3 Sicherheit von Kryptosystemen

3.1 Informationstheoretische Sicherheit

Claude E. Shannon untersuchte die Sicherheit kryptographischer Systeme auf informationstheoretischer Basis (1945, freigegeben 1949). Seinen Untersuchungen liegt das Modell einer Nachrichtenquelle zugrunde, die einzelne Nachrichten unter einer bestimmten Wahrscheinlichkeitsverteilung aussendet.

Bei der Betrachtung der informationstheoretischen Eigenschaften von Kryptosystemen gehen wir von einer Wahrscheinlichkeitsverteilung auf den Paaren $(k, x) \in K \times M$ aus, d. h. $p(k, x)$ gibt die Wahrscheinlichkeit an, dass der Klartext x mit dem Schlüssel k verschlüsselt wird. Dabei setzen wir voraus, dass nach jeder Verschlüsselung einer Nachricht x ein neuer Schlüssel gewählt wird. Dies bedeutet, dass beispielsweise bei der additiven Chiffre für $M = A^n$ zu setzen ist (und nicht $M = A$ wie in Definition 4), falls mit dem selben Schlüssel eine Folge von n Buchstaben chiffriert wird.

Weiterhin nehmen wir an, dass der Schlüssel unabhängig vom Klartext gewählt wird. D.h. es ist $p(k, x) = p(k)p(x)$, wobei

$$p(k) = \sum_{x \in M} p(k, x)$$

die Wahrscheinlichkeit für den Schlüssel k und

$$p(x) = \sum_{k \in K} p(k, x)$$

die Wahrscheinlichkeit für den Klartext x ist. O.B.d.A. sei $p(x) > 0$ für alle Klartexte $x \in M$, da wir andernfalls x aus M entfernen können. Für einen Kryptotext y berechnet sich die Wahrscheinlichkeit zu

$$p(y) = \sum_{k,x:E(k,x)=y} p(k, x)$$

und für einen fest vorgegebenen Kryptotext y ist

$$p(x|y) = \sum_{k:E(k,x)=y} \frac{p(k, x)}{p(y)}$$

die (bedingte) Wahrscheinlichkeit dafür, dass y aus dem Klartext x entstanden ist. Wir nehmen o.B.d.A. an, dass für alle $y \in C$ $p(y) > 0$ ist (andernfalls kann y aus C entfernt werden).

Definition 64 (**informationstheoretisch sicher**)

Ein Kryptosystem heißt ***absolut sicher*** (***informationstheoretisch sicher***), falls für alle $x \in M$ und alle $y \in C$ gilt:

$$p(x) = p(x|y).$$

Bei einem absolut sicheren Kryptosystem ist demnach die *a posteriori* Wahrscheinlichkeit $p(x|y)$ einer Klartextnachricht x gleich der *a priori* Wahrscheinlichkeit $p(x)$, d.h. die Wahrscheinlichkeit von x sie ist unabhängig davon, ob der Kryptotext y bekannt ist oder nicht. Die Kenntnis von y erlaubt somit keinerlei Rückschlüsse auf die gesendete Nachricht x. Dies bedeutet, dass es dem Gegner nicht möglich ist – auch nicht mit unbegrenzten Rechenressourcen – das System zu brechen. Wie wir sehen werden, lässt sich diese Art der Sicherheit nur mit einem extrem hohen Aufwand realisieren.

Wegen $p(x|y)p(y) = p(x,y) = p(y|x)p(x)$ ist

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

(Satz von Bayes) und daher ist die Bedingung $p(x) = p(x|y)$ gleichbedeutend mit $p(y) = p(y|x)$.

Beispiel 65 *Sei* (M, C, E, D, K) *ein Kryptosystem mit* $M = \{x_1, \ldots, x_4\}$, $K = \{k_1, \ldots, k_4\}$, $C = \{y_1, \ldots, y_4\}$ *und*

E	x_1	x_2	x_3	x_4
k_1	y_1	y_4	y_3	y_2
k_2	y_2	y_1	y_4	y_3
k_3	y_3	y_2	y_1	y_4
k_4	y_4	y_3	y_2	y_1

Weiter sei $p(x_1) = 1/2$, $p(x_2) = p(x_3) = p(x_4) = 1/6$, *sowie* $p(k_1) = 1/2$, $p(k_2) = 1/4$ *und* $p(x_3) = p(x_4) = 1/8$. *Dann ergibt sich folgende Verteilung auf* C*:*

$$\begin{aligned} p(y_1) &= 1/2 \cdot 1/2 + (1/4 + 1/8 + 1/8) \cdot 1/6 = 1/3 \\ p(y_2) &= 1/4 \cdot 1/2 + (1/8 + 1/8 + 1/2) \cdot 1/6 = 1/4 \\ p(y_3) &= 1/8 \cdot 1/2 + (1/8 + 1/2 + 1/4) \cdot 1/6 = 5/24 \\ p(y_4) &= 1/8 \cdot 1/2 + (1/2 + 1/4 + 1/8) \cdot 1/6 = 5/24 \end{aligned}$$

Die bedingten Wahrscheinlichkeiten $p(x|y_1)$ *berechnen sich wie folgt:*

$$\begin{aligned} p(x_1|y_1) &= p(k_1, x_1)/p(y_1) = (1/2)(1/2)/(1/3) = 3/4 \\ p(x_2|y_1) &= p(k_2, x_2)/p(y_1) = (1/4)(1/6)/(1/3) = 1/8 \\ p(x_3|y_1) &= p(k_3, x_3)/p(y_1) = (1/8)(1/6)/(1/3) = 1/16 \\ p(x_4|y_1) &= p(k_4, x_4)/p(y_1) = (1/8)(1/6)/(1/3) = 1/16 \end{aligned}$$

Wegen $p(x_1) = 1/2 \neq 3/4 = p(x_1|y_1)$ *ist das System unter dieser Verteilung nicht absolut sicher.*

Die Bedingung $p(x) = p(x|y)$ *ist nach dem Satz von Bayes genau dann erfüllt, wenn* $p(y) = p(y|x)$ *ist. Da jedoch für jedes Paar* (x, y) *genau ein Schlüssel* $k = k_{x,y} \in K$ *mit* $E(k, x) = y$ *existiert, also* $p(y|x) = p(k_{x,y})$ *ist, ist dies äquivalent zu* $p(y) = p(k_{x,y})$. *Für* $y = y_1$ *bedeutet dies, dass alle Schlüssel* $k_i = k_{x_i, y_1}$ *die gleiche Wahrscheinlichkeit* $p(k_i) = 1/4$ *haben müssen. Eine leichte Rechnung zeigt, dass unter dieser Schlüsselverteilung* $p(y_i) = 1/4$ *für* $i = 1, \ldots, 4$ *ist. Somit ist das Kryptosystem genau dann absolut sicher, wenn der Schlüssel unter Gleichverteilung gewählt wird (dies gilt unabhängig von der Klartextverteilung).*

Wie in diesem Beispiel lässt sich allgemein folgende hinreichende Bedingung für die absolute Sicherheit von Kryptosystemen zeigen.

Satz 66 *Ein Kryptosystem mit $\|M\| = \|C\| = \|K\|$, in dem es für jeden Klartext x und jeden Kryptotext y genau einen Schlüssel k mit $E(k,x) = y$ gibt, ist absolut sicher, wenn die Schlüssel unter Gleichverteilung gewählt werden.*

Beweis: Bezeichne $k_{x,y}$ den eindeutig bestimmten Schlüssel, der den Klartext x auf den Kryptotext y abbildet. Wegen $p(k_{x,y}) = \|K\|^{-1}$ für alle x, y folgt zunächst

$$p(y) = \sum_{k,x:E(k,x)=y} p(k,x) = \sum_x p(x) \cdot p(k_{x,y}) = p(k_{x,y}) \sum_x p(x) = p(k_{x,y})$$

und damit

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x) \cdot p(y|x)}{p(y)} = \frac{p(x) \cdot p(k_{x,y})}{p(y)} = p(x).$$

■

In den Übungen wird gezeigt, dass im Fall $p(x) > 0$ für alle $x \in M$ auch die Umkehrung dieses Satzes gilt.

Verwendet man beim *One-Time-Pad* nur Klartexte einer festen Länge n, d. h. $M \subseteq A^n$, so ist dieser nach obigem Satz absolut sicher (vorausgesetzt, der Schlüssel wird rein zufällig, also unter Gleichverteilung gewählt). Variiert die Klartextlänge, so kann ein Gegner aus y nur die Länge des zugehörigen Klartextes x ableiten. Wird jedoch derselbe Schlüssel k zweimal verwendet, so kann aus den Kryptotexten die Differenz der zugehörigen Klartexte ermittelt werden:

$$\left.\begin{array}{lcl} y_1 & = & E(x_1,k) = x_1 + k \\ y_2 & = & E(x_2,k) = x_2 + k \end{array}\right\} \rightsquigarrow y_1 - y_2 = x_1 - x_2$$

Sind die Klartexte natürlichsprachig, so können aus $y_1 - y_2$ die beiden Nachrichten x_1 und x_2 ähnlich wie bei der Analyse einer Lauftextverschlüsselung (siehe Abschnitt 2.5) rekonstruiert werden.

Da in einem absolut sicheren Kryptosystem mit $p(x) > 0$ für alle $x \in M$ der Schlüsselraum K mindestens die Größe des Klartextraumes M haben muss (siehe Übungen), ist der Aufwand extrem hoch. Vor der Kommunikation muss ein Schlüssel, dessen Länge der des zu übertragenden Klartextes entspricht, zufällig generiert und zwischen den Partnern auf einem sicheren Kanal ausgetauscht werden. Wird hingegen keine absolute Sicherheit angestrebt, so kann der Schlüsselstrom auch von einem Pseudo-Zufallsgenerator erzeugt werden. Dieser erhält als Eingabe eine Zufallsfolge s_0 (den sogenannten *Keim*) und erzeugt daraus eine lange Folge $v_0\, v_1 \ldots$ von Pseudo-Zufallszahlen. Als Schlüssel muss jetzt nur noch das Wort s_0 ausgetauscht werden.

In der Informationstheorie wird die Unsicherheit, mit der eine durch X beschriebene Quelle ihre Nachrichten aussendet, nach ihrer Entropie bemessen. Das heißt, die Unsicherheit über X entspricht genau dem Informationsgewinn, der sich aus der Beobachtung der Quelle X ziehen lässt. Dabei wird die in einer einzelnen Nachricht (***m**essage*)

m steckende Information um so höher bemessen, je seltener m auftritt. Tritt eine Nachricht m mit einer positiven Wahrscheinlichkeit $p(m) = Prob[X = m] > 0$ auf, dann ist

$$\mathit{Inf}_X(m) = \log_2(1/p(m))$$

der **Informationsgehalt** von m. Ist dagegen $p(m) = 0$, so sei $\mathit{Inf}_X(m) = 0$. Dieser Wert des Informationsgehalts ergibt sich zwangsläufig aus den beiden folgenden Forderungen:

- Der gemeinsame Informationsgehalt $\mathit{Inf}_{X,Y}(m, m')$ von zwei Nachrichten m und m', die aus stochastisch unabhängigen Quellen X und Y stammen, sollte gleich $\mathit{Inf}_X(m) + \mathit{Inf}_Y(m')$ sein;
- der Informationsgehalt einer Nachricht, die mit Wahrscheinlichkeit $1/2$ auftritt, soll genau 1 (bit) betragen.

Die Entropie von X ist nun der erwartete Informationsgehalt einer von X stammenden Nachricht.

Definition 67 (**Entropie**)

Sei X eine Zufallsvariable mit Wertebereich $W(X) = \{m_1, \ldots, m_n\}$ und sei $p_i = Prob[X = m_i]$. Dann ist die **Entropie** von X definiert als

$$H(X) = \sum_{i=1}^{n} p_i \mathit{Inf}_X(m_i).$$

Die Entropie nimmt also im Fall $p_1 = \cdots p_n = 1/n$ den Wert $\log_2(n)$ an. Für jede andere Verteilung $p_1, \ldots, p_n$ gilt dagegen $H(X) < \log_2(n)$ (Beweis unten). Generell ist die Unsicherheit über X um so kleiner, je ungleichmäßiger X verteilt ist. Bringt X nur einen einzigen Wert mit positiver Wahrscheinlichkeit hervor, dann (und nur dann) nimmt $H(X)$ den Wert 0 an. Für den Nachweis von oberen Schranken für die Entropie benutzen wir folgende Hilfsmittel aus der Analysis.

Definition 68 (**konkav**)

Eine reellwertige Funktion f ist **konkav** auf einem Intervall I, falls für alle $x \neq y \in I$ gilt:

$$f\left(\frac{x+y}{2}\right) \quad \geq \quad \frac{f(x)+f(y)}{2}$$

Gilt sogar „$>$“ anstelle von „$\geq$“, so heißt f **streng konkav** auf I.

Satz 69 (Jensensche Ungleichung) *Sei f eine streng konkave Funktion auf I und seien $0 \leq a_1, \ldots, a_n \leq 1$ reelle Zahlen mit $\sum_{i=1}^{n} a_i = 1$. Dann gilt*

$$\sum_{i=1}^{n} a_i f(x_i) \quad \leq \quad f\left(\sum_{i=1}^{n} a_i x_i\right).$$

Die beiden Seiten der Ungleichung sind genau dann gleich, wenn $x_1 = \cdots = x_n$ ist.

Beispiel 70 *Die Funktion* $f(x) = \log_2(x)$ *ist konkav auf* $(0, \infty)$.

Satz 71 *Sei* X *eine Zufallsvariable mit Wertebereich* $W(X) = \{x_1, \ldots, x_n\}$ *und Verteilung* $Prob[X = x_i] = p_i$, $i = 1, \ldots, n$. *Dann gilt* $H(X) \leq \log_2(n)$, *wobei Gleichheit genau im Fall* $p_i = 1/n$ *für* $i = 1, \ldots, n$ *eintritt.*

Beweis: Es gilt

$$
\begin{aligned}
H(X) &= \sum_{i=1}^{n} p_i \log_2(1/p_i) \\
&\leq \log_2 \sum_{i=1}^{n} p_i/p_i \\
&= \log_2 n.
\end{aligned}
$$

Nach obigem Satz tritt Gleichheit genau im Fall $1/p_1 = \cdots = 1/p_n$ ein, was mit $p_i = 1/n$ für $i = 1, \ldots, n$ gleichbedeutend ist. ■

Eine wichtige Eigenschaft der Entropie ist, dass sie eine untere Schranke für die mittlere Codewortlänge von Binärcodes bildet. Ein **Binärcode** für X ist eine (geordnete) Menge $C = \{x_1, \ldots, x_n\}$ von binären Codewörtern x_i für die Nachrichten m_i mit der Eigenschaft, dass die Abbildung $c : M^* \to \{0,1\}^*$ mit $c(m_{i_1} \cdots m_{i_k}) = x_{i_1} \cdots x_{i_k}$ injektiv ist. Die mittlere Codewortlänge von C unter X ist

$$
L(C) = \sum_{i=1}^{n} p_i \cdot |x_i|.
$$

C heißt **optimal**, wenn kein anderer Binärcode für X eine kürzere mittlere Codewortlänge besitzt. Für einen optimalen Binärcode C für X gilt

$$
H(X) \leq L(C) < H(X) + 1.
$$

Beispiel 72 (Entropie) *Sei* X *eine Zufallsvariable mit der Verteilung*

m_i	*sonnig*	*leicht bewölkt*	*bewölkt*	*stark bewölkt*	*Regen*	*Schnee*	*Nebel*
p_i	$1/4$	$1/4$	$1/8$	$1/8$	$1/8$	$1/16$	$1/16$

Dann ergibt sich die Entropie von X *zu*

$$
H(X) = 1/4 \cdot (2+2) + 1/8 \cdot (3+3+3) + 1/16 \cdot (4+4) = 2{,}625.
$$

Betrachten wir die beiden Codes $C_1 = \{\texttt{001}, \texttt{010}, \texttt{011}, \texttt{100}, \texttt{101}, \texttt{110}, \texttt{111}\}$ *und* $C_2 = \{\texttt{00}, \texttt{01}, \texttt{100}, \texttt{101}, \texttt{110}, \texttt{1110}, \texttt{1111}\}$, *so erhalten wir für die mittlere Codewortlänge von* C_1 *den Wert* $L(C_1) = 3$, *während* C_2 *wegen* $|x_i| = \log_2(1/p_i)$ *den Wert* $L(C_2) = H(X)$ *erreicht und somit optimal ist.*

Die Redundanz eines Codes für eine Zufallsvariable X ist um so höher, je größer seine mittlere Codewortlänge im Vergleich zur Entropie von X ist. Um auch Codes über unterschiedlichen Alphabeten miteinander vergleichen zu können, ist es notwendig, die Codewortlänge in einer festen Einheit anzugeben. Hierzu berechnet man die **Bitlänge** eines Wortes x über einem Alphabet A mit $m > 2$ Buchstaben zu $|x|_2 = |x| \log_2(m)$. Beispielsweise ist die Bitlänge von GOLD (über dem lateinischen Alphabet) $|\texttt{GOLD}|_2 = 4 \log_2(26) = 18,8$. Entsprechend berechnet sich für einen Code $C = \{x_1, \ldots, x_n\}$ unter einer Verteilung $p_1, \ldots, p_n$ die mittlere Codewortlänge (in bit) zu

$$L_2(C) = \sum_{i=1}^{n} p_i \cdot |x_i|_2.$$

Damit können wir die Redundanz eines Codes als den mittleren Anteil der Codewortbuchstaben definieren, die keine Information tragen.

Definition 73 (**Redundanz**)

Die **(relative) Redundanz** eines Codes C für X ist definiert als

$$\mathcal{R}(C) = \frac{L_2(C) - H(X)}{L_2(C)}.$$

Beispiel 72 (Entropie, Fortsetzung) *Während eine von X generierte Nachricht im Durchschnitt $H(X) = 2{,}625$ bit an Information enthält, haben die Codewörter von C_1 eine Bitlänge von 3. Der Anteil an „überflüssigen“ Zeichen pro Codewort beträgt also*

$$\mathcal{R}(C_1) = \frac{3 - 2{,}625}{3} = 12,5\%,$$

wogegen C_2 keine Redundanz besitzt.

Auch Schriftsprachen wie Deutsch oder Englisch und Programmiersprachen wie C oder PASCAL können als eine Art Code aufgefasst werden. Um die statistischen Eigenschaften einer solchen Sprache L zu erforschen, erweist es sich als zweckmäßig, die Textstücke der Länge n (n-Gramme) von L für unterschiedliche n getrennt voneinander zu betrachten. Sei also L_n die Zufallsvariable, die die Verteilung aller n-Gramme in L beschreibt. Interpretieren wir diese n-Gramme als Codewörter einer einheitlichen Codewortlänge n, so ist

$$\mathcal{R}(L_n) = \frac{n \log_2 m - H(L_n)}{n \log_2 m}$$

die Redundanz dieses Codes. Es ist zu erwarten, dass eine Sprache umso mehr Redundanz aufweist, je restriktiver die Gesetzmäßigkeiten sind, unter denen in ihr Worte und Sätze gebildet werden.

Definition 74 (**Entropie einer Sprache**)

Für eine Sprache L über einem Alphabet A mit $\|A\| = m$ und n-Gramm-Verteilung L_n ist $H(L_n)/n$ die ***Entropie von*** L_n (pro Buchstabe). Falls dieser Wert für n gegen ∞ gegen einen Grenzwert

$$H(L) = \lim_{n\to\infty} H(L_n)/n$$

konvergiert, so wird dieser Grenzwert als die ***Entropie von*** L bezeichnet. In diesem Fall konvergiert $\mathcal{R}(L_n)$ gegen den Grenzwert

$$\lim_{n\to\infty} \mathcal{R}(L_n) = \frac{\log_2 m - H(L)}{\log_2 m},$$

der als die ***(relative) Redundanz*** $\mathcal{R}(L)$ von L bezeichnet wird. Der Zähler $\log_2 m - H(L)$ in diesem Ausdruck wird auch als die ***absolute Redundanz*** $\mathcal{R}_{abs}(L)$ der Klartextsprache (gemessen in bit/Buchstabe) bezeichnet.

Für eine Reihe von natürlichen Sprachen wurden die Redundanzen $\mathcal{R}(L_n)$ der n-Gramme (für nicht allzu große Werte von n) empirisch bestimmt, so dass sich $\mathcal{R}(L)$ näherungsweise bestimmen lässt.

Beispiel 75 *Im Deutschen haben die Einzelbuchstaben eine Entropie von $H(L_1) = 4,1$ bit, während sich für die Bi- und Trigramme Entropiewerte von 3,86 und 3,61 bit pro Buchstabe ergeben. Mit wachsender Länge sinkt die Entropie von deutschsprachigen Texten weiter ab und strebt gegen einen Grenzwert $H(L)$ von 1,56 bit pro Buchstabe.*

n	$H(L_n)$	$H(L_n)/n$	$R(L_n)$
1	$4,10$	$4,10$	14%
2	$7,72$	$3,86$	19%
3	$10,82$	$3,61$	24%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
∞	∞	$1,56$	67%

Für die Redundanz $\mathcal{R}(L)$ ergibt sich hieraus (wegen log(26) = 4,76) ein Wert von ca. 67%, so dass sich ein deutscher Text bei optimaler Kodierung auf circa $1/3$ seiner ursprünglichen Länge komprimieren lässt.

Wir betrachten nun den Fall, dass mit einem Kryptosystem Klartexte der Länge n verschlüsselt werden, ohne dass dabei der Schlüssel gewechselt wird. D. h.

$$E_n : K \times A^n \to C_n$$

wobei C_n die Menge der zugehörigen Kryptotexte ist. Der Einfachheit halber nehmen wir $\|C_n\| = \|A^n\| = m^n$ an. Ist y ein abgefangener Kryptotext, so ist

$$K(y) = \{k \in K \mid \exists x \in A^n : E_n(k,x) = y \wedge p(x) > 0\}$$

die Menge aller in Frage kommenden Schlüssel für y. $K(y)$ besteht aus einem „echten" (d. h. dem zur Generierung von y tatsächlich benutzten) und $\|K(y)\| - 1$ so genannten „unechten" Schlüsseln. Aus informationstheoretischer Sicht ist das Kryptosystem desto unsicherer, je kleiner die erwartete Anzahl

$$\bar{s}_n = \sum_{y\in C_n} p(y) \cdot (\|K(y)\| - 1) = \sum_{y\in C_n} p(y) \cdot \|K(y)\| - 1$$

der unechten Schlüssel ist. Ist $\bar{s}_n$ gleich 0, so liefert der abgefangene Kryptotext y dem Gegner genügend Information, um den benutzten Schlüssel und somit den zu y gehörigen Klartext eindeutig bestimmen zu können (sofern er über unbegrenzte Ressourcen an Rechenkraft und Zeit verfügt).

Definition 76 (**Eindeutigkeitsdistanz**)

Die **Eindeutigkeitsdistanz** n_0 eines Kryptosystems ist der kleinste Wert von n, für den $\bar{s}_n = 0$ wird.

Als nächstes wollen wir eine untere Schranke für n_0 herleiten. Hierzu benötigen wir den Begriff der bedingten Entropie $H(X|Y)$ von X, wenn Y bereits bekannt ist.

Definition 77 (**bedingte Entropie**)

Seien X, Y Zufallsvariablen. Dann ist die **bedingte Entropie** von X unter Y definiert als

$$H(X|Y) = \sum_{y \in W(Y)} p(y) \cdot H(X|y),$$

wobei $X|y$ die Zufallsvariable mit der Verteilung $Prob[X|y = x] = p(x|y) = Prob[X = x \mid Y = y]$ ist (d.h. $H(X|y) = \sum_{x \in W(X)} p(x|y) \cdot \log_2(1/p(x|y))$).

Satz 78 $H(X, Y) = H(Y) + H(X|Y)$.

Beweis: s. Übungen. ■

Satz 79 $H(X, Y) \leq H(X) + H(Y)$, *wobei Gleichheit genau dann eintritt, wenn* X *und* Y *stochastisch unabhängig sind.*

Beweis: s. Übungen. ■

Korollar 80 $H(X|Y) \leq H(X)$, *wobei Gleichheit genau dann eintritt, wenn* X *und* Y *stochastisch unabhängig sind.*

Satz 81 *In jedem Kryptosystem gilt für die Klartextentropie* $H(X)$, *die Schlüsselentropie* $H(K)$ *und die Kryptotextentropie* $H(Y)$

$$H(K|Y) = H(K) + H(X) - H(Y).$$

Beweis: Zunächst ist $H(K|Y) = H(K, Y) - H(Y)$. Es reicht also zu zeigen, dass

$$H(K, Y) = H(K) + H(X)$$

ist. Da bei Kenntnis des Schlüssels der Wert von X bereits eindeutig durch Y und der Wert von Y eindeutig durch X festgelegt ist, folgt unter Berücksichtigung der gemachten Annahme, dass X und K unabhängig sind,

$$H(K, Y) = H(K, X, Y) = H(K, X) = H(K) + H(X).$$

■

Jetzt verfügen wir über alle Hilfsmittel, um die erwartete Anzahl

$$\bar{s}_n = \sum_{y \in C_n} p(y) \cdot \|K(y)\| - 1$$

unechter Schlüssel nach unten abschätzen zu können. Seien X_n und Y_n die Zufallsvariablen, die die Verteilungen der n-Gramme der Klartextsprache und der zugehörigen Kryptotexte beschreiben. Mit Satz 81 folgt

$$H(K|Y_n) = H(K) + H(X_n) - H(Y_n).$$

Die Klartextentropie $H(X_n)$ lässt sich durch

$$H(X_n) \geq nH(L) = n(1 - R(L)) \log_2 m$$

abschätzen, wobei $m = \|A\|$ ist. Zudem lässt sich die Kryptotextentropie $H(Y_n)$ wegen $W(Y_n) = C_n$ und $\|C_n\| = m^n$ durch

$$H(Y_n) \leq n \log_2 m$$

abschätzen. Somit ist

$$H(K|Y_n) = H(K) + \underbrace{H(X_n) - H(Y_n)}_{\geq -nR(L)\log_2 m}.$$

Andererseits gilt (unter Verwendung der Jensenschen Ungleichung)

$$\begin{aligned} H(K|Y_n) &= \sum_{y \in C_n} p(y) \cdot H(K|y) \\ &\leq \sum_{y \in C_n} p(y) \cdot \log_2 \|K(y)\| \\ &\leq \log_2 \sum_{y \in C_n} p(y) \cdot \|K(y)\| \\ &= \log_2(\bar{s}_n + 1). \end{aligned}$$

Zusammen ergibt sich also

$$\log_2(\bar{s}_n + 1) \geq H(K) - nR(L) \log_2 m.$$

Im Fall, dass der Schlüssel unter Gleichverteilung gezogen wird, erreicht $H(K)$ den maximalen Wert $\log_2 \|K\|$, was auf die gesuchte Abschätzung für $\bar{s}_n$ führt. Wir fassen zusammen.

Satz 82 *Werden mit einem Kryptosystem Klartexte $x \in A^n$ der Länge n mit einem unter Gleichverteilung gezogenen Schlüssel $k \in K$ verschlüsselt, und ist $\|C_n\| = \|A^n\| = m^n$ für den zugehörigen Kryptotextraum $C_n = \{E(k,x) \mid k \in K, x \in A^n\}$, so gilt für die erwartete Anzahl $\bar{s}_n$ der unechten Schlüssel,*

$$\bar{s}_n \geq \frac{\|K\|}{m^{nR(L)}} - 1.$$

Setzen wir in obiger Abschätzung $\bar{s}_n = 0$, so erhalten wir folgende untere Schranke für die Eindeutigkeitsdistanz n_0 des Kryptosystems:

$$n_0 \geq \frac{\log_2 \|K\|}{R(L)\log_2 m} = \frac{\log_2 \|K\|}{\log_2 m - H(L)} = \log_2 \|K\|/\mathcal{R}_{abs}(L).$$

Beispiel 83 *Für Substitutionen bei deutschsprachigem Klartext ergeben sich die folgenden Mindestwerte $\log_2 \|K\|/\mathcal{R}_{abs}(L)$ für die Eindeutigkeitsdistanz n_0 (wobei wir $\mathcal{R}_{abs}(L) = 3{,}2$ annehmen, was einer relativen Redundanz von $R(L) = 3,2/4,76 \approx 67\%$ entspricht):*

Kryptosystem	*Schlüsselanzahl* $\|K\|$	$\log_2 \|K\|$	$\log_2 \|K\|/\mathcal{R}_{abs}(L)$
Verschiebechiffre	26	4,7	$\frac{4{,}7}{3{,}2} \approx 1{,}5$
Affine Chiffre	$12 \cdot 26 = 312$	8,3	2,6
einfache Substitution	26!	88,4	27,6
Vigenère-Chiffre	26^d	$4{,}7 \cdot d$	$1{,}5 \cdot d$
DES-Algorithmus	2^{56}	56	17,5

Dagegen erhalten wir für Blocktranspositionen folgende Schätzwerte für die Menge an Kryptotext, die zur eindeutigen Bestimmung des Schlüssels benötigt wird:

Analyse auf Basis von	$\mathcal{R}_{abs}(L)$	*Blocklänge l* 10	20	50	100	1 000
n*-Grammen*	3, 20	*7*	*19*	*67*	*164*	*2 665*
Trigrammen	1, 15	*24*	*65*	*226*	*553*	*9 473*
Bigrammen	0, 90	*40*	*111*	*390*	*954*	*15 502*
Einzelzeichen	0, 66	*59*	*165*	*578*	*1415*	*22986*

Auch wenn die Schätzwerte für n_0 bei der Analyse auf Basis von Einzelzeichen endlich sind, ist in dieseen Fällen $n_0 = \infty$. Denn wie wir wissen, führt eine solche Analyse nicht zum Ziel, und zwar unabhängig davon, über wie viel Kryptotext der Gegner verfügt.

3.2 Komplexitätstheoretische Sicherheit

Wie wir gesehen haben, muss für die Benutzung eines informationstheoretisch sicheren Kryptosystems ein immenser Aufwand betrieben werden. Daher begnügt man sich in der Praxis meist mit schwächeren Sicherheitsanforderungen.

– Ein Kryptosystem gilt als **komplexitätstheoretisch sicher** oder als **berechnungssicher** (***computationally secure***), falls es dem Gegner nicht möglich ist, das System mit einem für ihn lohnenswerten Aufwand zu brechen. Das heißt, der Zeitaufwand und die Kosten für einen erfolgreichen Angriff übersteigen den potentiellen Nutzen bei weitem.

– Ein Kryptosystem gilt als **nachweisbar sicher** (***provably secure***), wenn seine Sicherheit mit bekannten komplexitätstheoretischen Hypothesen verknüpft werden kann, deren Gültigkeit gemeinhin akzeptiert ist.

– Als **praktisch sicher** (***practically secure***) werden dagegen Kryptosysteme eingestuft, die über mehrere Jahre hinweg jedem Versuch einer erfolgreichen Kryptoanalyse widerstehen konnten, obwohl sie bereits eine weite Verbereitung gefunden haben und allein schon deshalb ein lohnenswertes Ziel für einen Angriff darstellen.

Die komplexitätstheoretische Analyse eines Kryptosystems ist äußerst schwierig. Dies hängt damit zusammen, daß der Aufwand eines erfolgreichen Angriffs unabhängig von der vom Gegner angewandten Strategie abgeschätzt werden muss. Das heißt, es müssen nicht nur alle derzeit bekannten kryptoanalytischen Ansätze, sondern alle *möglichen* in Betracht gezogen werden. Dabei darf sich die Aufwandsanalyse nicht ausschließlich an einer vollständigen Rekonstruktion des Klartextes orientieren, da bereits ein geringfügiger Unterschied zwischen dem a posteriori und dem a priori Wissen für den Gegner einen Vorteil bedeuten kann.

Aus den genannten Gründen ist es bis heute noch für kein praktikables Kryptosystem gelungen, seine komplexitätstheoretische Sicherheit mathematisch zu beweisen. Damit ist auch nicht so schnell zu rechnen, zumindest nicht solange der Status fundamentaler komplexitätstheoretischer Fragen wie etwa des berühmten $\mathrm{P} \stackrel{?}{=} \mathrm{NP}$-Problems offen ist. Dagegen gibt es eine ganze Reihe praktikabler Kryptosysteme, die als nachweisbar sicher oder praktisch sicher gelten.

Wir schließen diesen Abschnitt mit einer Präzisierung des komplexitätstheoretischen Sicherheitsbegriffs. Hierzu ist es erforderlich, die Verletzung der Vertraulichkeit als ein algorithmisches Problem für den Gegner zu formulieren.

Definition 84 **(Vorteil eines Gegners)**

Sei $S = (M, C, E, D, K)$ ein Kryptosystem mit Schlüsselverteilung K. Ein Gegner ist ein Paar (G, V) von probabilistischen Algorithmen, wobei G zwei Klartexte $x_0 \neq x_1 \in M$ generiert und V bei Eingabe dieser Klartexte und eines Kryptotextes $y \in C$ ein Bit ausgibt. Der Vorteil von (G, V) ist

$$\alpha(G, V) = Prob[V(X_0, X_1, E(K, X_B)) = B] - 1/2,$$

wobei die Zufallsvariablen X_0, X_1 die von G generierte Ausgabe und B die zufällige Wahl eines Bits beschreibt, d.h. $Prob[B = 0] = Prob[B = 1] = 1/2$.

Satz 85 *Ein Kryptosystem ist unter einer Klartextverteilung $p(x) > 0$ für alle $x \in M$ genau dann absolut sicher, wenn kein Gegner mit einem Vorteil $\alpha(G, V) > 0$ existiert.*

Beweis: Falls ein Kryptosystem unter einer Klartextverteilung $p(x) > 0$ für alle $x \in M$ absolut sicher ist, so sind X und $E(K, X)$ für jede Klartextverteilung stochastisch unabhängig. Folglich ist

$$
\begin{aligned}
&Prob[V(X_0, X_1, E(k, X_B)) = B]\\
&= \underbrace{Prob[B = 0]}_{1/2} Prob[V(X_0, X_1, E(k, X_0)) = 0]\\
&\quad + \underbrace{Prob[B = 1]}_{1/2} \underbrace{Prob[V(X_0, X_1, E(k, X_1)) = 1]}_{\substack{= Prob[V(X_0, X_1, E(k, X)) = 1]\\ = Prob[V(X_0, X_1, E(k, X_0)) \neq 0]}}\\
&= 1/2
\end{aligned}
$$

Ist dagegen ein Kryptosystem nicht absolut sicher, so müssen ein Kryptotext $\hat{y}$ und zwei Klartexte x_0, x_1 mit $p(\hat{y}|x_0) > p(\hat{y}|x_1)$ existieren. Für die a priori Verteilung $p(x_0) = p(x_1) = 1/2$ ergibt sich dann wegen

$$
p(\hat{y}) = \underbrace{p(x_0)}_{1/2} p(\hat{y}|x_0) + \underbrace{p(x_1)}_{1/2} \underbrace{p(\hat{y}|x_1)}_{<p(\hat{y}|x_0)} < p(\hat{y}|x_0)
$$

eine a posteriori Wahrscheinlichkeit für x_0 von

$$
p(x_0|\hat{y}) = p(\hat{y}|x_0)p(x_0)/p(\hat{y}) = p(\hat{y}|x_0)/2p(\hat{y}) > 1/2.
$$

Folglich erzielt der Gegner (G, V), wobei G mit Wahrscheinlichkeit 1 das Paar (x_0, x_1) ausgibt und

$$
Prob[V(x_0, x_1, y) = 0] = \begin{cases} 1, & y = \hat{y},\\ 1/2, & y \neq \hat{y} \end{cases}
$$

ist, einen Vorteil von

$$
\begin{aligned}
\alpha(G, V) &= Prob[V(x_0, x_1, E(K, x_B)) = B]\\
&= \sum_{y \in C} p(y) \underbrace{Prob[V(x_0, x_1, E(K, x_B)) = B \mid E(K, x_B) = y]}_{= \begin{cases} p(x_0|\hat{y}), & y = \hat{y},\\ 1/2, & sonst \end{cases}}\\
&= (1 - p(\hat{y}))\frac{1}{2} + p(\hat{y}) \underbrace{p(x_0|\hat{y})}_{>1/2}\\
&> 1/2.
\end{aligned}
$$

■

Für die Präzisierung des komplexitätstheoretischen Sicherheitsbegriffs sind nun die beiden folgenden Fragen von entscheidender Bedeutung:

- Über welche Rechenressourcen kann ein Gegner realistischerweise verfügen?
- Wie groß darf der vom Gegner erzielte Vorteil höchstens sein, damit die Vertraulichkeit der Nachricht gewahrt bleibt?

Wir beantworten diese Fragen durch folgende Definitionen. Dabei gehen wir davon aus, dass die Sicherheit des Kryptosystems S von der Wahl eines prinzipiell beliebig groß wählbaren Sicherheitsparameters k abhängt. Alle legalen Operationen (wie die Schlüsselgenerierung oder die Chiffrierung) sollten effizient (d.h. in Zeit $k^{O(1)}$) durchführbar sein. Typischerweise werden Kryptosysteme nach der Schlüssellänge parameterisiert, wobei dann nur Klartexte der Länge $k^{O(1)}$ effizient verschlüsselt werden können.

Definition 86 (**komplexitätstheoretisch sicher**)

Sei S ein Kryptosystem mit Sicherheitsparameter k und bezeichne S_k das mit festem Parameterwert k betriebene Kryptosystem.

- Ein Gegner (G, V) für S heißt **effizient**, wenn sowohl G als auch V durch probabilistische Schaltkreise der Größe $k^{O(1)}$ berechenbar sind.
- Sei $\alpha(k)$ der von (G, V) gegenüber S_k erzielte Vorteil. $\alpha(k)$ heißt **vernachlässigbar**, wenn für jedes Polynom p eine Zahl n existiert, so dass $\alpha(k) < 1/p(k)$ für alle $k \geq n$ ist.
- S heißt **komplexitätstheoretisch sicher**, wenn der Vorteil jedes effizienten Gegners vernachlässigbar ist.