Datenverschlüsselung: Die Hill-Chiffre

André Nuber Leistungsnachweis Fachdidaktik GymInf 14.02.23

Bei dem vorliegenden Dokument handelt es sich um einen Leistungsnachweis im Rahmen der Fachdidaktik I, GymInf, 2022. Das folgende Unterrichtsmaterial richtet sich an eine MINT-Klasse, welche die lineare Algebra bereits behandelt hat. Wir befinden uns in einem Unterrichtsblock zum Thema Kryptographie. Monoalphabetische Kryptosysteme wurden bereits behandelt und mit stochastischer Analyse geknackt. Das vorliegende Unterrichtsmaterial zu einem polyalphabetischen Kryptosystem ist für die Dauer von mindestens zwei Lektionen konzipiert; es ist ähnlich aufgebaut wie die Themen im Buch INFORMATIK – Data Science und Sicherheit (Barot et al., 2022).

Voraussetzungen

- Die SuS wurden bereits in die lineare Algebra eingeführt. Insbesondere sind sie in der Lage, Matrix-Vektor-Multiplikationen auszuführen und Matrizen zu invertieren. Sie können die Determinante einer Matrix berechnen und beurteilen, ob die Matrix invertierbar ist.
- Die SuS sind mit der Modulo-Operation vertraut.
- Die SuS kennen den Begriff «inverses Element» bzgl. Addition bzw. Multiplikation.
- Wir befinden uns in einem Ausbildungsblock zum Thema Kryptographie. Die monoalphabetische Verschlüsselung wurde bereits mit stochastischer Analyse geknackt (Barot et al., S. 62ff.; ABZ, 2022). Der Bedarf nach polyalphabetischen Kryptosystemen (Barot et al, S. 66ff.) liegt daher auf der Hand.

Lernziele

Rahmenlehrplan Informatik

- «Verschiedene Codierungen und Darstellungen von Informationen kennen
- Sicherheitsaspekte der digitalen Kommunikation verstehen, z.B. Verschlüsselung [...]
- Eigene und fremde Lösungswege formal beschreiben und kritisch analysieren
- Algorithmen entwerfen, beurteilen und in einer Programmiersprache umsetzen
- Sicherheitsrisiken bei der digitalen Kommunikation einschätzen und angemessene Massnahmen treffen»

(EDK, 2017)

Eigene Lernziele

Die SuS ...

- verstehen die Hill-Chiffre als polyalphabetisches Kryptosystem (T2).
- können die Hill-Chiffre mit Schlüssellänge 4 auf einen kurzen Text anwenden (T3).
- können einen mit Hill-Chiffre verschlüsselten Text entschlüsseln (T3).
- können beschreiben, wie anhand einer langen Klartext-Nachricht und deren Chiffre ein Schlüssel der bekannten Länge 4 geknackt werden kann (T2).
- erkennen eine Schwäche der Hill-Chiffre und können eine Optimierung vorschlagen (T4-5).

T1-5: Taxonomiestufe nach Bloom et al. (1956)

Motivation

A1: Monoalphabetische Verschlüsselung knacken

Die folgende Nachricht wurde mit einem monoalphabetischen Kryptosystem verschlüsselt:

IUD QUFJIGIQ VIGTXJGID CXQUIGM XZT IUDIP KIJIUPID QFJBZIQQIB.

a) Finde die Klartext-Nachricht heraus. Als Starthilfe sei gegeben: Der Buchstabe C im Klartext wurde mit der Chiffre F verschlüsselt.

b) Welche Voraussetzungen müssen erfüllt sein, damit dein Hack funktioniert?

Fazit: Mit stochastischer Analyse lassen sich monoalphabetische Kryptosysteme sehr leicht knacken. Schwieriger wird es, wenn der gleiche Buchstabe nicht immer in die gleiche Chiffre übersetzt wird. Dies ist die Grundidee von polyalphabetischen Verfahren. So könnte man den Text z.B. in Zweierblöcke unterteilen und jedem 2er-Block einen anderen zuordnen, z.B.:

Klartext	GE	HE
Chiffre	EI	JC

Der Buchstabe E wird nun also einmal dem I, einmal dem C zugeordnet, abhängig davon, welcher Buchstabe vor dem E steht. Die Verschlüsselung ist kontextsensitiv. Dadurch verändert sich die relative Häufigkeit der Buchstaben in der verschlüsselten Nachricht. Eine Statistik über die relative Häufigkeit von Einzelbuchstaben wird somit nutzlos.

Historischer Kontext

Ein solches Verfahren hat der amerikanische Mathematiker Lester S. Hill im Jahre 1929 vorgeschlagen (Hill, 1929). Je länger die Blöcke gewählt werden, desto sicherer wird das Verfahren. Hills Kryptograph (Fig. 1) war damals der einzige, der Blöcke mit mehr als vier Zeichen gleichzeitig bearbeiten konnte! Um das Verfahren kennen zu lernen, beschränken wir uns hier aber vorerst auf Blöcke der Länge 2.

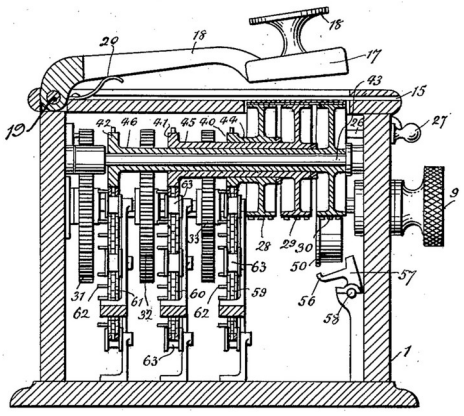

Figure 1: Kryptograph nach L. S. Hill, 1929 (Wikipedia, 2022).

Theorie

Die Hill-Chiffre basiert auf linearer Algebra, genauer gesagt auf Matrix-Vektor-Multiplikationen mit einer invertierbaren Matrix. Die Verschlüsselung läuft wie folgt ab:

Algorithmus: Verschlüsseln nach Hill

1. Unterteile die Nachricht in Blöcke der Länge $n = 2$.
2. Ordne mit der Tabelle jedem Buchstaben eine Zahl zu, so entstehen Vektoren $\vec{u}_i$ der Länge *n*.

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

3. Der Schlüssel entspricht einer invertierbaren $n \times n$-Matrix A. Jeder Block wird mit einer Matrix-Vektor-Multiplikation $\vec{v}_i = A\,\vec{u}_i$ verschlüsselt.
4. Rechne die Vektoren $\vec{v}_i$ komponentenweise mod 26 (Länge des Alphabets).
5. Jede Zahl wird nun gemäss Tabelle in Schritt 2 wieder einem Buchstaben zugeordnet.

Beispiel 1: Verschlüsseln nach Hill

Der Name TONY wird mit dem Schlüssel HILL gemäss obigem Rezept verschlüsselt:

1. TO | NY
2. Die zu verschlüsselnden Vektoren lauten also:
   $\vec{u}_1 = \begin{pmatrix} T \\ O \end{pmatrix} = \begin{pmatrix} 19 \\ 14 \end{pmatrix}$ und $\vec{u}_2 = \begin{pmatrix} N \\ Y \end{pmatrix} = \begin{pmatrix} 13 \\ 24 \end{pmatrix}$
   Soll ein Wort als Schlüssel verwendet werden, so muss auch dieses in Zahlen übersetzt und in eine Matrix eingetragen werden:
   $A = \begin{pmatrix} H & I \\ L & L \end{pmatrix} = \begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix}$
3. Die Matrix-Vektor-Multiplikation ergibt:
   $\vec{v}_1 = \begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix} \cdot \begin{pmatrix} 19 \\ 14 \end{pmatrix} = \begin{pmatrix} 245 \\ 363 \end{pmatrix}$, $\quad \vec{v}_2 = \begin{pmatrix} 7 & 8 \\ 11 & 11 \end{pmatrix} \cdot \begin{pmatrix} 13 \\ 24 \end{pmatrix} = \begin{pmatrix} 283 \\ 407 \end{pmatrix}$
4. Damit die Zahlen wieder Buchstaben zugeordnet werden können, rechnen wir «mod 26»:
   $\vec{v}_1 = \begin{pmatrix} 245 \\ 363 \end{pmatrix} \mod 26 = \begin{pmatrix} 11 \\ 25 \end{pmatrix}$, $\quad \vec{v}_2 = \begin{pmatrix} 283 \\ 407 \end{pmatrix} \mod 26 = \begin{pmatrix} 23 \\ 17 \end{pmatrix}$
5. Aus den Zahlen 11, 25, 23, 17 wird das Wort LZXR.

A2: Verschlüsseln nach Hill

a) Verschlüssle das Wort GEHEIM mit dem Schlüssel FAUL.

b) Beschreibe, wie du vorgehen würdest, um eine Chiffre zu *ent*schlüsseln. Teste dein Vorgehen anhand deiner Chiffre, bis du auf grössere Probleme stösst.

Beim Entschlüsseln dürftest du wohl auf ein Problem gestossen sein: Möchte man $\vec{u}_1$ mit der inversen Matrix von A berechnen, so erhält man Bruchzahlen als Resultate. Wie sollen diese nun Buchstaben zugeordnet werden? Was wir brauchen, ist nicht die übliche, sondern die modulare Inverse einer Matrix; wir nennen diese in der Folge A^{-1}.

Algorithmus: Entschlüsseln nach Hill

1. Unterteile die Chiffre in Blöcke der Länge $n = 2$.
2. Ordne jedem Buchstaben eine Zahl zu gemäss der oben angegebenen Tabelle.
3. Berechne die modulare Inverse A^{-1}.
4. Entschlüssle gemäss $\vec{u}_i = A^{-1}\,\vec{v}_i$.
5. Rechne die Vektoren komponentenweise mod 26 (Länge des Alphabets).
6. Jede Zahl wird nun gemäss Tabelle in Schritt 2 wieder einem Buchstaben zugeordnet.

Wir werfen einen Blick auf Schritt 3. Die übliche Inverse einer Matrix berechnen wir gemäss:

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} := (\det A)^{-1} B$$

Um die *modulare Inverse* einer Matrix zu berechnen, multiplizieren wir die Kofaktor-Matrix B (siehe oben) mit der modularen Inversen der Determinante statt mit ihrem Kehrwert. Die modulare Inverse j einer Zahl i ist so definiert, dass: $i \cdot j \mod 26 = 1$.

A3: Modulare Inverse

Ziel dieser Aufgabe ist es, die modularen Inversen (mod 26) der Zahlen 0 bis 25 zu finden. Dafür probieren wir einfach alle durch.

a) Schreibe ein Programm, das für alle Zahlen $i, j \in \{0, ..., 25\}$ das Produkt $i \cdot j \mod 26$ berechnet. Wenn es gleich 1 ist, soll das Programm die Zahlen ausgeben. Beispielausgabe:

```
1 * 1 mod 26 = 1
3 * 9 mod 26 = 1
...
```

b) Trage die Zahlen i und ihre modularen Inversen j in die Tabelle ein:

i	1	3
j	1	9

c) Wie du siehst, haben nicht alle Zahlen eine modulare Inverse. Kannst du einen Zusammenhang formulieren zwischen den Zahlen, die eine besitzen, und der Länge des Alphabets, 26?

Nun sind wir in der Lage, die modulare Inverse von Zahlen zu berechnen. Damit können wir auch die modulare Inverse einer Matrix berechnen:

$$A^{-1} = (\det A)^{-1}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \mod 26$$

Dabei ist $(\det A)^{-1}$ die modulare Inverse der Determinanten.

Beispiel 2: Entschlüsseln nach Hill

Wir entschlüsseln die Chiffre LZXR, die mit dem Schlüssel HILL chiffriert wurde.

1. LZ | XR
2. Die zu entschlüsselnden Vektoren lauten:

   $$\vec{v}_1 = \begin{pmatrix} L \\ Z \end{pmatrix} = \begin{pmatrix} 11 \\ 25 \end{pmatrix} \text{ und } \quad \vec{v}_2 = \begin{pmatrix} X \\ R \end{pmatrix} = \begin{pmatrix} 23 \\ 17 \end{pmatrix}$$

3. Wir berechnen zunächst die modulare Inverse der Determinanten:

   $$(\det A)^{-1} = ((7 \cdot 11 - 8 \cdot 11) \mod 26)^{-1} = (-11 \mod 26)^{-1} = 15^{-1} = 7$$

   ... und daraus die modulare Inverse der Schlüsselmatrix A:

   $$\begin{aligned} A^{-1} &= (\det A)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \mod 26 \\ &= 7 \cdot \begin{pmatrix} 11 & -8 \\ -11 & 7 \end{pmatrix} \mod 26 \\ &= \begin{pmatrix} 77 & -56 \\ -77 & 49 \end{pmatrix} \mod 26 \\ &= \begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} \end{aligned}$$

4. Die Matrix-Vektor-Multiplikation ergibt:

   $$\vec{u}_1 = \begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} \cdot \begin{pmatrix} 11 \\ 25 \end{pmatrix} = \begin{pmatrix} 825 \\ 586 \end{pmatrix}, \qquad \vec{v}_2 = \begin{pmatrix} 25 & 22 \\ 1 & 23 \end{pmatrix} \cdot \begin{pmatrix} 23 \\ 17 \end{pmatrix} = \begin{pmatrix} 949 \\ 414 \end{pmatrix}$$

5. Damit die Zahlen wieder Buchstaben zugeordnet werden können, rechnen wir «mod 26»:

   $$\vec{u}_1 = \begin{pmatrix} 825 \\ 586 \end{pmatrix} \mod 26 = \begin{pmatrix} 19 \\ 14 \end{pmatrix}, \qquad \vec{v}_2 = \begin{pmatrix} 949 \\ 414 \end{pmatrix} \mod 26 = \begin{pmatrix} 13 \\ 24 \end{pmatrix}$$

6. Wir erhalten die richtigen Vektoren zurück, die TONY zugeordnet werden können.

A4: Entschlüsseln nach Hill

a) Berechne die modulare Inverse der Schlüssel-Matrix aus A2.

b) Überprüfe dein Resultat mit diesem Online-Tool.

c) Entschlüssle nun deine Chiffre aus A2. Erhältst du das ursprüngliche Wort zurück?

Nicht jede Matrix ist als Schlüssel für die Hill-Chiffre geeignet. Natürlich muss sie invertierbar sein, mehr noch, sie muss modular invertierbar sein. Daher muss gelten:

- $\det(A) \neq 0$
- $\det(A)$ hat eine modulare Inverse, sprich, $\det(A)$ und die Länge des Alphabets sind teilerfremd.

Die zweite Bedingung lässt sich besonders leicht erfüllen, wenn die Länge *n* des Alphabets einer Primzahl entspricht.

A5: Geeignete Schlüssel

a) Welche dieser Wörter sind als Schlüssel geeignet (Länge des Alphabets: 26)?
MAMA, TELL, BAUM

b) Finde zwei weitere geeignete Schlüssel.

Stochastische Analyse der Hill-Chiffre

Unser Ziel war, ein Verfahren zu finden, dass resistent ist gegen stochastische Angriffe. Doch ist uns dies wirklich gelungen? Woran denkst du, wenn du diese Tabelle siehst (aus Barot et al., 2022)?

ER	3,89 %
EN	3,74 %
CH	2,97 %
TE	2,21 %
ND	2,11 %
DE	2,06 %

A6: Stochastisches Knacken der Hill-Chiffre

a) Formuliere deine Überlegungen anhand der obigen Tabelle. Wie können Bigramme genutzt werden, um die Hill-Chiffre zu knacken? Notiere deine Überlegungen, bevor du weiterliest.

b) In einem langen, verschlüsselten Text, haben wir die Bigramme gezählt. Wir vermuten die folgende Zuordnung:

Klartext	ER	CH
Chiffre	MZ	GL

Die Schlüsselmatrix sei $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Verwende die vermutete Zuordnung, um den Schlüssel herauszufinden.

c) DGZL hat die Nachricht verschlüsselt. Wer ist das?

d) Überarbeite deine Überlegungen aus Teilaufgabe (a).

e) Wie kann das Verfahren sicherer gemacht werden?

Zusammenfassung

Monoalphabetische Kryptoverfahren sind mit einer stochastischen Analyse sehr leicht zu knacken. Eine mögliche Verbesserung bieten polyalphabetische Verfahren. L. S. Hill hat ein solches vorgeschlagen, das auf linearer Algebra basiert. Die Idee ist: Die Nachricht wird in gleich lange Blöcke unterteilt. Jeder dieser Blöcke wird, nachdem die Buchstaben Zahlen zugeordnet wurden, mit einer Matrix (= Schlüssel) multipliziert. Dadurch wirkt sich jeder Buchstabe auf die Verschlüsselung des ganzen Blockes aus (sog. Diffusion), sodass ein Buchstabe nicht mehr eindeutig einem anderen zugeordnet werden kann. Die Verschlüsselung ist kontextsensitiv, da die Chiffre jedes Buchstabens auch von seinen Nachbarn abhängt. Die stochastische Analyse wird somit nutzlos, wenn die Blöcke und die Schlüssel genügend gross gewählt werden. Die Schlüsselmatrix muss modular invertierbar sein, damit man die Chiffre auch wieder entschlüsseln kann. Ihre Determinante darf also nicht 0 sein und muss zusätzlich teilerfremd zur Länge des Alphabets sein. Dies gelingt am einfachsten, wenn man das Alphabet so wählt, dass seine Länge einer Primzahl entspricht.

Teste dich selbst

Konzepte und Begriffe

1. In welchem Sinne ist die Hill-Chiffre monoalphabetischen Verfahren überlegen?
2. Beschreibe das Verschlüsseln mit der Hill-Chiffre in eigenen Worten.
3. Welche Voraussetzungen muss der Schlüssel für die Hill-Chiffre erfüllen?
4. Eignet sich ein Alphabet der Länge 36 für die Hill-Chiffre?
5. Beschreibe das Entschlüsseln mit der Hill-Chiffre in eigenen Worten.
6. Wodurch unterscheidet sich die Berechnung der modularen Inverse einer Matrix von der üblichen Berechnung einer Inversen?
7. Unter welchen Bedingungen ist die Hill-Chiffre mit stochastischer Analyse angreifbar?
8. Wie würdest du vorgehen, wenn die Länge des zu verschlüsselnden Textes nicht durch die Länge der Blöcke teilbar ist?

Zusatzaufgaben

1. Verschlüssle das Wort SAHARA mit dem Schlüssel BUNT. Entschlüssle die Chiffre, um zu kontrollieren, ob du die Nachricht rekonstruieren kannst.
2. Geeignete Länge des Alphabets
   a) Modifiziere dein Programm aus A3a so, dass für alle Zahlen $i, j \in \{0, ..., 28\}$ das Produkt $i \cdot j \mod 29$ berechnet wird. Welche Zahlen haben nun eine modulare Inverse?
   b) Die Vermutung liegt nahe, dass man einfach die nächstgrössere Primzahl als Länge des Alphabets wählt, sodass $\det(A) \neq 0$ stets eine modulare Inverse hat. Welche Probleme könnten dabei auftauchen?
   c) Wie lässt sich das Problem aus (b) beheben?
3. Wähle eine 3×3-Matrix als Schlüssel und stelle sicher, dass sie die notwendigen Bedingungen erfüllt. Teile dein Schlüsselpaar, sprich, die Schlüssel zum Ver- und Entschlüsseln, mit jemandem aus der Klasse, sodass ihr verschlüsselte Nachrichten austauschen könnt. Schickt euch eine kurze, verschlüsselte Nachricht (z.B. ein Wort).

Literaturverzeichnis

ABZ (2022): *einfach INFORMATIK – Entschlüssle ohne Schlüssel*. Abgerufen am 18.11.2022 unter https://einfachinformatik.inf.ethz.ch/application/ (Aufgaben – Datenschutz und Geheimschriften – Entschlüssle ohne Schlüssel).

Barot, M., Dorn, B., Fourny, G., Gallenbacher, J., Hromkovič, J., Lacher, R. (2022): *INFORMATIK – Data Science und Sicherheit*. Baar: Klett und Balmer.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. & Kratwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals.* Vol. Handbook I: Cognitive domain. New York: David McKay Company.

EDK (2017): Rahmenlehrplan für die Maturitätsschulen: Informatik. Abgerufen am 18.11.2022 unter https://www.edk.ch/de/themen/gymnasium.

Hill, L. S. (1929): *Cryptography in an Algebraic Alphabet*. The American Mathematical Monthly, Vol. 36, No. 6, pp. 306-312.

Iyer, S. R. (2018): *Inverse of a modular matrix*. Abgerufen am 18.11.2022 unter https://math.stackexchange.com/questions/2686150/inverse-of-a-modular-matrix.

Städeli, C. (2010): *Das AVIVA-Modell für den kompetenzorierntierten Unterricht – Die fünf Säulen der guten Unterrichtsvorbereitung*. Folio, 6/2010.

Wikipedia (2022): *Hill cipher*. Abgerufen am 18.11.2022 unter https://en.wikipedia.org/wiki/Hill_cipher.