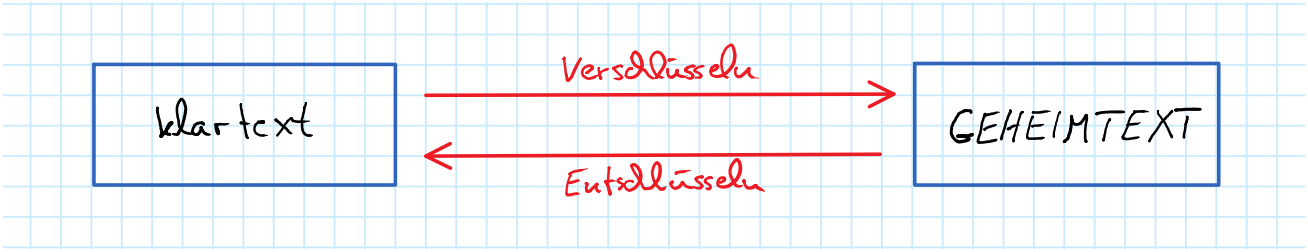

Abbildung 1.1: In der Kryptologie wird der Klartext in Kleinbuchstaben und der Geheimtext in Großbuchstaben geschrieben.

Ein **Verschlüsselungsverfahren** legt neben den möglichen Schlüsseln auch den eigentlichen Algorithmus zum Verschlüsseln fest. Dabei ist durch den Verschlüsselungsalgorithmus und der Wahl eines Schlüssels auch das Entschlüsseln klar definiert, der Algorithmus ist in diesem Sinne umkehrbar. Dies wird jedoch in der Bezeichnung „Verschlüsselungsverfahren“ vernachlässigt. Wir nennen ein Verschlüsselungsverfahren **symmetrisch**, wenn der Schlüssel zum Entschlüsseln mit dem Schlüssel zum Verschlüsseln identisch ist bzw. leicht aus diesem bestimmt werden kann. Wir nennen ein Verfahren **asymmetrisch**, wenn diese Schlüssel in

keiner erkennbaren Beziehung zueinander stehen.

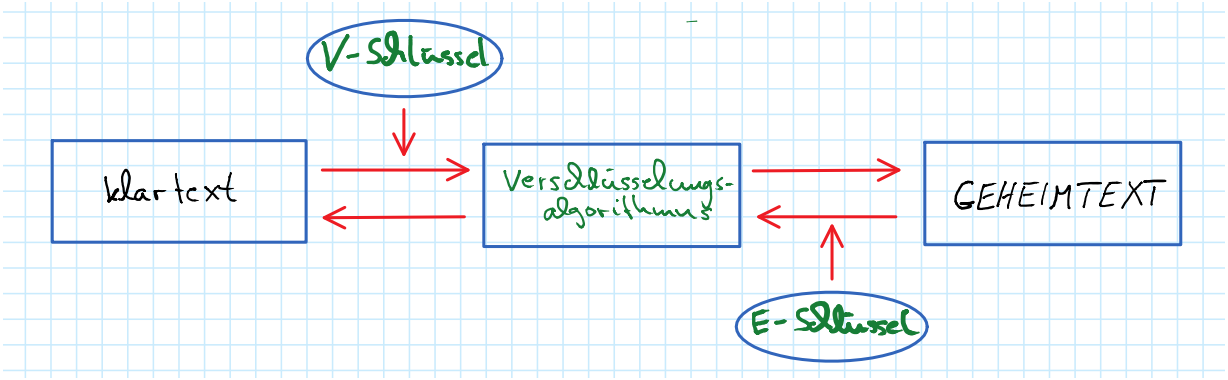

Abbildung 1.2: Verschlüsselungsverfahren: Der entsprechende Schlüssel wird dem Algorithmus als Parameter übergeben.

Auguste Kerckhoffs (1835-1903) niederländischer Linguist und Kryptologe schrieb: „Die Sicherheit eines Kryptosystems [Verschlüsselungsverfahren] darf nicht von der Geheimhaltung des Algorithmus abhängen. Die Sicherheit gründet sich nur auf die Geheimhaltung des Schlüssels“ (siehe Zitat in [2] S. 27). Zu ergänzen ist hier: Die Sicherheit eines Verschlüsselungsverfahrens wird nicht allein durch die Geheimhaltung des Schlüssels gewährleistet, nötig ist auch eine große Anzahl von möglichen Schlüsseln (vgl. [2] S. 27). Wir kommen bereits im Rahmen der Caesar-Substitution darauf zurück.

Mathematik spielt in allen Teilgebieten der Kryptologie eine wesentliche Rolle. Durch sie erhält man z. B. die Möglichkeit, die Sicherheit eines Verschlüsselungsverfahrens abzuschätzen. Zudem kann man mit ihrer Hilfe sicherere Verfahren entwickeln und verbessern. Welche mathematischen Methoden und Werkzeuge dabei in der Kryptologie zum Einsatz kommen, wollen wir in der Schnuppervorlesung vorstellen. Dabei werden wir ganz unterschiedliche Gebiete aus der Mathematik kennenlernen.

Kapitel 2

Klassische Verschlüsselungsverfahren

2.1 Transposition und monoalphabetische Substitution

Nachrichten, die wir verschlüsseln wollen, bestehen aus Zeichen. Die Menge der Zeichen, die in einem Text auftreten, bezeichnen wir als **Alphabet**. Meist wird unser **Klartextalphabet** das uns wohlbekannte Alphabet {a,b,c, . . . , x,y,z} sein, wobei Umlaute wie ä als ae geschrieben werden, ß wird durch ss ersetzt. Hier ist es üblich Kleinbuchstaben zu wählen. Das **Geheimtextalphabet** kann dasselbe sein, man kann aber Zahlen verwenden oder eigene Zeichen erfinden z. B. $\{0, 1, 2, \ldots, 25\}$ oder $\{*\sharp\circ\star\divideontimes\bigtriangledown\rtimes\ltimes\amalg\otimes\odot\circledcirc\circledast\bigtriangleup\cup\wr\sqcap\vee\infty\lhd\flat\heartsuit\diamondsuit\perp\ni\blacksquare\}$. Wir werden in der Regel die Großbuchstaben {A,B,C, . . . , X,Y,Z} als Geheimtextalphabet nutzen.

Bei sogenannten **Transpositionen** als Verschlüsselungsverfahren werden die Buchstaben einer Nachricht einfach in einer anderen Reihenfolge angeordnet. Es entsteht ein Anagramm. So besitzt der sehr kurze Klartext „tor“ z. B. nur die Geheimtexte „TOR“ „OTR“ „ORT“ „RTO“ „ROT“ und „TRO“, wobei der erste wohl nicht wirklich als geheim angesehen werden kann. Doch schon der einfache Satz „betrachten wir zum beispiel diesen satz“ mit 34 Buchstaben besitzt mehr als 14830 000 000 000 000 000 000 000 000 Anagramme, wobei wir die Leerzeichen ignorieren.

Bei einer monoalphabetischen (griechisch *μόνος* (mónos) einzig, allein) Substitution (lateinisch *substituere* ersetzen) wird jeder Klartextbuchstabe durch einen Geheimtextbuchstaben ersetzt. In einem engeren Sinne wird jedem Buchstaben des Klartextalphabets genau ein fester Buchstabe des Geheimalphabets zugeordnet wird. Fasst man den Begriff etwas weiter, so kann ein- und derselbe Klartextbuchstabe auch durch mehrere Geheimtextbuchstaben repräsentiert werden. Dies ist natürlich nur möglich, wenn das Geheimtextalphabet eine größere Anzahl von Buchstaben enthält als das Klartextalphabet. Wichtig im Hinblick einer eindeutigen Entschlüsselung ist jedoch, dass jeder Geheimtextbuchstabe nur für genau einen Klartextbuchstaben steht.

2.1.1 „Gartenzaun“-Transposition

Wir haben oben beim Beispielsatz bereits bemerkt, dass die Anzahl von Transpositionen also die Anordnung von z. B. 34 Buchstaben, von denen einige mehrfach vorkommen, sehr sehr groß ist. Was hinter einer solchen Transposition aus mathematischer Sicht steckt, schauen wir uns weiter unten genauer an. Bei der Vielzahl von Anordnungen, die eine Nachricht verschlüsseln können, ist zu bedenken, dass auch der Empfänger in der Lage sein muss, die Anordnung rückgängig zu machen, um so den Geheimtext zu entschlüsseln. Daher sind einfache Transpositionen wie z. B. die „Gartenzaun“-Transposition sehr populär:

eine transposition ist aus mathematischer sicht eine permutation der positionen

E N T A S O I I N S A S A H M T S H R I H E N P R U A I N E P S T O E
I E R N P S T O I T U M T E A I C E S C T I E E M T T O D R O I I N N

Der Geheimtext lautet: ENTASOIINSASAHMTSHRIHENPRUAINEPSTOEIERNPSTOITUMTEAICESCTIEEMTTODROIINN

Erklärung: Die Buchstaben des Klartextes werden abwechselnd auf zwei Zeilen geschrieben. Dann wird die Buchstabenfolge der zweiten Zeile an die der erste angehängt.

Die „Gartenzaun“-Transposition kann variiert werden, indem mehr als zwei Zeilen verwendet werden.

2.1.2 Caesar-Substitution

Eines der ältesten Beispiele für eine monoalphabetische Substitution ist die Caesar-Substitution, die von Julius Caesar (100 bis 44. v. Chr.) verwendet wurde. Die Zeichen des Geheimtextalphabets sind dabei dieselben wie die des Klartextalphabets. Man schreibt das Geheimtextalphabet einfach um einige Stellen verschoben unter das Klartextalphabet - bei Caesar waren es drei Stellen.

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
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

Insgesamt kann man aus 26 Verschiebungen wählen, wobei die Verschiebungen mit dem Schlüsselbuchstaben A trivial ist, es gibt also nur 25 sinnvolle Schlüssel. Im Jahre 1470 erfand Leon Battista Alberti ein simples Hilfsmittel für die Caesar-Verschlüsselung: Die Caesarscheibe.
Die Caesar-Verschlüsselung ist sehr schnell zu knacken. Es genügt, einfach systematisch die Schlüssel durchzuprobieren. Spätestens im 25. Versuch ist man erfolgreich. Man kann jedoch geschickter vorgehen (vgl. Abschnitt 2.1.4).

Beispiel 1.

Klartext:	q	u	a	n	t	e	n	c	o	m	p	u	t	e	r
Geheimtext:	T	X	D	Q	W	H	Q	F	R	P	S	X	W	H	U

Man könnte nun also einfach versuchen, eine größere Anzahl an möglichen Schlüsseln zu erzeugen, um eine höhere Sicherheit zu gewährleisten. Dies könnte man z. B. dadurch erreichen, dass man das Alphabet nicht nur verschiebt, sondern die 26 Buchstaben des Geheimtextalphabets in beliebiger Reihenfolge unter das Klartextalphabet schreibt (Anagramm des Geheimtextalphabets), z. B.:

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

In der Mathematik nennt man solch eine Umordnung von Objekten eine Permutation.

Zusammenfassung: Bei der Transposition bleibt sich jeder Buchstabe gleich, er wechselt jedoch seine Position, während bei der Substitution jeder Buchstabe seine Gestalt wechselt, doch seine Position behält.

2.1.3 Die Mathematik dahinter: Permutationen

Definition 1. Sei $\mathcal{M}$ eine Menge von N verschiedenen Objekten (z. B. $\mathcal{M} = \{1, \ldots, N\}$ oder $\mathcal{M} = \{\mathrm{a}, \mathrm{b}, \ldots, \mathrm{z}\}$ also hier $N = 26$). Wir nennen eine Funktion $\pi \colon \mathcal{M} \to \mathcal{M}$ eine Permutation auf $\mathcal{M}$, falls für zwei verschiedene Objekte $m, n \in \mathcal{M}$ auch $\pi(m)$ und $\pi(n)$ verschieden sind, d. h. für alle $m, n \in \mathcal{M}$ gilt

$$m \neq n \Rightarrow \pi(m) \neq \pi(n).$$

Bemerkung 1. Mit anderen Worten: Wir nennen $\pi \colon \mathcal{M} \to \mathcal{M}$ eine Permutation auf $\mathcal{M}$, falls keine zwei Elemente aus $\mathcal{M}$ auf das gleiche Element abgebildet werden.

Beispiel 2. Permutationen auf der Menge $\{1, \ldots, N\}$ können auf die folgende Art notiert werden

$$\pi = \begin{pmatrix} 1 & 2 & \ldots & N \\ \pi(1) & \pi(2) & \ldots & \pi(N) \end{pmatrix}$$

wobei in der unteren Zeile jede der Zahlen $1, \ldots, N$ genau einmal vorkommt.

Ordnet man N verschiedene Objekte in einer Reihe an, so kann man sich überlegen, wie viele Möglichkeiten es dafür gibt: $N \cdot (N-1) \cdot (N-2) \cdots 2 \cdot 1$. Man schreibt für das Produkt kurz $N!$ und sagt dazu N-Fakultät. Es gibt also $N!$ viele verschiedene Permutationen auf einer N-elementigen Menge.

Anwendung auf die Transposition

Jeder Schlüssel ist eine Permutation $\pi\colon \{1,\dots,N\} \to \{1,\dots,N\}$, wobei N die Anzahl der Buchstaben der Nachricht ist:

Klartext: $b_1 b_2 \dots b_N$

Geheimtext: $b_{\pi(1)} b_{\pi(2)} \dots b_{\pi(N)}$

Bemerkung 2. Nachteil: Ist der Schlüssel als Permutation der Zahlen $\{1,\dots,N\}$ genauso lang wie die Nachricht selbst, so wird das Problem der sicheren Übermittlung einer Nachricht komplett auf die ebenso aufwendige Übermittlung des Schlüssels verschoben. Stattdessen verwendet man, wie bei der „Gartenzaun"-Transposition von der Länge N unabhängige Schlüssel.

Anwendung auf die monoalphabetische Substitution

Jeder Schlüssel ist eine Permutation $\pi\colon \{\mathrm{a},\mathrm{b},\dots,\mathrm{z}\} \to \{\mathrm{a},\mathrm{b},\dots,\mathrm{z}\}$.

Klartext: $b_1 b_2 \dots b_N$

Geheimtext: $\pi(b_1)\pi(b_2)\dots\pi(b_N)$

In diesem Fall besteht die Menge, das Alphabet, aus 26 Buchstaben. Der Schlüssel besteht aus der Reihenfolge in der wir die Buchstaben des Geheimtextalphabets unter das Klartextalphabet schreiben. Nach obiger Überlegung gibt also $26! = 403291461126605635584000000 \approx 4 \cdot 10^{26}$ mögliche Anordnungen der Buchstaben im Geheimtextalphabet und damit ebensoviele Schlüssel. Einfaches Ausprobieren von Hand wie im Falle der Caesar-Verschiebung ist hier also nicht mehr zu realisieren.
Dass man diese Art der Verschlüsselung mit Methoden der Statistik dennoch leicht knacken kann, werden wir im Folgenden sehen.

2.1.4 Kryptoanalyse der monoalphabetischen Substitution: Häufigkeitsanalyse

In Texten, die in natürlichen Sprachen verfasst sind, kommen die Buchstaben des zugrundeliegenden Alphabets unterschiedlich häufig vor. Tatsächlich hat jede Sprache eine eigene Charakteristik: Zwar kommt sowohl im Deutschen, Englischen, Französischen, Spanischen, Italienischen und Türkischen der Buchstabe E am häufgsten vor, jedoch findet man bei den zweithäufigsten Buchstaben stärkere Unterschiede zwischen den Sprachen. Hat man einen Geheimtext vorliegen und vermutet, dass dieser monoalphabetisch verschlüsselt ist, zählt man, wie oft ein bestimmtes Geheimtextzeichen vorkommt, notiert die Anzahl und setzt sie in Verhältnis zur Gesamtanzahl der Buchstaben des vorliegenden Textes. So erhält man relative Häufigkeiten für das Vorkommen der verschiedenen Buchstaben. Nun kann man diese mit bekannten Werten vergleichen: Im Deutschen sticht z. B. der Buchstabe E mit einer besonders hohen relativen Häufigkeit von ca. 17% heraus, er ist damit mit großem Abstand der häufigste Buchstabe. Der zweithäufigste Buchstabe im Deutschen ist – mit ca. 9% – das N.

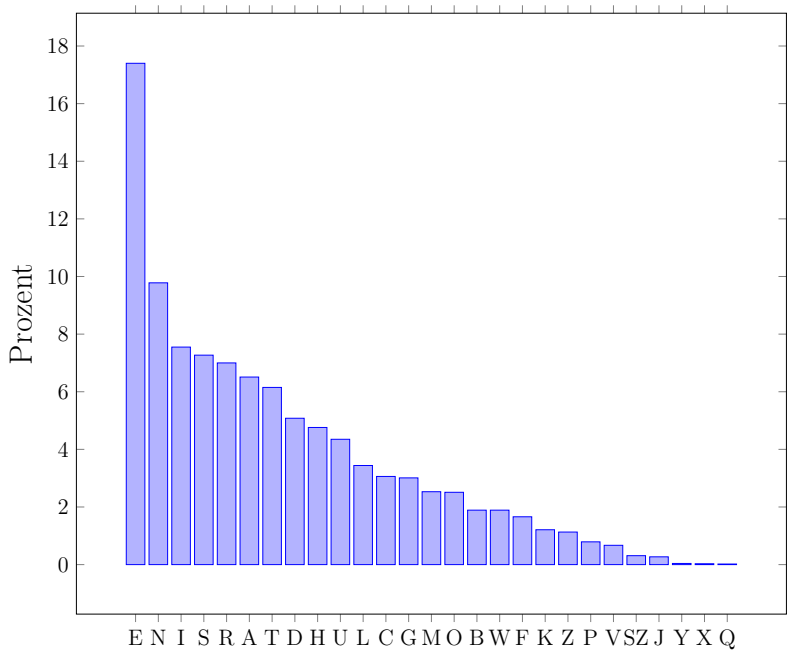

Abbildung 2.1: Relative Häufigkeiten der Buchstaben im Deutschen

Die für die deutsche Sprache typische Häufigkeitsverteilung ist in Abbildung 2.1 zu sehen. Dieser Häufigkeitsanalyse liegt ein Text mit 1000 Zeichen zugrunde.
Die Häufigkeitsanalyse ist ein nützliches Werkzeug beim Knacken einer verschlüsselten Nachricht. Geht man davon aus, dass ein Geheimtext in einer bestimmten Sprache z. B. in deutsch verfasst und dann monoalphabetisch verschlüsselt wurde, so kann man durch einen Vergleich der Häufigkeitsanalyse des Geheimtextes mit der typischen Häufigkeitsverteilung im Deutschen, einzelne Buchstaben, wie z. B. e, n, i , s , r, a und t identifizieren. Auch sehr hilfreich ist die Häufigkeitsanalyse von bestimmten Buchstabenfolgen. Eine Buchstabenfolge der Länge N wird als N-Gramm bezeichnet, oft werden als Vorsilben auch die griechischen Zahlwörter verwendet. Buchstabenpaare werden demnach Bigramme genannt. Im Deutschen besonders häufig auftretende Bigramme sind z. B. ER, EN. Die Bigramme IE und EI haben eine Ausnahmestellung, da I und E das einzige Buchstabenpaar ist, für die beide Kombinationen in etwa gleich oft zu finden sind. Während der Buchstabe C alleine sehr selten bis gar nicht auftritt, kommt er als CH in der Kombination mit H, einem etwas häufiger auftretenden Buchstaben, sehr oft vor, vergleiche hierzu auch Abbildung 2.2.
In der Regel kommt man beim Knacken einer monoalphabetischen Verschlüsselung schon sehr weit, wenn man die relativen Häufigkeiten der einzelnen Buchstaben und der wichtigen Bigramme kennt. Man sollte aber immer im Hinterkopf behalten, dass relative Häufigkeiten von der Länge und Art eines Textes abhängig sind. Die Nachricht „Tarnung intakt. Missionsstart nach Plan am Montag acht Uhr“. kommt ohne ein einziges E aus. Der französische Schriftsteller Georges Perec schrieb sogar seinen 300-seitigen Roman „La Disparition“, ohne einmal den Buchstaben E zu verwenden, der im Französischen, genau wie im Deutschen statistisch gesehen der häufigste ist.

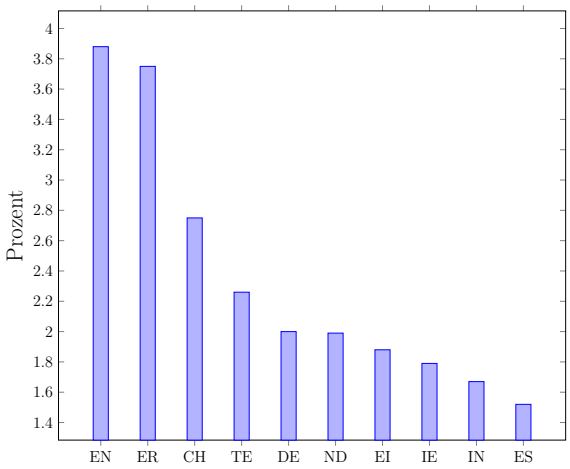

Abbildung 2.2: Relative Häufigkeiten der meistauftretenden Bigramme im Deutschen

2.2 Polyalphabetische Verschlüsselungsverfahren

2.2.1 Vigenère-Verschlüsselung

Die Häufigkeitsanalyse soll schon im 7. Jahrhundert benutzt worden sein, um monoalphabetische Verschlüsselungen zu knacken. Man suchte also eine Verschlüsselungstechnik, bei der ein Angriff mittels Häufigkeitsanalyse nicht mehr funktionieren sollte. Eine naheliegende Idee wäre z. B. den häufiger auftretenden Buchstaben mehrere Geheimtextbuchstaben zuzuordnen (*homophone Substitution*). Eine andere Idee hatte Blaise de Vigenère, ein französischer Diplomat, im Jahre 1586. Er schlug eine Verschlüsselung vor, die mehrere monoalphabetische Verschlüsselungen miteinander kombiniert.

Bei der Vigenère-Verschlüsselung werden mehrere Caesar-Verschiebungen angewendet. Ein wiederholt über den Klartext geschriebenes Schlüsselwort gibt an, welcher Caesar-Schlüssel jeweils verwendet werden soll:

Schlüsselwort:	GEH	EIMGEHE	IMG	EH	EIM	GEHEIM
Klartext:	wir	treffen	uns	an	der	kirche.
Geheimtext:	CMY	XZQLJLR	CZY	EU	HMD	QMYGPQ.

Beim Ver-und Entschlüsseln ist das Vigenère-Quadrat sehr hilfreich (siehe Tabelle 2.1).

Dass hier die Häufigkeitsanalyse nicht mehr funktioniert, kann man im Beispiel unten sehen (siehe Abbildung 2.3). In dieser Abbildung sieht man links die Häufigkeitsverteilung der Buchstaben im Klartext, der in deutscher Sprache verfasst wurde, rechts wurde der Text mit der Technik von Vigenère und dem Schlüsselwort *Schokolade* verschlüsselt.

Ungefähr 300 Jahre lang wurde der Vigenère-Code nicht geknackt. Im Jahre 1854 schließlich fand der Mathematiker Charles Babbage eine Methode, die er allerdings geheim hielt. Neun

Schlüsselwortbuchstabe \ Klartextbuchstabe	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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

Tabelle 2.1: Vigenère-Quadrat

Jahre später kam der preusische Offizier Friedrich Kasiski auf dieselbe Idee, er veröffentlichte seine Methode, die deshalb den Namen *Kasiski-Test* trägt: Die wesentliche Aufgabe besteht darin, die richtige Schlüsselwortlänge zu ermitteln. Kennt man die Schlüsselwortlänge, so kann man in den Fällen, bei denen der Geheimtext entsprechend länger als das Schlüsselwort ist, mit Hilfe von Häufigkeitsanalysen auch das Schlüsselwort herausbekommen und so den Code knacken.

Bestimmung der Schlüsselwortlänge

Ist der Geheimtext im Vergleich zum Schlüsselwort lang, so kann es passieren, dass ein Wort oder N-Gramm mehrmals mit denselben Buchstaben verschlüsselt wird. Bei einem ungünstigen Zusammenspiel von Schlüsselwortlänge und Klartext, kann dies sogar in einem relativ kurzen Text passieren:

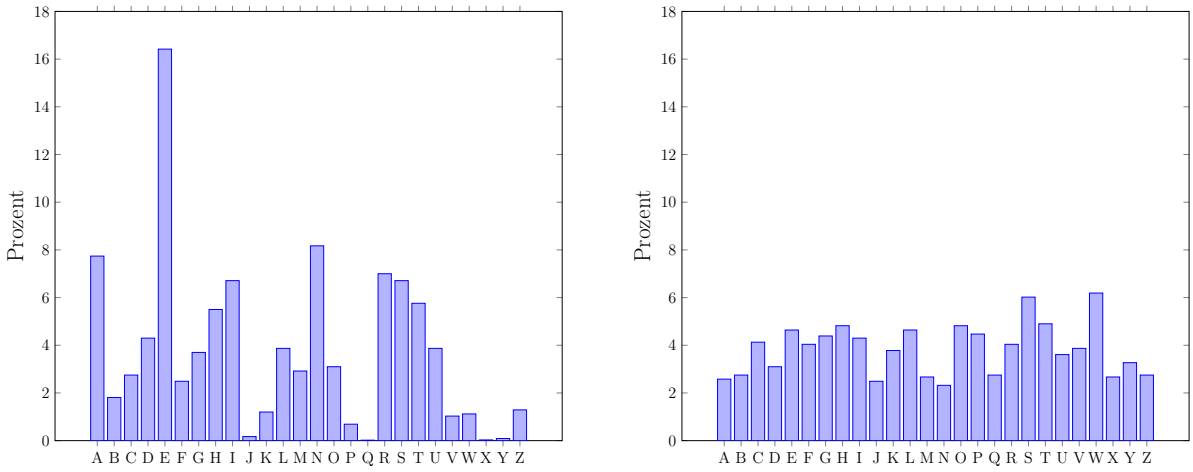

Abbildung 2.3: Vergleich der relativen Buchstabenhäufigkeiten, links Häufigkeitsverteilung im Klartext, rechts Häufigkeitsverteilung des Vigenère-verschlüsselten Textes.

```
wirtreffenunsanderkirchedirtraueichzunichtsanderenzuverraten
CMYXZQLJLRCZYEUHMDQMYGPQJMYXZMAIPGPLARPGPFYEUHMDKRGYDQXVHXMZ
```

Die Klartextbuchstaben irtr stehen hier jeweils an denselben Stellen des Schlüsselwortes:

```
GEHEIMGEHEIMGEHEIMGEHEIMGEHEIMGEHEIMGEHEIMGEHEIMGEHEIMGEHEIM
wirtreffenunsanderkirchedirtraueichzunichtsanderenzuverraten
CMYXZQLJLRCZYEUHMDQMYGPQJMYXZMAIPGPLARPGPFYEUHMDKRGYDQXVHXMZ
```

Das passiert dann, wenn der Abstand dieser Buchstabengruppe ein Vielfaches der Schlüsselwortlänge ist.
Der Kasiski-Test macht sich nun genau dies zu Nutze: Man sucht nach bestimmten Zeichenfolgen im Geheimtext, die sich wiederholen und zählt jeweils ihren Abstand zwischen den Anfängen. In unserem Beispiel beträgt der Abstand der Folge MYXZ $24 = 2 \cdot 2 \cdot 2 \cdot 3$ Zeichen, für EUHMD zählt man $30 = 2 \cdot 3 \cdot 5$ Zeichen und für PGP sind es $6 = 2 \cdot 3$ Zeichen. Man nimmt dann an, dass die Abstände jeweils ein Vielfaches der Schlüsselwortlänge sind. Mit großer Wahrscheinlichkeit ist dann die Schlüsselwortlänge ein gemeinsamer Teiler der Abstände. In unserem Fall ist es sogar der größte gemeinsame Teiler: $\mathrm{ggT}(6, 24, 30) = 6$.

Bestimmung des Schlüsselworts
Kennt man die Länge des Schlüsselworts und ist dieses im Verhältnis zum Text nicht zu lang, so kann man nun das Schlüsselwort selbst herausbekommen. In unserem Fall wissen wir, dass die Buchstaben Nr. 1,7,13,19,25,... immer mit derselben Caesar-Substitution verschlüsselt wurden. Gleiches gilt für die Buchstaben Nr. 2,8,14,20,26,.... Wir können nun also die Buchstaben in Gruppen zusammenfassen und auf diese Häufigkeitsanalysen anwenden, um jeweils die zugrundeliegende Caesar-Substitution zu bestimmen. Ist das Schlüsselwort ein lexikalisches und nicht zu lang, hat man in der Regel mit dieser Methode Erfolg. Schwieri-

ger wird es schon, wenn das Schlüsselwort aus einer zufälligen Buchstabenfolge besteht. Wir werden im nächsten Abschnitt hieraus eine unknackbare Verschlüsselungsmethode herleiten.

Friedman-Test

Wir betrachten eine Folge der Buchstaben $\{\mathrm{a}, \mathrm{b}, \ldots, \mathrm{z}\}$ von insgesamt N Buchstaben. Wir wollen zunächst klären, wieviele ungeordnete Buchstabenpaare aus diesen N Buchstaben gebildet werden können: Für den ersten Buchstaben haben wir N viele Möglichkeiten. Für jede dieser Möglichkeiten gibt es für die Wahl des zweiten Buchstabens noch $N-1$ Möglichkeiten, also insgesamt $N(N-1)$ viele (geordnete) Paare. Verzichten wir auf die Anordnung der beiden gewählten Buchstaben, gibt es insgesamt

$$\frac{N(N-1)}{2} \text{ ungeordnete Paare.}$$

Seien nun n_0 viele Buchstaben der Folge gleich a, n_1 viele gleich b usw. Dann gibt es, analog zu den Überlegungen oben, z. B. $n_0(n_0-1)/2$ viele Paare (a, a) und z. B. $n_{25}(n_{25}-1)/2$ viele Paare (z, z). Die Wahrscheinlichkeit solcher Paare gleicher Buchstaben ergibt sich nach dem Motto „Anzahl der günstigen Fälle“ durch „Anzahl der möglichen Fälle“, also z. B. für das Paar (c, c) die Wahrscheinlichkeit

$$\frac{n_2(n_2-1)}{2} : \frac{N(N-1)}{2} = \frac{n_2(n_2-1)}{N(N-1)}.$$

Die Gesamtwahrscheinichkeit, dass ein Paar gleicher Buchstaben auftritt heißt **Koinzidenzindex** der Buchstabenfolge und wird mit I (manchal auch κ) bezeichnet:

$$I = \sum_{k=0}^{25} \frac{n_i(n_i-1)}{N(N-1)}.$$

Bezeichne z. B. p_0 die Wahrscheinlichkeit, dass in einem deutschen Text der Buchstabe a auftritt, p_1 die Wahrscheinlichkeit für b usw., so lässt sich der Koinzidenzindex der deutschen Sprache näherungsweise mit Hilfe von Häufigkeitsanalysen wie folgt bestimmen:

$$I = \sum_{k=0}^{25} p_k^2 \approx 0.0762.$$

Ein hinreichend langer Text in deutscher Sprache weist also einen Koinzidenzindex von ungefähr 0.0762 auf, während für einen „Text“, bei dem die 26 Buchstaben $\{\mathrm{a}, \ldots, \mathrm{z}\}$ gleichverteilt auftreten ($p_k = \frac{1}{26}$), gilt

$$I = \sum_{k=0}^{25} p_k^2 = 26\frac{1}{26^2} = \frac{1}{26} \approx 0.0385.$$

Bemerkung 3. Bei einer monoalphabetischen Substitution ist der Koinzidenzindex vom Klartext mit dem vom Geheimtext identisch, da ja nicht die Art der Buchstaben, sondern nur die Anzahl des Auftretens gleicher Buchstaben in die Berechnung des Koinzidenzindexes eingeht. Ist ein Geheimtext jedoch durch ein polyalphabetische Verfahren verschlüsselt worden, so liegt der Koinzidenzindex deutlich näher bei 0.0385 als bei 0.0762. Man kann also mit dem Koinzidenzindex testen, ob eine mono- oder polyalphabetische Verschlüsselung vorliegt.

Wir wollen nun mit dem Koinzidenzindex eine Näherung für die Schlüsselwortlänge ermitteln. Wir nehmen dazu an, dass die Schlüssellänge l sei. Teilen wir die Buchstaben eines Geheimtextes entsprechend ihrer Position in l Klassen (siehe Tabelle unten und den Abschnitt über Kongruenzarithmetik **??**), so sind die Buchstaben innerhalb einer Klasse monoalphabetisch verschlüsselt, besitzen also einen Koinzidenzindex ungefähr gleich 0.0762. In einer Klasse liegen ungefähr N/l viele Buchstaben.

$\equiv 1 \mod l$	$\equiv 2 \mod l$	$\equiv 3 \mod l$	$\cdots$	$\equiv 0 \mod l$
1	2	3		l
$l+1$	$l+2$	$l+3$		$2l$
$2l+1$	$2l+2$	$2l+3$		$3l$
$3l+1$	$3l+2$	$3l+3$		$4l$
$\vdots$	$\vdots$	$\vdots$		$\vdots$

Wie groß ist die Wahrscheinlichkeit, dass ein Buchstabenpaar des Textes in einer gemeinsamen Spalte liegt (genauer die Nummer ihrer Position) und aus gleichen Buchstaben besteht? Diese Wahrscheinlichkeit ist das Produkt aus der Wahrscheinlichkeit, dass die Buchstaben in einer Spalte liegen $N(N/l-1)/(N(N-1))$, und dem Faktor 0.762, also der Wahrscheinlichkeit, dass es sich um ein Paar gleicher Buchstaben bei einer monoalphabetischen Verschlüsselung handelt:

$$\frac{N(N/l-1)}{N(N-1)} \cdot 0.0762.$$

Beachte: Ist der erste Buchstabe gewählt (N Möglichkeiten), so ist die Spalte mit verbleibenden $N/l-1$ Buchstaben festgelegt.
Wie groß ist die Wahrscheinlichkeit, dass die Buchstaben eines Buchstabenpaares aus zwei unterschiedlichen Spalten stammen und gleich sind? Diese Wahrscheinlichkeit ist das Produkt aus der Wahrscheinlichkeit, dass die Buchstaben in unterschiedlichen Spalten liegen $N(N-N/l)/(N(N-1))$, und dem Faktor ungefähr 0.385, also der Wahrscheinlichkeit, dass es sich um ein Paar gleicher Buchstaben bei einer polyalphabetischen Verschlüsselung handelt:

$$\frac{N(N-N/l)}{N(N-1)} \cdot 0.0385.$$

Beachte: Ist der erste Buchstabe gewählt (N Möglichkeiten), so können $N-N/l$ Buchstaben aus den übrigen Spalten gewählt werden.
Insgesamt folgt:

$$I \approx \frac{N(N/l-1)}{N(N-1)} \cdot 0.0762 + \frac{N(N-N/l)}{N(N-1)} \cdot 0.0385 = \frac{0.0377N}{l(N-1)} + \frac{0.0385N - 0.0762}{N-1}$$

Beachte: $0.0762 - 0.0385 = 0.0377$. Berechnen wir für den vorliegenden Geheimtext I und lösen wir diese Gleichung nach l auf,

$$l = \frac{0.0377N}{I(N-1) - 0.0385N + 0.0762},$$

so können wir die Schlüsselwortlänge näherungsweise ermitteln.

2.2.2 One-Time-Pad - eine unknackbare Verschlüsselung

Wählt man bei der Vigenère-Verschlüsselung das Schlüsselwort nach drei Regeln, bekommt man eine unknackbare Verschlüsselung:

1. Das Schlüsselwort muss aus genauso vielen Zeichen wie der Klartext bestehen.
2. Die Zeichenfolge des Schlüsselworts muss zufällig sein, sie darf nicht vorhergesagt werden können.
3. Jedes Schlüsselwort darf nur ein einziges Mal verwendet werden.

Doch diese unknackbare Verschlüsselung hat auch einen Nachteil: Sie ist wenig praktikabel. Sender und Empfänger müssen irgendwann zuvor den eventuell sehr langen Schlüssel ausgetauscht haben. Das One-Time-Pad wird vor allem dann eingesetzt, wenn allerhöchste Sicherheit gefordert ist, wie z. B. beim so genannten roten Telefon, dass 1962 als Verbindung zwischen den obersten Regierungsstellen der USA und der UdSSr eingerichtet wurde.

2.3 Die Enigma-Verschlüsselungsmaschine

Kapitel 3

Moderne Verschlüsselungsverfahren

3.1 Zahlentheoretische Grundlagen

Satz 1 (Division mit Rest)**.** *Seien $a \in \mathbb{Z}$ und $b \in \mathbb{N}$. Dann existieren eindeutig bestimmte $q, r \in \mathbb{Z}$ mit*

$$a = q \cdot b + r, \qquad 0 \leq r < b.$$

Man nennt r den Rest, der bei Division von a durch b entsteht.

Beweis. Zur Existenz: Wir setzen $q = 0$ und $r = a$.
Im Fall $a \geq 0$ führen wir folgende Schleife aus:

while $r \geq b$:
    $q = q + 1$
    $r = r - b$

Im Fall $a < 0$ führen wir folgende Schleife aus:

while $r < 0$:
    $q = q - 1$
    $r = r + b$

Zur Eindeutigkeit: Sei $a = qb + r = \tilde{q}b + \tilde{r}$ mit $0 \leq r, \tilde{r} < b$, wobei ohne Einschränkung $r \geq \tilde{r}$. Dann folgt nach Bildung der Differenz beider Darstellungen:

$$0 = (q - \tilde{q})b + r - \tilde{r},$$

also $0 \leq r - \tilde{r} = (\tilde{q} - q)b$. Somit ist $0 \leq r - \tilde{r} < b$ ein Vielfaches von b. Dies ist nur für $r = \tilde{r}$, also auch $q = \tilde{q}$ möglich. □

Definition 2 (Teiler)**.** Eine Zahl $d \in \mathbb{N}$ heißt Teiler von $a \in \mathbb{Z}$, falls $q \in \mathbb{Z}$ mit $a = qd$ existiert ($r = 0$ im Satz oben). In Zeichen $d \mid a$.

Zu $a, b \in \mathbb{Z}$ sei $D \coloneqq \{d \in \mathbb{N} \colon d \mid a \text{ und } d \mid b\}$ die Menge der gemeinsamen Teiler von a und b. Es gilt einerseits $1 \in D$ und im Fall $ab \neq 0$ gilt für $d \in D$ die Abschätzung $d \leq \min\{|a|, |b|\}$.

Definition 3 (**g**rößter **g**emeinsamer **T**eiler)**.** Zu $a, b \in \mathbb{Z}$ nennen wir

$$\mathrm{ggT}(a, b) \coloneqq \max\{d \in \mathbb{N} \colon d \mid a \text{ und } d \mid b\}$$

den gröten gemeinsamen Teiler von a und b.

Definition 4 (teilerfremd)**.** Wir nennen $a, b \in \mathbb{Z}$ teilerfremd, falls $\mathrm{ggT}(a, b) = 1$ gilt.

Beobachtungen zum größten gemeinsamen Teiler:

- Für $a, b \in \mathbb{Z}$ gilt

  $$\mathrm{ggT}(a, b) = \mathrm{ggT}(-a, b) = \mathrm{ggT}(a, -b) = \mathrm{ggT}(-a, -b).$$

  Sei also im Folgenden ohne Einschränkung $a, b \in \mathbb{N}$.

- Für $a, b \in \mathbb{N}$ gilt

  $$\mathrm{ggT}(a, b) = \mathrm{ggT}(b, a)$$

  Sei also im Folgenden ohne Einschränkung $0 < b \leq a$.

Basierend auf dem Satz oben kann man mit dem sogenannten Euklidischen Divisionsalgorithmus den größten gemeinsamen Teiler $\mathrm{ggT}(a, b)$ zweier ganzer Zahlen $a, b \in \mathbb{Z}$ bestimmen: Seien dazu ohne Einschränkung $a, b \in \mathbb{N}$ mit $0 < b < a$ (siehe Beobachtungen oben).

1. Setze $r_0 \coloneqq a$, $r_1 \coloneqq b$ und $k = 1$.
2. Bestimme $r_{k-1} = q_k r_k + r_{k+1}$ mit $0 \leq r_{k+1} < r_k$.
3. Falls $r_{k+1} = 0$, so ist r_k der $\mathrm{ggT}(a, b)$.
   Sonst erhöhe k um eins und gehe zu 2.

Die entscheidende Beobachtung ist $\mathrm{ggT}(a, b) = \mathrm{ggT}(r_k, r_{k+1})$. Um dies einzusehen, betrachten wir die Division mit Rest in Schritt 2 des Algorithmus etwas genauer. Es ist

$$r_{k-1} = q_k r_k + r_{k+1}.$$

Ist q ein Teiler von r_{k-1} und r_k, so muss q aufgrund der Summe auch r_{k+1} teilen. Ist umgekehrt q ein Teiler von r_k und r_{k+1}, so teilt q auch r_{k-1}. Mit dieser leicht nachvollziehbaren Argumentation erhalten wir

$$\mathrm{ggT}(r_k, r_{k+1}) = \mathrm{ggT}(r_{k-1}, r_k) = \cdots = \mathrm{ggT}(r_0, r_1).$$

Somit wird das Bestimmen des größten gemeinsamen Teilers zweier Zahlen auf das Bestimmen des größten gemeinsamen Teilers zweier kleinerer Zahlen zurückgeführt. Dabei spielt die Ungleichung $r_{k+1} < r_k$ im zweiten Schritt des Algorithmus eine wichtige Rolle. Die Reste werden immer echt kleiner. Schließlich muss für einen Index k_0 die Gleichheit $r_{k_0+1} = 0$ gelten. Einerseits gilt dann

$$r_{k_0-1} = q_k r_{k_0},$$

also $\mathrm{ggT}(r_{k_0}, 0) = \mathrm{ggT}(r_{k_0-1}, r_{k_0}) = r_{k_0}$, andererseits ist nach den Überlegungen oben auch $\mathrm{ggT}(a, b) = r_{k_0}$.

Beispiele 3.

(a) Der ggT von 1632 und 164 ist 4. Denn

$$\begin{aligned} 1632 &= 9 \cdot 164 + 156 \\ 164 &= 156 + 8 \\ 156 &= 19 \cdot 8 + 4 \\ 8 &= 2 \cdot 4 + 0. \end{aligned}$$

(b) Der ggT von -1535 und 123 ist 1. Denn

$$\begin{aligned} 1535 &= 12 \cdot 123 + 59 \\ 123 &= 2 \cdot 59 + 5 \\ 59 &= 11 \cdot 5 + 4 \\ 5 &= 1 \cdot 4 + 1 \\ 4 &= 4 \cdot 1 + 0. \end{aligned}$$

Satz 2. *Seien $a, b \in \mathbb{Z}$. Dann existieren Zahlen $x, y \in \mathbb{Z}$ mit*

$$\mathrm{ggT}(a, b) = ax + by$$

(man nennt eine solche Gleichung lineare diophantische Gleichung).

Beweis. Der erweiterte Euklidische Divisionsalgorithmus liefert eine Lösung der Gleichung $ax + by = \mathrm{ggT}(a, b)$. Sei $r_0 \coloneqq a$ und $r_1 \coloneqq b$.

$$\begin{aligned} r_0 &= q_1 r_1 + r_2 \\ r_1 &= q_2 r_2 + r_3 \\ &\vdots \\ r_{n-3} &= q_{n-2} r_{n-2} + r_{n-1} \\ r_{n-2} &= q_{n-1} r_{n-1} + \mathrm{ggT}(a, b) \\ r_{n-1} &= q_n \,\mathrm{ggT}(a, b) + 0 \end{aligned} \qquad \begin{aligned} \mathrm{ggT}(a, b) &= r_{n-2} - q_{n-1} r_{n-1} \\ &= r_{n-2} - q_{n-1}[r_{n-3} - q_{n-2} r_{n-2}] \\ &\vdots \\ &= \tilde{y} r_1 + \tilde{x} r_2 \\ &= \tilde{y} r_1 + \tilde{x}[r_0 - q_1 r_1] \\ &= \tilde{x} r_0 + (\tilde{y} - \tilde{x} q_1) r_1 \end{aligned}$$

Man führt, wie oben beschrieben, einen Euklidischen Divisionsalgorithmus durch und erhält nach endlichen vielen Schritten:

$$r_{n-2} = q_{n-1} r_{n-1} + \mathrm{ggT}(a, b).$$

Aufgelöst nach dem größten gemeinsamen Teiler lautet diese Gleichung

$$\mathrm{ggT}(a, b) = r_{n-2} - q_{n-1} r_{n-1}.$$

Die Idee besteht nun darin mit Hilfe der bereits durch den Euklidischen Algorithmus berechneten Reste, Reste mit höherem Index durch Reste mit kleinerem Index auszudrücken und letztlich nur noch r_0 und r_1 zu benutzen:

$$\begin{aligned} \mathrm{ggT}(a, b) &= r_{n-2} - q_{n-1} r_{n-1} = r_{n-2} - q_{n-1}[r_{n-3} - q_{n-2} r_{n-2}] \\ &= -q_{n-1} r_{n-3} + (1 + q_{n-1} q_{n-2}) r_{n-2} \\ &= -q_{n-1} r_{n-3} + (1 + q_{n-1} q_{n-2})[r_{n-4} - q_{n-3} r_{n-3}] \\ &\vdots \end{aligned}$$

□

Das Vorgehen wird an folgenden Beispielen leicht deutlich:

Beispiele 4.

(a) Löse $1632x + 164y = 4$.

$$\begin{aligned} 1632 &= 9 \cdot 164 + 156 \\ 164 &= 156 + 8 \\ 156 &= 19 \cdot 8 + 4 \\ 8 &= 2 \cdot 4 + 0. \end{aligned} \qquad \begin{aligned} 4 &= 156 - 19 \cdot 8 \\ &= 156 - 19[164 - 156] \\ &= -19 \cdot 164 + 20 \cdot 156 \\ &= -19 \cdot 164 + 20[1632 - 9 \cdot 164] \\ &= 20 \cdot 1632 - 199 \cdot 164 \end{aligned}$$

(b) Wir lösen $-1535x + 123y = 2$. Mit Beispiel 3 (b) erhalten wir

$$\begin{aligned} 1 &= 5 - 4 \\ &= 5 - [59 - 11 \cdot 5] = -59 + 12 \cdot 5 \\ &= -59 + 12[123 - 2 \cdot 59] = 12 \cdot 123 - 25 \cdot 59 \\ &= 12 \cdot 123 - 25[1535 - 12 \cdot 123] = 25 \cdot (-1535) + 312 \cdot 123. \end{aligned}$$

Die Lösung ist somit $x = 50$ und $y = 624$.

3.1.1 Restklassen

Definition 5. Zwei ganze Zahlen a, b heißen **kongruent modulo** $m \in \mathbb{N}$, wenn m beide Zahlen mit gleichem Rest teilt. Wir schreiben in diesem Fall

$$a \equiv b \mod m.$$

Wichtige Beobachtung: Es gilt: $(a \equiv b \mod m) \Leftrightarrow m|(a-b)$

Begründung dieser Äquivalenz: Zwei Zahlen besitzen beim Teilen durch m genau dann denselben Rest, wenn m die Differenz $(a-b)$ teilt, d. h., wenn $m|(a-b)$ gilt.
„⇒“: Es gilt $a = q_a m + r$ und $b = q_b m + r$. Somit finden wir für die Differenz $a - b = (q_a - q_b)m$, welche offenbar durch m teilbar ist.
„⇐“: Sei nun $a = q_a m + r_a$ und $b = q_b m + r_b$ mit $0 \leq r_a, r_b < m$. Da m die Differenz $a - b = (q_a - q_b)m + r_a - r_b$ teilt, folgt mit $-m < r_a - r_b < m$, dass $r_a - r_b = 0$ ist. Die Reste sind also identisch.

Lemma 1 (Rechenregeln der Kongruenzarithmetik)**.**
Für $(a \equiv a' \mod m)$ und $(b \equiv b' \mod m)$ gilt:

(i) $(a \pm b) \equiv (a' \pm b') \mod m$.

(ii) $(a \cdot b) \equiv (a' \cdot b') \mod m$.

(iii) Für $c \neq 0$ gilt: $\big((ca) \equiv (cb) \mod m\big) \Leftrightarrow \big(a \equiv b \mod \frac{m}{\mathrm{ggT}(c,m)}\big)$.

Beweis. Mit $a - a'$ und $b - b'$ ist auch folgender Ausdruck durch m teilbar
zu (i): $(a \pm b) - (a' \pm b') = (a - a') \pm (b - b')$
zu (ii): $ab - a'b' = ab - ab' + ab' - a'b' = a(b - b') + (a - a')b'$.
Für die Behauptung in (iii) ist folgende Gleichheit interessant:

$$\frac{ca - cb}{m} = \frac{c}{\mathrm{ggT}(c,m)} \cdot \frac{a-b}{\frac{m}{\mathrm{ggT}(c,m)}},$$

wobei $\frac{c}{\mathrm{ggT}(c,m)}, \frac{m}{\mathrm{ggT}(c,m)} \in \mathbb{Z}$ gilt. Somit gilt

$$\frac{ca - cb}{m} \in \mathbb{Z} \Leftrightarrow \frac{a-b}{\frac{m}{\mathrm{ggT}(c,m)}} \in \mathbb{Z}.$$

□

Definition 6. Für eine Zahl $a \in \mathbb{Z}$ definieren wir

$$\overline{a} = \{b \in \mathbb{Z} \colon b \equiv a \mod m\},$$

die **Restklasse** von a bzgl. m.

Beobachtungen: Es gilt: $\mathbb{Z} = \overline{0} \cup \overline{1} \cup \cdots \cup \overline{m-1}$ und $\overline{a} \neq \overline{b}$ für $0 \leq a < b \leq m-1$.
Zu $m \in \mathbb{N}$ gibt es also genau m paarweise disjunkte Restklassen. Zwei Mengen heißen disjunkt, wenn sie kein gemeinsames Element besitzen.

Beispiele 5.

(a) Sei $m = 2$. Dann ist $\overline{1}$ die Menge der ungeraden Zahlen.

(b) Sei $m = 5$. Dann ist $\overline{2}$ die Fünferreihe um „plus zwei verschoben“: $\{\ldots, -8, -3, 2, 7, 12, \ldots\}$

(c) Es gilt $\overline{m} = \overline{0}$ und $\overline{m+1} = \overline{1}$ usw.

Wir definieren über Repräsentanten eine Addition und eine Multiplikation auf der Menge der Restklassen wie folgt:

$$\overline{a} + \overline{b} := \overline{a+b}, \qquad \overline{a} \cdot \overline{b} := \overline{a \cdot b}.$$

Mit Worten: Wir wählen von zwei Restklassen Repräsentanten a und b, verknüpfen die Repräsentanten durch die Addition bzw. Multiplikation der ganzen Zahlen, die wir bereits kennen, und ordnen dem Ergebnis, dem neuen Repräsentant, seine entsprechende Restklasse zu. Klingt eigentlich ganz plausibel. Bei genauerer Betrachtung stellt sich jedoch die Frage, ob die resultierende Restklasse von der Wahl der Repräsentanten abhängig ist. Das wäre ein echtes Problem, welches die neu definierten Verknüpfungen als nicht wohldefiniert entlarven würde. Jedoch gilt für $a, a' \in \overline{p}$ und $b, b' \in \overline{q}$, also $(a \equiv a' \mod m)$ und $(b \equiv b' \mod m)$:

$$\overline{a+b} = \overline{a'+b'}, \qquad \overline{a \cdot b} = \overline{a' \cdot b'}.$$

Die Gleichheiten lassen sich mit den Beobachtungen oben und den Rechenregeln $((a \pm b) \equiv (a' \pm b') \mod m)$ und $((a \cdot b) \equiv (a' \cdot b') \mod m)$ folgern.

3.1.2 Eulersche φ-Funktion

Definition 7. Bezeichne $\varphi(m)$ die Anzahl von zu m teilerfremden Zahlen zwischen 1 und m, also

$$\varphi(m) = |\{n \in \mathbb{N} : 1 \leq n \leq m \wedge \mathrm{ggT}(m,n) = 1\}|.$$

Die Funktion heißt **Eulersche φ-Funktion**.

Beispiele 6. Es gilt:

(a) $\varphi(4) = 2$. Die teilerfremden Zahlen sind $1, 3$.

(b) $\varphi(5) = 4$. Die teilerfremden Zahlen sind $1, 2, 3, 4$.

(c) $\varphi(12) = 4$. Die teilerfremden Zahlen sind $1, 5, 7, 11$.

Satz 3. *Für eine Primzahl p und $k \in \mathbb{N}$ gilt $\varphi(p^k) = p^{k-1}(p-1)$.*

Beweis. Für $k = 1$ ist $\varphi(p) = p - 1$ klar. Die teilerfremden Zahlen sind $1, 2, \ldots, p-1$. Im Fall $k > 1$ gibt es unter den Zahlen $1, 2, \ldots, p^k - 1$ auch Zahlen, die durch p geteilt werden können, zum Beispiel $p, 2p, 3p$. Tatsächlich sind es genau die Zahlen $p, 2p, \ldots, (p^{k-1}-1)p$, also $p^{k-1} - 1$ viele Zahlen. Somit gilt

$$\varphi(p^k) = p^k - 1 - (p^{k-1} - 1) = p^k - p^{k-1} = p^{k-1}(p-1).$$

□

Zusammen mit dem folgenden Satz können wir $\varphi(m)$ für $m \in \mathbb{N}$ mit Hilfe einer Primfaktorzerlegung berechnen.

Satz 4. *Für zwei teilerfremde natürliche Zahlen a, b gilt $\varphi(a \cdot b) = \varphi(a)\varphi(b)$.*

Beweis. Beobachtung: Eine Zahl $n \in \mathbb{N}$ hat genau dann einen gemeinsamen Teiler mit $a \cdot b$, wenn sie einen gemeinsamen Teiler mit a oder einen gemeinsamen Teiler mit b hat. Wir streichen daher im folgenden Beweis von den Zahlen $1, \ldots, a \cdot b$ zunächst alle Zahlen weg, die einen gemeinsamen Teiler mit a haben, und dann die Zahlen, die einen gemeinsamen Teiler mit b besitzen. Übrig bleiben dann die Zahlen, die keinen gemeinsamen Teiler mit a und b, also auch nicht mit $a \cdot b$ besitzen.

Wir betrachten dazu die Anordnung der ersten $a \cdot b$ Zahlen in einem zweidimensionalen Schema

$$\begin{array}{llll} 1 & 2 & \cdots & a \\ 1+a & 2+a & \cdots & 2a \\ \vdots & \vdots & & \vdots \\ 1+(b-1)a & 2+(b-1)a & \cdots & b \cdot a. \end{array}$$

In der ersten Zeile gibt es genau $a-\varphi(a)$ Zahlen, die einen gemeinsamen Teiler mit a haben. Die darunter in einer Spalte stehenden Zahlen besitzen dann ebenfalls einen gemeinsamen Teiler mit a, da alle Zahlen in einer Spalte zueinander kongruent modulo a sind. Wir streichen somit im Schema die $a-\varphi(a)$ Spalten und benennen die zu a teilerfremden Zahlen der Größe nach geordnet mit $x_1, x_2, \ldots, x_{\varphi(a)}$. Es bleibt das folgende Schema mit $\varphi(a)$ Spalten und b Zeilen

$$\begin{array}{llll} x_1 & x_2 & \cdots & x_{\varphi(a)} \\ x_1+a & x_2+a & \cdots & x_{\varphi(a)}+a \\ \vdots & \vdots & & \vdots \\ x_1+(b-1)a & x_2+(b-1)a & \cdots & x_{\varphi(a)}+(b-1)a, \end{array}$$

in welchem nur teilerfremde Zahlen zu a enthalten sind. Wir zeigen nun abschließend, dass jede Spalte modulo b genau die Zahlen 0 bis $b-1$ enthält. Dann gäbe es in jeder Spalte genau $\varphi(b)$ zu b teilerfremde Zahlen und somit $\varphi(a)\cdot\varphi(b)$ viele teilerfremde Zahlen zu a und b, also zu $a\cdot b$. Damit wäre der Satz bewiesen.
Seien also x_k+ja und x_k+la zwei verschiedene Elemente, $l\neq j$, aus der k-ten Spalte und $(j+1)$-ten bzw. $(l+1)$-ten Zeile des Schemas. Wir finden

$$(x_k+ja)-(x_k+la)=(j-l)a$$

mit $0<|j-l|<b$. Somit ist die Differenz nicht durch b teilbar und die Zahlen x_k+ja und x_k+la sind nicht kongruent modulo b, sie haben einen unterschiedlichen Rest. Da dies paarweise für alle Spaltenelemente der k-ten Spalte gilt, müssen alle Reste von 0 bis $b-1$ auftreten. □

Bemerkung 4. Der Wert von $\varphi(n)$ kann also aus der Primfaktorzerlegung der Zahl n berechnet werden. Für sehr große n gibt es vermutlich keinen anderen praktikablen Weg $\varphi(n)$ zu berechnen als über die Primfaktorzerlegung. Diese Vermutung bildet eine wichtige Grundlage des RSA-Algorithmus (siehe weiter unten).

Beispiel 7. Die Primfaktorzerlegung von 1364 ist

$$1364=2^2\cdot 11\cdot 31.$$

Damit folgt aus den beiden vorangegangenen Sätzen

$$\varphi(1364)=\varphi(2^2\cdot 11\cdot 31)=\varphi(2^2)\varphi(11)\varphi(31)=2\cdot 10\cdot 30=600.$$

Zur Primfaktorzerlegung: Offensichtlich ist 2 Primfaktor von 1364. Zudem teilt $2^2=4$ die Zahl 64, jedoch $2^3=8$ nicht 364. Somit enthält die Primfaktorzerlegung von 1364 den Faktor 2 genau zweimal. Es gilt $1364=4\cdot 341$. Die Zahl 3 ist kein Primfaktor, da die Quersumme von 341 nicht durch 3 teilbar ist, 5 ist kein Primfaktor, da sie 1 nicht teilt. Da die alternierende Quersumme von 341 (nämlich 3-4+1=0) durch 11 teilbar ist, ist es auch die Zahl 341. Wir finden schließlich $2^2\cdot 11\cdot 31=1364$.

3.2 Asymmetrische Verschlüsselungsverfahren

Sender und Empfänger besitzen unterschiedliche Schlüssel und der Schlüssel zum Entschlüsseln kann nicht aus dem Verschlüsselungsschlüssel ermittelt werden. Typischerweise wird einer der Schlüssel, wie zum Beispiel beim weit verbreiteten RSA-Algorithmus (dazu später mehr), sogar öffentlich gemacht und ist somit für alle zugänglich. Der private Schlüssel hingegen muss sicher aufbewahrt werden.
Szenario der Nachrichtenübermittlung bei einem asymmetrischen Verschlüsselungsverfahren:

- Alice will Bob eine Nachricht schicken.
- Sie besorgt sich den öffentlichen Schlüssel von Bob (z. B. auf dessen Webseite).
- Sie verschlüsselt mit diesem Schlüssel die Nachricht.
- Sie sendet die verschlüsselte Nachricht an Bob.
- Jemand fängt die Nachricht ab, kann sie aber ohne den privaten Schlüssel von Bob nicht lesen.
- Bob erhält die Nachricht und kann sie mit seinem privaten Schlüssel lesen.

Bemerkungen zur Sicherheit:

1. Aus dem öffentlichen Schlüssel kann der private Schlüssel nicht ermittelt werden.
2. Aus dem Geheimtext kann der private Schlüssel nicht ermittelt werden.
3. Es sollte eine große Anzahl an privaten Schlüsseln geben.
4. Der private Schlüssel sollte niemals übermittel werden müssen.

Problematik: Absenderverifikation. Dies kann durch eine digitale Unterschrift gelöst werden:

- Alice will Bob eine Nachricht schicken.
- Sie verschlüsselt die Nachricht zunächst mit ihrem privaten Schlüssel.
- Sie besorgt sich den öffentlichen Schlüssel von Bob.
- Sie verschlüsselt mit diesem Schlüssel die Nachricht zusätzlich.
- Sie sendet die doppelt verschlüsselte Nachricht an Bob.
- Bob erhält die Nachricht und entschlüsselt sie mit seinem privaten Schlüssel.
- Bob entschlüsselt die Nachricht zusätzlich mit dem öffentlichen Schlüssel von Alice und kann sie lesen. Auf diese Weise ist er sicher, dass Alice die Nachricht gesendet hat.

Literaturverzeichnis

[1] A. Beutelspacher: Kryptologie, Eine Einführung in die Wissenschaft vom Verschlüsseln, Verbergen und Verheimlichen, Springer Spektrum, Wiesbaden, 2015[10].

[2] S. Singh: Geheime Botschaften, Die Kunst der Verschlüsselung von der Antike bis in die Zeit des Internet, dtv Verlagsgesellschaft mbH & Co. KG, München, 2017[14].