Arbeit mit klassischen Chiffrierverfahren

Inhaltsverzeichnis

1. Anliegen und Grundaufbau des Versuchs

Dieser Versuch dient der Vertiefung des Lehrstoffs der Lehrveranstaltung Kryptographie und -analyse, kann aber auch ohne diese Vorkenntnisse absolviert werden. Es werden einige der klassischen Chiffrierverfahren und deren Kryptoanalyse mit dem Rechner erarbeitet. Es soll klargemacht werden, wie das Herangehen bei der Kryptoanalyse von Chiffraten mit unbekannten Schlüsseln und unbekannter Herkunft ist. Zur Durchführung des Versuches gibt es ein Programm, mit dessen Hilfe Texte ver- bzw. entschlüsselt und Analysen an Chiffretexten vorgenommen werden können.

Im Folgenden werden die im Versuch behandelten klassische Chiffrierverfahren kurz erläutert. Danach werden wesentliche Aspekte der Kryptoanalyse dieser drei Verfahren dargestellt.

Die Aufgabe für die Praktikumsteilnehmer besteht dann darin, nach Erarbeitung der Verfahren und Analysetechniken diese praktisch am Beispiel von drei gegebenen Schlüsseltexten durchzuführen.

2. Die klassischen Chiffrierverfahren

Die im Praktikumsversuch behandelten Verfahren sind:

MM-Substitutionen: vertreten durch die Tauschchiffren mit den Schlüsseln k und s, die auch Verschiebechiffren und multiplikative Chiffren umfassen.

PM-Substitutionen: vertreten durch die bekannteste PM-Chiffre, die Vigenère-Chiffre, mit Schlüsselworten bis 30 Buchstaben.

Transposition: vertreten durch einfache Matrixtranspositionen, bei denen die Größe der Matrizen angegeben werden kann und eine Permutation, in der die Spalten ausgelesen werden.

Die Verfahren sollen im folgenden kurz beschrieben werden. Grundlage aller behandelten Verfahren ist ein Alphabet $A = \{a_0, \ldots, a_{n-1}\}$, aus dessen Zeichen Klar- und Schlüsseltexte bestehen. Für MM-Substitutionen und PM-Substitutionen ist neben den im Alphabet enthaltenen Zeichen auch dessen Länge n (d.h. die Anzahl der im Alphabet enthaltenen Zeichen) sowie die Position der Zeichen im Alphabet, hier bezeichnet durch $i = v(a_i)$, wobei gilt: $a_i = v^{-1}(v(a_i))$, bedeutsam.

2.1. MM-Substitution – Tauschchiffre

Ein Schlüsseltextzeichen c_i einer Tauschchiffre wird gebildet, indem ein Schlüssel k zur Abbildung eines Buchstabens a_i (d.h. zu dessen Position im Alphabet) multipliziert und ein Schlüssel s zu diesem Produkt addiert wird:

$$c_i := v^{-1}((k \cdot v(a_i) + s) \bmod n)$$

Dabei muß die Bedingung $(k, n) = 1$ erfüllt sein, also der Schlüssel k und die Alphabetlänge n müssen teilerfremd sein.

Die Entschlüsselung erfolgt entsprechend durch:

$$a_i := v^{-1}(((v(c_i) - s) \cdot k^{-1} \bmod n)$$

wobei k^{-1} das multiplikative Inverse von k modulo n ist.

2.2. PM-Substitution – Vigenère-Chiffre

Die Vigenère-Chiffre benutzt ein Schlüsselwort $K = \{k_0, k_1, \ldots, k_{l-1}\}$ der Länge l zur Chiffrierung eines Textes. Die Verschlüsselung erfolgt, indem ein Klartextzeichen a_i mit dem Schlüsselzeichen $k_{(i \bmod l)}$ verknüpft wird, wobei die Verknüpfung darin besteht, dass die Position

des Klartextzeichens $v(a_i)$ zur Position des Schlüsselzeichens $v(k_{(i \bmod l)})$ modulo n addiert wird:

$$c_i := v^{-1}((v(a_i) + v(k_{(i \bmod l)})) \bmod n)$$

Dies bedeutet für Klartexte, die nicht länger als der Schlüssel sind, dass jedes Klartextzeichen mit einem anderen Schlüsselzeichen verschlüsselt wird. Ist der Klartext allerdings länger als der Schlüsseltext werden Schlüsselzeichen wiederverwendet, indem nach dem letzten Zeichen des Schlüsselwortes wieder mit dem ersten begonnen wird.

Die Entschlüsselung erfolgt entsprechend:

$$a_i := v^{-1}((v(c_i) - v(k_{(i \bmod l)})) \bmod n)$$

2.3. Transposition – Matrixtransposition

Eine Matrixtransposition vertauscht die Positionen der Zeichen des Klartextes, indem der Text zunächst zeilenweise in Matrizen der Höhe h und Breite b geschrieben wird. Danach werden die Spalten der Matrizen entsprechend einer Permutation P vertauscht. Der Schlüsseltext wird erzeugt, indem die Zeichen spaltenweise aus den resultierenden Matrizen ausgelesen werden. Der Schlüssel einer Matrixtransposition besteht also aus der Höhe h und der Breite b der Matrix sowie der Spaltenpermutation P. Im Praktikumsversuch werden nur *quadratische Matrizen* verwendet, d.h. es gilt $h = b$.

Die Entschlüsselung erfolgt durch die umgekehrte Anwendung der Schritte der Verschlüsselung, d.h. der Schlüsseltext wird zunächst spaltenweise in Matrizen der Höhe h und Breite b eingelesen. Danach wird die zu P inverse Permutation P^{-1} auf die Spalten der Matrizen angewendet, um deren Vertauschung rückgängig zu machen. Der Klartext wird dann zeilenweise aus den resultierenden Matrizen ausgelesen.

Zur Angabe der Permutationen wird im Praktikumsversuch die Zyklenschreibweise verwendet, wobei die Positionen der Spalten von 0 bis $(b - 1)$ nummeriert werden. Z.B. werde für eine Blocklänge (Breite der Matrix) von 6 die Permutation $P = (4, 0, 2, 5, 1, 3)$ verwendet, d.h., die Spalte 4 wird die Spalte 0, die Spalte 1 wird die Spalte 3, usw. bis zur Spalte 3, welche die Spalte 4 wird. Für eine Permutation der Art $P = (3, 5, 2)(0)(1, 4)$ schreibt man nur $P = (3, 5, 2)(1, 4)$.

3. Kryptoanalyse eines Chiffretextes

Ausgangspunkt ist immer ein Chiffrat unbekannter Herkunft. Das heißt, zuerst muß versucht werden herauszufinden, um welche Art Chiffrierung es sich handelt. Dazu können verschiedene statistische Betrachtungen des Chiffrates durchgeführt werden.

Anhand dieser Auswertungen kann entschieden werden, für welche Chiffrierart eine weitergehende Analyse durchgeführt werden soll, d.h. ob eine MM-Substitution, PM-Substitution oder Transposition analysiert werden soll.

3.1. Auswertungen der statistischen Eigenschaften

3.1.1. Untersuchungen zu Buchstabenhäufigkeiten

Die Verteilung der Buchstaben in einem Text sagt viel darüber aus, um was für einen Text es sich handelt. Ein Klartext in deutscher Sprache zum Beispiel weißt eine ganz bestimmte Verteilung der Buchstaben des Alphabets auf. Wenn man diese Verteilung graphisch mit einer Kurve ausdrückt, entsteht folgende Abbildung. Die häufigsten Buchstaben in einem normalen Klartext mit ihren prozentualen Häufigkeiten sind in der Tabelle angegeben.

E	18,10%
N	10,42%
R	8,08%
I	7,52%
S	6,35%
T	5,57%

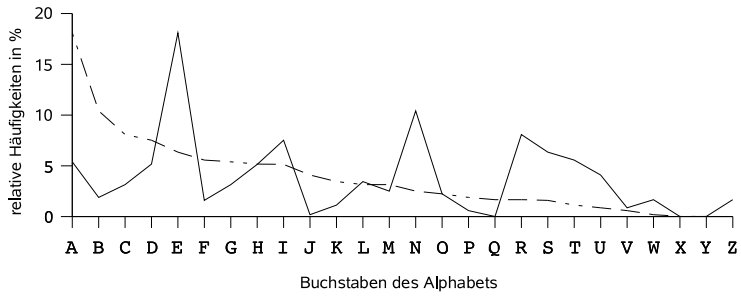


Abbildung 1: Relative Buchstabenhäufigkeiten in einem Klartext

Die gestrichelte Linie im Diagramm zeigt die Verteilung der Häufigkeiten absteigend geordnet, sie bezieht sich nicht auf die Beschriftung der x-Achse.

Wenn man nun einen Text mit dem Alphabet $A = [A : Z]$ chiffriert, ändert sich diese Kurve in irgend einer Weise. Wie sie sich ändert, sagt allerdings viel darüber aus, welches Chiffrierverfahren möglicherweise verwendet wurde.

Im folgenden sollen die Veränderungen betrachtet werden, die bei den verschiedenen, im Praktikum behandelten Chiffrierverfahren auftreten.

Buchstabenhäufigkeiten bei Tauschchiffren

Was passiert mit Buchstaben, wenn ein Text durch eine MM-Substitution, speziell durch eine Tauschchiffre, chiffriert wird? Jeder Buchstabe erhält eine eineindeutige Zuordnung eines anderen Buchstabens des Alphabets. Also wird zum Beispiel ein 'A' immer durch ein 'O' und keinen anderen Buchstaben dargestellt. Das heißt aber, dass sich die Kurve in Abbildung 1 nur so verändern würde, dass die einzelnen Wahrscheinlichkeiten der Buchstaben anders über das Alphabet verteilt werden, sich aber im einzelnen nicht ändern. Die Verteilung der Buchstabenhäufigkeiten für eine Tauschchiffre mit den Schlüsseln $(k, s) = (11, 19)$ und dem Alphabet $A = [A : Z]$ würde folgende Kurve bringen. Die häufigsten Buchstaben in diesem Chiffrat einer Tauschchiffre mit ihren prozentualen Häufigkeiten sind in der Tabelle angegeben.

Die gestrichelte Linie im Diagramm zeigt die Verteilung der Häufigkeiten absteigend geordnet, sie bezieht sich nicht auf die Beschriftung der x-Achse. Man sieht, dass es die selbe Kurve ist, wie in der Abbildung zuvor für die Verteilung der Buchstaben in einem Klartext (Abb. 1).

Man sieht, dass sich die Wahrscheinlichkeiten nur anders auf die Buchstaben verteilt haben, sie haben sich aber sonst nicht verändert.

L	18,10%
G	10,42%
Y	8,08%
D	7,52%
J	6,35%
U	5,57%

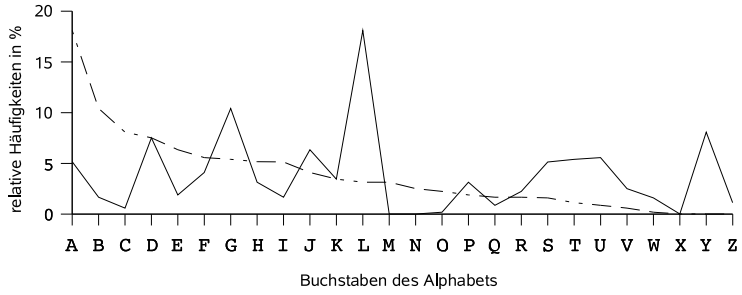


Abbildung 2: Relative Buchstabenhäufigkeiten in einer Tauschchiffre mit den Schlüsseln $(k, s) = (11, 19)$

Einer der häufigsten Buchstaben der Tauschchiffre in Abbildung 2 ist wahrscheinlich die Entsprechung zum Buchstaben 'E', denn dieser ist in einem normalen Text mit dem Alphabet $A = [A : Z]$ fast immer der häufigste. Anhand ihrer Häufigkeiten kann man bei der Kryptoanalyse von Tauschchiffren die Buchstaben im Chiffrat denen zuordnen, denen sie höchstwahrscheinlich im Klartext entsprechen.

Buchstabenhäufigkeiten bei Vigenère-Chiffren (PM-Substitutionen)

Wenn ein Text nach Vigenère chiffriert wird, ändern sich die Buchstabenhäufigkeiten mehr oder weniger stark, je nach Länge des verwendeten Schlüsselworts. Bei einer PM-Substitution wird ein bestimmter Buchstabe nicht immer durch das selbe Zeichen verschlüsselt. Jeder Buchstabe des Alphabets kann durch alle anderen Buchstaben dargestellt sein, je nachdem mit welchem Schlüsselbuchstaben er verknüpft wird. Die relevanten Häufigkeiten der Buchstaben sind gegenüber denen des Klartextes verfälscht. Man beobachtet, dass sie sich einander annähern; je länger das Schlüsselwort ist, desto mehr gleichen sich die relativen Buchstabenhäufigkeiten. Bei einer Schlüssellänge $l = 1$ entsprechen sie exakt denen eines Klartextes, denn es handelt sich in diesem Falle um eine Verschiebechiffre, also eine spezielle Tauschchiffre. Wird der Schlüssel aber länger, verändert sich die Kurve gegenüber Abbildung 1 deutlich.

Die folgenden zwei Abbildungen sind Kurven der relativen Buchstabenhäufigkeiten von Vigenère-Chiffraten mit den Schlüssellängen $l = 5$ und $l = 14$. Es wurde ein Text der Länge 2550 Buchstaben verschlüsselt, das Alphabet ist $A = [A : Z]$. Die gestrichelten Linien machen wieder eine Aussage über die Verteilung der Häufigkeiten absteigend geordnet, sie beziehen sich nicht auf die Beschriftung der x-Achse. Die häufigsten Buchstaben in diesem Chiffrat einer Vigenère-Chiffre mit ihren prozentualen Häufigkeiten sind in Abbildung 3 gezeigt.

Die häufigsten Buchstaben in diesem Chiffrat einer Vigenère-Chiffre mit $l = 14$ und deren prozentuale Häufigkeiten zeigt Abbildung 4.

Anhand dieser Aussagen kann man also deutlich das Chiffrat einer Tauschchiffre von dem einer Vigenère-Chiffre unterscheiden.

Buchstabenhäufigkeiten einer Matrixtransposition

Wies sieht diese Kurve bei einer Matrixtransposition aus? Ganz einfach – genauso wie die eines Klartextes. Die Buchstaben wurden nur in ihrer Reihenfolge verändert, sie sind weder durch einen noch durch mehrere Buchstaben ersetzt worden. Eine Permutation von Buchstaben ändert natürlich nicht deren prozentuale Häufigkeit im Text. Das Diagramm für die

V	9,49%
Z	8,08%
M	6,35%
Y	6,04%
W	5,73%

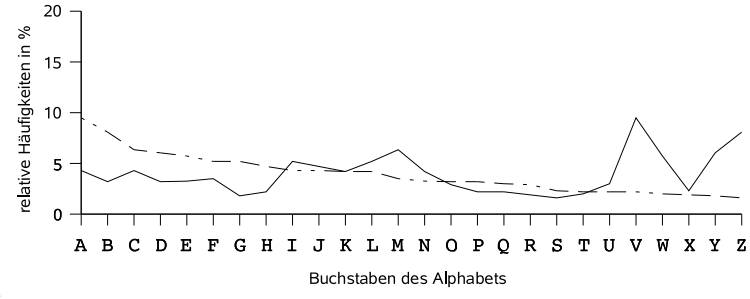


Abbildung 3: Relative Buchstabenhäufigkeiten in einem Vigenère-Chiffrat mit der Schlüsselwortlänge 5

I	5,50%
W	5,50%
V	5,38%
M	5,00%
R	4,92%

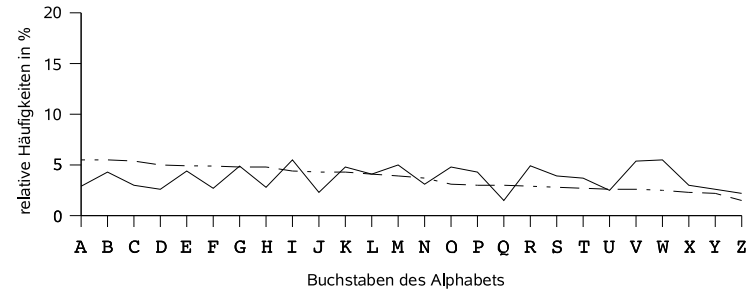


Abbildung 4: Relative Buchstabenhäufigkeiten in einem Vigenère-Chiffrat mit der Schlüsselwortlänge 14

Darstellung der relativen Häufigkeiten bei einer Transposition sieht also genauso aus, wie das eines Klartextes. Die häufigsten Buchstaben in dieser Chiffre sind in Abbildung 5 gezeigt.

E	18,10%
N	10,42%
R	8,08%
I	7,52%
S	6,35%
T	5,57%

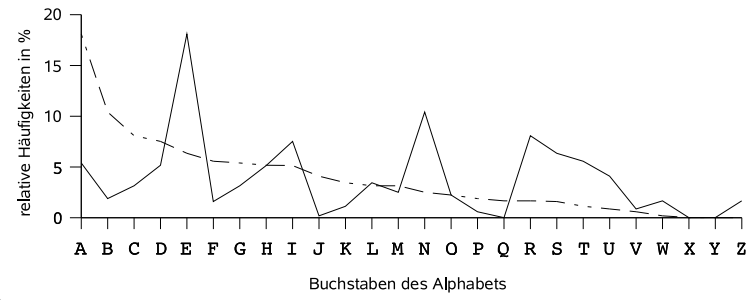


Abbildung 5: Relative Buchstabenhäufigkeiten in einem Chiffrat einer Transposition

Die gestrichelte Linie im Diagramm zeigt die Verteilung der Häufigkeiten absteigend geordnet, sie bezieht sich nicht auf die Beschriftung der x-Achse.

Diese erste Auswertung eines Chiffrates zeigt schon, dass man das Verfahren der Verschlüsselung anhand der Häufigkeiten der Buchstaben bestimmen kann.

Die in diesem Kapitel gemachten Aussagen sind natürlich nur anwendbar, wenn der Text, den es zu analysieren gilt, auch lang genug ist, um statistisch wirken zu können. Wie lang ein

Text sein muß, damit er bezüglich der Verteilung der Buchstabenhäufigkeiten aussagekräftig ist, kann man nicht pauschal sagen. Ein bestimmter Fachtext kann zum Beispiel durch häufig vorkommende Ausdrücke die Buchstabenverteilung beeinflussen und gegenüber denen eines normalen Klartextes verfälschen. Der Text dieser Arbeit beispielsweise enthält ungewöhnlich oft den Buchstaben 'Y', denn die Worte 'Kryptologie', 'Kryptoanalyse' usw. kommen recht häufig vor.

3.1.2. Auswertung von Bi- und Trigrammhäufigkeiten

In [1] findet man folgende kleine Tabelle (Abbildung 6, die eine Aussage darüber macht, welche prozentuale Häuigkeit bestimmte Bigramme in einem deutschen Text haben.

Buchstabenpaar	Häufigkeit
EN	3,88%
ER	3,75%
CH	2,75%
TE	2,26%
DE	2,00%

Buchstabenpaar	Häufigkeit
ND	1,99%
EI	1,88%
IE	1,79%
IN	1,67%
ES	1,52%

Abbildung 6: Häufigkeiten der häufigsten Bigramme der deutschen Sprache [1]

Wenn man die Bi- und Trigrammhäufigkeiten eines Chiffrates betrachtet, kann man Rückschlüsse auf die Chiffrierart ziehen. Es ist aber bei den Betrachtungen von n-Grammen sehr wichtig, dass ein Text nicht zu kurz ist, denn dann kann man diesbezüglich keine statistisch signifikanten Aussagen machen.

Zum Vergleich der Veränderungen von Bi- und Trigrammhäufigkeiten bei verschiedenen Chiffrierverfahren wurde für alle hier aufgeführten Chiffrearten als Ausgangspunkt ein Text der Länge 2550 Buchstaben aus einem Roman gewählt. Die Abbildungen 7 und 8 zeigen die Verteilung der Bi- und Trigramme im Ausgangstext.

Bigramm	Häufigkeit
EN	4,70%
TE	2,47%
IE	2,27%
CH	2,16%
ES	1,88%
SS	1,25%

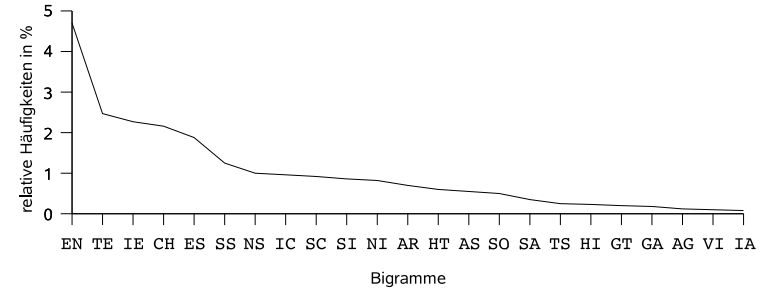


Abbildung 7: Häufigkeiten der Bigramme im Klartext

Dieser Klartext wurde nun durch jede der drei im Praktikum möglichen Arten chiffriert. Die Ergebnisse bezüglich der Verteilung der Bi- und Trigrammhäufigkeiten sind im einzelnen aufgeführt.

Trigramm	Häufigkeit
TEN	0,94%
ICH	0,82%
SCH	0,82%
CHT	0,63%
ENS	0,59%
IES	0,39%

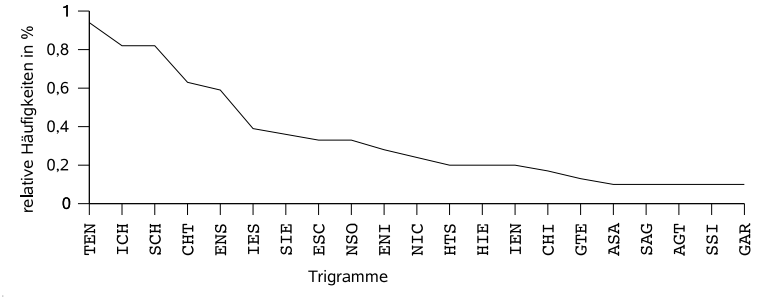


Abbildung 8: Häufigkeiten der Trigramme im Klartext

Veränderungen der n-Grammhäufigkeiten bei Tauschchiffren

Der Ausgangstext wurde als Tauschchiffre mit den Schlüsseln $k = 3$ und $s = 17$ chiffriert. Es ist egal, welche Schlüssel man verwendet, die prozentualen Häufigkeiten der Bi- und Trigramme des Klartextes bleiben bei einer Tauschchiffre erhalten, sie setzen sich nur aus anderen Buchstaben zusammen. Die beiden Abbildungen 9 und 10 zeigen, dass die Kurve exakt der eines Klartextes entspricht, nur die n-Gramme selbst sind andere.

Bigramm	Häufigkeit
DE	4,70%
WD	2,47%
PD	2,27%
XM	2,16%
DT	1,88%
TT	1,25%

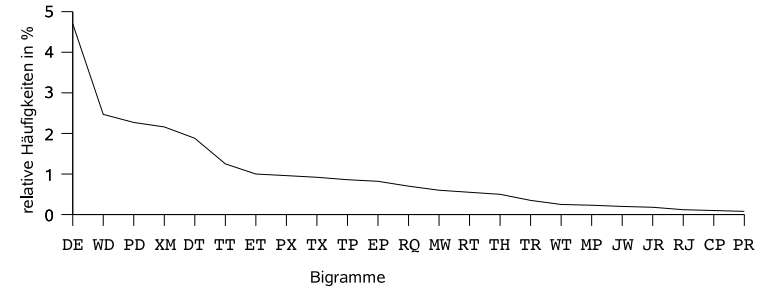


Abbildung 9: Verteilung der Bigramme in einem Beispieltext einer Tauschchiffre

Veränderungen der n-Grammhäufigkeiten bei Vigenère-Chiffren

Als Schlüssel für die Vigenère-Chiffre wurde zum einen ein Wort der Länge 5 Buchstaben und zum anderen ein Wort der Länge 14 Buchstaben verwendet. Wie schon im Kapitel zuvor erläutert, nähern sich die Häufigkeiten der einzelnen Buchstaben bei einer PM-Substitution immer stärker an, je länger der Schlüssel ist. Das heißt aber für die Verteilung der n-Gramme in einem Vigenère-Chiffrat, dass die ursprünglichen Häufigkeiten verlorengehen und sich ebenfalls gleichmäßiger verteilen müßten, je länger der Schlüssel wird. Die Abbildungen 11 und 12 zeigen die Ergebnisse für die Schlüssellängen 5 und 14.

Veränderungen der n-Grammhäufigkeiten bei einer Matrixtransposition

Der Ausgangstext wurde als Transpositionschiffre chiffriert. Dazu wurde der Text in Segmente der Länge 7 Buchstaben eingeteilt und über jedem Segment wurde die Permutation (31)(625)(74) angewendet.

Eine Transposition verändert zwar nicht die Buchstabenhäufigkeiten, aber dafür werden

Trigramm	Häufigkeit
WDE	0,94%
PXM	0,82%
TXM	0,82%
XMW	0,63%
DET	0,59%
PDT	0,39%

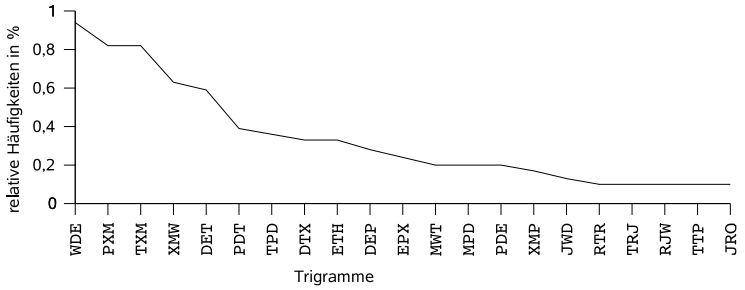


Abbildung 10: Verteilung der Trigramme in einem Beispieltext einer Tauschchiffre

$l = 5$		$l = 14$	
Bigr.	Häufig.	Bigr.	Häufig.
ZV	1,92%	VG	0,55%
LZ	1,14%	TK	0,35%
WI	1,06%	KV	0,35%
VZ	0,90%	LA	0,35%
KY	0,82%	BV	0,35%

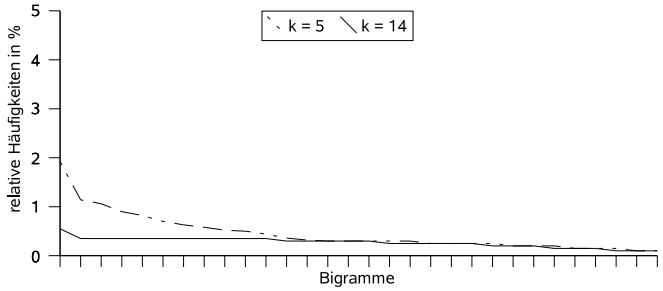


Abbildung 11: Verteilung der Bigramme in einem Beispieltext einer Vigenère-Chiffre mit den verschiedenen Schlüssellängen $l = 5$ und $l = 14$

die Buchstaben grundsätzlich umgeordnet. So verschwinden die ursprünglichen Bi- und Trigramme. Zum Vergleich zeigen die Abbildungen 13 und 14, welche Ergebnisse das Chiffrat der Matrixtransposition liefert.
Dieser Abschnitt sei kurz zusammengefaßt. Die Verteilung von Bi- und Trigrammen eines Chiffrates läßt Rückschlüsse auf die Art der Chiffrierung zu. Wenn die relativen Häufigkeiten eines Chiffrates denen eines vergleichbaren Klartextes entsprechen, so kann man davon ausgehen, dass es sich um eine MM-Substitution, speziell um eine Tauschchiffre handelt. Eine PM-Substitution dagegen zeigt eine viel ausgeglichenere Verteilung der Bi- und Trigramme im Chiffrat. Bei einer Transpositionschiffre findet man ebenfalls eine ausgeglichenere Kurve der relativen Häufigkeiten. Die Bi- und Trigramme des Klartextes wurden auseinandergerissen oder ihre Buchstaben wurden umgestellt, d.h. man findet sie oft in vertauschter Anordnung in den n-Grammen wieder.

Die Diagramme in den Abbildungen 15 und 16 enthalten zum Vergleich alle Kurven der letzten Abbildungen, eines für Bigrammhäufigkeiten und eines für Trigrammhäufigkeiten. Man sieht deutlich, wie sich die Verteilungen bei den einzelnen Chiffrierverfahren ändern.

3.1.3. Bewertung von Koinzidenzindexen des Chiffrates

Der Koinzidenzindex (zur Berechnung des Koinzidenzindexes siehe Abschnitt B.2) eines deutschen Klartextes mit dem Alphabet $A = [A : Z]$ beträgt 0,0762, wenn man von den stati-

$l = 5$		$l = 14$	
Trigr.	Häufig.	Trigr.	Häufig.
LZV	0,31%	CBV	0,12%
WZO	0,19%	VGV	0,12%
NKY	0,19%	GVL	0,12%
VZW	0,16%	LAA	0,12%
ZWZ	0,16%	BRG	0,08%

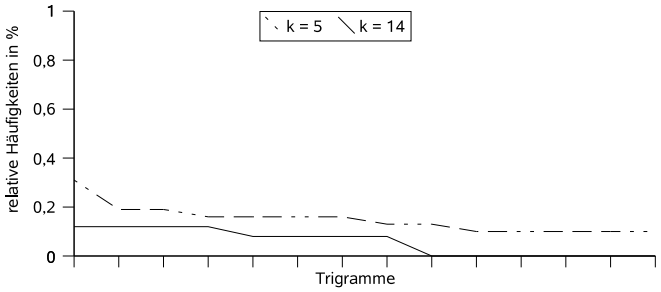


Abbildung 12: Verteilung der Trigramme in einem Beispieltext einer Vigenère-Chiffre mit den verschiedenen Schlüssellängen $l = 5$ und $l = 14$

Bigramm	Häufigkeit
NE	2,35%
ES	1,25%
SE	1,14%
ER	0,98%
EH	0,82%
RN	0,82%

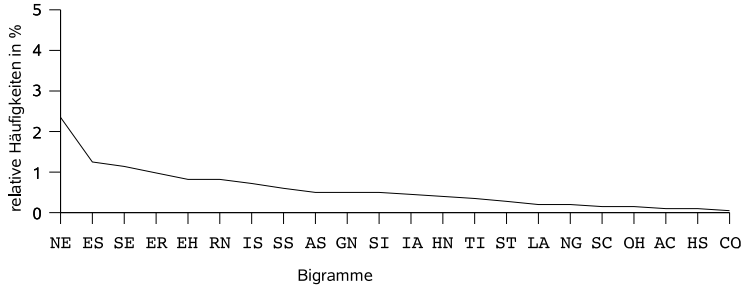


Abbildung 13: Verteilung der Bigramme in einem Beispieltext einer Transposition mit der Blocklänge 7

stischen Buchstabenhäufigkeiten ausgeht. In der Praxis liegt er auf meist geringfügig unter dieser Zahl, etwa im Bereich 0,072 bis 0,0762. Bei einem Text, bei dem alle 26 Buchstaben mit der gleichen Wahrscheinlichkeit $\frac{1}{26}$ auftreten, hat der Koinzidenzindex den Wert 0,0385 , sein Minimum.

Die Wahrscheinlichkeiten, mit denen die Buchstaben im Chiffrat auftreten, ändern sich gegenüber dem Klartext nur bei PM-Substitutionen, wie in den beiden vorangehenden Abschnitten herausgestellt wurde. Das heißt, der Koinzidenzindex einer Tauschchiffre und einer Transposition entspricht exakt dem des Klartextes, aus dem die Chiffrate gebildet wurden.

Für das Beispiel, das schon bei den letzten Abschnitten Verwendung fand, beträgt der Wert des Koinzidenzindex für den Normaltext, die Tauschchiffre und die Transposition $I = 0,0734$. Für die PM-Substitution (Vigenère-Chiffre) beträgt er bei einer Schlüssellänge von 5 Buchstaben $I = 0,0482$, bei einer Schlüssellänge 14 ist $I = 0,0420$. Ein deutlicher Unterschied also, an dem man die PM-Substitution eindeutig von den anderen Chiffrierverfahren und dem Klartext unterscheiden kann.

Wendet man nun die in Abschnitt B.2 beschriebene Methode an und bestimmt den Koinzidenzindex für alle möglichen angenommenen Schlüssellängen, so erhält man für die Tauschchiffre und die Transposition bei allen angenommenen Schlüssellängen den Koinzidenzindex eines Klartextes. Bei einer Vigenère-Chiffre erhält man einen solchen Wert nur in den Fällen, in denen die angenommene Schlüssellänge der tatsächlichen oder deren Vielfachen entspricht.

Trigramm	Häufigkeit
ESE	0,27%
ERN	0,19%
EHS	0,16%
ASS	0,12%
GNE	0,12%
NES	0,12%

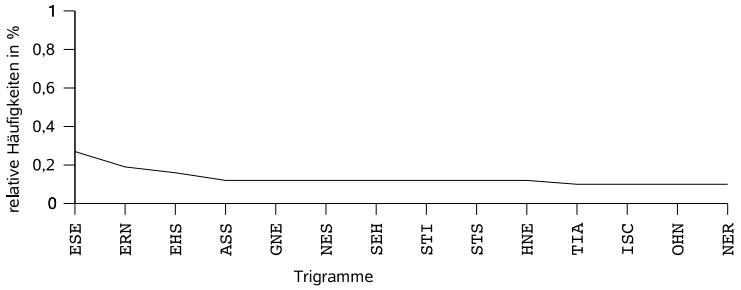


Abbildung 14: Verteilung der Trigramme in einem Beispieltext einer Transposition mit der Blocklänge 7

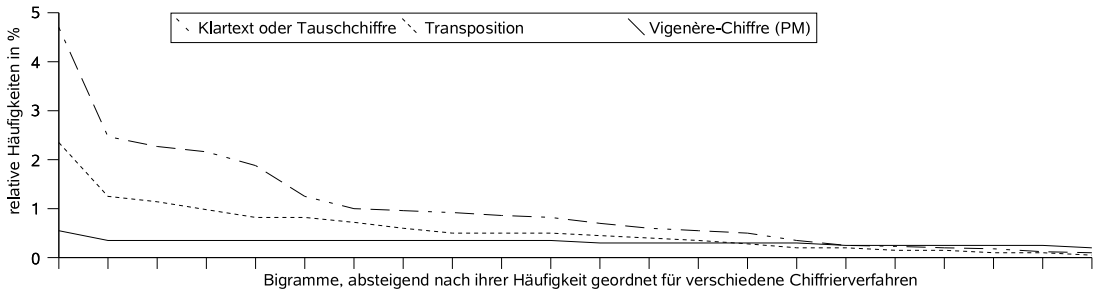


Abbildung 15: Bigrammverteilungen bei verschiedenen Chiffrierverfahren (für einen deutschen Beispieltext)

In diesem Fällen wurde nämlich der Koinzidenzindex einer MM-Substitution, speziell einer Verschiebechiffre, berechnet.

Abbildung 17 zeigt die Veränderung des Koinzidenzindex bei verschiedenen angenommenen Schlüssellängen $l = 1 \ldots 30$ für die oben angegebenen Beispiele.
Bei dieser Auswertung kann man nicht nur die PM-Substitution von den anderen beiden Chiffren unterscheiden, man sieht auch welche Schlüssellänge zur Chiffrierung verwendet wurde.

3.2. Kryptoanalyse einer MM-Chiffrierung

Wenn die statistischen Auswertungen durchgeführt sind und man aufgrund der Ergebnisse festgestellt hat, dass es sich wahrscheinlich um eine Tauschchiffre handelt, kann man nun die verwendeten Schlüssel analysieren.

Die Schlüssel einer Tauschchiffre können mit Hilfe eines Gleichungssystems errechnet werden. Dazu sucht man die zwei häufigsten Buchstaben im Chiffrat und setzt sie mit den zwei Buchstaben gleich, die statistisch gesehen in einem Klartext die häufigsten sind. Mit diesen vier Buchstaben kann man ein Gleichungssystem aufstellen, das die beiden Schlüssel k und s als Unbekannte enthält.

In den statistischen Auswertungen werden die häufigsten Buchstaben des Chiffrates angegeben. Diese können nun in verschiedenen Kombinationen mit den statistisch häufigsten Buchstaben eines Klartextes, also z.B. 'E', 'R', 'N', 'I' oder das Leerzeichen, in das Glei-

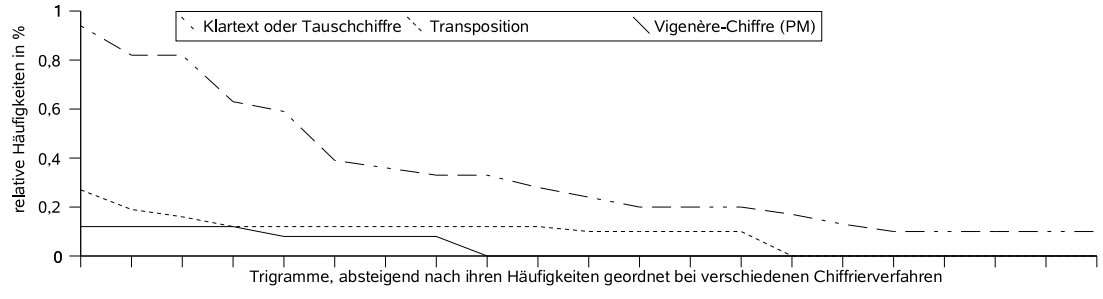


Abbildung 16: Trigrammverteilungen bei verschiedenen Chiffrierverfahren (für einen deutschen Beispieltext)

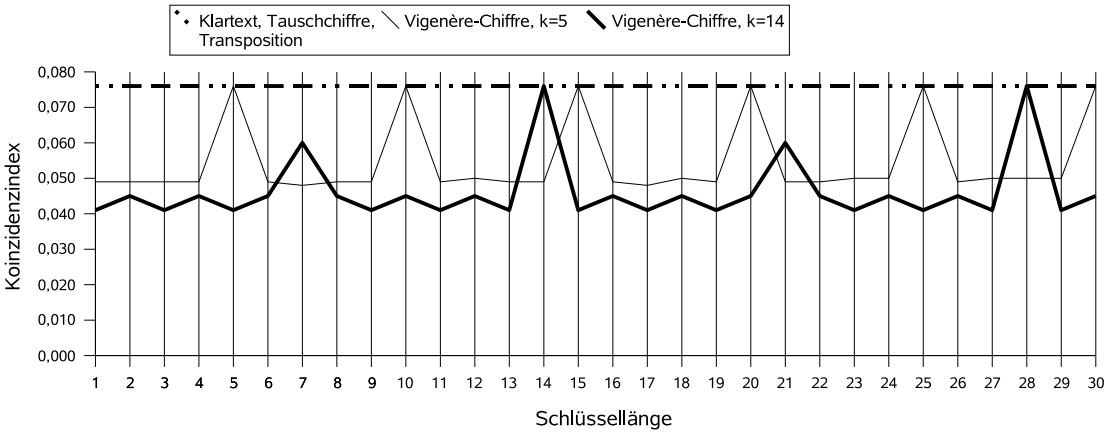


Abbildung 17: Veränderung des Koinzidenzindex

chungssystem eingesetzt werden. Im Programm werden die Buchstaben eingegeben, intern wird mit ihren Abbildungen gerechnet.

Das Programm bestimmt nun k und s für das aufgestellte Gleichungssystem, wobei es auch mehrere Lösungen geben kann. Jede ermittelte Lösung kann in einer Auswahlbox zum Entschlüsseln ausgewählt werden.

Es ist auch möglich, dass das Gleichungssystem keine Lösung hat, dann waren die eingegebenen Buchstaben in ihrer Zuordnung falsch. Andererseits kann das Gleichungssystem Lösungen haben, die die Bedingung $(k, n) = 1$ nicht erfüllen, auch in diesem Fall könnte falsch zugeordnet worden sein. Wenn k und s akzeptable Lösungen haben, kann es auch sein, dass mit diesen Schlüsseln kein sinnvoller Text entsteht.

Wenn alle Fälle nicht zum Ergebnis führen, kann eine andere Zuordnung der Buchstaben getroffen werden und andere Gleichungssysteme ausprobiert werden. Wenn gar keine Lösung gefunden werden kann, so handelt es sich wahrscheinlich nicht um eine Tauschchiffre.

3.3. Kryptoanalyse einer PM-Chiffrierung

Der erste Schritt bei der Kryptoanalyse der Vigenère-Chiffre ist die Ermittlung der Schlüssellänge. Einen ersten wichtigen Ansatz dazu erhält man schon über die Auswertungen, bei denen verschiedene Koinzidenzindexe für angenommene Schlüssellängen über das Chiffrat gebildet werden.

In Abschnitt B werden zwei Methoden zur Berechnung der Schlüssellänge einer Vigenère-Chiffre ausführlich dargestellt. Dies sind der Kasiski-Test und der Friedmann-Test.

Bei Kasiskis Methode werden gleiche Buchstabenfolgen im Chiffrat gesucht und der größte gemeinsame Teiler aller Abstände der Folgen entspricht der Schlüssellänge oder einem Teiler dieser. Der Kasiski-Test wird auch vom Programm durchgeführt. Es erscheint eine Tabelle der folgenden Form:

Folge	Abstand	Primfaktorzerlegung
LZV	255	3 3 5 5
JCW	1602	2 3 3 89
.	.	.
.	.	.
.	.	.

Abbildung 18: Auswertung eines Chiffrates mit dem Kasiski-Test

Bei der Betrachtung der Primfaktorzerlegungen kann man mit Ausnahme von sogenannten Ausreißern einen bestimmten Faktor feststellen, der bei allen Buchstabenfolgen erscheint.

Das Programm wertet die Primfaktorzerlegung aus und bestimmt diesen größten gemeinsamen Faktor. Da es meist einige gleiche Buchstabenfolgen im Chiffrat gibt, die zufällig entstanden sind und nicht durch die Verschlüsselung von gleichen Klartext-Schlüsselfolgen, werden bei der Bestimmung des größten gemeinsamen Teilers die Buchstabenfolgen nach ihrer Länge eingeteilt. Für alle Buchstabenfolgen mit mehr als zwei Buchstaben kann man meist keinen ggT bestimmen, der größer als 1 ist, da fast immer ein Ausreißer dabei ist. Wenn man aber den gemeinsamen Teiler nur für längere Folgen bestimmt, kann man schon sicherer sein, die Schlüssellänge oder einen Teiler zu errechnen, da es sehr selten sein wird, dass sich eine Folge von 5 oder mehr Buchstaben zufällig wiederholt.

Das Programm bestimmt also die größten gemeinsamen Teiler

- für alle Folgen von 3 und mehr Buchstaben,
- für alle Buchstabenfolgen, deren Länge 4 und größer ist und
- für alle Folgen mit 5 und mehr gleichen Buchstaben.

W. Friedman stellte eine Formel auf, mit der die Schlüssellänge einer Vigenère-Chiffre berechenbar ist. Man benötigt lediglich den Koinzidenzindex des Chiffrates und die Textlänge (siehe auch Abschnitt B.2). Das Ergebnis dieses Testes wird ebenfalls vom Programm bereitgestellt.

Durch Kombination von Kasiski- und Friedman-Test kann man in den meisten Fällen die Schlüssellänge eines Vigenère-Chiffrates sicher bestimmen.

Nachdem man eine wahrscheinliche Schlüssellänge bestimmt hat, kann man versuchen, den Text mit Schlüsseln dieser Länge zu entschlüsseln. Das Programm bestimmt nun für die Länge mögliche Schlüsselworte.

Es wird ein Schlüsselwort ausgegeben, dessen Buchstaben die höchste Wahrscheinlichkeit haben, die tatsächlichen Schlüsselwortbuchstaben für die jeweilige Stelle zu sein. Darunter stehen die Buchstaben, die ebenfalls sehr wahrscheinlich sind. Das Schlüsselwort mit den wahrscheinlichsten Buchstaben wird außerdem als Vorschlag zum Entschlüsseln voreingestellt.

Mit diesen Hilfen und dem jeweiligen Entschlüsselungsergebnis kann nun versucht werden, für jede Position im Schlüssel das korrekte Zeichen zu ermitteln.

Führt die Bestimmung des Schlüsselwortes zu keinem positiven Ergebnis, kann das an verschiedenen Faktoren liegen. Ein wichtiger Gesichtspunkt ist, dass genügend viel Chiffretext vorhanden sein muß, damit die Tests zur Bestimmung der Schlüssellänge nach Kasiski und Friedman erfolgreich sein können. Auch die Bestimmung der Schlüsselwortbuchstaben benötigt genügend viel Text, denn die entsprechenden Buchstaben werden durch statistische Betrachtungen von relativen Buchstabenhäufigkeiten gewonnen, die bei sehr kurzen Texten nicht aussagekräftig sind. Die letzte Möglichkeit ist, dass es sich bei dem Chiffrat nicht um eine Vigenère-Chiffre handelt.

3.4. Kryptoanalyse einer Transposition

Ob es sich bei einem Chiffrat um eine Transposition handelt, ist anhand der statistischen Auswertungen leicht zu erkennen. Die relativen Buchstabenhäufigkeiten des gesamten Textes haben sich gegenüber denen eines Klartextes nicht geändert. Die Buchstabenzusammensetzungen, also Bi- und Trigramme, sind dagegen nicht mehr mit denen eines Klartextes vergleichbar.

Nachdem man die Chiffre als Transposition identifiziert hat, gestaltet es sich mehr oder weniger schwierig, den Originaltext zu gewinnen.

Handelt es sich um eine sehr einfache Transposition, so kann ein Betrachter mit einigem Probieren den Klartext schnell gewinnen. Zum Beispiel wurden bei dem Schlüsseltext

CESVLRHEESUUSLLGANOSGMIHRSTU

jeweils sieben Buchstaben des Textes nach einer festen Permutation getauscht. Das häufige Trigramm „SCH“ führt dabei schnell zur verwendeten Blocklänge und Permutation. (Die Entschlüsselung dieses Textes verbleibt als kleine Übung für den Leser.)

Komplizierter wird es dagegen, wenn der Text in Matrizen vorgegebener Größe geschrieben und danach spaltenweise nach vorgeschriebener Permutation wieder aus den Matrizen herausgelesen wird. Diese sogenannten Matrixtranspositionen können mit dem Praktikumsprogramm erstellt werden.

Wie gewinnt man nun den Originaltext einer solchen Matrixtransposition? Eine Methode wäre, es einfach zu probieren. Man nimmt verschiedene Blockgrößen und probiert, ob man zusammenhängende Buchstaben oder Worte bilden kann. Diese Art, den Klartext einer Transposition „herauszuknobeln“, ist für einen Menschen mit seiner Intuition und seinem Wortschatz durchaus durchführbar.

Im Praktikum geht es aber nun darum, die Kryptoanalyse mit dem Computer durchzuführen. Ein Computer braucht immer einen vorgegebenen Algorithmus, er kann nicht „knobeln“. Man müßte ihm den Wortschatz eines Menschen geben, damit er alle möglichen Worte bilden kann. So zeigt sich am Ende, dass kein allgemeingültiges Verfahren angegeben werden kann, das aussagt, wie das Chiffrat einer Transposition analysiert werden kann, um den Klartext zu erhalten; so wie es das z.B. für die Bestimmung der Schlüssel k und s einer Tauschchiffre gibt.

In der zu diesem Thema verfügbaren Literatur gab es nur eine Quelle [3], die auf die Kryptoanalyse von Transpositionen mit Hilfe von Algorithmen eingeht. Die dort beschriebenen Methoden werden vom Praktikumsprogramm als Hilfe bei der Kryptoanalyse von Transpositionschiffren in folgenden drei Punkten angeboten:

- Bestimmung von Vokalhäufigkeiten für verschiedene Blocklängen,
- Auswertung unter der Voraussetzung, dass Worte im Originaltext bekannt sind oder ihr Vorkommen vermutet wird,
- Ausgabe des Textes in beliebigen Matrixgrößen.

Wie diese Möglichkeiten der Analyse zu werten sind und welche Probleme dabei auftreten, soll in den folgenden Abschnitten beschrieben werden.

3.4.1. Bestimmung der Blockgröße anhand von Vokalhäufigkeiten

Unter diesem Punkt erhält man eine Auswertung des Chiffrates in einer Tabelle, in der für verschiedene angenommen Blockgrößen (d.h. $(h{\cdot}b)$ die Vokalhäufigkeiten der Blöcke angegeben sind. Für verschiedene prozentuale Vokalhäufigkeiten wird jeweils angeben, wieviele Blöcke die entsprechende Vokalhäufigkeit aufweisen.

Nach [3] müßte die Mehrzahl der Blöcke in einem Klartext ca. 35-55% Vokale aufweisen. Für das Chiffrat einer Transposition wäre das nur für die Zeile der Fall, die die tatsächlich verwendete Matrixgröße enthält. Diese Auswertung funktioniert für die meisten Texte nicht so gut. Der vermutete prozentuale Anteil von Vokalen ist besonders bei kleinen Blocklängen nicht für die Mehrzahl feststellbar, die Werte schwanken stark, und bei großen Blocklängen kommen mehrere in Frage.

Wird also der Prozentanteil von Vokalen für verschiedene Blockgrößen ermittelt so ist anhand der Werte nicht eindeutig, welche Blockgröße die tatsächlich verwendete ist. Es ist aber möglich, trotzdem eine Aussage zu treffen, wenn angenommen wird, dass die Matrixgröße die Textlänge teilen muß. In diesem Fall braucht man nur Blocklängen zu betrachten, die dafür in Frage kommen.

Die Methode, anhand von Vokalhäufigkeiten die Blockgröße einer Transposistion zu errechnen, ist nicht sehr aussagekräftig und eigentlich nur anwendbar, wenn angenommen wird, dass die Matrixgröße die Textlänge teilt, d.h. der letzte Block wurde mit Blendern aufgefüllt. Für einige Texte kann dieses Vorgehen daher durchaus zum Erfolg führen. Man kann sich aber nicht ausschließlich darauf stützen.

3.4.2. Auswertung mit eventuell vorkommenden Worten

Diese Methode ist bei der Bestimmung der Blockgröße, speziell der Blockbreite, sehr aussagekräftig. Um dieses Verfahren anwenden zu können, sollte man eine ungefähre Vorstellung vom Thema des Chiffrates haben. In der Praxis könnte so ein Fall aber durchaus eintreten.

Man erhält ein Chiffrat, von dem das Thema des Klartextes bekannt ist, z.B. handelt der Text über Architektur oder Gentechnologie oder Kryptoanalyse. Für jedes Gebiet gibt es Worte, die möglicherweise in einem Fachtext über dieses Gebiet vorkommen. Es ist wichtig, besonders lange Worte zu wählen, da sie nur hilfreich sind, wenn sie die Blockgröße überschreiten und damit Sequenzen bilden, die aus Buchstaben eines Wortes entstehen.

Der Ablauf läßt sich wie folgt kurz erläutern: Der Benutzer gibt also ein Wort ein, dessen Vorkommen er im Klartext vermutet. Das Programm sucht daraufhin Sequenzen aus Buchstaben dieses Wortes im Chiffrat und gibt den Abstand der Buchstaben der gefundenen Sequenz im Wort aus. Es erscheint eine Tabelle, aus der man herauslesen kann, dass es für einige Buchstaben jeweils eine Sequenz gibt, bei der der Abstand der Buchstaben im Wort gleich ist. Dieser gemeinsame Abstand entspricht der Blocklänge. Dabei können natürlich nicht alle Buchstaben des Wortes diesen Abstand haben. Es sind nur diejenigen Buchstaben relevant, die im Chiffrat auch wirklich Überlappungen bilden.

Der Benutzer hat die Möglichkeit, mehrere Worte zu untersuchen. Das Programm gibt die am häufigsten vorkommenden Abstände aus. Wenn man weiß oder vermutet, dass die untersuchten Worte im Klartext vorkommen, kann man sicher sein, die Blocklänge unter diesen häufigsten Abständen zu finden.

3.4.3. Auswertung des Textes in vorgegebener Matrixgröße

Mit den beiden vorangegangenen Auswertungen zur Blockgröße bzw. Blockbreite von Matrixtranspositionen kann man eventuell die Matrixgröße ermitteln.

Anhand dieser Matrixgröße kann man nun versuchen, die verwendete Spaltenpermutation zu ermitteln. Nach Eingabe der Matrixhöhe bzw. Breite wird der Text in diesem Matrixformat ausgegeben.

Gibt man ein Chiffrat einer Transposition in der tatsächlich benutzen Matrixgröße aus, kann die Permutation, mit der die Spalten vertauscht wurden, leichter bestimmt werden. Dazu benutzt man wieder Betrachtungen über statistische Häufigkeiten von Klartexten.

Die besonders markanten Eigenschaften der deutschen Sprache, dass 'Q' nur vor 'U' steht und 'C' immer in Verbindung mit 'H' oder 'K', können dabei besonders hilfreich sein. Auch die häufigsten Bi- und Trigramme kann man in Betracht ziehen.

Der Benutzer kann am Ende dieser Auswertung verschiedene Permutationen in Zyklenschreibweise eingeben, mit denen und den eingegebenen Blockgrößen das Chiffrat ausgelesen wird.

A. Dokument zum Praktikum

Die folgenden Seiten sind als Hilfe für den Benutzer des Programmes bei der Durchführung des Praktikums vorgesehen.

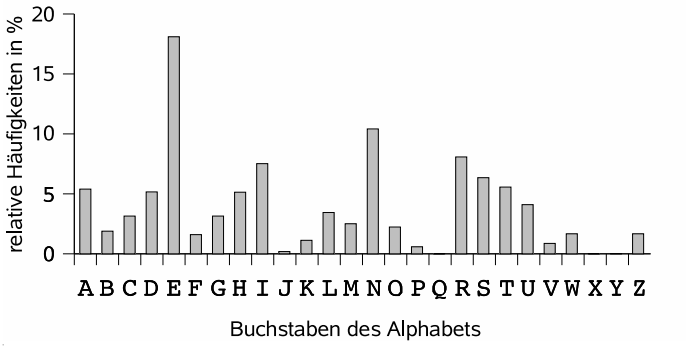
Es soll davon ausgegangen werden, dass Chiffrate unbekannter Herkunft zu analysieren sind. Das Programm macht statistische Auswertungen der Geheimtexte. Der Nutzer muß die Ergebnisse mit denen eines Klartextes vergleichen, um das Chiffrierverfahren bestimmen zu können. Dazu sind die Tabellen und Diagramme über prozentuale Häufigkeiten zu benutzen.

A.1. Statistische Eigenschaften deutscher Klartexte

A.1.1. Buchstabenhäufigkeiten

Buchstabe	Häufigkeit in einem Klartext in %
A	5,40
B	1,89
C	3,15
D	5,17
E	18,1
F	1,60
G	3,15
H	5,14
I	7,52
J	0,19
K	1,13
L	3,45
M	2,51
N	10,42
O	2,24
P	0,59
Q	0,01
R	8,08
S	6,35
T	5,57
U	4,10
V	0,87
W	1,67
X	0,01
Y	0,02
Z	1,67

Die Tabelle zeigt die relativen Buchstabenhäufigkeiten in einem Klartext deutscher Sprache. Das Diagramm bringt die Werte graphisch zum Ausdruck.

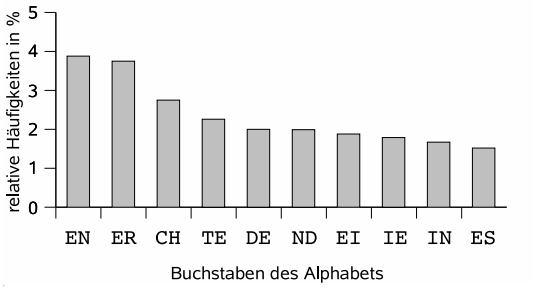


A.1.2. Bigrammhäufigkeiten

Rang	Bigramm
1	EN
2	ER
3	CH
4	ND
5	EI
6	DE
7	IN
8	ES
9	TE
10	IE
11	UN
12	GE
13	ST
14	IC
15	HE
16	NE
17	SE
18	NG
19	RE
20	AU
21	DI
22	BE
23	SS
24	NS
25	AN
26	SI
27	UE
28	DA
29	AS
30	NI

Bigramm	Häufigkeit
EN	3,88%
ER	3,75%
CH	2,75%
TE	2,26%
DE	2,00%
ND	1,99%
EI	1,88%
IE	1,79%
IN	1,67%
ES	1,52%

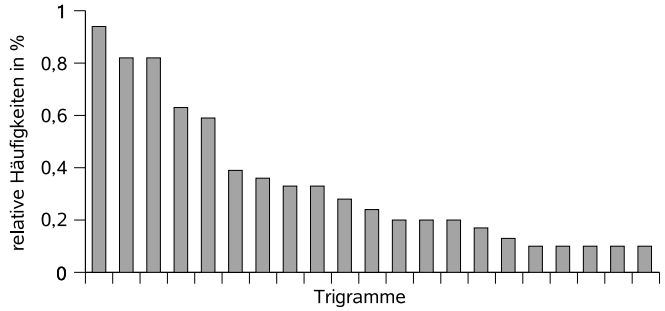
Die Tabelle links zeigt die Bigrammhäufigkeiten deutscher Klartexte, entnommen aus [2].
Die kleine Tabelle zeigt, wie häufig, statistisch gesehen, die Bigramme auftreten, die am meisten in einem deutschen Klartext vorkommen. Das Diagramm zeigt diese Häufigkeiten graphisch. Für diese Tabelle und das Diagramm wurden Angaben aus [1] zugrundegelegt, deshalb ergeben sich für manche Bigramme Unterschiede zur Tabelle links.



A.1.3. Trigrammhäufigkeiten

Rang	Trigramm
1	EIN
2	ICH
3	NDE
4	DIE
5	UND
6	DER
7	CHE
8	END
9	GEN
10	SCH
11	CHT
12	DEN
13	INE
14	NGE
15	NUN
16	UNG
17	DAS
18	HEN
19	IND
20	ENW
21	ENS
22	IES
23	STE
24	TEN
25	ERE
26	LIC
27	ACH
28	NDI
29	SSE
30	AUS

Die häufigsten Trigramme eines Klartextes kommen mit ca. 0,9 - 0,6% Wahrscheinlichkeit vor.
Das Diagramm zeigt eine Kurve, die bei der Analyse eines Klartextes in deutscher Sprache entsteht. (Entnommen aus einem Beispiel.)



B. Analyseverfahren für PM-Substitutionen

Hier sollen zwei Methoden zur Analyse von PM-Substitutionen, der Kasiski-Test und der Friedman-Test, detailierter erläutert werden. Grundgedanke der Analyse ist es, bei periodischen Schlüsseln zunächst die Periode p, die Schlüssellänge, zu ermitteln. Gelingt dies, besteht nur noch das einfach lösbare Problem der Analyse von p MM-Substitutionen.

B.1. Der Kasiski-Test

In jedem Schlüsseltext gibt es Folgen gleicher Zeichen. Wiederholungen treten auf, wenn gleiche Klartextzeichen mit gleichen Schlüsselzeichen verschlüsselt werden. Diese Folgen müssen logischerweise einen Abstand haben, der ein Vielfaches der Schlüssellänge ist. Selbstverständlich können auch andere Klartext-Schlüsselpaare zu solchen Schlüsseltextfolgen führen. Diese jedoch werden im allgemeinen nicht mit der gleichen Periodizität auftreten und je länger die periodischen Schlüsseltextfolgen sind, desto geringer ist die Auftrittswahrscheinlichkeit dafür, dass diese Folgen aus anderen Klartext/Schlüsselpaaren entstanden sind.

An einem Beispiel soll nachvollzogen werden, wie die Schlüssellänge ermittelt wird.

```
Klartext:     T O B E O R N O T T O B E T H A T I S T H E . . .
Schlüssel:    H A M
Schlüsseltext:A O N L O D U O F A O N L T T H T U Z T T L
```

In unserem Beispiel beträgt die Schlüssellänge 3. Die Abstände der sich wiederholenden Folgen betragen für die Folge

A O N L	9 Zeichen und für
T T	6 Zeichen.

Bei längeren Texten ließen sich noch weitere Folgen finden. Sind für alle Folgen die Abstände ermittelt, zerlegt man diese in ihre Primfaktoren. Alle periodischen Abstände enthalten einen größten gemeinsamen Teiler, der die Periode des Schlüssels ist. Ausgenommen sind hier die Folgen, die zufällig aus anderen Klartext/Schlüsselfolgen entstanden sind. Diese können das Ergebnis verfälschen. In der Praxis sind diese einfach zu finden, da die periodischen Folgen gemeinsame Primfaktoren (oder einen) aufweisen. Die nichtperiodischen Folgen treten seltener auf und können durch die Anwendungen dieser Gesetzmäßigkeit eliminiert werden. In unserem Beispiel existieren keine Folgen, die von der Periodizität abweichen. Die Zerlegung liefert:

$$9 = 3 \cdot 3$$
$$6 = 2 \cdot 3$$

Der größte gemeinsame Teiler (ggT) ist 3. Das ist das erwartete Ergebnis für die Periode. Zur Ermittlung des Schlüssels wird im nächsten Abschnitt eine Aussage getroffen.

B.2. Friedman-Test und Koinzidenzindex

Der Friedman-Test bietet eine weitere Möglichkeit, die Periode in einer Vigenère-Chiffre zu bestimmen. Dazu dient der durch Friedman definierte **Koinzidenzindex** I. Er dient zur Bewertung der Häufigkeitsverteilung in Schlüsseltexten. Der Koinzidenzindex I gibt die Wahrscheinlichkeit an, dass zwei unabhängig voneinander gewählte Schlüsseltextzeichen übereinstimmen. Er nimmt bei PM-Substitutionen in Abhängigkeit von der Schlüssellänge ab. Dieser Umstand wird zur Ermittlung der Schlüssellänge benutzt. Es seien

A ein Alphabet mit $A = \{a_0, ..., a_{n-1}\}$
p_i die Wahrscheinlichkeit für das Auftreten des Zeichens a_i

Der Koinzidenzindex ist definiert als:

$$I = \sum_{i=0}^{n-1} p_i^2$$

Der Koinzidenzindex erreicht seinen minimalen Wert, wenn alle Zeichen gleichwahrscheinlich sind. D.h.

$$p_i = \frac{1}{n} \qquad \text{und daraus folgt} \qquad I_{\min} = \frac{1}{n}$$

Für das lateinische Alphabet (nur Großbuchstaben) gilt:

$$I_{\min} = \frac{1}{26} = 0,038$$

Für eine Nachricht in deutscher Sprache wird $I = 0,0762$ und für eine Nachricht in englischer Sprache wird $I = 0,065$. Diesen Wert haben auch Schlüsseltexte, die durch MM-Substitutionen entstanden sind.

Betrachten wir nun die Kryptoanalyse einer Vigenère-Chiffre:

$$\begin{aligned} A &= \{a_0, ..., a_{n-1}\} \\ K &= k_0, ..., k_{l-1} \qquad 1 \leq l < \text{length}(m) \end{aligned}$$

$$f_K : A^* \to A^*$$

Länge der Nachricht: $\text{length}(m) = N$

Zwei Schritte können zur Lösung unterschieden werden:

1. Ermittlung von $\text{length}(K) = l$
2. Ermittlung von K nach Kenntnis von l

Wir ordnen zunächst den Schlüsseltext in ein Schema ein, in dem alle Schlüsseltextzeichen c_j, die mit dem gleichen s_i verschlüsselt wurden, in einer Spalte stehen.

Schlüssel	k_0	k_1	k_2	...	k_{l-1}
Schlüsseltext	c_0	c_1	c_2	...	c_{l-1}
	c_r	c_{l+1}	c_{l+2}	...	$c_{l+(l-1)}$
	c_{2l}	c_{2l+1}	c_{2l+2}	...	$c_{2l+(l-1)}$
	c_{3l}	c_{3l+1}	c_{3l+2}	...	$c_{3l+(l-1)}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$

Die Zeichenfolge einer Spalte kann als MM-Substitution (Verschiebechiffre) interpretiert werden. Wenn K eine zufällige Folge von Zeichen k_i ist, dann ist die Wahrscheinlichkeit in einer Spalte ein Paar gleicher Buchstaben zu finden $I_{\max}$. Ein Paar aus verschiedenen Spalten liefert $I_{\min}$. Die Anzahl von Buchstabenpaaren aus gleichen und verschiedenen Spalten ergibt folgendes:

$$\begin{aligned} \text{gleiche Spalte:} \qquad \text{Anz}_{\text{gl}} &= \frac{N(\frac{N}{l}-1)}{2} = \frac{N(N-l)}{2l} \\ \text{verschiedene Spalten:} \qquad \text{Anz}_{\text{v}} &= \frac{N(N-\frac{N}{l})}{2} = \frac{N^2(l-1)}{2l} \end{aligned}$$

Die Wahrscheinlichkeit ein Paar gleicher Buchstaben auszuwählen ist:

$$\begin{aligned} \sum_{i=0}^{n-1} p(a_i, a_i) &= \sum_{i=0}^{n-1} p(a_i) \cdot p(a_i) = \frac{\text{Anz}_{\text{gl}} \cdot I_{\max} + \text{Anz}_{\text{v}} \cdot I_{\min}}{\frac{N(N-1)}{2}} \\ \sum_{i=0}^{n-1} p(a_i) \cdot p(a_i) &= \frac{N-l}{l(N-1)} I_{\max} + \frac{N(l-1)}{l(N-1)} I_{\min} \end{aligned}$$

Durch Einsetzen von $I_{\max} = 0,0762$ und $I_{\min} = 0,0385$ und Auflösen nach l erhält man:

$$\begin{aligned} l &= \frac{0,0377N}{I(N-1) - 0,0385N + 0,0762} \\ l &= \frac{0,0377}{I - 0,0385} \qquad \text{für } N \gg 1 \\ I &= \sum_{i=0}^{n-1} p(a_i) \cdot p(a_i) \approx \sum_{i=0}^{n-1} \frac{n_i^2}{N^2} = \frac{1}{N^2} \sum_{i=0}^{n-1} n_i^2 \\ \frac{n_i}{N} &\qquad \text{relative Häufigkeit des Buchstabens } a_i \text{ im Schlüsseltext} \end{aligned}$$

Der Wert für l ist eine Näherung. Mit Hilfe des Kasiski-Tests läßt sich durch Vergleich beider Ergebnisse die Schlüssellänge mit hoher Genauigkeit angeben. Nachdem l bestimmt wurde, kann jetzt das Schema konkret mit den Schlüsseltextzeichen aufgefüllt und der Schlüsselbuchstabe für jede Zeile bestimmt werden.

Literatur

[1] Beutelspacher, Albrecht: Kryptologie. Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig, 1991

[2] Fumy, Walter; Rieß, Hans Peter: Kryptographie: Entwurf und Analyse symmetrischer Kryptosysteme. Oldenbourg; München, Wien, 1988

[3] Gaines, Helen Fouchè: Cryptoanalysis – a study of ciphers and their solution. Dover Publications Inc.; New York, 1940

[4] Heider, Franz-Peter; Kraus, Detlef; Welschenbach, Michael: Mathematische Methoden der Kryptoanalyse. Vieweg; Braunschweig, Wiesbaden, 1985

[5] Horster, Patrik: Kryptologie, (Reihe Informatik; 47). Bibliographisches Institut; Mannheim, Wien, Zürich, 1985

[6] Klam, Andrea: Erarbeitung eines Praktikums zu klassischen Chiffrierverfahren. Diplomarbeit, TU Dresden, 1993

[7] Ryska, N.; Herda, S.: Kryptographische Verfahren in der Datenverarbeitung. Informatik-Fachberichte Band 24, Springer Verlag; Berlin, Heidelberg, New York, 1985