Blockchiffren auf Basis der Substitution

Reto Bader

Klassenstufe: 10 – 12

Fachliche Einordnung

Die Aufgabe dient dazu, das Prinzip von Blockchiffren, die auch Bestandteil moderner symmetrischer Verschlüsselungsverfahren sind, am Beispiel der Hill-Chiffre mit Blocklänge $n = 2$ (d.h. Bigrammen) einzuführen.

Voraussetzungen

- Die SchülerInnen sollten mit dem Mechanismus einfacher Substitutionschiffren vertraut sein. Beispielsweise sollte die Caesar-Verschlüsselung und die Ver- und Entschlüsselung mit dem Kryptosystem LINCAESAR bekannt und eingeübt worden sein (siehe Aufgabe 2.21 in «Informatik: Daten verwalten, schützen und auswerten», Klett-Verlag, 2022).
- Die SchülerInnen sollten die Bedeutung und Verwendung der Frequenzanalyse einzelner Buchstaben und von Bigrammen in der Kryptoanalyse kennen.
- Aus der Mathematik sollte den SchülerInnen bekannt sein, wie man lineare Gleichungssyteme löst.

Aufgabe

Ausgangslage

Einfache monoalphabetische Substitutionschiffren, wie sie in den Aufgaben 2.16 – 2.21 vorgestellt wurden, sind mit statistischen Methoden vergleichsweise leicht zu knacken. Dazu analysiert man die Häufigkeitsverteilung der einzelnen Geheimtextzeichen und vergleicht diese mit den Häufigkeiten von Buchstaben in Klartexten verschiedener Sprachen. Ist einmal klar, um welche Sprache es sich handelt, können die Häufigkeiten der Buchstaben und Muster in Buchstabenfolgen dazu genutzt werden, den Geheimtext sukzessive in Klartext zurückzuführen.

Um monoalphabetische Substitutionschiffren weniger angreifbar durch die stochastische Kryptonanalyse zu machen, darf ein bestimmtes Klartextzeichen nicht immer auf dasselbe Zeichen im Geheimtextalphabet abgebildet werden. Im Laufe der Geschichte gab es verschiedene Ideen, um dieses Ziel zu erreichen, unter anderem das polyalphabetische Kryptosystem Vigenère (siehe Beispiel 2.7 in «Informatik: Daten verwalten, schützen und auswerten», Klett-Verlag, 2022). In dieser Aufgabe lernen Sie eine weitere Möglichkeit zur Verschleierung von Buchstabenhäufigkeiten kennen. Es handelt sich dabei um sogenannte Blockchiffren, die auch in modernen Chiffren weiterhin zur Anwendung kommen.

Hintergrundwissen: Blockchiffre

In einer **Blockchiffre** wird der Klartext in Blöcke fester Länge aufgeteilt. Jeder Klartextblock X wird auf einen Geheimtextblock Y (in der Regel der gleichen Länge) abgebildet. Die Zeichenfolge von Geheimtextblock Y ist dabei in eindeutiger Weise durch die exakte Abfolge jedes einzelnen Buchstabens von Klartextblock X festgelegt. Durch die Änderung eines einzelnen Buchstabens im Klartextblock ändern sich deshalb in der Regel mehrere oder sogar alle Zeichen des Geheimtextblocks. Die (injektive) Abbildung, mit der die Zeichenfolge des Geheimtextblocks Y aus der Zeichenfolge des Klartextblocks X bestimmt wird, ist von einem Schlüssel abhängig. Unter Kenntnis des Schlüssels ist die Entschlüsselung eines Geheimtexts gleich aufwändig wie die Verschlüsselung.

Blockchiffren sind Bestandteil von modernen symmetrischen Verschlüsselungsverfahren wie z.B. DES (Data Encryption Standard) und AES (Advanced Encryption Standard).

In dieser Aufgabe erweitern wir das in Aufgabe 2.21 vorgestellte monoalphabetische Kryptosystem LINCAESAR zu einer Blockchiffre. Das resultierende Verfahren heisst HILL-CHIFFRE. Es wurde 1929 von Lester Hill entworfen.

Wie funktioniert die Hill-Chiffre?

Zunächst wird der Klartext in Blöcke gleicher Länge n aufgeteilt. Im folgenden Beispiel bilden wir für das Wort TEXT zwei Blöcke der Länge $n = 2$. Wir müssen also die beiden Blöcke TE und XT chiffrieren.

Zur Chiffrierung von Blöcken der Länge n benötigt man einen Schlüssel bestehend aus n^2 Zahlen aus der Menge $\{1, 2, \ldots, 25\}$. Für $n = 2$ könnte man etwa folgende $2^2 = 4$ Zahlen wählen: $a = 1$, $b = 8$, $c = 6$ und $d = 25$

Die beiden Buchstaben $\square_1$ und $\square_2$ eines Blockes werden gleichzeitig verschlüsselt. Die Buchstaben $\triangle_1$ und $\triangle_2$ des Geheimtextes werden mit dem Schlüssel wie folgt ermittelt:

$$\text{Ordnung}(\triangle_1) = a \cdot \text{Ordnung}(\square_1) + b \cdot \text{Ordnung}(\square_2) \bmod 26$$

$$\text{Ordnung}(\triangle_2) = c \cdot \text{Ordnung}(\square_1) + d \cdot \text{Ordnung}(\square_2) \bmod 26$$

Damit erhält man zum Beispiel die Zuordnung TE → ZG.

1 Überprüfen Sie die Chiffrierung von TE nach obigem Schema.

2 Vervollständigen Sie die Chiffrierung des Wortes TEXT.

3 Chiffrieren Sie den Textblock TT. Vergleichen Sie die Chiffrentextzeichen, auf die der Buchstabe T bei der Chiffrierung der Blöcke TE, XT und TT abgebildet wird. Was schliessen Sie daraus?

4 Ist die Hill-Chiffre mono- oder polyalphabetisch?

5 Der folgende Text wurde mit einer Hill-Chiffre basierend auf Blöcken der Länge 2 codiert. Der Schlüssel ist «leider» unbekannt.

xff aisna cajf gwofg gybsx xfm okur yhumo hu zszea qg ta nhzuexp fk vumehuhozey of xkj sfph la xfi gjffqdmyyi vjf xfwy wy sfphagfmzycj bnkw la qonvji gwof ihssacp ukjwv zym ol aym wmxweqw me pprvueealjw lwywivdxi gq txfji gwtpl qkj lwjisku ena lgxk kj ecet ehdq sv vgdkgf g mngkjf nce qtxfhuh mu sxqadgf rkjw kj hugffxg gjf xfjii gt gm cml sv onpj whj oylx fwy hugffxkay je ct zitgzngut gbjfwvqf la ugzyw mrn xfq wngkjmb la xfp uitxfuek kwv xfn qcewy la uklq la xfu efojic rbelqzedikag zngkjmb

Der Text stammt aus dem «Manual for the Solution of Military Ciphers» und ist 1916 publiziert worden. Der Autor heisst Parker Hitt.

Welche der folgenden Informationen helfen bei der Entschlüsselung?

- Die vier häufigsten Buchstaben im Klartext sind E (48), T (39), I (38) und O(34).
- Die zwei häufigsten Bigramme im (englischen) Klartext sind TH(13) und HE(9).
- Die zwei häufigsten Bigramme mit Schrittweite 2 im Klartext sind TH(13) und HE(8).

6 Entschlüsseln Sie den Text! Hinweis: Die Entschlüsselung funktioniert in der Hill-Chiffre nach dem gleichen Prinzip wie die Verschlüsselung. Um Chiffrentextbigramme $\triangle_1\triangle_2$ auf Klartextbigramme $\square_1\square_2$ zurückzuführen, benötigt man im entsprechenden Rechenschema jedoch die passenden Entschlüsselungszahlen *e*, *f*, *g* und *h*.

$$\text{Ordnung}(\square_1) = e \cdot \text{Ordnung}(\triangle_1) + f \cdot \text{Ordnung}(\triangle_2) \bmod 26$$

$$\text{Ordnung}(\square_2) = g \cdot \text{Ordnung}(\triangle_1) + h \cdot \text{Ordnung}(\triangle_2) \bmod 26$$

Wer die Zahlen *a*, *b*, *c* und *d* kennt, die bei der Verschlüsselung zur Anwendung kamen, kann daraus die Zahlen *e*, *f*, *g* und *h* mathematisch sofort herleiten.

Da Ihnen der Schlüssel jedoch nicht bekannt ist, müssen Sie die Zahlen *e*, *f*, *g* und *h* mit Hilfe der im Geheimtext beobachteten Häufigkeiten von Bigrammen herleiten.

7 Stimmen die Buchstabenhäufigkeiten in Klar- und Geheimtext überein? Was schliessen Sie daraus?

8 Wie liesse sich das Verfahren demnach noch besser gegen stochastische Kryptoanalyse absichern?

Hintergrundwissen: Diffusion

Angriffe, die auf der Zuordnung von Klartextteilen zu bestimmten Chiffrentextteilen basieren, werden mit zunehmender Blocklänge immer schwieriger. Dies liegt an einem Phänomen, das den Namen «Diffusion» trägt. Diffusion bedeutet, dass eine kleine Änderung im Klartext weitreichende Änderungen im Chiffrentext herbeiführt. Wird nämlich im Klartext eines Blocks nur ein einzelnes Zeichen durch ein anderes ersetzt, können sich bei der Chiffrierung dennoch mehrere oder gar alle Chiffrentextzeichen ändern.

Lösungen:

Die Ordnungen der Buchstaben können folgender Tabelle entnommen werden:

0	1	2	3	4	5	6	7	8	9	10	11	12
A	B	C	D	E	F	G	H	I	J	K	L	M
13	14	15	16	17	18	19	20	21	22	23	24	25
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

1 Die Buchstaben $\square_1$=T und $\square_2$=E haben Ordnung ($\square_1$) = 19 bzw. Ordnung ($\square_2$) = 4. Mit $a = 1$, $b = 8$, $c = 6$ und $d = 25$ wird das erste Chiffrentextzeichen $\triangle_1$ wie folgt ermittelt:

Ordnung ($\triangle_1$) $= a \cdot$ Ordnung ($\square_1$) + $b \cdot$ Ordnung ($\square_2$) mod 26 = 1· 19 + 8· 4 mod 26 = 25

Somit ist $\triangle_1$ =Z.

Für das zweite Chiffrentextzeichen $\triangle_2$ gilt entsprechend:

Ordnung ($\triangle_2$) $= c \cdot$ Ordnung ($\square_1$) + $d \cdot$ Ordnung ($\square_2$) mod 26 = 6· 19 + 25· 4 mod 26 = 6

Somit ist $\triangle_2$ = G.

2 Für die Chiffrierung des Bigramms XT gilt analog zu Aufgabe 1:

Ordnung ($\triangle_1$) = 1· 23 + 8· 19 mod 26 = 19 d.h. $\triangle_1$ =T

Ordnung ($\triangle_2$) = 6· 23 + 25· 19 mod 26 = 15 d.h. $\triangle_2$ =P

Der Chiffrentext zum Klartext TEXT lautet somit ZGTP.

3 Das Bigramm TT wird wie folgt chiffriert:

Ordnung ($\triangle_1$) = 1· 19 + 8· 19 mod 26 = 15 d.h. $\triangle_1$ =P

Ordnung ($\triangle_2$) = 6· 19 + 25· 19 mod 26 = 17 d.h. $\triangle_2$ =R

Wir erhalten also folgende Chiffrentexte für die Bigramme TE, XT und TT:

Klartext	Chiffrentext
TE	ZG
XT	TP
TT	PR

Beobachtung:

Die Chiffrentextzeichen eines Bigramms können sich beide ändern, auch wenn im Klartext nur ein Zeichen durch ein anderes ersetzt wird (vergleiche z.B. TE → ZG und TT → PR). Es ist also nicht möglich, für ein bestimmtes Zeichen an einer bestimmten Position

innerhalb eines Blocks vorherzusagen, auf welches Geheimtextsymbol es abgebildet wird, ohne die übrigen Klartextzeichen des Blocks ebenfalls zu kennen.

4 Da bei Blockchiffren die Chiffrierung blockweise erfolgt (bei Bigrammen z.B. paarweise), kann keine sinnvolle Aussage über Abbildungen einzelner Zeichen innerhalb eines Blocks gemacht werden. Die Abbildung ganzer Blöcke hingegen ist injektiv. Die Hill-Chiffre ist also bzgl. eines ganzen Blocks «monoalphabetisch». Derselbe Klartextblock wird immer auf denselben Chiffrentextblock abgebildet.

5 Wegen der minimalen Blockgrösse $n = 2$, die in diesem Beispiel zur Anwendung kommt, kann eine Häufigkeitsanalyse von Bigrammen mit Schrittweite 2 die zur Entschlüsselung notwendigen Informationen liefern. In der Praxis werden bei Blockchiffren sehr viel grössere Blocklängen (64 Bit oder 128 Bit) gewählt und mit Transpositionen kombiniert, so dass eine stochastische Kryptoanalyse nicht mehr praktikabel ist.

6 Zunächst müssen die Bigramme mit Schrittweite 2 ausgezählt werden. Für die 4 häufigsten Bigramme erhält man:

Bigramm	Anzahl	Rel. Häufigkeit
XF	13	6.5%
KJ	8	4.0%
LA	7	3.5%
HU	6	3.0%

Vom Klartext wissen wir, dass die Bigramme TH und HE mit den Häufigkeiten 6.5% und 4% die am häufigsten vorkommenden sind, da es sich offenbar um einen englischsprachigen Text handeln muss.

Somit schliessen wir, dass TH auf XF und HE auf KJ abgebildet wird. Zur Entschlüsselung müssen wir also Zahlen e, f, g und h so finden, dass das Chiffrentextpaar $\triangle_1$= X und $\triangle_2$=F auf das Klartextzeichenpaar $\square_1$= T und $\square_2$= H und das Chiffrentextpaar $\triangle_1$= K und $\triangle_2$=J auf das Klartextzeichenpaar $\square_1$= H und $\square_2$= E abgebildet wird:

Ordnung (T) = $e \cdot$ Ordnung(X) + $f \cdot$ Ordnung(F) mod 26

Ordnung (H) = $g \cdot$ Ordnung(X) + $h \cdot$ Ordnung(F) mod 26

Ordnung (H) = $e \cdot$ Ordnung(K) + $f \cdot$ Ordnung(J) mod 26

Ordnung (E) = $g \cdot$ Ordnung(K) + $h \cdot$ Ordnung(J) mod 26

Nach Einsetzen der entsprechenden Ordnungszahlen für die einzelnen Buchstaben erhalten wir ein Gleichungssystem mit 4 Gleichungen und 4 Unbekannten:

$$\begin{cases} 19 = 23\,e + 5\,f \bmod 26 \\ 7 = 23\,g + 5\,h \bmod 26 \\ 7 = 10\,e + 9\,f \bmod 26 \\ 4 = 10\,g + 9\,h \bmod 26 \end{cases}$$

Da die Variablen e und f nur in der ersten und dritten Gleichung auftauchen sowie die Variablen g und h nur in der zweiten und vierten Gleichung, können wir das 4x4-Gleichungssystem auch lösen, indem wir die beiden entsprechenden 2x2-Gleichungssysteme getrennt betrachten:

$$\begin{cases} 19 = 23\,e + 5\,f \bmod 26 \\ 7 = 10\,e + 9\,f \bmod 26 \end{cases}$$

Durch Subtraktion des 5-fachen der zweiten Gleichung vom 9-fachen der ersten Gleichung erhalten wir:

$$136 = 157e \bmod 26$$

Wegen 136 mod 26 = 6 und 157 mod 26 = 1 ist diese Gleichung äquivalent zur Gleichung:

$$6 = e \pmod{26}$$

Durch Einsetzen von $e = 6$ in der ersten Gleichung $19 = 23\,e + 5\,f \bmod 26$ folgt für f:

$$19 = 138 + 5\,f \bmod 26$$

Die Subtraktion von 138 auf beiden Seiten ergibt:

$$-119 = 5\,f \bmod 26$$

Wegen $-119 \bmod 26 = -119 + 5 \cdot 26 \bmod 26 = 11$ ist diese Gleichung äquivalent zur Gleichung:

$$11 = 5\,f \bmod 26$$

Beispielsweise durch Probieren findet man, dass mit $f = 23$ die Gleichung erfüllt ist:

$$5 \cdot 23 \bmod 26 = 115 \bmod 26 = 11$$

Für die Variablen g und h betrachten wir analog folgendes 2x2-Gleichungssystem:

$$\begin{cases} 7 = 23\,g + 5\,h \bmod 26 \\ 4 = 10\,g + 9\,h \bmod 26 \end{cases}$$

Durch Subtraktion des 5-fachen der zweiten Gleichung vom 9-fachen der ersten Gleichung erhalten wir:

$$43 = 157g \bmod 26$$

Wegen 38 mod 26 = 12 und 157 mod 26 = 1 ist diese Gleichung äquivalent zur Gleichung:

$$17 = g \pmod{26}$$

Durch Einsetzen von $g = 17$ in der ersten Gleichung $7 = 23\,g + 5\,h \mod 26$ folgt für h:

$$7 = 391 + 5\,h \mod 26$$

Die Subtraktion von 391 auf beiden Seiten ergibt:

$$-384 = 5\,h \mod 26$$

Wegen $-384 \mod 26 = -384 + 15 \cdot 26 \mod 26 = 6$ ist diese Gleichung äquivalent zur Gleichung-

$$6 = 5\,h \mod 26$$

Wiederum durch Probieren findet man, dass $h = 22$ die Gleichung erfüllt:

$$5 \cdot 22 \mod 26 = 110 \mod 26 = 6$$

Mit Hilfe der so erhaltenen Werte $e = 6, f = 23, g = 17$ und $h = 22$ können wir nun Chiffrentextbigramme $\triangle_1\triangle_2$ mit folgendem Schema in Klartextbigramme $\square_1\square_2$ entschlüsseln :

$$\text{Ordnung}(\square_1) = 6 \cdot \text{Ordnung}(\triangle_1) + 23 \cdot \text{Ordnung}(\triangle_2) \mod 26$$

$$\text{Ordnung}(\square_2) = 17 \cdot \text{Ordnung}(\triangle_1) + 22 \cdot \text{Ordnung}(\triangle_2) \mod 26$$

Der Klartext lautet:

the human mind works along the same lines in spite of an attempt at originality on the part of the individual and this is particularly true of cipher work because there are so few sources of information available in other words the average man when he sits down to evolve a cipher has nothing to improve upon he invents and there is no one to tell him that his invention is in principle hundreds of years old the ciphers of the tritheme are the basis of most of the modern substitution ciphers

Geschichtliche Anmerkung

Der Autor des Texts, Parker Hitt, bezieht sich hier auf die sogenannten «Trithemius-Chiffren», die der Benediktiner Johannes Trithemius zu Beginn des 16. Jahrhunderts veröffentlichte. Unter diesen Chiffren war auch eine der ersten polyalphabetischen Geheimschriften (siehe «Informatik: Daten verwalten, schützen und auswerten», Klett-Verlag, 2022, S. 26).

7 Häufigkeitsanalysen von Zeichen in Klar- und Geheimtext bestätigen, dass Klartextzeichen nicht einzeln durch jeweils dasselbe Geheimtextzeichen ersetzt worden sind:

Klartextzeichen	Anzahl	Rel. Häufigkeit
E	48	11.9%
T	39	9.7%
I	38	9.5%
O	34	8.5%
N	32	8.0%
H	28	8.0%

Geheimtextzeichen	Anzahl	Rel. Häufigkeit
F	36	9.0%
G	28	7.0%
J	24	6.0%
X	22	5.5%
W	22	5.5%
K	20	5.0%

Da wir in diesem Beispiel Blocklänge $n = 2$ gewählt haben, sehen wir hingegen in der Häufigkeitsanalyse von Bigrammen mit Schrittweite 2 eine klare 1 : 1 Korrespondenz:

Klartextbigramm	Anzahl	Rel. Häufigkeit
TH	13	6.5%
HE	8	4.0%
OF	7	3.5%
IN	6	3.0%

Geheimtextbigramm	Anzahl	Rel. Häufigkeit
XF	13	6.5%
KJ	8	4.0%
LA	7	3.5%
HU	6	3.0%

8 Durch Vergrösserung der Blocklängen wird ein auf der Häufigkeitsanalyse basierender Angriff wegen der sog. «Diffusion» von Klartextänderungen auf alle Geheimtextzeichen eines Blockes sehr viel schwieriger.

Allerdings werden in der Hill-Chiffre lineare Gleichungen zur Verschlüsselung benutzt. Diese sind grundsätzlich sehr anfällig für sogenannte Klartextangriffe. Wie wir bei der Entschlüsselung gesehen haben, reicht die Kenntnis von n Klartextblöcken in Kombination mit den entsprechenden Geheimtextblöcken aus, um die für die Entschlüsselung benötigten Zahlen mit Hilfe eines linearen Gleichungssystems zu ermitteln. Deshalb ist die Hill-Chiffre heute nicht mehr von praktischer Bedeutung. Verwendet man hingegen bei der Verschlüsselung von Blöcken nichtlineare Gleichungen, werden Klartextangriffe sehr viel schwieriger.