Matheschülerzirkel
Universität Augsburg
Schuljahr 2014/2015
Klasse 6/7

Zirkelzettel vom 10. Februar 2015

Die sogenannte *Skytale* ist das älteste bekannte Verschlüsselungsverfahren und stammt aus der Zeit um 500 v. Chr. In Sparta wurde damals ein Streifen Pergament oder Leder um einen Holzstab mit festem Durchmesser gewickelt. Die Nachricht, die geheim gehalten werden sollte (der sogennante *Klartext*) wurde längs des Stabs auf das Pergament geschrieben. Beim Abwickeln ergab sich nur noch eine Folge von durcheinandergewürfelten Buchstaben, der *Geheimtext*. Beim Entschlüsseln wurde das Band auf eine Skytale des selben Durchmessers gewickelt und konnte so wieder gelesen werden.

Man kann dieses Verschlüsselungsverfahren auch beschreiben, wenn man keine Holzstäbe zur Hand hat. Der geheime *Schlüssel* des Verschlüsselungsverfahrens ist der Umfang des Holzstabs, gemessen als Anzahl der Buchstaben n, die einmal um den Stab passen. Man verschlüsselt nun, indem man den Text zeilenweise in n Zeilen schreibt. Der Geheimtext ergibt sich, indem man spaltenweise liest. Entschlüsselt wird, indem man einen Text spaltenweise in n Zeilen schreibt und zeilenweise liest.

Das sogenannte *Kerckhoffs'sche Prinzip* in der Kryptographie besagt, dass die Sicherheit eines Verschlüsselungsverfahrens nur auf der Geheimhaltung des Schlüssels, nicht aber auf der Geheimhaltung des Verschlüsselungsverfahrens beruht. In der heutigen Zeit sprechen viele Gründe für das Kerckhoffs'sche Prinzip. Die wichtigsten beiden sind folgende: Erstens bestehen die Algorithmen aus deutlich mehr Information als der Schlüssel und können daher schwerer geheimgehalten werden. Zweitens werden Fehler in öffentlichen Verfahren leichter entdeckt als in nicht-öffentlichen, da sich mehr Menschen damit befassen.

Akzeptiert man das Kerckhoffs'sche Prinzip (was wir tun werden), dann ist die Skytale kein sehr sicheres Verschlüsselungsverfahren. Ein wesentlicher Grund dafür ist, dass die Anzahl der möglichen Schlüssel sehr klein sind: Als Schlüssel kommen nur Zahlen zwischen 1 und der Länge des Textes in Frage. Also kann man relativ schnell alle möglichen Schlüssel durchprobieren.

Ein Beispiel: Wir haben folgenden Text gegeben, von dem wir wissen, dass er mit einer Skytale verschlüsselt wurde:

D E E H E T A R I R I E S I N G M X H S S E E T I T E H R

Wir können jetzt ausprobieren, welche Längen in Frage kommen. Länge 2 würde den Anfang „DEEA…“ ergeben; das kann nicht sein. Bei Länge 3 hat man „DH…“, für Länge 4 „DEISM…“. Länge 5 ergibt „DT…“, für Länge 6 hätte man „DASHIERI…“. Das klingt schon nach einem vernünftigen Beginn. Jetzt kann man den Text einmal probehalber mit Länge 6 entschlüsseln und erhält tatsächlich die Nachricht

D A S H I
E R I S T
E I N S E
H R G E H
E I M E R
T E X T

Auch längere Texte können so relativ leicht entschlüsselt werden.

Ein zweites schon in der Antike bekanntes Verfahren ist die *Cäsar-Verschlüsselung*. Hier wird jeder Buchstabe des Alphabets durch einen Buchstaben, der drei Plätze weiter hinten im Alphabet liegt, ersetzt. Dabei ist die Zahl 3 ziemlich willkürlich; allgemeiner kann man als Cäsar-Verschlüsselung das folgende Verfahren bezeichnen: Der geheime Schlüssel ist ein Buchstabe. Zum Verschlüsseln und Entschlüsseln schreibt man das Alphabet zweimal untereinander, einmal normal und einmal beginnend mit dem Schlüsselbuchstabe. Ist der Schlüsselbuchstabe beispielsweise D wie in der ursprünglichen Cäsar-Verschlüsselung, dann sieht das folgendermaßen aus:

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

Die Verschlüsselung erfolgt, indem jeder Buchstabe aus dem Text durch den darunterstehenden Buchstaben ersetzt wird. Um den so entstandenen Geheimtext wieder zu entschlüsseln, muss man die Buchstaben von unten nach oben ersetzen.

Diese Art von Verschlüsselung ist ein Spezialfall der sogenannten *monoalphabetischen Substitution*. Dabei ist der Schlüssel ein beliebiges Zeichen für jeden Buchstaben des Alphabets. Die Ver- und Entschlüsselung funktioniert wieder wie bei der Cäsar-Verschlüsselung. Ein Beispiel wäre folgendes:

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**
ℚ	∨	H	8	ł	n	&	Γ	=	%	∞	¡	√	€	≃	W	2	←	ℍ	[	ʏ	]	↑	ℏ	ı	æ

Eine etwas praktischere Form ist die sogenannte *Schlüsselwort-Methode*. Hier ist der Schlüssel die Kombination aus einem Schlüsselbuchstaben und einem Schlüsselwort. Das Schlüsselwort wird jetzt, beginnend unter dem Schlüsselbuchstaben, unter das ursprüngliche Alphabet geschrieben. Dabei werden Buchstaben, die im Schlüsselwort mehrfach vorkommen, nur einmal aufgeschrieben. Unter das restliche Alphabet werden jetzt die Buchstaben des Alphabets in der üblichen Reihenfolge untereinander geschrieben, wobei die Buchstaben des Schlüsselworts ausgelassen werden. Besteht der Schlüssel beispielsweise aus dem Schlüsselwort „Handcreme“ und dem Schlüsselbuchstaben F, dann sieht das Verschlüsselungsschema folgendermaßen aus:

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

Die Cäsar-Verschlüsselung kann wieder leicht durch Ausprobieren geknackt werden, da es nicht sehr viele Schlüssel gibt. Aber auch andere monoalphabetische Substitutionsmethoden können ohne Kenntnis des Schlüssels vergleichsweise einfach entschlüsselt werden. Dazu verwendet man, dass nicht alle Buchstaben in deutschen Texten mit denselben Häufigkeiten auftreten. Beispielsweise ist das E mit einem Anteil von über 17 Prozent der mit Abstand häufigste Buchstabe, gefolgt vom N mit knapp 10 Prozent. Das Zeichen, das im Geheimtext also am Häufigsten vorkommt, entspricht mit hoher Wahrscheinlichkeit dem E, und die Zeichen mit den nächstmeisten Häufigkeiten dürften den Buchstaben N, I, S, R, A und T entsprechen (möglicherweise aber nicht genau in dieser Reihenfolge, da die Häufigkeiten dieser Buchstaben recht nahe beieinander liegen). Oft kann man jetzt schon weitere Buchstaben erraten und so den Text entschlüsseln.

Ein Verschlüsselungsverfahren, das diesem Mangel abhilft, ist die *homophone Chiffre*. Hier werden Buchstaben nicht immer mit demselben Zeichen verschlüsselt, sondern jeder Buchstabe bekommt mehrere Zeichen, je nach seiner Häufigkeit. Beispielsweise kann man als Geheimtext-„Alphabet“ alle zweistelligen Ziffernfolgen von 00 mit 99 nehmen. Da das E in der deutschen Sprache vier mal so häufig vorkommt wie das U, werden dem E auch viermal so viele Ziffernfolgen zugeteilt.

Allerdings kann auch die homophone Chiffre geknackt werden, wenn man die Häufigkeit von Buchstabenpaaren betrachtet: Manche Buchstabenpaare sind deutlich häufiger als andere, was als Ansatz zur Analyse eines Geheimtextes verwendet werden kann.

Ein anderer Ansatz, bei dem gleiche Buchstaben nicht immer mit dem gleichen Zeichen verschlüsselt werden, ist die sogenannte *Vigenère-Verschlüsselung*. Der Schlüssel ist hier ein Wort. Dieses Wort wird – immer wiederholt – über den Klartext geschrieben. Ist beispielsweise das Wort „Kaugummi“, dann sieht das Schema folgendermaßen aus:

KAU GUMM IKAU GUMM IKAUGUM MIK AUGUMMI KAUGUM
MAN KANN FUER JEDE LOESUNG EIN PROBLEM FINDEN.

Nun verschlüsselt man jeden Buchstaben mit der Cäsar-Verschlüsselung, wobei man als Schlüsselbuchstabe den Buchstaben nimmt, der direkt drüber steht. Also verschlüsselt man das erste M mit der Cäsar-Verschlüsselung zum Buchstaben K, das A mit Cäsar-Verschlüsselung mit Buchstabe A und so weiter.

Der Text von oben würde dann als

WAH EMZV FOKL VMNE RIQECXG KCZ XBOVRYY NSNXKH.

verschlüsselt werden. Doch selbst dieser raffinierte Verschlüsselungsmechanismus kann geknackt werden. Der wesentliche Schritt dabei ist, die Länge des Schlüsselworts zu bestimmen. Kennt man nämlich bereits die Länge des Schlüsselworts, dann kann man eine Häufigkeitsanalyse für die Teilportionen durchführen. Weiß man beispielsweise, dass das gewählte Schlüsselwort die Länge 4 hat, dann kann man die Buchstaben des Geheimtexts in vier Portionen aufteilen: Diejenigen, die mit dem ersten Buchstaben verschlüsselt wurden, diejenigen, die mit dem zweiten Buchstaben verschlüsselt wurden, und so fort. Bei jedem dieser Stapel zählt man die Buchstabenhäufigkeiten. Der häufigste Buchstabe ist dann wieder mit großer Wahrscheinlichkeit der Buchstabe, der für das E steht, und der nächsthäufige steht möglicherweise für das N. Falls die Texte nicht sehr lang sind, muss das zwar nicht immer der Fall sein, aber oft kommt man durch genaues Vergleichen trotzdem sehr bald auf die richtige Lösung.

Die erste Methode, um die Schlüssellänge zu bestimmen, ist der sogenannte *Kasiski-Test*. Er lässt sich am Besten anhand eines Beispiels beschreiben:

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

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

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

Hier sieht man einige Zeichenfolgen, die mehrfach vorkommen. Höchstwahrscheinlich entstehen solche Zeichenfolgen, indem dieselbe Zeichenfolge im Originaltext mit denselben Schlüsselbuchstaben verschlüsselt wird. Das ist nur dann der Fall, wenn der Abstand zwischen zwei Vorkommnissen der Zeichenfolgen ein Vielfaches der Schlüssellänge ist. Man kann also die Abstände zwischen den mehrfach vorkommenden Zeichenfolgen bestimmen und ihren größten gemeinsamen Teiler berechnen, Dieser Wert oder ein Teiler von diesem Wert wird höchstwahrscheinlich die Schlüssellänge sein. Im Beispiel ergibt sich:

Folge	**Abstand**	**Primfaktorzerlegung**
TNE	50	$2 \cdot 5^2$
FCRV	265	$5 \cdot 53$
NEVHJAOVWU	90	$2 \cdot 3^2 \cdot 5$
VWU	75	$3 \cdot 5^2$

Der größte gemeinsame Teiler der Abstände ist in diesem Fall 5. Allerdings kann auch etwas schiefgehen: Beispielsweise kommt zweimal im Text die Buchstabenfolge OIS vor, und zwar mit einem Abstand von $26 = 2 \cdot 13$. Zählt man das dazu, dann wäre der größte gemeinsame Teiler 1, es würde sich also um eine normale Cäsar-Verschlüsselung handeln. Dass das nicht der Fall ist überprüft man leicht. Solche Ausreißer kann es immer geben; allerdings ist es extrem unwahrscheinlich, dass so eine lange Zeichenkette wie NEVHJAOVWU einen Ausreißer darstellt.

Ein Nachteil dieser Methode ist, dass man die Schlüsselwortlänge nur bis auf Teiler und Vielfache ausrechnen kann. Die nächste Methode liefert eine Größenordnung für die Schlüsselwortlänge; mit der Kombination aus beiden bekommt man zumindest bei längeren Texten sehr wahrscheinlich das richtige Ergebnis für die Schlüsselwortlänge. Diese zweite Methode ist der *Friedman-Test*. Er besteht aus zwei auf den ersten Blick kompliziert aussehenden Formeln:

Tabelle 1: Formeln für den Friedman-Test

$$I = \frac{\sum_{i=1}^{26} n_i(n_i - 1)}{n(n-1)} \qquad \text{(Koinzidenzindex)}$$

$$h = \frac{0{,}0377n}{I(n-1) - 0{,}0385n + 0{,}0762} \qquad \text{(Schlüsselwortlänge)}$$

Mit der ersten Formel wird der sogenannte *Koinzidenzindex* des Textes berechnet. Das ist die Wahrscheinlichkeit dafür, dass zwei zufällig herausgegriffene Buchstaben aus dem Text gleich sind. Dabei sind die Zahlen $n_1, n_2, n_3, \ldots$ die Anzahlen des ersten Buchstabens (A), des zweiten Buchstabens (B), des dritten Buchstabens (C) und so weiter. Das Symbol $\sum_{i=1}^{26}$ bedeutet,

dass man über die Werte von $n_i(n_i-1)$ summieren soll, wobei i jede Zahl zwischen 1 und 26 einmal annimmt. Der Nenner des Bruchs ist also einfach eine Abkürzung:

$$\sum_{i=1}^{26} n_i(n_i-1) = n_1 \cdot (n_1-1) + n_2 \cdot (n_2-1) + n_3 \cdot (n_3-1) + \cdots .$$

Das Symbol n im Nenner ist die Anzahl der Buchstaben des Textes. Um also den Koinzidenzindex zu berechnen, zählt man mit einer Strichliste die Häufigkeiten jedes Buchstabens, also die Zahlen $n_1, n_2, \ldots$, und berechnet dann daraus die Zahlen $n_1(n_1-1), n_2(n_2-1), n_3(n_3-1), \ldots$. Diese Zahlen werden alle addiert und das Ergebnis durch $n(n-1)$ geteilt. Ein Beispiel, bei dem nur die Buchstaben A, B, C und D vorkommen, wäre folgendes:

		n_i	$n_i(n_i-1)$
A	𝍸	8	56
B	𝍸	4	12
C	𝍸	6	30
D	𝍸	7	42

Die Summe im Zähler der Formel für I ist dann $56 + 12 + 30 + 42 = 140$. Die Zahl n ergibt sich als Summe der einzelnen Buchstabenhäufigkeiten, also $n = 8 + 4 + 6 + 7 = 25$, und daher $n(n-1) = 25 \cdot 24 = 600$. Insgesamt ist in diesem Fall also

$$I = \frac{140}{600} = \frac{7}{30} \approx 0{,}23333333.$$

Jetzt kann man die Zahlen I und n in die zweite Formel einsetzen und erhält für die Schlüssellänge

$$h = \frac{0{,}0377 \cdot 25}{0{,}2333333 \cdot 24 - 0{,}0385 \cdot 25 + 0{,}0762} = \frac{0{,}9425}{4.7136992} \approx 0.1999.$$

Das zeigt auch gleich, dass unser Beispiel nicht sehr realistisch ist, da die Häufigkeiten der einzelnen Buchstaben stark voneinander abweichen (22 Buchstaben kommen kein einziges Mal vor, das A dagegen sogar 8 Mal!). Allerdings zeigt das Beispiel auch gleich zwei Schwachstellen des Friedman-Tests: Erstens kommt mit hoher Wahrscheinlichkeit bei dem Test keine ganze Zahl als Ergebnis heraus, und zweitens muss das Ergebnis nicht unbedingt stimmen; es stimmt nur *wahrscheinlich* relativ gut, und die Wahrscheinlichkeit dafür, dass es gut stimmt, steigt mit der Gesamtlänge des Textes. Eine Schlüssellänge von etwa $\frac{1}{5}$ ist jedenfalls sicher nicht möglich.