1 Einmalige symmetrische Verschlüsselung und klassische Verfahren

Wir diskutieren hier eine einführende, einfache Situation, für symmetrische Konzelationssysteme und Sicherheitsmodelle. In einer Fallstudie betrachten wir Methoden zum „Brechen“ eines klassischen Kryptosystems.

1.1 Kryptosysteme und possibilistische Sicherheit

Szenarium 1 (Einmalige Verschlüsselung, COA): Alice möchte Bob *einen* Klartext *vorher bekannter* Länge schicken, Alice und Bob haben sich auf einen Schlüssel geeinigt, Eva hört den Chiffretext mit.

„bekannte Länge“: Klartexte entstammen einer bekannten endlichen Menge X, z. B. $X = \{0,1\}^\ell$.

Fragen: Wie soll man vorgehen, damit das verwendete Verfahren als „sicher“ gelten kann? Was soll „sicher“ überhaupt bedeuten? Wie kann man „Sicherheit“ *beweisen*? Später: Was sind die Risiken von Varianten (mehrere Nachrichten, längere Nachrichten usw.)?

Definition 1.1 *Ein* Kryptosystem *ist ein Tupel* $\mathcal{S} = (X, K, Y, e, d)$*, wobei*

- X *und* K *nichtleere endliche Mengen sind [Klartexte bzw. Schlüssel],*
- Y *eine Menge ist [Chiffretexte], und*
- $e\colon X \times K \to Y$ *und* $d\colon Y \times K \to X$ *Funktionen sind [Verschlüsselungsfunktion bzw. Entschlüsselungsfunktion],*

so dass Folgendes gilt:

(1) $\forall x \in X \forall k \in K\colon d(e(x,k),k) = x$ *(Dechiffrierbedingung)*

(2) $\forall y \in Y \exists x \in X, k \in K\colon y = e(x,k)$ *(Surjektivität)*

Bemerkung: Surjektivität kann immer hergestellt werden, indem man Y auf das Bild $\mathrm{Bi}(e) = e(X \times K)$ einschränkt. Die Forderung ist für unsere Analysen aber bequem.

Für festes $k \in K$ wird die Funktion $e(\,.\,,k)\colon X \to Y, x \mapsto e(x,k)$ als *Chiffre* bezeichnet.

Beispiel 1.2 Sei $n > 0$, $X = \{a_i, b_i \mid 1 \leq i \leq n\}$, $K = \{k_0, k_1\}$, $Y = \{A_i, B_i \mid 1 \leq i \leq n\}$. Die Funktionen e und d sind als Tabellen gegeben:

e	a_1	b_1	a_2	b_2	$\ldots$	a_n	b_n
k_0	A_1	B_1	A_2	B_2	$\ldots$	A_n	B_n
k_1	B_1	A_1	B_2	A_2	$\ldots$	B_n	A_n

d	A_1	B_1	A_2	B_2	$\ldots$	A_n	B_n
k_0	a_1	b_1	a_2	b_2	$\ldots$	a_n	b_n
k_1	b_1	a_1	b_2	a_2	$\ldots$	b_n	a_n

Dann gelten Dechiffrierbedingung und Surjektivität, (X, K, Y, e, d) ist also ein Kryptosystem (wenn auch auf den ersten Blick ein nicht sehr intelligentes).

Man kann Kryptosysteme auch durch eine mathematische Beschreibung angeben. Im Wesentlichen genau dasselbe Kryptosystem wie in Beispiel 1.2 ist das folgende: $X = Y = \{0,1\}^\ell$, $n = 2^{\ell-1}$. Die Elemente dieser Mengen fassen wir als Binärdarstellungen von Zahlen in $\{0, 1, \ldots, 2^\ell - 1\}$ auf. $A_1, \ldots, A_n$ sind die geraden Zahlen $0, 2, \ldots, 2^\ell - 2$, $B_1, \ldots, B_n$ die ungeraden Zahlen $1, 3, \ldots, 2^\ell - 1$ in dieser Menge. Genauso sind $a_1, \ldots, a_n$ und $b_1, \ldots, b_n$ definiert. Die Schlüssel sind $k_0 = 0$ und $k_1 = 1$, und $e(x, k_0) = d(x, k_0) = x$ und $e(x, k_1) = d(x, k_1)$ ist das Binärwort, das man erhält, wenn man in x das letzte Bit kippt: $e(x, k) = d(x, k) = x \oplus_\ell k$. (Dabei steht k für die Binärdarstellung von k mit ℓ Bits und $\oplus_\ell$ steht für das bitweise XOR.)

Beispiel 1.3 $X = \{\mathtt{a}, \mathtt{b}\}$, $K = \{k_0, k_1, k_2\}$, $Y = \{\mathtt{A}, \mathtt{B}, \mathtt{C}\}$. Die Funktion e ist gegeben durch die erste, die Funktion d durch die zweite der folgenden Tabellen. Dann ist (X, K, Y, e, d) Kryptosystem, denn die Dechiffrierbedingung und die Surjektivität sind erfüllt.

e	a	b
k_0	A	B
k_1	B	A
k_2	A	C

d	A	B	C
k_0	a	b	a
k_1	b	a	a
k_2	a	a	b

Beispiel 1.4 $X = \{\mathtt{a}, \mathtt{b}\}$, $K = \{k_0, k_1, k_2\}$, $Y = \{\mathtt{A}, \mathtt{B}, \mathtt{C}\}$. Die Funktion e ist durch die folgende Tabelle gegeben:

e	a	b
k_0	A	B
k_1	B	B
k_2	A	C

Wegen $e(\mathsf{a}, k_1) = e(\mathsf{b}, k_1)$ existiert keine Funktion d, so dass (X, K, Y, e, d) ein Kryptosystem ist, die Dechiffrierbedingung kann also nicht erfüllt werden.

Merke: Jede Chiffre $e(\,.\,, k)$ eines Kryptosystems muss injektiv sein. (Sonst kann es keine Entschlüsselungsfunktion d geben. Anschaulich: Die Einträge in jeder Zeile der Tabelle für e müssen verschieden sein.)

Beispiel 1.5 $X = K = Y = \{0\}$, $e(0,0) = d(0,0) = 0$ (auch $X = \{x\}$, $K = \{k\}$, $Y = \{y\}$). Dies ist das „triviale“ minimale Kryptosystem. Dechiffrierbedingung und Surjektivität gelten offensichtlich.

Beispiel 1.6 Sei $\oplus\colon \{0,1\}\times\{0,1\} \to \{0,1\}$ die Funktion $(b,c) \mapsto b+c-2bc$ ($= b \,\mathrm{XOR}\, c$). Für $\ell > 0$ sei $\oplus_\ell\colon \{0,1\}^\ell \times \{0,1\}^\ell \to \{0,1\}^\ell$ die komponentenweise Anwendung von $\oplus = \oplus_1$:

$$(b_1, b_2, \ldots, b_\ell) \oplus_\ell (c_1, c_2, \ldots, c_\ell) = (b_1 \oplus c_1, b_2 \oplus c_2, \ldots, b_\ell \oplus c_\ell)\,.$$

Sei $\ell > 0$. Das *Vernam-Kryptosystem*[10] oder *one-time pad der Länge* ℓ ist das Kryptosystem $(\{0,1\}^\ell, \{0,1\}^\ell, \{0,1\}^\ell, \oplus_\ell, \oplus_\ell)$.

In diesem Fall ist es für nicht ganz kleine ℓ offensichtlich unbequem, wenn nicht ganz unmöglich, die Ver- und Entschlüsselungsfunktion durch Tabellen anzugeben. Man benutzt hier und auch üblicherweise mathematische Beschreibungen.

Beispiel: $x = 1011001$, $k = 1101010$. Dann ist $y = e(x,k) = 1011001 \oplus_7 1101010 = 0110011$. Zur Kontrolle: $d(y,k) = 0110011 \oplus_7 1101010 = 1011001 = x$.

Wir kontrollieren, dass das Vernam-System tatsächlich ein Kryptosystem ist.

(1) Für $x \in X$ und $k \in K$ gelten $d(e(x,k),k) = (x \oplus_\ell k) \oplus_\ell k = x \oplus_\ell (k \oplus_\ell k) = x \oplus_\ell 0^\ell = x$, d.h. die Dechiffrierbedingung ist erfüllt.

(2) Für $y \in Y$ gilt $e(y, 0^\ell) = y$ und $y \in X$, $0^\ell \in K$. Also gilt Surjektivität.

Wann soll ein Kryptosystem als sicher betrachtet werden?

Erste Idee: Wenn Eva den Chiffretext $e(x,k)$ abhört und den Schlüssel k nicht kennt, so soll sie nicht in der Lage sein, x zu *berechnen*.

[10] Benannt nach Gilbert S. Vernam (1890–1960), der im Jahr 1918 dieses System für fünf Bits in der Sprache einer Relais-Schaltung beschrieben und zum US-Patent angemeldet hat. Siehe `http://www.cryptomuseum.com/crypto/files/us1310719.pdf`.

Beispiel 1.2 (Fortsetzung) Wenn Eva den Chiffretext A_1 abhört, so weiß sie, dass der Klartext a_1 oder b_1 ist; sie kann aber nicht sagen, welcher von beiden es ist. Allerdings hat sie (signifikante) nichttriviale Information gewonnen, nämlich dass $a_2, b_2, \ldots, a_n, b_n$ nicht in Frage kommen.

Die Anforderung, dass x aus y nicht eindeutig bestimmt werden kann, führt also zu keinem befriedigenden Sicherheitsbegriff.

Zweite Idee: Wenn Eva den Chiffretext y abhört und den Schlüssel k nicht kennt, so kann sie keinen Klartext *ausschließen*. Dies führt zu der folgenden Definition.

Definition 1.7 *Ein Kryptosystem* $\mathcal{S} = (X, K, Y, e, d)$ *heißt* possibilistisch sicher*, wenn*

$$\forall y \in Y \, \forall x \in X \, \exists k \in K : e(x,k) = y \,.$$

Bemerkung 1.8

1. Sei $\mathcal{S} = (X, K, Y, e, d)$ Kryptosystem. Dann sind äquivalent:
   - $\mathcal{S}$ ist possibilistisch sicher.
   - $\forall x \in X : e(x, K) = \{e(x,k) \mid k \in K\} = Y$.
     (Anschaulich: Jede Spalte der e-Tabelle enthält alle Chiffretexte aus Y.)
2. Für $n \geq 2$ ist das Kryptosystem aus Beispiel 1.2 nicht possibilistisch sicher, denn A_1 kann nicht Chiffretext zu a_2 sein.
3. Das Kryptosystem aus Beispiel 1.3 ist nicht possibilistisch sicher, denn C kann nicht Chiffretext zu 0 sein.
4. Das Vernam-Kryptosystem der Länge ℓ ist possibilistisch sicher: Seien $x \in X$ und $y \in Y$. Setze $k = x \oplus_\ell y$. Dann gilt $e(x,k) = x \oplus_\ell (x \oplus_\ell y) = (x \oplus_\ell x) \oplus_\ell y = 0^\ell \oplus_\ell y = y$.

In der Einführung wurde die Verschiebechiffre betrachtet, bei der Buchstaben des alten lateinischen Alphabets auf Chiffrebuchstaben abgebildet wurden, indem man das Bild von `A` angab und jeder andere Buchstabe um dieselbe Distanz verschoben wurde. Auch die allgemeineren Substitutionschiffren wurden erwähnt, bei der man für jeden Buchstaben x einen beliebigen Bildbuchstaben $\pi(x)$ angibt, auf injektive Weise. Beispiel für eine Substitutionschiffre:

x	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	V	X
$\pi(x)$	F	Q	M	P	K	V	T	A	E	L	H	N	I	G	D	S	B	X	O	R	C

Man überlege: Es gibt 21! (= 51 090 942 171 709 440 000) viele solche Chiffren. Wir können für ganz beliebige Mengen X die Menge aller Substitutionschiffren auf X betrachten.

Definition 1.9 *Für eine endliche, nichtleere Menge X sei $K = \mathcal{P}_X$ die Menge der Permutationen*[11] *auf X. Das* Substitutionskryptosystem auf X *ist das Tupel $(X, \mathcal{P}_X, X, e, d)$ mit*[12]

$$e(x, \pi) = \pi(x) \text{ und } d(y, \pi) = \pi^{-1}(y)\,.$$

Man sieht leicht, dass dies tatsächlich ein Kryptosystem ist (Dechiffrierbedingung und Surjektivität).

Proposition 1.10 *Ist X eine endliche und nichtleere Menge, so ist das Substitutionskryptosystem auf X possibilistisch sicher.*

Beweis: Seien $y \in Y = X$ und $x \in X$. Definiere $\pi\colon X \to X$ wie folgt:

$$\pi(z) = \begin{cases} y & \text{falls } z = x \\ x & \text{falls } z = y \\ z & \text{falls } z \notin \{x, y\} \end{cases}$$

Dann ist $\pi \in \mathcal{P}_X$ mit $e(x, \pi) = \pi(x) = y$. □

Beobachtung: K ist in diesem Fall sehr groß, es gibt $|X|!$ Schlüssel. Im Fall des Vernam-Kryptosystems ist $|X| = |Y| = |K|$.

Proposition 1.11 *Ist $\mathcal{S} = (X, K, Y, e, d)$ ein possibilistisch sicheres Kryptosystem, so gilt $|X| \leq |Y| \leq |K|$.*

Beweis: Wähle $k \in K$ beliebig. Da $\mathcal{S}$ ein Kryptosystem ist, erfüllt es die Dechiffrierbedingung. Also ist die Chiffre $e(\,.\,, k)\colon X \to Y$ injektiv, d. h. es gilt $|X| \leq |Y|$.

Sei nun $x \in X$ beliebig. Da $\mathcal{S}$ possibilistisch sicher ist, gibt es für jedes $y \in Y$ ein $k \in K$ mit $e(x, k) = y$. Also ist die Abbildung $K \ni k \mapsto e(x, k) \in Y$ surjektiv, und es folgt $|Y| \leq |K|$. □

Folgerung: Bei possibilistischer Sicherheit und Klartexten und Schlüsseln, die Zeichenreihen über einem Alphabet sind, müssen Schlüssel mindestens so lang sein wie der zu übermittelnde Text. Wenn man etwa den Inhalt einer Festplatte verschlüsseln will, benötigt man eine weitere Festplatte für den Schlüssel. In solchen Fällen extrem langer Klartexte wird possibilistische Sicherheit unrealistisch. Possibilistisch sichere Systeme kommen daher nur als Bausteine in größeren Systemen vor, siehe Kapitel 2.

[11] Eine Permutation auf X ist eine bijektive Funktion $\pi\colon X \to X$.

[12] Wenn $\pi\colon X \to X$ eine Permutation ist, dann ist $\pi^{-1}\colon X \to X$ die Permutation mit $\pi^{-1}(\pi(x)) = \pi(\pi^{-1}(x)) = x$ für alle $x \in X$.

Beispiel 1.12 Sei $X = \{\mathtt{a}, \mathtt{b}\}$, $K = \{0, 1, 2\}$, $Y = \{\mathtt{A}, \mathtt{B}\}$ und die Verschlüsselungsfunktion sei durch

e	a	b
0	A	B
1	A	B
2	B	A

gegeben. Dann ist $\mathcal{S} = (X, K, Y, e, d)$ ein possibilistisch sicheres Kryptosystem. Fängt Eva den Chiffretext $e(x, k) = \mathtt{A}$ ab, so nimmt sie an, dass $x = \mathtt{a}$ „wahrscheinlicher“ ist als $x = \mathtt{b}$. Das ist zum Beispiel dann sinnvoll, wenn die Schlüssel 0, 1, 2 dieselbe Wahrscheinlichkeit haben. (Das Kerckhoffs-Prinzip würde sagen, dass Eva auch die verwendete Wahrscheinlichkeitsverteilung auf K kennt. Dazu unten mehr.)

Um formal auszudrücken, warum dieses Kryptosystem nicht „sicher“ ist, wenn Schlüssel 0, 1 und 2 gleich wahrscheinlich sind, beziehungsweise um einen passenden Sicherheitsbegriff überhaupt zu formulieren, benötigen wir etwas Wahrscheinlichkeitsrechnung.

1.2 Wiederholung: Elementare Wahrscheinlichkeitsrechnung

Die in dieser Vorlesung benötigten Konzepte aus der Wahrscheinlichkeitsrechnung wurden in den Veranstaltungen „Grundlagen und Diskrete Strukturen“ und „Stochastik für Informatiker“ im Prinzip behandelt. Wir erinnern hier kurz an die für unsere Zwecke wichtigen Konzepte und legen Notation fest.

Beispiel: Ein Wahrscheinlichkeitsraum, mit dem man das Zufallsexperiment „Einmaliges Werfen eines fairen Würfels“ modellieren kann, ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ mit der Idee, dass jede „Augenzahl“ $a \in \Omega$ Wahrscheinlichkeit $\frac{1}{6}$ hat. Die Wahrscheinlichkeit, 5 oder 6 zu würfeln, schreibt man dann als $\Pr(\{5, 6\}) = \frac{1}{3}$, die Wahrscheinlichkeit für eine gerade Augenzahl als $\Pr(\{2, 4, 6\}) = \frac{1}{2}$. Allgemein gilt jede Menge $A \subseteq \Omega$ als „Ereignis“ mit Wahrscheinlichkeit $\Pr(A) = |A|/|\Omega|$.

Wir fassen unsere Grundbegriffe etwas allgemeiner insofern, als wir auch verschiedene Wahrscheinlichkeiten für Elementarereignisse $a \in \Omega$ zulassen und es erlaubt ist, dass Ω abzählbar unendlich ist. Wir beschränken uns aber auf den Fall endlicher oder abzählbarer Wahrscheinlichkeitsräume, so genannter *diskreter* W-Räume. (Damit vermeiden wir die technisch etwas kompliziertere Diskussion von σ-Algebren. Wahrscheinlichkeitsrechnung wird im Modul „Stochastik für Informatiker“ genauer behandelt.)

Definition (Erinnerung)
Ein (*diskreter*) *Wahrscheinlichkeitsraum* ist ein Paar $(\Omega, \Pr)$, wobei

- Ω eine nichtleere endliche oder abzählbar unendliche Menge und
- $\Pr : \mathfrak{P}(\Omega) \to [0,1]$ eine Abbildung ($\mathfrak{P}(\Omega) = \{A \mid A \subseteq \Omega\}$ ist die Potenzmenge)

ist, so dass Folgendes gilt:

1. $\Pr(\Omega) = 1$
2. für alle $A \subseteq \Omega$ gilt $\Pr(\overline{A}) = 1 - \Pr(A)$, für $\overline{A} = \Omega \setminus A$
3. für alle $A_1, A_2, \ldots \in \mathfrak{P}(\Omega)$ gilt, falls die Mengen A_i paarweise disjunkt sind:[13]
   $\Pr(\bigcup_{i\geq 1} A_i) = \sum_{i\geq 1}^{\infty} \Pr(A_i)$. („$\sigma$-Additivität“)

Man nennt

- die Elemente von Ω *Ergebnisse* oder *Elementarereignisse*,
- die Elemente von $\mathfrak{P}(\Omega)$ (also die Teilmengen von Ω) *Ereignisse* und
- $\Pr$ die *Wahrscheinlichkeitsverteilung*

des Wahrscheinlichkeitsraums $(\Omega, \Pr)$. Für $A \in \mathfrak{P}(\Omega)$ heißt $\Pr(A)$ die *Wahrscheinlichkeit* von A.

Bemerkung 1.13 $\Pr(A) = \sum_{a\in A} \Pr(\{a\})$, d. h., die Wahrscheinlichkeitsverteilung $\Pr$ ist durch die *Wahrscheinlichkeitsfunktion* $\Omega \to [0,1]$, $a \mapsto p_a = \Pr(\{a\})$, eindeutig gegeben.[14]

Wir schreiben auch für diese Funktion $\Pr$ und damit $\Pr(a)$ anstelle von $\Pr(\{a\})$. Es gilt dann:

$$\Pr(A) = \sum_{a\in A} \Pr(a), \text{ für jedes Ereignis } A,$$

und insbesondere $\sum_{a\in\Omega} \Pr(a) = 1$.

Sei nun Ω sogar endlich. Dann ist die *uniforme Verteilung* (oder *Gleichverteilung*) die Wahrscheinlichkeitsverteilung $A \mapsto \frac{|A|}{|\Omega|}$, für Ereignisse $A \in \mathfrak{P}(\Omega)$, mit der Wahrscheinlichkeitsfunktion $a \mapsto \frac{1}{|\Omega|}$, für $a \in \Omega$.

[13] Mit 1., 2., 3. zusammen erhält man leicht, dass die Summationsformel auch für endliche Summen gilt („endliche Additivität“).

[14] Die Mathematik zeigt, dass Summen wie $\sum_{a\in A} \Pr(\{a\})$ sinnvoll sind, auch wenn keine Summationsreihenfolge festgelegt ist. Das liegt daran, dass die Summanden nichtnegativ sind und $\Pr(\Omega) = 1$ gilt.

Lemma 1.14 *Sei* $(\Omega, \Pr)$ *ein Wahrscheinlichkeitsraum und seien* $A, B \subseteq \Omega$ *Ereignisse. Dann gilt* $\Pr(A \setminus B) \geq \Pr(A) - \Pr(B)$.

Beweis: Für Ereignisse $C \subseteq D$ gilt stets $\Pr(C) = \sum_{a \in C} \Pr(a) \leq \sum_{a \in D} \Pr(a) = \Pr(D)$. Daher gilt $\Pr(A \setminus B) + \Pr(B) = \Pr((A \setminus B) \cup B) = \Pr(A \cup B) \geq \Pr(A)$. □

Sei $(\Omega, \Pr)$ Wahrscheinlichkeitsraum, B Ereignis mit $\Pr(B) > 0$. Definiere

$$\Pr_B \colon \mathfrak{P}(\Omega) \to [0,1], A \mapsto \frac{\Pr(A \cap B)}{\Pr(B)}\,.$$

Dann ist $(\Omega, \Pr_B)$ selbst ein Wahrscheinlichkeitsraum, wie man leicht nachrechnet. Intuitiv ist $\Pr_B(A)$ die Wahrscheinlichkeit für das Eintreten von A, wenn schon bekannt ist, dass B eingetreten ist. Daher nennt man $\Pr_B$ die *bedingte Wahrscheinlichkeit* bzgl. B und schreibt für $\Pr_B(A)$ auch $\Pr(A \mid B)$. Aus der Definition folgt die Grundformel

$$\Pr(A \cap B) = \Pr(A \mid B) \cdot \Pr(B).$$

Achtung: die bedingte Wahrscheinlichkeit $\Pr(A \mid B)$ ist nur definiert, wenn $\Pr(B) > 0$ gilt.

Lemma 1.15 *Sei* $(\Omega, \Pr)$ *ein Wahrscheinlichkeitsraum.*

(a) *(„Formel von der totalen Wahrscheinlichkeit“.)*
Seien $B_1, \ldots, B_t$ *disjunkte Ereignisse mit* $\Pr(B_1 \cup \cdots \cup B_t) = 1$.
Dann gilt

$$\Pr(A) = \sum_{1 \leq s \leq t} \Pr(A \mid B_s)\Pr(B_s).$$

(b) *Seien A, B, C Ereignisse mit* $\Pr(B \cap C), \Pr(C \setminus B) > 0$. *Dann gilt*

$$\begin{aligned}\Pr(A \mid C) &= \Pr(A \cap B \mid C) + \Pr(A \setminus B \mid C)\\ &= \Pr(A \mid B \cap C)\Pr(B \mid C) + \Pr(A \mid C \setminus B)\Pr(\overline{B} \mid C)\,.\end{aligned}$$

Beweis: (a) Wir beweisen dies für $t = 2$. (Die allgemeine Aussage erhält man dann durch vollständige Induktion.) Sei $N = \overline{B_1 \cup B_2}$. Dann ist $\Pr(N) = 0$. Wegen der Additivität gilt:

$$\Pr(A) = \Pr(A \cap B_1) + \Pr(A \cap B_2) + \underbrace{\Pr(A \cap N)}_{=0} = \Pr(A \mid B_1)\Pr(B_1) + \Pr(A \mid B_2)\Pr(B_2).$$

(b) Beachte, dass $C \setminus B = \overline{B} \cap C$ gilt. Weil $\Pr(C) \geq \Pr(B \cap C) > 0$ und $\Pr(C \setminus B) > 0$, gibt es die bedingten Wahrscheinlichkeiten $\Pr(\,\cdot \mid C)$, $\Pr(\,\cdot \mid B \cap C)$ und $\Pr(\,\cdot \mid \overline{B} \cap C)$. Mit der Additivität erhalten wir

$$\begin{aligned}\Pr(A \cap C) &= \Pr(A \cap B \cap C) + \Pr(A \cap \overline{B} \cap C)\\ &= \Pr(A \mid B \cap C)\Pr(B \cap C) + \Pr(A \mid \overline{B} \cap C)\Pr(\overline{B} \cap C).\end{aligned}$$

Division durch $\Pr(C)$ liefert die behaupteten Gleichungen. □

Beispiel: In unserem Würfel-Wahrscheinlichkeitsraum mit $\Omega = \{1, \ldots, 6\}$ und der uniformen Verteilung betrachten wir die Ereignisse $A = \{3, 6\}$ (durch 3 teilbare Augenzahl) und $B = \{2, 4, 6\}$ (gerade Augenzahl). Wir haben:

$$\Pr(A \cap B) = \Pr(\{6\}) = \frac{1}{6} = \frac{1}{3} \cdot \frac{1}{2} = \Pr(A) \cdot \Pr(B).$$

Damit sind die Ereignisse {Augenzahl ist gerade} und {Augenzahl ist durch 3 teilbar} (*stochastisch*) *unabhängig* im folgenden Sinn:

Definition 1.16 *Sei* $(\Omega, \Pr)$ *ein Wahrscheinlichkeitsraum und seien A, B Ereignisse. Dann heißen A und B* unabhängig, *wenn* $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$ *gilt.*

Bemerkung: Wenn $\Pr(B) > 0$ gilt, dann sind A und B genau dann unabhängig, wenn $\Pr(A) = \frac{\Pr(A \cap B)}{\Pr(B)} = \Pr(A \mid B)$ gilt. Das bedeutet, dass sich durch die Information, dass B eingetreten ist, nichts an der Wahrscheinlichkeit für A ändert. (Im Beispiel: Wenn wir wissen, dass die Augenzahl beim Würfeln gerade ist, dann ist die Wahrscheinlichkeit für eine Augenzahl, die durch 3 teilbar ist, genau $\frac{1}{3}$, genau dieselbe wie im gesamten Wahrscheinlichkeitsraum.)

Zufallsvariable bzw. Zufallsgrößen Zufallsvariable ordnen den Ergebnissen eines Experiments (d. h. eines Wahrscheinlichkeitsraums) „Werte“ aus einer Menge R zu. (Diese Werte können Zahlen oder andere „Eigenschaften“ sein.)

Definition 1.17 *Sei* $(\Omega, \Pr)$ *ein Wahrscheinlichkeitsraum und R eine endliche oder abzählbare Menge. Eine* Zufallsvariable *ist eine Abbildung*[15] $X \colon \Omega \to R$.

Zufallsvariable mit $R \subseteq \mathbb{R}$ *heißen* reelle Zufallsvariable.

Beispiel 1.18 Zu $\Omega = \{1, 2, \ldots, N\}^q$ $(q, N \geq 1)$ betrachten wir den Wahrscheinlichkeitsraum $(\Omega, \Pr)$ mit der Gleichverteilung Pr. Beispiele für Zufallsvariablen sind:

- $R = \mathbb{N}$ und $X \colon \Omega \to R, (a_1, \ldots, a_q) \mapsto a_5$ (eine *Projektion*, definiert für $q \geq 5$)
- $R = \{-1, 0, 1\}$ und $Y_{ij}((a_1, \ldots, a_q)) = \begin{cases} -1 & \text{falls } a_i < a_j \\ 0 & \text{falls } a_i = a_j \\ 1 & \text{falls } a_i > a_j \end{cases}$, für $1 \leq i < j \leq n$

[15] Die Bezeichnungen X, X_1, X_2, …, Y, usw. für Zufallsvariablen haben sich so eingebürgert, dass man bei dieser Bezeichnung bleiben sollte. Aber Vorsicht: Nicht mit Klartextmenge oder Chiffretextmenge verwechseln!

- $R = \mathbb{N}$ und $Z\colon \Omega \to R, (a_1, \ldots, a_q) \mapsto \sum_{1 \leq i \leq q} a_i$

Sei $X\colon \Omega \to R$ eine Zufallsvariable. Für $S \subseteq R$ setze

$$\Pr^X(S) := \Pr(X^{-1}(S)) = \Pr(\{a \in \Omega \mid X(a) \in S\}).$$

Dann ist $(R, \Pr^X)$ ein Wahrscheinlichkeitsraum (nachrechnen!). Dieser heißt *der von X auf R induzierte* Wahrscheinlichkeitsraum. $\Pr^X$ heißt auch die *Verteilung von X*.

Schreibweisen: Für $S \subseteq R$ ist

$$X^{-1}(S) = \{a \in \Omega \mid X(a) \in S\}$$

ein Ereignis, für das wir „$X \in S$“ oder „$\{X \in S\}$“ schreiben. Für $X^{-1}(r) = \{a \in \Omega \mid X(a) = r\}$ schreiben wir analog „$X = r$“ oder „$\{X = r\}$“. Insbesondere schreiben wir:

$$\Pr(X = r) = P^X(r) = \Pr(X^{-1}(r)) \text{ und } \Pr(X \in S) = P^X(S) = \Pr(X^{-1}(S)).$$

Sind $X_i\colon \Omega \to R_i$ Zufallsvariable und $S_i \subseteq R_i$, für $i = 1, 2$, dann schreiben wir „$\{X_1 \in S_1, X_2 \in S_2\}$“ für das Ereignis $X^{-1}(S_1) \cap X^{-1}(S_2)$. Die beiden Zufallsvariablen X_1 und X_2 heißen *unabhängig*, wenn $\Pr(X_1 \in S_1, X_2 \in S_2) = \Pr(X_1 \in S_1) \cdot \Pr(X_2 \in S_2)$ gilt, für alle $S_i \subseteq R_i$, $i = 1, 2$. Dies ist gleichbedeutend mit der Forderung

$$\Pr(X_1 = r_1, X_2 = r_2) = \Pr(X_1 = r_1) \cdot \Pr(X_2 = r_2) \text{ für alle } r_i \in R_i, i = 1, 2.$$

(Der Beweis der letzten Aussage besteht in recht langweiligem Nachrechnen.)

1.3 Informationstheoretische Sicherheit

Man erinnere sich an Beispiel 1.12. Eine naheliegende Annahme[16] ist, dass jeder Klartextbuchstabe mit Wahrscheinlichkeit $\frac{1}{2}$ und jeder Schlüssel mit Wahrscheinlichkeit $\frac{1}{3}$ auftritt, und zwar unabhängig voneinander. Dann ist $\Pr(\text{Klartext } x \text{ ist } \mathtt{a} \wedge \text{Chiffretext } y \text{ ist } \mathtt{A}) = \frac{1}{3}$, $\Pr(\text{Chiffretext } y \text{ ist } \mathtt{A}) = \frac{1}{2}$, also $\Pr(\text{Klartext } x \text{ ist } \mathtt{a} \mid \text{Chiffretext } y \text{ ist } \mathtt{A}) = \frac{2}{3} \neq \frac{1}{2} = \Pr(\text{Klartext } x \text{ ist } \mathtt{a})$. Wenn Eva also `A` beobachtet, ändert sich ihre Ansicht über die Verteilung auf den Klartextbuchstaben.

Für das Konzept der informationstheoretischen Sicherheit nehmen wir an, dass Klartexte mit bestimmten Wahrscheinlichkeiten auftreten. Was diese Wahrscheinlichkeiten sind, kann der Anwender normalerweise nicht kontrollieren. Die konsequente Anwendung des Kerckhoffs-Prinzips besagt aber, dass man annehmen muss, dass Eva die relevante Wahrscheinlichkeitsverteilung auf X kennt. (Zum Beispiel würde sie wissen, dass $\Pr_X(x_0) = \frac{1}{2}$ ist, für ein bestimmtes $x_0 \in X$.) Nun betrachten wir ein Kryptosystem $\mathcal{S} = (X, K, Y, e, d)$. Wir nehmen an, dass Alice und Bob ihren gemeinsamen Schlüssel k durch ein Zufallsexperiment wählen. Hierzu gehört ein zweiter Wahrscheinlichkeitsraum $(K, \Pr_K)$. Es ist sinnvoll anzunehmen, dass $\Pr_X$ und $\Pr_K$ nichts miteinander zu tun haben. Es wird verschlüsselt und Chiffretext y wird gesendet. Dieser wird von Eva beobachtet. Wenn sich dadurch die Meinung von Eva über die Wahrscheinlichkeiten der verschiedenen Klartexte von der ursprünglichen Verteilung unterscheidet (etwa jetzt: „mit 90%iger Wahrscheinlichkeit ist es Klartext x_0“), hat Eva aus der Beobachtung von y eine gewisse *Information erhalten.*

Wir geben nun ein mathematisches Modell an, innerhalb dessen man über Begriffe wie „Eva erhält Information“ sprechen und argumentieren kann. Dazu konstruieren wir einen W-Raum mit $\Omega = X \times K$. In das Modell bauen wir die Vorstellung ein, dass $x \in X$ und $k \in K$ nach den Verteilungen $\Pr_X$ und $\Pr_K$ zufällig und unabhängig gewählt werden.

Man beachte, dass die Verteilung $\Pr_K$ „Teil des Kryptosystems“ ist, also der Kontrolle von Alice and Bob unterliegt, während $\Pr_X$ „Teil der Anwendung“ oder „Teil der Realität“ ist, also von den Teilnehmern normalerweise nicht beeinflusst werden kann. Die Verteilung $\Pr_X$ braucht beim Entwurf des Kryptosystems nicht einmal bekannt zu sein. (Alice und Bob sollten ihr Kryptosystem ohne Kenntnis von $\Pr_X$ planen können. Die Annahme, dass Eva $\Pr_X$ kennt, ist eine worst-case-Annahme, sie muss in der Realität nicht unbedingt erfüllt sein.)

[16] Dies ist eine Anwendung des Laplaceschen *Indifferenzprinzips*: Wenn über die Wahrscheinlichkeitsverteilung auf einer endlichen Menge überhaupt keine Information vorliegt, dann sollte man annehmen, dass jedes Elementarereignis dieselbe Wahrscheinlichkeit hat. Siehe `https://de.wikipedia.org/wiki/Indifferenzprinzip`.

Definition 1.19 *Ein* Kryptosystem mit Schlüsselverteilung (KSV) *ist ein 6-Tupel* $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$*, wobei*

- $\mathcal{S} = (X, K, Y, e, d)$ *ein Kryptosystem* (*das* zugrundeliegende Kryptosystem) *ist und*
- $\Pr_K \colon K \to (0, 1]$ *eine Wahrscheinlichkeitsfunktion* (*die* Schlüsselverteilung) *ist.*

Für $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ *schreiben wir auch* $\mathcal{S}[\Pr_K]$.

Achtung: Die Definition verlangt $\Pr_K(k) \in (0, 1]$, also $\Pr_K(k) > 0$ für alle $k \in K$. (Hat man Schlüssel mit Wahrscheinlichkeit 0, kann man sie aus K einfach weglassen.)

Sei weiter $\Pr_X \colon X \to [0, 1]$ eine Wahrscheinlichkeitsfunktion auf der Menge der Klartexte. Das heißt: $\sum_{x \in X} \Pr_X(x) = 1$. Diese Wahrscheinlichkeitsfunktion definiert natürlich eine W-Verteilung auf X, die wir wieder $\Pr_X$ nennen. (*Achtung*: Es kann Klartexte x mit $\Pr(x) = 0$ geben. Solche Klartexte heißen *passiv*, die anderen, mit $\Pr_X(x) > 0$, *aktiv*.) Wir definieren die *gemeinsame Wahrscheinlichkeitsfunktion* $\Pr \colon X \times K \to [0, 1]$ durch

$$\Pr((x, k)) := \Pr_X(x) \cdot \Pr_K(k).$$

Dies definiert einen Wahrscheinlichkeitsraum auf $X \times K$, für den

$$\Pr(X' \times K') = \Pr_X(X') \cdot \Pr_K(K'), \text{ für alle } X' \subseteq X, K' \subseteq K$$

gilt. (Dies zeigt man durch eine einfache Rechnung.) Durch diese Definition wird die Annahme modelliert, dass der Schlüssel k *unabhängig vom Klartext* durch ein von $\Pr_K$ gesteuertes Zufallsexperiment gewählt wird.[17]

Beispiel 1.20 Sei $X = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}$, $K = \{0, 1, 2, 3\}$, $Y = \{\mathtt{A}, \mathtt{B}, \mathtt{C}\}$ und die Verschlüsselungsfunktion sei durch die folgende Tabelle gegeben:

e	a (0,4)	b (0)	c (0,6)
0 ($\frac{1}{4}$)	A	B	C
1 ($\frac{1}{8}$)	B	C	A
2 ($\frac{1}{2}$)	C	A	B
3 ($\frac{1}{8}$)	C	B	A

Die Wahrscheinlichkeiten $\Pr_X(x)$ sind bei den Klartexten, die Wahrscheinlichkeiten $\Pr_K(k)$ bei den Schlüsseln in Klammern notiert. Klartexte a und c sind aktiv, Klartext b ist passiv. Die Wahrscheinlichkeit für einen Punkt $(x, k) \in X \times K$ erhält man durch Multiplikation: $\Pr((\mathtt{c}, 2)) = 0{,}6 \cdot \frac{1}{2} = 0{,}3$ und $\Pr((\mathtt{b}, k)) = 0 \cdot \Pr_K(k) = 0$ für alle $k \in K$.

[17] Es handelt sich hier um den Spezialfall einer allgemeinen Konstruktion, die in der Wahrscheinlichkeitsrechnung unter dem Namen *Produktraum* behandelt wird. Häufige Bezeichnung hierfür: $(X \times K, \Pr_X \otimes \Pr_K)$.

Der Chiffretext y ist dann eine Zufallsvariable auf diesem Wahrscheinlichkeitsraum:

$$X_3((x,k)) := e(x,k).$$

Auch die beiden Komponenten x und k werden als Zufallsvariable betrachtet (Projektionen):

$$X_1 \colon X \times K \to X, (x,k) \mapsto x$$
$$X_2 \colon X \times K \to K, (x,k) \mapsto k.$$

Wir beobachten einige einfache Zusammenhänge, für $x_0 \in X$, $k_0 \in K$, $y_0 \in Y$:

$$\Pr(x_0) := \Pr(X_1 = x_0) = \Pr(\{x_0\} \times K) = \Pr_X(x_0) \cdot \Pr_K(K) = \Pr_X(x_0).$$
$$\Pr(k_0) := \Pr(X_2 = k_0) = \Pr(X \times \{k_0\}) = \Pr_X(X) \cdot \Pr_K(k_0) = \Pr_K(k_0).$$

(Man erhält also die ursprünglichen Wahrscheinlichkeiten für Klartexte und Schlüssel zurück. Dies ist eine einfache Grundeigenschaft von Produkträumen.)

$$\begin{aligned}\Pr(y_0) := \Pr(X_3 = y_0) &= \Pr(\{(x,k) \mid x \in X, k \in K, e(x,k) = y_0\}) \\ &= \sum_{x \in X, k \in K, e(x,k) = y_0} \Pr((x,k)) \\ &= \sum_{x \in X, k \in K, e(x,k) = y_0} \Pr_X(x) \cdot \Pr_K(k).\end{aligned}$$

(In Beispiel 1.20 gilt $\Pr(\mathtt{A}) = \frac{1}{4} \cdot 0{,}4 + \frac{1}{8} \cdot 0{,}6 + \frac{1}{2} \cdot 0 + \frac{1}{8} \cdot 0{,}6 = 0{,}25$ und $\Pr(\mathtt{B}) = \frac{1}{4} \cdot 0 + \frac{1}{8} \cdot 0{,}4 + \frac{1}{2} \cdot 0{,}6 + \frac{1}{8} \cdot 0 = 0{,}35$.)

$$\begin{aligned}\Pr(x_0, y_0) := \Pr(X_1 = x_0, X_3 = y_0) &= \Pr(\{x_0\} \times \{k \in K \mid e(x_0,k) = y_0\}) \\ &= \Pr_X(x_0) \cdot \sum_{k \in K \colon e(x_0,k) = y_0} \Pr_K(k).\end{aligned}$$

(In Beispiel 1.20 gilt $\Pr(\mathtt{c}, \mathtt{A}) = 0{,}6 \cdot (\frac{1}{8} + \frac{1}{8}) = 0{,}15$ und $\Pr(\mathtt{a}, \mathtt{C}) = 0{,}4 \cdot (\frac{1}{2} + \frac{1}{8}) = 0{,}25$.)

$$\begin{aligned}\Pr(x_0 \mid y_0) := \Pr(X_1 = x_0 \mid X_3 = y_0) &= \frac{\Pr(x_0, y_0)}{\Pr(y_0)} \\ &= \Pr_X(x_0) \cdot \frac{\sum_{k \in K \colon e(x_0,k) = y_0} \Pr_K(k)}{\sum_{x \in X, k \in K \colon e(x,k) = y_0} \Pr_X(x) \cdot \Pr_K(k)}.\end{aligned}$$

(In Beispiel 1.20 gilt $\Pr(\mathtt{c} \mid \mathtt{A}) = 0{,}15/0{,}25 = 0{,}6$.)

Die letzte Formel ist nur für y_0 mit $\Pr(y_0) > 0$ definiert.

Definition 1.21 *Sei $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ ein Kryptosystem mit Schlüsselverteilung.*

(a) *Sei $\Pr_X$ eine Wahrscheinlichkeitsfunktion auf den Klartexten. Dann heißt $\mathcal{V}$* informationstheoretisch sicher bezüglich $\Pr_X$*, wenn für alle $x \in X$, $y \in Y$ mit $\Pr(y) > 0$ gilt:*

$$\Pr(x) = \Pr(x \mid y)\,.$$

(b) *Das KSV $\mathcal{V}$ heißt* informationstheoretisch sicher*, wenn es bezüglich jeder beliebigen Klartextverteilung $\Pr_X$ informationstheoretisch sicher ist.*

Bemerkungen: Hinter Definition (a) steckt die folgende Vorstellung: Eva kennt (im schlimmsten Fall) die Wahrscheinlichkeitsfunktion $\Pr_X$. Das System gilt als sicher, wenn sich durch Abfangen eines Chiffretextes y aus Evas Sicht die Wahrscheinlichkeiten der einzelnen Klartexte x nicht ändern. Die Bedingung $\Pr(y) > 0$ in (a) ist nötig, damit $\Pr(x \mid y)$ definiert ist. Sie bedeutet aber keine Einschränkung, da Chiffretexte y mit $\Pr(y) = 0$ nie vorkommen, also auch nicht abgefangen werden können. Das Konzept in (b) ist relevant, weil man beim Entwurf eines Kryptosystems meistens die Klartextverteilung nicht oder nicht genau kennt.

Man beachte, dass in der Definition der informationstheoretischen Sicherheit die Fähigkeiten von Eva überhaupt nicht eingeschränkt werden. Auf welche Weise sie eventuell ermittelt, dass sich Wahrscheinlichkeiten geändert haben, wird gar nicht diskutiert. (Eva könnte zum Beispiel für jedes $y \in Y$ eine Tabelle haben, in der die Wahrscheinlichkeiten $\Pr(x \mid y)$ für alle $x \in X$ aufgelistet sind. Oder sie fängt beim Vorliegen von y an, eine solche Tabelle zu berechnen. Beides ist natürlich für nicht ganz kleine X und Y völlig unrealistisch.)

Beispiel 1.22

e	a $(\frac{1}{4})$	b $(\frac{3}{4})$
k_0 $(\frac{1}{3})$	A	B
k_1 $(\frac{2}{3})$	B	A

(Notation: In der Tabelle stehen neben den Namen von Klartexten und Schlüsseln in Klammern deren Wahrscheinlichkeiten.) Dieses Kryptosystem ist possibilistisch sicher. Es gilt aber:

$$\Pr(\mathtt{a} \mid \mathtt{A}) = \frac{\Pr(\mathtt{a}, \mathtt{A})}{\Pr(\mathtt{A})} = \frac{\frac{1}{3} \cdot \frac{1}{4}}{\frac{1}{3} \cdot \frac{1}{4} + \frac{2}{3} \cdot \frac{3}{4}} = \frac{1}{7}$$

und

$$\Pr(\mathtt{a}) = \frac{1}{4}$$

Nach dem Abhören von A sieht also Eva den Klartext a als weniger wahrscheinlich an als vorher. Also ist dieses Kryptosystem mit Schlüsselverteilung bzgl. $\Pr_X$ nicht informationstheoretisch sicher.

Beispiel 1.23

e	$\mathtt{a}$ $(\frac{1}{4})$	$\mathtt{b}$ $(\frac{3}{4})$
k_0 $(\frac{1}{2})$	$\mathtt{A}$	$\mathtt{B}$
k_1 $(\frac{1}{2})$	$\mathtt{B}$	$\mathtt{A}$

Dieses System ist bezüglich $\Pr_X$ informationstheoretisch sicher. Zum Beispiel gilt

$$\Pr(\mathtt{a} \mid \mathtt{A}) = \frac{\Pr(\mathtt{a}, \mathtt{A})}{\Pr(\mathtt{A})} = \frac{\frac{1}{4} \cdot \frac{1}{2}}{\frac{1}{4} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{2}} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}.$$

und

$$\Pr(\mathtt{a}) = \frac{1}{4}.$$

Die anderen drei verlangten Gleichheiten rechnet man analog nach.

Satz 1.24 (Informationstheoretische Sicherheit des Vernam-Systems) *Sei $\ell > 0$ und $\mathcal{S} = (X, K, Y, e, d)$ mit $X = K = Y = \{0,1\}^\ell$ und $e = d = \oplus_\ell$ das Vernam-System der Länge ℓ. Sei weiter $\Pr_K \colon K \to [0,1]$ die Gleichverteilung. Dann ist $\mathcal{V} = \mathcal{S}[\Pr_K]$ informationstheoretisch sicher.*

Beweis: Sei $\Pr_X \colon X \to [0,1]$ eine beliebige Wahrscheinlichkeitsfunktion. Wir müssen zeigen, dass $\mathcal{V}$ bezüglich $\Pr_X$ informationstheoretisch sicher ist. Wir beginnen mit folgender Beobachtung: Zu $x \in X$ und $y \in Y$ existiert genau ein $k_{x,y} \in K$ mit $e(x, k_{x,y}) = y$, nämlich $k_{x,y} = x \oplus_\ell y$. Damit gilt für jedes $y \in Y$:

$$\Pr(y) = \sum_{\substack{x \in X, k \in K \\ e(x,k)=y}} \Pr(x)\Pr(k) = \sum_{x \in X} \Pr(x) \underbrace{\Pr(k_{x,y})}_{=2^{-\ell}} = 2^{-\ell} \cdot \underbrace{\sum_{x \in X} \Pr(x)}_{=1} = 2^{-\ell}.$$

(D. h.: Jeder Chiffretext y hat dieselbe Wahrscheinlichkeit $2^{-\ell}$, ganz gleich was $\Pr_X$ ist.)

Sei nun $x \in X$ und $y \in Y$ beliebig gewählt. Dann gilt

$$\Pr(x, y) = \Pr(x) \cdot \sum_{\substack{k \in K \\ e(x,k)=y}} \Pr(k) = \Pr(x) \cdot \Pr(k_{x,y}) = \Pr(x) \cdot 2^{-\ell} = \Pr(x) \cdot \Pr(y).$$

Damit folgt

$$\Pr(x) = \frac{\Pr(x, y)}{\Pr(y)} = \Pr(x \mid y),$$

wie bei der informationstheoretischen Sicherheit verlangt. □

Bemerkung 1.25
1. Der Beweis und damit das Vernamsystem kommt mit jeder beliebigen Klartextverteilung zurecht.
2. Im KSV $\mathcal{V}$ wird die Gleichverteilung $\Pr_K$ auf den Schlüsseln benutzt.

Wir wollen nun überlegen, dass diese beiden Sachverhalte nicht zufällig sind. Es wird sich herausstellen, dass informationstheoretische Sicherheit in bestimmten Fällen (nämlich wenn y und K möglichst „sparsam“ gebaut sind) Gleichverteilung auf den Schlüsseln erzwingt, und dass die informationstheoretische Sicherheit eines KSV nichts mit den konkreten Wahrscheinlichkeiten der Klartextverteilung $\Pr_X$ zu tun hat, sondern nur die Menge $\{x \in X \mid \Pr_X(x) > 0\}$ der “aktiven” Klartexte relevant ist.

Lemma 1.26 *Sei $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ ein KSV. Sei $\mathcal{V}$ informationstheoretisch sicher bezüglich einer Klartextverteilung $\Pr_X$ mit $\Pr(x) > 0$ für alle $x \in X$. Dann gilt:*

(a) $\Pr(y) > 0$ *für alle $y \in Y$, und $\mathcal{S} = (X, K, Y, e, d)$ ist possibilistisch sicher.*

(b) *Gilt zusätzlich $|X| = |Y| = |K|$, so gilt $\Pr_K(k) = \frac{1}{|K|}$ für alle $k \in K$.*

Beweis: (a) Sei $y \in Y$ beliebig. Nach Definition 1.1(2) gibt es $x_0 \in X$ und $k_0 \in K$ mit $e(x_0, k_0) = y$. Da $\Pr_X(x_0) > 0$ (nach Vor.) und $\Pr_K(k_0) > 0$ (nach Def. 1.19), erhalten wir $\Pr(y) \geq \Pr_X(x_0)\Pr_K(k_0) > 0$. Sei nun zusätzlich auch $x \in X$ beliebig. Dann gilt:

$$\sum_{k \in K\,:\, e(x,k)=y} \Pr(x)\Pr(k) = \Pr(x, y) = \Pr(x \mid y)\Pr(y) \overset{(*)}{=} \Pr(x)\Pr(y) > 0.$$

($(*)$ gilt, da $\mathcal{V}$ informationstheoretisch sicher bzgl. $\Pr_X$ ist.) Also existiert $k \in K$ mit $e(x, k) = y$. Da x und y beliebig waren, ist $\mathcal{S}$ possibilistisch sicher.

(b) Nun nehmen wir zusätzlich $|X| = |Y| = |K|$ an. Wir beobachten zuerst zwei Dinge:

(i) Für jedes $x \in X$ ist die Abbildung $K \ni k \mapsto e(x, k) \in Y$ bijektiv.
(Dass diese Abbildung surjektiv ist, ist eine Umformulierung der possibilistischen Sicherheit, die nach (a) gegeben ist. Aus Surjektivität folgt Bijektivität, wegen $|K| = |Y|$.)

(ii) Für jedes $k \in K$ ist die Abbildung $X \ni x \mapsto e(x, k) \in Y$ bijektiv.
(Dass die Abbildung injektiv ist, folgt aus der Dechiffrierbedingung. Aus Injektivität folgt Bijektivität, wegen $|X| = |Y|$.)

Nun seien $k_1, k_2 \in K$ beliebig. Unser Ziel ist zu zeigen, dass $\Pr(k_1) = \Pr(k_2)$ gilt. (Dann ist gezeigt, dass $\Pr_K$ die uniforme Verteilung ist.) Wähle $x \in X$ beliebig und setze $y := e(x, k_1)$. Beachte, dass es wegen (i) keinen Schlüssel $k \neq k_1$ mit $y = e(x, k)$ gibt. Wegen (ii) gibt es ein $x' \in X$ mit $e(x', k_2) = y$. Auch hier gibt es kein $k' \neq k_2$ mit $e(x', k') = y$.

Es gilt also:

$$\Pr(x)\Pr(k_1) = \sum_{k \in K:\, e(x,k)=y} \Pr(x)\Pr(k) = \Pr(x, y) = \Pr(x \mid y)\Pr(y) \overset{(*)}{=} \Pr(x)\Pr(y),$$

und daher $\Pr(k_1) = \Pr(y)$, wegen $\Pr(x) > 0$. ($(*)$ gilt, weil $\mathcal{V}$ informationstheoretisch sicher ist.) Analog gilt $\Pr(x')\Pr(k_2) = \Pr(x')\Pr(y)$, und daher $\Pr(k_2) = \Pr(y)$. Es folgt $\Pr(k_1) = \Pr(k_2)$, wie gewünscht. □

Teil (b) dieses Lemmas hat eine Umkehrung.

Lemma 1.27 *Sei $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ KSV mit $|X| = |Y| = |K|$. Wenn $\mathcal{S} = (X, K, Y, e, d)$ possibilistisch sicher ist und $\Pr_K$ die Gleichverteilung ist, dann ist $\mathcal{V}$ informationstheoretisch sicher.*

Beweis: Es sei eine beliebige Klartextverteilung $\Pr_X$ gegeben. Da $\mathcal{S}$ possibilistisch sicher ist und $|X| = |Y| = |K|$ gilt, existiert für jedes Paar $(x, y) \in X \times Y$ *genau ein* $k_{x,y} \in K$ mit $e(x, k_{x,y}) = y$ (vgl. Aussage (i) im Beweis des vorherigen Lemmas).

Damit gilt für jedes $y \in Y$:

$$\Pr(y) = \sum_{\substack{x \in X, k \in K \\ e(x,k)=y}} \Pr(x)\Pr(k) = \sum_{x \in X} \Pr(x) \underbrace{\Pr(k_{x,y})}_{=1/|K|} = \frac{1}{|K|} \cdot \sum_{x \in X} \Pr(x) = \frac{1}{|K|}.$$

Wir haben benutzt, dass $\Pr_K$ die uniforme Verteilung ist und dass $\sum_{x \in X} \Pr(x) = 1$ gilt.

Seien nun $x \in X$ und $y \in Y$ beliebig. Wenn $\Pr(x) = 0$ ist, gilt auf jeden Fall $\Pr(x \mid y) = 0 = \Pr(x)$. Wir können also $\Pr(x) > 0$ annehmen und rechnen:

$$\Pr(x \mid y) = \frac{\Pr(x, y)}{\Pr(y)} = \frac{\Pr(y \mid x) \cdot \Pr(x)}{\Pr(y)} = \frac{\Pr_K(k_{x,y}) \cdot \Pr(x)}{\Pr(y)} \overset{(*)}{=} \frac{\frac{1}{|K|} \cdot \Pr(x)}{\frac{1}{|K|}} = \Pr(x).$$

(Für $(*)$ benutzen wir die Annahme über $\Pr_K$ und die Gleichheit $\Pr(y) = 1/|K|$ von oben.) Das heißt, dass $\mathcal{V}$ für $\Pr_X$ informationstheoretisch sicher ist. □

Aus den beiden Lemmas erhalten wir den folgenden Satz, der die informationstheoretisch sicheren KSVs für den Fall $|X| = |Y| = |K|$ vollständig beschreibt.

Satz 1.28 *Sei* $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ *ein KSV mit* $|X| = |Y| = |K|$.
Dann sind äquivalent:

(a) $\mathcal{V}$ *ist* informationstheoretisch sicher.

(b) (X, K, Y, e, d) *ist possibilistisch sicher und* $\Pr_K(k) = \frac{1}{|K|}$ *für alle* $k \in K$.

Beweis: „(a) $\Rightarrow$ (b)“: Wenn $\mathcal{V}$ informationstheoretisch sicher ist, dann auch bezüglich einer Klartextverteilung $\Pr_X$, in der alle Klartexte aktiv sind. Lemma 1.26 liefert (b).
„(b) $\Rightarrow$ (a): Lemma 1.27. □

Der Satz besagt, dass man informationstheoretisch sichere Systeme mit $|X| = |Y| = |K|$ daran erkennt, dass in der Verschlüsselungstabelle (für e) in jeder Spalte alle Chiffretexte vorkommen (possibilistische Sicherheit) und dass die Schlüsselverteilung $\Pr_K$ uniform ist. Auch in jeder *Zeile* kommen natürlich alle Chiffretexte vor: das liegt aber einfach an der Dechiffrierbedingung. Wir geben ein Beispiel für ein solches informationstheoretisch sicheres Kryptosystem mit $|X| = |Y| = |K| = 6$ an. Die Klartextverteilung ist irrelevant. (Die Verschlüsselungsfunktion ist übrigens mit Hilfe der Multiplikationstabelle der multiplikativen Gruppe $\mathbb{Z}_7^*$ des Körpers $\mathbb{Z}_7$ konstruiert worden. Solche Tabellen haben die Eigenschaft, dass jeder mögliche Eintrag in jeder Zeile und in jeder Spalte genau einmal vorkommt.)

Beispiel 1.29 Wir betrachten $X = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{f}\}$, $K = \{k_0, \ldots, k_5\}$, $Y = \{\mathtt{A}, \ldots, \mathtt{F}\}$.

e	a	b	c	d	e	f
k_0 $(\frac{1}{6})$	A	B	C	D	E	F
k_1 $(\frac{1}{6})$	B	D	F	A	C	E
k_2 $(\frac{1}{6})$	C	F	B	E	A	D
k_3 $(\frac{1}{6})$	D	A	E	B	F	C
k_4 $(\frac{1}{6})$	E	C	A	F	D	B
k_5 $(\frac{1}{6})$	F	E	D	C	B	A

Nun betrachten wir allgemeinere Situationen, und fragen auch nach informationstheoretischer Sicherheit für spezifische Klartextverteilungen $\Pr_X$ und für Mengen K und Y, die größer als X sind. Die Bedingung „uniforme Verteilung auf den Schlüsseln“ verschwindet dann komplett! Wir erinnern uns: Klartexte x mit $\Pr_X(x) > 0$ heißen *aktiv* (bzgl. $\Pr_X$), die anderen *passiv*. Es wird sich herausstellen, dass sich informationstheoretische Sicherheit für $\Pr_X$ mit dem Verhalten von $e(x, k)$ auf den aktiven Klartexten entscheidet, wobei es auf die tatsächlichen Wahrscheinlichkeiten für die aktiven Klartexte nicht ankommt.

Technisch hilfreich sind die folgenden Größen, die nur von der Verschlüsselungsfunktion und der Schlüsselverteilung abhängen (nicht von irgendeiner Klartextverteilung):

$$P^x(y) := \sum_{\substack{k\in K\\ e(x,k)=y}} \Pr(k), \text{ für } x \in X, y \in Y. \tag{1.1}$$

Man beobachtet sofort die folgenden Gleichungen, die aus der Unabhängigkeit der Verteilungen $\Pr_X$ und $\Pr_K$ folgen:

$$\text{Für alle } x \in X\text{:} \quad \Pr(x,y) = \Pr(x) \cdot P^x(y). \tag{1.2}$$

$$\text{Wenn } \Pr(x) > 0\text{:} \quad \Pr(y \mid x) = \frac{\Pr(x,y)}{\Pr(x)} = P^x(y). \tag{1.3}$$

Umgekehrt wie bei der Definition der informationstheoretischen Sicherheit stellt man sich hier vor, dass ein Klartext x gegeben ist und man fragt nach der resultierenden Verteilung auf den Chiffretexten.

Das nächste Lemma besagt, dass man die Wahrscheinlichkeiten aktiver Klartexte beliebig ändern kann (auch auf 0, also sie weglassen), ohne dass eine bestehende informationstheoretische Sicherheit zerstört wird.

Lemma 1.30 *Sei $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ KSV und seien $\Pr_X$ und $\Pr'_X$ Klartextverteilungen mit $\Pr'_X(x) > 0 \Rightarrow \Pr_X(x) > 0$. Dann gilt: Ist $\mathcal{V}$ informationstheoretisch sicher bzgl. $\Pr_X$, so ist $\mathcal{V}$ informationstheoretisch sicher bzgl. $\Pr'_X$.*

Beweis: Sei $\mathcal{V}$ informationstheoretisch sicher bzgl. $\Pr_X$. Wir haben es jetzt mit zwei Wahrscheinlichkeitsräumen zu tun, einem zu $\Pr_X$ und $\Pr_K$ (bezeichnet mit $(X \times K, \Pr)$) und einem zu $\Pr'_X$ und $\Pr_K$ (bezeichnet mit $(X \times K, \Pr')$).

Wir zeigen nacheinander vier Aussagen (i), (ii), (iii), (iv).

(i) $\Pr_X(x) > 0 \Rightarrow P^x(y) = \Pr(y \mid x) = \Pr(y)$ für alle $y \in Y$.

(Die Verteilungen $\Pr^x(\,\cdot\,) = \Pr(\,\cdot \mid x)$ auf den Chiffretexten sind für alle (Pr-)aktiven Klartexte x gleich und sind auch gleich der globalen Verteilung auf den Chiffretexten.)

Beweis hierzu: Sei $\Pr(x) > 0$. Dann gilt $P^x(y) = \Pr(y \mid x)$, siehe (1.3). Wenn $\Pr(y) = 0$ gilt, folgt auch $\Pr(y \mid x) = 0$. Sei also $\Pr(y) > 0$. Dann gilt:

$$\Pr(y \mid x) = \frac{\Pr(x,y)}{\Pr(x)} = \frac{\Pr(x \mid y)\Pr(y)}{\Pr(x)} \overset{(*)}{=} \frac{\Pr(x)\Pr(y)}{\Pr(x)} = \Pr(y).$$

($(*)$ gilt, weil $\mathcal{V}$ informationstheoretisch sicher bzgl. $\Pr_X$ ist.)

(ii) $\Pr'_X(x) > 0 \Rightarrow \Pr'(y \mid x) = \Pr(y)$ für alle $y \in Y$.

Beweis hierzu: Aus $\Pr'(x) > 0$ folgt $\Pr(x) > 0$, nach Voraussetzung. Wir wenden (1.3) für $\Pr'$ und für $\Pr$ an und erhalten für alle $y \in Y$:

$$\Pr'(y \mid x) = P^x(y) = \Pr(y \mid x) \overset{\text{(i)}}{=} \Pr(y).$$

(iii) $\Pr'(y) = \Pr(y)$ für alle $y \in Y$.

Beweis hierzu: Mit Lemma 1.15(a) (Formel von der totalen Wahrscheinlichkeit):

$$\Pr'(y) = \sum_{x \in X\colon \Pr'(x)>0} \Pr'(y \mid x)\Pr'(x) \overset{\text{(ii)}}{=} \sum_{x \in X\colon \Pr'(x)>0} \Pr(y)\Pr'(x) = \Pr(y).$$

(iv) $\Pr'(x) = \Pr'(x \mid y)$ für alle $x \in X$, $y \in Y$ mit $\Pr'(y) > 0$.
(D. h.: $\mathcal{V}$ ist bzgl. $\Pr'_X$ informationstheoretisch sicher.)

Beweis hierzu: Wenn $\Pr'(x) = 0$ gilt, dann folgt $\Pr'(x \mid y) = 0 = \Pr'(x)$. Sei nun $\Pr'(x) > 0$. Dann:

$$\Pr'(x \mid y) = \frac{\Pr'(x,y)}{\Pr'(y)} = \frac{\Pr'(y \mid x)\Pr'(x)}{\Pr'(y)} \overset{\text{(ii)},\text{(iii)}}{=} \frac{\Pr(y)\Pr'(x)}{\Pr(y)} = \Pr'(x).$$

□

Satz 1.31 *Sei $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ KSV und sei $\Pr_X$ eine Klartextverteilung. Dann sind äquivalent:*

(a) *$\mathcal{V}$ ist informationstheoretisch sicher für $\Pr_X$.*

(b) *Für jedes $x \in X$ und jedes $y \in Y$ gilt: $\Pr(x,y) = \Pr(x)\Pr(y)$ (das Eintreten von x und das Eintreten von y sind unabhängig).*

(c) *Für alle $x \in X$ mit $\Pr(x) > 0$ und alle $y \in Y$ gilt $\Pr(y) = \Pr(y \mid x)$ (andere Formulierung der Unabhängigkeit).*

(d) *Für alle $x, x' \in X$ mit $\Pr(x), \Pr(x') > 0$ und alle $y \in Y$ gilt $P^x(y) = P^{x'}(y)$.*

Bemerkung: Bedingung (a) fragt nach der Situation bei gegebenem Chiffretext y mit $\Pr(y) > 0$. Bedingung (b) ist die wahrscheinlichkeitstheoretisch klarste Charakterisierung von informationstheoretischer Sicherheit, ohne bedingte Wahrscheinlichkeiten zu verwenden. Bedingungen (c) und (d) machen deutlich, dass es nur auf das Verhalten des Kryptosystems (mit seiner Verteilung $\Pr_K$) auf den *aktiven* Klartexten ankommt, nicht auf die Klartextverteilung. Sie sagen auch, worauf genau es ankommt: Für jeden beliebigen aktiven Buchstaben ist die von $e(x, \cdot)$ und der Schlüsselverteilung erzeugte Verteilung auf den

Chiffretexten gleich, und zwar gleich der absoluten Verteilung auf den Chiffretexten. Informationstheoretische Sicherheit von $\mathcal{V}$ (also für alle Klartextverteilungen) heißt also, dass alle Funktionen $P^x\colon Y \to [0,1]$, für $x \in X$, gleich sind (weil man als $\Pr_X$ eine Verteilung wählen kann, bei der *alle* Klartexte aktiv sind, zum Beispiel die Gleichverteilung).

Beweis: „(a) ⇒ (b)“: Wenn $\Pr(y) = 0$, gilt $\Pr(x,y) = 0 = \Pr(x)\Pr(y)$. Sei jetzt $\Pr(y) > 0$. Dann gilt $\Pr(x,y) = \Pr(y)\Pr(x \mid y) = \Pr(y)\Pr(x)$, nach (a).

„(b) ⇒ (c)“: Wegen (b) gilt $\Pr(y)\Pr(x) = \Pr(x,y)$. Andererseits ist $\Pr(y \mid x)\Pr(x) = \Pr(x,y)$, also folgt (c) durch Kürzen mit $\Pr(x) > 0$.

„(c) ⇒ (d)“: Verwende (1.3) für x und x' und benutze (c).

„(d) ⇒ (a)“: (Dies ist natürlich der entscheidende und schwierigste Beweisschritt!) Nach Voraussetzung (d) gibt es für jedes $y \in Y$ ein p_y mit $P^x(y) = p_y$ für alle aktiven $x \in X$. Nach Lemma 1.15(a) (Formel von der totalen Wahrscheinlichkeit) gilt dann für jedes y:

$$\Pr(y) = \sum_{x\in X\colon \Pr(x)>0} \Pr(y \mid x)\cdot\Pr(x) = \sum_{x\in X\colon \Pr(x)>0} P^x(y)\cdot\Pr(x) = \sum_{x\in X\colon \Pr(x)>0} p_y\cdot\Pr(x) = p_y.$$

Sei nun $y \in Y$ mit $\Pr(y) > 0$, und sei $x \in X$. Wenn $\Pr(x) = 0$ gilt, folgt auch $\Pr(x \mid y) = 0$. Wenn x aktiv ist, dann dann gilt

$$\Pr(x \mid y) = \frac{\Pr(x,y)}{\Pr(y)} = \frac{\Pr(y \mid x)\Pr(x)}{p_y} = \frac{P^x(y)\Pr(x)}{p_y} = \Pr(x),$$

wie gewünscht. □

Beispiel 1.32 Wir geben noch ein Beispiel für ein informationstheoretisch sicheres Kryptosystem mit $|X| = 4$, $|Y| = 6$ und $|K| = 8$ an. Die Klartextverteilung ist irrelevant. Sei $X = \{\mathtt{a},\mathtt{b},\mathtt{c},\mathtt{d}\}$, $K = \{k_0,\dots,k_7\}$, $Y = \{\mathtt{A},\mathtt{B},\mathtt{C},\mathtt{D},\mathtt{E},\mathtt{F}\}$, und e durch die folgende Tabelle gegeben. (Sie entsteht durch Zusammensetzen zweier informationstheoretisch sicherer Kryptosysteme mit jeweils vier Schlüsseln und vier Chiffretexten.)

e	a	b	c	d
k_0 $(\frac{1}{6})$	A	B	C	D
k_1 $(\frac{1}{6})$	B	C	D	A
k_2 $(\frac{1}{6})$	C	D	A	B
k_3 $(\frac{1}{6})$	D	A	B	C
k_4 $(\frac{1}{12})$	A	B	E	F
k_5 $(\frac{1}{12})$	B	A	F	E
k_6 $(\frac{1}{12})$	E	F	A	B
k_7 $(\frac{1}{12})$	F	E	B	A

Offensichtlich ist die Schlüsselverteilung nicht uniform. Jeder Schlüssel k hat eine andere Chiffre $x \mapsto e(x,k)$. Die (absoluten) Wahrscheinlichkeiten für die Chiffretexte sind ebenfalls nicht uniform ($\Pr(\mathtt{A}) = \Pr(\mathtt{B}) = \frac{1}{4}$, $\Pr(\mathtt{C}) = \Pr(\mathtt{D}) = \frac{1}{6}$, $\Pr(\mathtt{E}) = \Pr(\mathtt{F}) = \frac{1}{12}$). Die informationstechnische Sicherheit drückt sich dadurch aus, dass diese Chiffretextwahrscheinlichkeiten auch für jeden Klartext (also jede Spalte) separat auftreten.

1.4 Fallstudie für Cyphertext-only-Angriffe: Vigenère-Chiffre

In der Einleitung wurde schon kurz die sogenannte *Vigenère-Chiffre* angesprochen. Dies ist ein klassisches Verfahren zur Verschlüsselung natürlichsprachiger Texte. Üblicherweise nimmt man dabei den zu verschlüsselnden Text, lässt alle Satzzeichen und alle Leerzeichen weg und wandelt Groß- in Kleinbuchstaben um. Umlaute und andere Sonderzeichen werden umschrieben. Resultat ist eine Folge $x = (x_0, \ldots, x_{\ell-1}) = x_0 \ldots x_{\ell-1}$ von Buchstaben im Klartextalphabet $\{\mathtt{a}, \ldots, \mathtt{z}\}$ der Größe 26. Wir betrachten hier nur den Fall, wo die Klartextlänge von vornherein beschränkt ist (gemäß Szenario 1), also ist $\ell \leq L$ für ein festes L. Nun möchte man x verschlüsseln. Ein informationstheoretisch sicheres Verfahren ist, für jede Buchstabenposition $0 \leq i < L$ rein zufällig einen Schlüssel $k_i \in \{\mathtt{A}, \ldots, \mathtt{Z}\}$ zu wählen und an Position i die Verschiebechiffre mit Schlüssel k_i anzuwenden. Der Schlüssel $k_0, \ldots, k_{L-1}$ ist dann aber mindestens so lang wie die Klartextfolge. Allerdings ist das nach unseren bisherigen Ergebnissen auch unvermeidlich: Wenn $\mathcal{V} = (X, K, Y, e, d, \Pr_K)$ informationstheoretisch sicher ist, ist (X, K, Y, e, d) possibilistisch sicher, also $|X| \leq |K|$.

Es liegt nahe, zu versuchen, mit nur einem Schlüsselbuchstaben oder mit einem kürzeren Schlüssel auszukommen. Dies führt zur einfachen (wiederholten) Verschiebechiffre und zur Vigenère-Chiffre. Wir zeigen, dass man diese mit einfachen Mitteln „brechen“ kann.[18]

1.4.1 Die Vigenère-Chiffre und Angriffe bei bekannter Schlüssellänge

Es ist bequem, anstelle von Buchstaben mit Zahlen zu rechnen. Mit $\mathbb{Z}_n$ bezeichnen wir den Ring $\mathbb{Z}/n\mathbb{Z}$, also (etwas vereinfachend gesagt) den Ring der Zahlen $\{0, 1, \ldots, n-1\}$ mit Addition und Multiplikation modulo n als Operationen.

Definition: Eine *Verschiebechiffre* ist ein Kryptosystem $\mathcal{S} = (\mathbb{Z}_n, \mathbb{Z}_n, \mathbb{Z}_n, e, d)$ mit $e(x,k) = (x+k) \bmod n$. (Offensichtlich ist dann $d(y,k) = (y-k) \bmod n$.)

Unser zentrales Beispiel ist der Fall $n = 26$, also $X = Y = K = \{0, 1, 2, \ldots, 25\}$. Wir

[18] Die Vigenère-Chiffre ist nach Blaise de Vigenère (1523–1596) benannt, einem französischen Diplomaten und Kryptographen. Er betrachtete eigentlich die Variante, bei der man zum Verschlüsseln eines Buchstabens nicht eine Verschiebechiffre, sondern eine Substitutionschiffre (s. Def. 1.9) benutzt. Viele der folgenden Überlegungen lassen sich auf diese allgemeinere Situation übertragen.

identifizieren die Elemente dieser Menge mit den Buchstaben `a`, ..., `z` (bei X) bzw. `A`, ..., `Z` (bei K und Y). Die Konvention ist nach wie vor, Klartextbuchstaben klein und Schlüsselbuchstaben und Chiffretextbuchstaben groß zu schreiben.

Die einfachste Methode ist folgende Version der Cäsar-Chiffre (s. Einleitung): Wähle einen Schlüssel k aus $K = \{0, 1, \ldots, 25\} \hat{=} \{\texttt{A}, \ldots, \texttt{Z}\}$ zufällig. Um „Texte“ (d. h. Wörter über $\mathbb{Z}_n$) zu verschlüsseln, wird $\mathcal{S}$ buchstabenweise angewandt: Aus $x_0x_1 \ldots x_{\ell-1}$ wird $e(x_0, k)\, e(x_1, k) \ldots e(x_{\ell-1}, k)$.

Diese Methode ist allerdings sehr leicht zu brechen, sogar „von Hand“, also ohne massiven Einsatz von Computern. Es gibt mindestens die folgenden naheliegenden Möglichkeiten, einen gegebenen Chiffretext $y_0 \ldots y_{\ell-1}$, der aus einem natürlichsprachigen Text entstanden ist, zu entschlüsseln:

1. probiere die 26 möglichen Schlüssel aus, oder
2. zähle, welche Buchstaben am häufigsten im Chiffretext vorkommen und teste die Hypothese, dass einer von diesen für „`e`“ steht.

Betrachte beispielsweise den Chiffretext `RYFWAVSVNPLVOULHUZAYLUNBUN`.

Zählen liefert folgende Häufigkeiten für die häufigsten Buchstaben: `U`: 4, `L`: 3, `N`: 3, `V`: 3.

Vermutung: Einer dieser Buchstaben entspricht dem „`e`“.

Der Schlüssel k mit $e(\texttt{e}, k) = \texttt{U}$ ist $k = \texttt{Q}$. Entschlüsselung mit `Q` liefert das Wort `bipgkfcfxzvfyevrejkivexlex`, das nicht sehr sinnvoll erscheint.

Der Schlüssel k mit $e(\texttt{e}, k) = \texttt{L}$ ist $k = \texttt{H}$. Entschlüsselung mit `H` liefert `kryptologieohneanstrengung`, und wir sind fertig.

Als Basis für solche Entschlüsselungsansätze benutzt(e) man Häufigkeitstabellen für Buchstaben, wie die folgende (Angaben in Prozent):

Englisch		Deutsch		Italienisch	
`E, e`	12,31	`E, e`	18,46	`E, e`	11,79
`T, t`	9,59	`N, n`	11,42	`A, a`	11,74
`A, a`	8,05	`I, i`	8,02	`I, i`	11,28
`O, o`	7,94	`R, r`	7,14	`O, o`	9,83

(Dass das „`e`“ im Deutschen deutlich häufiger als im Englischen ist, liegt auch daran, dass bei der Umschreibung der Umlaute ä,ö und ü als ae, oe ue jeweils ein „`e`“ entsteht.)

Man kann auch die Häufigkeiten von „Digrammen“ (zwei Buchstaben, z. B. `ng`) oder „Trigrammen“ (drei Buchstaben, z. B. `ung` oder `eit`) heranziehen, auch um unterschiedliche Sprachen zu unterscheiden.

Eine unangenehme Eigenschaft bei der wiederholten Anwendung von reinen Verschiebechiffren ist, dass identische Buchstaben stets gleich verschlüsselt werden. Zum Beispiel hat unabhängig vom Schlüssel der Klartext `otto` stets zu einem Chiffretext mit dem Muster *abba*.

Die Grundidee der Vigenère-Chiffre ist es nun, verschiedene Verschiebechiffren in festgelegter zyklischer Reihenfolge zu verwenden.

Schlüssel: $k = k_0 k_1 k_2 \dots k_{s-1} \in \mathbb{Z}_n^s$, $s \in \mathbb{N}$. (Eine Folge von Verschiebewerten.)

Klartext: $x = x_0 x_1 \dots x_{\ell-1} \in \mathbb{Z}_n^\ell$, $\ell \in \mathbb{N}$.

Man verschlüsselt x_0 mit k_0, x_1 mit k_1, und so weiter. Wenn irgendwann der Schlüssel „aufgebraucht“ ist, weil $s < \ell$ gilt, fängt man mit dem Schlüssel wieder von vorne an. Wir verschlüsseln also x_0 mit k_0, …, x_{s-1} mit k_{s-1}, x_ℓ mit k_0, …, x_{2s-1} mit k_{s-1}, usw. Zusammengefasst:

Der Chiffretext ist: $y = y_0 y_1 \dots y_{\ell-1} \in (\mathbb{Z}_n)^*$ mit $y_i := e(x_i, k_{i \bmod s})$, für $0 \le i < \ell$.

Man kann dieses Verfahren mit einem festen Schlüssel k nun natürlich auf beliebig lange Klartexte anwenden. Damit liegt hier kein Kryptosystem im (technischen) Sinn des letzten Abschnitts vor!

Beispiel: Wir benutzen der einfacheren Lesbarkeit halber Buchstaben anstelle der Zahlen $0, \dots, 25$. Der Schlüssel ist `VENUS`.

wiederholter Schlüssel:	V	E	N	U	S	V	E	N	U	S	V	E	N	U	S	V
Klartext:	p	o	l	y	a	l	p	h	a	b	e	t	i	s	c	h
Chiffretext:	K	S	Y	S	S	G	T	U	U	T	Z	X	V	M	U	C

Um Ver- und Entschlüsselung von Hand schnell auszuführen, hilft Tabelle 1.

Wir werden die Längen s des Schlüssels und ℓ des Klartextes „sinnvoll“ beschränken:

Definition 1.33 *Das* Vigenère-Kryptosystem *(mit Parametern $(n, S, L) \in \mathbb{N}^3$) ist das Kryptosystem $((\mathbb{Z}_n)^{\le L}, (\mathbb{Z}_n)^{\le S}, (\mathbb{Z}_n)^{\le L}, e, d)$, so dass für alle $s \le S$, $\ell \le L$, $x_i, k_j \in \mathbb{Z}_n$ gilt:*

$$e(x_0 \dots x_{\ell-1}, k_0 \dots k_{s-1}) = y_0 \dots y_{\ell-1}$$

mit $y_i = (x_i + k_{i \bmod s}) \bmod n$, für alle $0 \le i < \ell$.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_	_
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Tabelle 1: Alle 26 Verschiebechiffren auf einen Blick

Für Anwendungen sollte man L „fast unendlich“ wählen, um die unendliche Menge der möglichen Klartexte zu approximieren. Hingegen wird S nicht sehr groß sein, da man die Anzahl der Schlüssel klein halten will.

Nun betrachten wir einen Angriff von Eva im Szenarium 1, bei dem sie nur einen Chiffretext y der Länge ℓ hat. Nehmen wir zunächst an, dass sie auch die Schlüssellänge $s \ll \ell$ und die zugrundeliegende (natürliche) Sprache kennt. Dann kann sie den Chiffretext durch Häufigkeitsanalysen zu entschlüsseln versuchen. Die zentrale Idee ist, dass für die Verschlüsselung des „Teiltextes“ $y^i = y_i\, y_{i+s}\, y_{i+2s}\, \ldots$, für $0 \leq i < s$, der Buchstabe k_i benutzt wurde, genau wie bei der einfachen Verschiebechiffre. Für i, $0 \leq i < s$, bestimmt Eva also die in diesem Teiltext $y^i = y_i\, y_{i+s}\, y_{i+2s}\, \ldots$ am häufigsten vorkommenden Buchstaben und testet die Hypothesen, dass diese für „e“ oder einen anderen häufigen Buchstaben stehen.

Wir betrachten ein Beispiel für eine solche Analyse an einem Chiffretext. (In der klassischen Kryptographie war es üblich, die Texte in Fünfergruppen einzuteilen, um das Abzählen von Buchstabenpositionen zu erleichtern).

```
EYRYC FWLJH FHSIU BHMJO UCSEG TNEER FLJLV SXMVY SSTKC MIKZS
JHZVB FXMXK PMMVW OZSIA FCRVF TNERH MCGYS OVYVF PNEVH JAOVW
UUYJU FOISH XOVUS FMKRP TWLCI FMWVZ TYOIS UUIIS ECIZV SVYVF
PCQUC HYRGO MUWKV BNXVB VHHWI FLMYF FNEVH JAOVW ULYER AYLER
VEEKS OCQDC OUXSS LUQVB FMALF EYHRT VYVXS TIVXH EUWJG JYARS
ILIER JBVVF BLFVW UHMTV UAIJH PYVKK VLHVB TCIUI SZXVB JBVVP
VYVFG BVIIO VWLEW DBXMS SFEJG FHFVJ PLWZS FCRVU FMXVZ MNIRI
GAESS HYPFS TNLRH UYR
```

y^0 =EFFBUTFSSMJFPOFTMOPJUFXFTFTUESPHMBVFFJUAVOOLFEVTEJIJBUUPVTSJVBVDSFPFFMGHTU.

Buchstaben in y^0 mit Häufigkeiten > 1: AB(4)DE(4)F(14)GH(2)IJ(5)LM(3)O(4)P(5)S(5)T(7)U(7)V(6)X

Mögliches Bild von „e“: F. Schlüsselbuchstabe wäre: B

y^1 =YWHHCNLXSIHXMZCNCVNAUOOMWMYUCVCYUNHLNALYECUUMYYIUYLBLHAYLCZBYVWBFHLCMNAYNY.

Buchstaben in y^1: A(4)B(3)C(8)EFH(6)I(2)L(7)M(5)N(7)O(2)SU(5)V(3)W(2)X(2)Y(19)Z(2)

Mögliches Bild von „e“: Y Schlüsselbuchstabe wäre: U

y^2 =RLSMSEJMTKZMMSREGYEOYIVKLWOIIYQRWXHMEOYLEQXQAHVVWAIVFMIVHIXVVILXEFWRXIEPLR.

Buchstaben in y^2: A(2)E(7)F(2)GH(3)I(8)JK(2)L(5)M(6)O(2)PQ(3)R(5)S(3)TV(7)W(4)X(5)Y(4)Z

Mögliche Bilder von „e“: I, V. Schlüsselbuchstaben wären: E, R

y^3 =YJIJEELVKZVXVIVRYVVVJSURCVIIZVUGKVWYVVEEKDSVLRXXJREVVTJKVUVVFIEMJVZVVRSFR.
Buchstaben in y^3: CDE(6)F(2)GI(5)J(6)K(4)L(2)MR(6)S(3)TU(3)V(21)WX(3)Y(3)Z(3)
Mögliches Bild von „e“: V. Schlüsselbuchstabe wäre: R

y^4 =CHUOGRVYCSBKWAFHSFHWUHSPIZSSVFCOVBIFHWRRSCSBFTSHGSRFWVHKBIBPGOWSGJSUZISSH.
Buchstaben in y^4: AB(5)C(4)F(6)G(4)H(8)I(4)JK(2)O(3)P(2)R(4)S(13)TU(3)V(4)W(5)YZ(2)
Mögliches Bild von „e“: S. Schlüsselbuchstabe wäre: O

Man versucht Schlüssel BURRO und erhält keinen sinnvollen Text. Mit BUERO ergibt sich:

```
denho echst enorg anisa tions stand erfuh rdiek rypto logie
inven edigw osiei nform einer staat liche nbuer otaet igkei
tausg euebt wurde esgab schlu essel sekre taere dieih rbuer
oimdo genpa lasth atten undfu erihr etaet igkei trund zehnd
ukate nimmo natbe kamen eswur dedaf uerge sorgt dasss iewae
hrend ihrer arbei tnich tgest oertw urden siedu rften ihreb
ueros abera uchni chtve rlass enbev orsie eineg estel lteau
fgabe geloe sthat ten
```

Ohne Gruppierung erhält man:

```
denhoechstenorganisationsstanderfuhrdiekryptologie
invenedigwosieinformeinerstaatlichenbuerotaetigkei
tausgeuebtwurdeesgabschluesselsekretaeredieihrbuer
oimdogenpalasthattenundfuerihretaetigkeitrundzehnd
ukatenimmonatbekameneswurdedafuergesorgtdasssiewae
hrendihrerarbeitnichtgestoertwurdensiedurftenihreb
uerosaberauchnichtverlassenbevorsieeinegestellteau
fgabegeloesthatten
```

Nun muss man nur noch die Wortzwischenräume und Satzzeichen ergänzen, um zu erhalten:

> *Den höchsten Organisationsstand erfuhr die Kryptologie in Venedig, wo sie in Form einer staatlichen Bürotätigkeit ausgeübt wurde. Es gab Schlüsselsekretäre, die ihr Büro im Dogenpalast hatten und für ihre Tätigkeit rund zehn Dukaten im Monat bekamen. Es wurde dafür gesorgt, dass sie während ihrer Arbeit nicht gestört wurden. Sie durften ihre Büros aber auch nicht verlassen, bevor sie eine gestellte Aufgabe gelöst hatten.*

```
AWMCJ IENAW NMOZV EYJOK HPXNK TFKQC JPJSJ NTIVT TCOJA AWKBS          50
NHKBV UYMJG NNUAH UEKFF DLNSJ SZRZL EUKRW IYLCJ MLZWC ECOBM         100
NOPSV ECOBX OCSOL IVKFC EYTHF IDYSM EMKFV IPCSK EYZZA CSKBS         150
LRUFA TSSWK CSKBO ECQNW URKVS BPTIW BPXGL RFQHM RPTRA EPYSJ         200
LLAPW NOGHW NPLTA ZTKBL ZFUFY AYOGA ECKBM NOGIX ZFLWF DPTIW         250
BPXVS EFLWY BPTIL ZEKOD GZXWL HXKBM NOAST ECJWW SEGBV ACJHW         300
CSTWC EYSWL DPTSF MLTOD GZXWL HXOGU HPVFG BWKAW MZJSD LTKFW         350
NGKFK TPNSF UYJZG EDKBC AYT
```

Abbildung 1: Vigenère-Chiffretext

1.4.2 Der Kasiski-Test

Das bisher betrachtete Verfahren setzt voraus, dass die Schlüssellänge s bekannt ist. Ist die maximale Schlüssellänge S klein, dann kann man die Schlüssellängen 1 bis S einzeln durchprobieren. Ist S groß, möchte man die Suche nach der richtigen Schlüssellänge abkürzen. (Besonders vor dem Computerzeitalter, wo die Dechiffrierung per Hand durchgeführt werden musste, war eine solche Zeitersparnis wichtig.) Die Schlüssellänge kann oft durch den *Kasiski-Test* näherungsweise bestimmt werden. (Der Test ist benannt nach Friedrich Wilhelm Kasiski (1805–1881), einem preußischen Infanteriemajor. Der Test wurde von ihm 1863 veröffentlicht. Er war aber bereits 1854 von Charles Babbage entwickelt, aber nicht veröffentlicht worden.)

Die zentrale Idee des Tests ist die folgende einfache Beobachtung: Gleiche Klartextfragmente, die eventuell mehrfach vorkommen (z. B. das Wort „ein“) werden in gleiche Chiffretexte übersetzt, wenn sie unter dem gleichen Schlüsselfragment liegen. Genauer: Stimmt der Klartext im Abschnitt $i+s\cdot\ell$ bis $j+s\cdot(\ell+h)$ mit dem Klartext im Abschnitt von $i+s\cdot\ell'$ bis $j+s\cdot(\ell'+h)$ überein, so gilt dies auch für den Chiffretext ($1\le i,j\le s$, $\ell,\ell',h\in\mathbb{N}$). Anders ausgedrückt: Kommt ein Teilwort im Klartext an zwei Positionen i und j und ist $j-i$ ein Vielfaches von s, so werden die beiden Vorkommen des Wortes gleich verschlüsselt.

Diese Beobachtung wird in die folgende Idee für einen Angriff umgemünzt: Für möglichst viele „lange“ Wörter, die im *Chiffretext* mehrfach auftreten, notiere die Abstände des Auftretens. („lang“ sollte wenigstens 3 sein.) Dann suche ein großes s, das viele dieser Abstände teilt (nicht unbedingt alle, denn einige Mehrfachvorkommen im Chiffretext könnten zufällig entstanden sein).

Beispiel 1.34 Im Chiffretext von Abbildung 1 kommen (mindestens) die folgenden Wörter der Länge 3 mehrfach vor. Wir geben die Positionen und die Abstände an.

Wort	Positionen	Abstände
ODGZXWLHX	269, 319	50
DPT	246, 311	65
BPT	176, 261	115
ECOB	96, 106	10
CSK	138, 146, 161	8*, 15
AWM	1, 339	338*
PTIWBPX	177, 247	70
BMNO	99, 234, 279	135, 45

Wir vermuten:

- Periode ist 5 (dann wären Wiederholungen von AWM und CSK durch Zufall entstanden)

Das Ergebnis der Entschlüsselung wie oben beschrieben mit vermuteter Schlüssellänge 5 und versuchten Schlüsseln ALGXS (erfolglos) und ALGOS (erfolgreich) ergibt den Text aus Abbildung 2.

```
algor ithme nbild endas herzs tueck jeder nicht trivi alena   50
nwend ungvo ncomp utern daher sollt ejede infor matik erinu  100
ndjed erinf ormat ikerk enntn isseu eberd iewes entli chena  150
lgori thmis chenw erkze ugeha benue berst ruktu rendi eeser  200
laube ndate neffi zient zuorg anisi erenu ndauf zufin denue  250
berha eufig benut zteal gorit hmenu ndueb erdie stand ardte  300
chnik enmit denen manal gorit hmisc hepro bleme model liere  350
nvers tehen undlo esenk ann
```

Mit Wortzwischenräumen und Satzzeichen:

> *Algorithmen bilden das Herzstück jeder nichttrivialen Anwendung von Computern. Daher sollte jede Informatikerin und jeder Informatiker Kenntnisse über die wesentlichen algorithmischen Werkzeuge haben: über Strukturen, die es erlauben, Daten effizient zu organisieren und aufzufinden, über häufig benutzte Algorithmen und über die Standardtechniken, mit denen man algorithmische Probleme modellieren, verstehen und lösen kann.*

Abbildung 2: Entschlüsselter Text mit Schlüssel ALGOS

Bemerkungen: (i) Der Test funktioniert nur gut, wenn die Schlüssellänge s gering im Verhältnis zur Chiffretextlänge ℓ ist. (ii) Um ihn anwenden zu können, muss die Klartextsprache bekannt sein. (iii) Der Test kann auch in der viel allgemeineren Situation benutzt werden, in der Schlüssel nicht s Verschiebungen, sondern s beliebige Substitutionschiffren auf X bestimmen (z. B. $X = Y$ und Schlüssel ist Tupel $(\pi_0, \ldots, \pi_{s-1})$ von Permutationen von X).

Was passiert im Extremfall $s = \ell$?

- Grundsätzlich hat man dann ein informationstheoretisch sicheres one-time pad vor sich ...
- ... aber nur dann, wenn die Schlüssel *gleichverteilt* gewählt werden.
  Wenn der Schlüssel selbst ein deutscher Text ist (z. B. ein Textstück aus einem Buch), so weist der Chiffretext wieder statistische Merkmale auf, die zum Brechen ausgenutzt werden können. (Beispiel: Wenn Schlüssel und Klartext beides deutsche Texte sind, werden ca. 7,6% der Buchstaben mit sich selbst verschlüsselt, d. h. Chiffretextbuchstabe = 2 · Klartextbuchstabe modulo 26.)

Effektive Verfahren der Schlüsselverlängerung (die aber keine informationstheoretische Sicherheit bringen):

- Autokey-Vigenère: Schlüssel k, Klartext x. Dann wird die klassische Vigenère-Chiffre mit der Konkatenation kx als Schlüssel auf x angewendet.
- Pseudozufallszahlen: Geheimer Schlüssel ist *seed* eines (Pseudo-)Zufallszahlengenerators, mit dem eine lange Schlüsselfolge $k_0, \ldots, k_{\ell-1}$ erzeugt wird.

1.4.3 Koinzidenzindex und Friedman-Methode

Wir betrachten noch eine andere interessante Methode zur Abschätzung der Schlüssellänge, die bei der Verwendung einer Vigenère-Chiffre oder anderen Substitutionschiffren mit fester Schlüssellänge s helfen können, diese zu ermitteln. Die Methode beruht darauf, dass die Buchstabenhäufigkeiten (zu einer gegebenen Sprache) feststehen und sich bei der Verschlüsselung mit einer einfachen Substitutionschiffre nicht ändert. Ebenso ändert sich nicht die Wahrscheinlichkeit, bei der zufälligen Wahl eines Buchstabenpaars zwei identische Buchstaben zu erhalten. Die Methode stammt von William F. Friedman (1891–1969), einem amerikanischen Kryptographen.

Sei $x = x_0 \ldots x_{\ell-1}$ ein Klartext, sei $y = y_0 \ldots y_{\ell-1}$ der zugehörige Chiffretext, bei $s = 1$ (an jeder Stelle derselbe Schlüssel). Seien $n_0, \ldots, n_{25}$ die Anzahlen der Buchstaben a, ...,

z in x, $n'_0, \ldots, n'_{25}$ die in y. Wir wählen zufällig ein Paar von zwei Positionen in x (ohne „Zurücklegen“). Dafür gibt es $\binom{\ell}{2}$ Möglichkeiten. Genau $\binom{n_i}{2}$ viele davon führen dazu, dass man zweimal den Buchstaben Nummer i zieht, und $\sum_{0\le i<26}\binom{n_i}{2}$ viele führen dazu, dass man an den beiden Positionen denselben Buchstaben sieht. Wir setzen

$$\mathrm{IC}(x) := \frac{\sum_{0\le i<26}\binom{n_i}{2}}{\binom{\ell}{2}} = \frac{\sum_{0\le i<26} n_i(n_i-1)}{\ell(\ell-1)}.$$

Diese Zahl nennt man den *Koinzidenzindex* von x. Sie ist die Wahrscheinlichkeit dafür, dass an den beiden zufällig gewählten Positionen derselbe Buchstabe steht. Weil die Verschlüsselung auf den Buchstaben eine Bijektion ist, also sich die vorkommenden Häufigkeiten durch die Verschlüsselung nicht ändern, gilt für

$$\mathrm{IC}(y) := \frac{\sum_{0\le i<26} n'_i(n'_i-1)}{\ell(\ell-1)}$$

die Gleichung $\mathrm{IC}(x) = \mathrm{IC}(y)$.

Für lange Texte mit (sprachtypischer) Häufigkeitsverteilung der Buchstaben nähert sich $\mathrm{IC}(x)$ einem bestimmten Wert an. Wenn p_i die Häufigkeit von Buchstabe i in der verwendeten Sprache ist, wird für lange Texte x die Näherung $\frac{n_i}{\ell} \approx \frac{n_i-1}{\ell-1} \approx p_i$ gelten, also

$$\mathrm{IC}(x) \approx \sum_{0\le i<26} p_i^2$$

sein. Die Summe $\sum_{0\le i<26} p_i^2$ hat beispielsweise einen Wert von etwa 0,076 für deutsche und 0,066 für englische Texte. Wenn (in einer fiktiven Sprache) jeder Buchstabe dieselbe Wahrscheinlichkeit hat, ist

$$\sum_{0\le i<26} p_i^2 = 26 \cdot \left(\frac{1}{26}\right)^2 = \frac{1}{26} \approx 0{,}0385;$$

dies ist zugleich der minimal mögliche Wert.

Für die Ermittlung eines Schätzwertes für die Schlüssellänge s gehen wir wie folgt vor. Wir nehmen an, die zugrundeliegende Sprache ist Deutsch. Wir berechnen zunächst $\mathrm{IC}(y)$ für den Chiffretext y. Die unbekannte Schlüssellänge nennen wir s. Dann berechnen wir eine Näherung für $\mathrm{IC}(y)$, auf eine zweite Weise. Dies wird uns eine (Näherungs-)Gleichung für s liefern.

Wir überlegen: Bilde die Teilwörter $y^0, \ldots, y^{s-1}$ wie in Abschnitt 1.4, jedes mit der Länge ℓ/s. Innerhalb jedes Teilworts kommen Kollisionen ebenso häufig vor wie in einem gewöhnlichen Text mit nur einem Schlüssel, also erwarten wir zusammen

$$\binom{\ell/s}{2}\mathrm{IC}(y^0) + \cdots + \binom{\ell/s}{2}\mathrm{IC}(y^{s-1}) \approx s\binom{\ell/s}{2} \cdot 0{,}076 = \frac{1}{2}\ell(\ell/s-1) \cdot 0{,}076$$

viele „Kollisionen“ (Paare identischer Chiffretextbuchstaben) aus den einzelnen Teilwörtern. Zwischen zwei Teilwörtern y^u und y^v erwarten wir $(\ell/s)^2 \cdot \frac{1}{26} \approx 0{,}0385(\ell/s)^2$ Kollisionen, aus allen $\binom{s}{2}$ Paaren von Teilwörtern zusammen also

$$\binom{s}{2} 0{,}0385(\ell/s)^2 = \frac{s(s-1)}{2} \cdot 0{,}0385(\ell/s)^2 = \frac{1}{2} \cdot 0{,}0385\ell^2 \left(1 - \frac{1}{s}\right)$$

viele. Zusammen ist die erwartete Anzahl an Kollisionen in y gleich

$$\frac{1}{2}\ell \left(0{,}076(\ell/s - 1) + 0{,}0385\ell \left(1 - \frac{1}{s}\right)\right).$$

Diese Zahl sollte näherungsweise gleich $\frac{1}{2}\ell(\ell-1)\mathrm{IC}(y)$ sein. Wir können die resultierende Gleichung

$$(\ell - 1)\mathrm{IC}(y) = 0{,}076(\ell/s - 1) + 0{,}0385\ell \left(1 - \frac{1}{s}\right)$$

nach s auflösen und erhalten:

$$s \approx \frac{(0{,}076 - 0{,}0385)\ell}{(\ell - 1)\mathrm{IC}(y) - 0{,}0385\ell + 0{,}076}.$$

(Wenn man anstelle der Konstanten 0,076 den Wert 0,066 einsetzt, erhält man die entsprechende Formel für englischsprachige Texte.)

Eine tatsächliche Durchführung des Verfahrens mit Chiffretexten wie im vorigen Kapitel erfordert viel Geduld (oder den Einsatz eines Computers).

Beim „venezianischen“ Chiffretext `EYRYC...UYR` von oben ergibt sich:

a_i	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
n'_i	8	12	13	2	18	25	7	19	20	14	8	15	16	7	12	8	3	15	25	10	19	41	13	11	19	8

Dies liefert $\mathrm{IC}(y) \approx 0{,}048024$ und $\ell = 368$. Damit erhalten wir

$$s \approx \frac{0{,}0375 \cdot 368}{367 \cdot 0{,}048024 - 0{,}0385 \cdot 368 + 0{,}076} \approx 3{,}9.$$

Das ist nicht zu nahe am tatsächlichen Wert 5, aber auch nicht ungeheuer weit weg. (Die Formel reagiert sehr empfindlich auf kleine Änderungen in $\mathrm{IC}(y)$. Mit $\mathrm{IC}(y) = 0{,}05$ ergibt sich $s \approx 3{,}24$, mit $\mathrm{IC}(y) = 0{,}046$ ergibt sich $s \approx 4{,}95$.)