wohl nur ein „D“ stehen und so weiter. Der Rest ist ein reines Puzzlespiel und führt zum Klartext

> *Voyl ging also ins Hospital. Man schafft ihn in’n Saal mit achtundzwanzig Schlafcouchs, …zwanzig davon mit Inhalt, wo krank ist und sogar halbtot. Man gibt ihm Tranquilitantia mit Mordswirkung (Largactyl, Procalmadiol, Atarax). Am Tag sah Anton Voyl ’n Big Boss, wo’n Gang macht von Saal zu Saal; mit ihm Hofstaat und Tross vom Hilfsarzt bis zum Arztstift; Big Boss trinkt Milch und spricht dazu, grinst und lacht manchmal … Dann und wann ging Big Boss bis zur Schlafstatt von’m Mann, wo todkrank war und aufm ultimo Loch pfiff …*

Der Text stammt aus der deutschen Übersetzung des Romans *La disparaition* von GEORGE PEREC *(Edition Denoël, Paris,* 1969*)*. Dieser französische Autor zahlreicher Romane, hatte nach einem feuchtfröhlichen Abend gewettet, daß er einen Roman schreiben könne, ohne den Buchstaben „E“ zu verwenden; er gewann. Die deutsche Übersetzung von EUGEN HELMLÉ mit dem Titel *Anton Voyls Fortgang (Zweitausendeins,* [3]1998*)* hält sich auch an die Vorgabe, verwendet aber Umlaute – daher die Auslassungspunkte im obigen Text.

Wie man sieht, lassen sich die einzelnen Buchstaben also selbst dann noch recht gut anhand ihrer Umgebung identifizieren, wenn ihre Rangfolge nicht mehr ganz der für deutsche Texte üblichen entspricht.

Eine hier nicht behandelte Möglichkeit, sich Buchstaben zu verschaffen, liegt auch im Erraten von Wörtern aus der Botschaft. Da man beim Abhören meist weiß, wonach man sucht, gibt es oft Wörter, die mit hoher Wahrscheinlichkeit im Text vorkommen sollten. Falls ein solches Wort einen oder gar mehrere Buchstaben mehrfach enthält, kann man im Chiffretext nach demselben Muster suchen. Ein ideales Wort wäre beispielsweise *Substitution:* Es muß von jeder monoalphabetischen Substitution in eine Folge der Art

$$XYpXZWZYZWqr$$

transformiert werden, und eine Folge dieser Art kann praktisch nur für das Wort *Substitution* stehen. Damit sind sieben Buchstaben identifiziert.

Völlig unsicher sind monoalphabetische Substitutionen (wie auch VIGENÈRE- oder gar CAESAR-Chiffren natürlich gegen Attacken mit bekanntem Klartext: Sobald ein auch nur kurzer Teil des Chiffretexts einem Klartext zugeordnet werden kann, hat man den Schlüssel praktisch gefunden.

§ 6: Permutationschiffren

Bei monoalphabetischen Substitutionen werden die Buchstaben des Alphabets permutiert; bei den hier betrachteten Permutationschiffren dagegen werden Permutationen angewandt auf die Reihenfolge der Buchstaben im Text: Der Text wird aufgeteilt in Blöcke einer vorgegebenen Länge N, wobei ein eventuell unvollständiger letzter Block durch sogenannte „Nullen“, d.h. offensichtlich sinnlose Buchstaben, aufgefüllt wird. Sodann werden in jeden dieser Blöcke die Buchstaben gemäß einer festen Regel permutiert. Schlüssel ist hier also die Zahl N zusammen mit einer der $N!$ möglichen Permutationen von N Elementen. Permutationschiffren waren beispielsweise Teil des Verschlüsselungsalgorithmus der deutschen Wehrmacht im ersten Weltkrieg.

Der Kryptanalytiker kann Permutationschiffren leicht erkennen: Die Verteilung der 26 Buchstaben ist natürlich identisch mit der im Klartext. Bezüglich der Blocklänge N gibt es immerhin den Hinweis, daß sie ein Teiler der Nachrichtenlänge sein muß – es sei denn, der Anwender war intelligent genug, ans Ende noch einige „Nullen“ anzuhängen.

Zur Identifikation der Permutation (und zum Ausschließen falscher Blocklängen) arbeitet man mit charakteristischen Buchstabenpaaren: Im Deutschen etwa folgt auf ein „C“ fast immer ein „H“ und auf ein „Q“ ein „U“. Wenn also in einem Block „C“ und „H“ oder „Q“ und „U“ vorkommen, kann man fast sicher sein, daß die entsprechenden Spalten benachbart sein müssen. Andere häufige Buchstabenpaare wie „ND“ oder „BE“ geben zumindest Anhaltspunkte für Versuche, und natürlich hilft dem Kryptanalytiker vor allem sein Sprachgefühl: Da jede Identifikation eines Worts Konsequenzen in allen anderen Zeilen hat, ist die Kryptanalyse bei hinreichend langen Chiffretexten nicht sonderlich schwer. Sehr leicht ist sie bei bekanntem Klartext am Anfang oder Ende, denn wenn der Klartext zu nur einer Zeile bekannt ist, kennt man den Schlüssel. Aus diesem Grund wurden Permutationschiffren praktisch nie für sich allein angewandt, sondern nur als Teil eines zwei- oder mehrstufigen Verschlüsselungsverfahrens.

§ 7: Polyalphabetische Substitutionen

Als nächste Stufe der Komplexität bietet sich natürlich an, aufeinanderfolgende Buchstaben mit verschiedenen Permutationen zu verschlüsseln, d.h. also eine Art von VIGENÈRE-Verfahren mit allgemeinen monoalphabetischen Substitutionen anstelle der doch sehr speziellen CAESAR-Substitutionen.

Leider steigt dadurch, sofern die Anzahl der Permutationen nicht in die Größenordnung der Nachrichtenlänge kommt, die Sicherheit nur unwesentlich an, denn die Verfahren aus dem vorigen Paragraphen funktionieren auch für diesen Fall: Man stellt zunächst, genau wie bei den VIGENÈRE-Chiffren, die Periode fest; danach weiß man, welche Buchstaben mit derselben Permutation verschlüsselt wurden. Jede dieser Buchstabenmengen wird nach Häufigkeit geordnet, und man kann wieder Kontaktdiagramme aufstellen. Dabei muß nur beachtet werden, daß die rechten und linken Nachbarn nicht aus derselben Gruppe kommen, sondern aus je einer anderen festen Gruppe, so daß Balken nach oben und nach unten nicht mehr für denselben Buchstaben stehen müssen. Abgesehen von dieser kleinen Erschwernis ändert sich nichts, und ein erfahrener Kryptanalytiker kann ab etwa 30–40 Buchstaben pro Alphabet praktisch sicher sein, daß er die Nachricht entschlüsseln kann.

Polyalphabetische Chiffren mit sehr großen Periodenlängen erzeugten die sogenannten *Rotormaschinen,* die in den Zwanzigerjahren aufkamen und die Kryptographie des zweiten Weltkriegs dominierten. Ein Rotor ist eine Scheibe, die auf jeder Seite in gleichmäßigen Abständen 26 Kontakte hat. Jeder Kontakt auf der Vorderseite ist mit einem der Kontakte auf der Rückseite verbunden, wobei nur darauf zu achten ist, daß diese Verbindungen spiegelsymmetrisch angebracht werden, so daß die so realisierte Permutation ein Produkt von dreizehn Transpositionen ist.

Zum Einsatz der Maschine wählt man, je nach Maschine, drei bis fünf von mehreren verfügbaren Rotoren aus und setzt diese in einer festzulegenden Reihenfolge und Anfangsstellung in die Maschine ein. Der erste Buchstabe wird dann von einem festen Kontakt auf der einen Seite der Maschine durch die Rotoren in dieser Stellung an einen der 26 festen Kontakte auf der anderen Seite der Maschine übertragen; das Chiffrierergebnis ist der diesem Kontakt zugeordnete Buchstabe. Danach drehen sich einer oder mehrere der Rotoren um eine Position weiter. Bei den ersten Rotormaschinen wie etwa der von HEBERN geschah dies in einer völlig regelmäßigen Weise: Der fünfte Rotor bewegte sich nach jedem Buchstaben um eins weiter, der erste nach je 26 und der dritte nach je 676 Buchstaben. Der zweite und der vierte Rotor änderten ihre Position nicht. Bei anderen Maschinen wie der Enigma wurde die Bewegung der Rotoren durch unregelmäßig auf den einzelnen Rotoren angeordneten Haken realisiert und war damit auch von der Rotorkonfiguration abhängig.

Zur Entschlüsselung wird der Chiffretext einfach in eine identisch aufgebaute Maschine mit gleicher Anfangsstellung aber umgekehrter Rotorreihenfolge gegeben.

Bei der Enigma, im Gegensatz zu anderen Rotormaschinen, war der letzte Rotor fest und leitet den Strom, nach einer weiteren Permutation, zurück durch die beweglichen Rotoren. Dadurch kann die Entschlüsselung mit derselben Rotorenkonfiguration erfolgen.

Im WWW sind verschiedene Simulatoren sowohl der vielen deutschen Enigma-Modelle als auch anderer Rotor-Maschinen zu finden; eine Enigma der deutschen Abwehr mit vier Rotoren findet man beispielsweise unter

`http://home.caiway.nl/~antonh/enigma_ga.html`

Die mit solchen Maschinen erzielbaren Perioden liegen im Bereich über 100 000; trotzdem wurden sie schon im zweiten Weltkrieg häufig geknackt.

Ein Grund dafür ist, daß während der Gültigkeitsperiode eines Schlüssels alle Nachrichten mit derselben Anfangsstellung verschlüsselt werden müssen (oder aber die Anfangsstellung auf eine kryptographisch unsichere Weise übermittelt werden muß, was sich gerade bei der deutschen Enigma als noch schlimmer herausstellte), so daß damit zumindest für die ersten Buchstaben relativ viele Chiffrebuchstaben zusammenkommen. Das weitere Problem, daraus die Anfangsstellung der Maschine zu ermitteln, lösten die Engländer im zweiten Weltkrieg mit speziell gebauten Maschinen, den sogenannten „Bomben“, die mechanisch alle möglichen Anfangsstellungen durchprobierten. Die Verdrahtung der einzelnen Rotoren war schon vor dem Krieg von polnischen Kryptanalytikern geknackt worden, und während des gesamten zweiten Weltkriegs wurde nie ein Rotor aus dem Verkehr gezogen. Zwar wurden in manchen Netzwerken wie etwa bei den U-Booten ein oder zwei neue Rotoren eingeführt, aber da diese zusammen mit den alten benutzt wurden, hatten die Kryptanalytiker dann nur noch das Problem, die Verdrahtung *eines* unbekannten Rotors zu ermitteln, was angesichts des großen Nachrichtenvolumens stets gelang.

Seit etwa 1970 werden Rotormaschinen nicht mehr eingesetzt – außer in UNIX. Das dortige `crypt(1)`-Kommando simuliert eine Rotormaschine mit *einem* Rotor, der 256 Ein- und Ausgänge hat. Sehr sicher ist das nichts; selbst in der `man`-Seite dazu heißt es *Methods of attack on such machines are widely known; thus crypt provides minimal security.* Für Paßwörter verwendet UNIX daher auch nicht dieses Kommando, sondern das auf einem „gesalzenen“ DES beruhende deutlich bessere `crypt(3)`-Unterprogramm. In LINUX ist `crypt(1)` inzwischen aufgegangen in einem universelleren Kommando `mcrypt`, das sich zwar mit Option `--enigma` genauso verhält wie `crypt(1)`, das aber auch eine ganze Reihe von Algorithmen anbietet, die nach heutigen Standards wirklich sicher sind. Von den Verfahren, mit denen wir uns später beschäftigen werden, sind unter anderem Triple DES und Rijndael implementiert.

§8: Literaturhinweise

Das (dickleibige) Standardwerk zur alten Kryptographie mit Schwerpunkt auf der geschichtlichen Darstellung ist

DAVID KAHN: *The Codebreakers – the comprehensive history of secret communication from ancient time to the internet,* Scribner, New York, ^{2}1996

Abgesehen von einem Anhang über public key Kryptographie ist das Buch weitgehend identisch mit der ersten Auflage von 1967, die für zahlreiche ältere Kryptologen der Einstieg in ihr Arbeitsgebiet war.

Deutlich kürzer und billiger und verfahrensorientierter ist

HELEN FOUCHÉ GAINES: *Cryptanalysis – a study of ciphers and their solution,* Dover, New York, 1956 *(Originalausgabe 1939)*

L. SACCO: *Manuel de cryptographie,* Payot, Paris, 1951

beschreibt die klassischen Verfahren und ihre Kryptanalyse aus seiner Sicht als Chef des Chiffrierdienstes der italienischen Armee. Eine Reihe entsprechender Bücher gibt es auch von WILLIAM FRIEDMAN, jedoch sind diese in Deutschland wenn überhaupt dann nur schwer zu finden. Rotormaschinen und ihrer Kryptanalyse gewidmet ist das Buch

CIPHER A. DEAVOURS, LOUIS KRUH: *Machine cryptography and modern cryptanalysis,* Artech House, Dedham MA, 1985

Das Wort „modern“ bezeichnet hier den Zeitraum von etwa 1920–1970. Weitere Informationen zur Kryptanalyse von Rotormaschinen im zweiten Weltkrieg findet man im Buch von KAHN sowie bei

BENGT BECKMAN: *Codebreakers – Arne Beurling and the Swedish crypto program during Word War II,* American Mathematical Society, Providence, R.I., 2002

Ein wirklich modernes Lehrbuch mit Kapitel über klassische Kryptographie sowie über statistische und informationstheoretische Ansätze zur Kryptanalyse ist

JAN C.A. VAN DER LUBBE: *Basic Methods of cryptography,* Cambridge University Press, 1998

Kapitel 2
Informationstheoretische Ansätze

Im vorigen Kapitel haben wir anhand einiger Beispiele gesehen, wie die statistischen Eigenschaften eines Kryptogramms zu Informationen über den Klartext und/oder den Schlüssel führen können. Dabei wurde klar, daß gewisse Verfahren trotz auf den ersten Blick gewaltiger Komplexität bereits für relativ kurze Nachrichten unsicher sind.

Bei der Wahl eines Kryptosystems interessiert vor allem die umgekehrte Frage: Wie *sicher* ist das System? Es hieße, den Gegner möglicherweise gewaltig zu unterschätzen, wenn man dabei nur die Sicherheit gegenüber den Angriffsmöglichkeiten betrachtet, die man selbst kennt: Ein ernstzunehmender Gegner wird häufig über Methoden verfügen, die nicht in der offenen Literatur dokumentiert sind.

In diesem Kapitel wollen wir daher untersuchen, wieviel Information über Schlüssel und Klartext in einem Kryptogramm enthalten ist – unabhängig davon, ob ein Gegner in der Lage ist, diese Information mit realistischem Aufwand nutzbar zu machen.

Man kann das auch so formulieren, daß wir den Gegner deutlich *über*schätzen, indem wir annehmen, daß ihm unbegrenzte Resourcen zur Verfügung stehen; er kann also *jeden* endlichen Algorithmus ausführen, unabhängig von physikalischen und sonstigen Einschränkungen. Insbesondere könnte er also eine Substitutionschiffre einfach dadurch lösen, daß er alle 26! Möglichkeiten ausprobiert; sofern dabei in genau einem Fall sinnvoller Klartext entsteht, hat er die Lösung gefunden. Es ist klar, daß von einem realen Gegner keine größere Gefahr ausgehen kann, als von diesem idealisierten Gegner; ein Kryptosystem ist also sicher, wenn es gegen einen solchen Gegner sicher ist.

In der Kryptologie bezeichnet man diesen idealisierten Gegner als den BAYESschen Gegner; sein Name ist abgeleitet von dem englischen Theologen THOMAS BAYES, der als erster das Entscheidungsprinzip formulierte, wonach unter mehreren möglichen Hypothesen diejenige als wahr anzusehen sei, die vor dem Hintergrund seines vorhandenen Wissens die größte Wahrscheinlichkeit hat; auf die Kryptanalyse angewandt bedeutet dies, daß man sich unter allen möglichen Schlüsseln für den entscheidet, der angesichts der Information, die der Chiffretext sowie das Vorwissen über die Quelle bieten die größte Wahrscheinlichkeit hat.

THOMAS BAYES (1702–1761) wurde in London geboren als ältestes von sieben Kindern eines der ersten nonkonformistischen Pastoren Englands. Da die englischen Universitäten Oxford und Cambridge keine Nonkonformisten akzeptieren, mußte er zum Studium 1719 nach Schottland an die Universität Edinburgh, wo er sich für Logik und Theologie immatrikulierte. Nach seinen späteren Äußerungen muß er sich auch bereits damals oder kurz danach mit Mathematik beschäftigt haben. Wie sein Vater wurde er Geistlicher; seine mathematischen Arbeiten, z.B. über die Grundlagen der Analysis, erschienen zu seinen Lebzeiten nur anonym. Trotzdem wurde er 1742 fellow der Royal Society, die 1764 auch posthum seinen *Essay towards solving a problem in the doctrine of chances* veröffentlichte.

§1: Die Entropie einer Quelle

Wenn wir das BAYESsche Prinzip in einer mathematisch fundierten Weise auf die Kryptographie anwenden wollen, müssen wir uns zunächst überlegen, wie wir die Information, die in einer Nachricht steckt, quantifizieren können. Tatsächlich brauchen wir sogar etwas mehr: Da wir bei der Entschlüsselung sowohl den Chiffretext haben als auch Informationen über den Klartext, müssen wir auch dir Information quantifizieren können, die uns der Chiffretext *vor dem Hintergrund unseres Wissens über die Quelle* gibt.

Wir betrachten Nachrichten als Folgen von „Buchstaben", wobei es sich hier keinesfalls – wie im Fall der klassischen Kryptoverfahren – um die Buchstaben von „A" bis „Z" handeln muß, sondern um Elemente irgendeiner endlichen Menge, die wir in Analogie zum klassischen Fall als „Alphabet“ bezeichnen. Ein in der Informationsverarbeitung wichtiges Alphabet ist beispielsweise die Menge $\{0, 1\}$, aber auch die Menge aller ASCII-Zeichen. Andere Beispiele sind Blöcke aus z.B. 32 ASCII-Zeichen, Tripel von Buchstaben oder um die Codewörter eines der alten Telegrammcodes für Geschäftsleute.

Wir wollen annehmen, daß jeder Buchstabe x_i aus dem Alphabet A eine bekannte Wahrscheinlichkeit p_i hat, wie wir sie im letzten Kapitel beispielsweise für $A = \{„A“, „B“, \dots, „Z“\}$ durch Auszählen geschätzt haben. Gegen Ende des Kapitels werden wir diese Bedingung etwas formaler betrachten und daraus allgemeine Sätze über die Sicherheit von Codes ableiten.

Zunächst aber wollen wir uns überlegen, wieviel Information uns ein Buchstabe der betrachteten Nachrichtenquelle liefert. Diese Information soll quantifiziert werden durch eine Funktion H, die sogenannte Entropie.

Da der Informationsgehalt einer Nachricht sicherlich nicht von den Namen der Buchstaben abhängt, können wir H einfach als Funktion der Buchstabenwahrscheinlichkeiten p_i betrachten; wir suchen also für jede natürliche Zahl n eine Funktion

$$H(p_1, \dots, p_n) \quad \text{für} \quad 0 \le p_i \le 1 \quad \text{und} \quad \sum_{i=1}^{n} p_i = 1\,.$$

Kurz, wenn auch etwas schlampig, werden wir gelegentlich auch einfach $H(A)$ schreiben, wobei A eine Quelle bezeichnet, d.h. ein Alphabet *mit* gegebenen Buchstabenwahrscheinlichkeiten (und gegebenenfalls auch Kontaktwahrscheinlichkeiten *usw.*).

Um geeignete Kandidaten für H zu finden, wollen wir zunächst einige Eigenschaften notieren, die eine solche Familie von Funktionen haben sollte.

Am wenigsten Information, nämlich überhaupt keine, liefert uns eine Quelle, die immer denselben Buchstaben aussendet; falls also ein $p_i = 1$ ist und alle anderen p_j verschwinden, sollte $H = 0$ sein. Im anderen Extrem, wenn alle Buchstaben gleich wahrscheinlich sind, ist unsere

Ungewißheit über den jeweils nächsten Buchstaben am größten. Daher liegt es nahe, zu fordern, daß stets

$$H(p_1, \dots, p_n) \le H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \tag{1}$$

sein soll.

Ebenfalls ziemlich offensichtlich ist folgende Bedingung: Angenommen, wie erweitern unser Alphabet $A = \{x_1, \dots, x_n\}$ um einen weiteren Buchstaben x_{n+1}, der nie vorkommt, d.h. $p_{n+1} = 0$ und alle anderen p_i bleiben unverändert. Es ist klar, daß wir durch eine solche formale Manipulation am Modell weder Information gewinnen noch verlieren können; deshalb sollte gelten

$$H(p_1, \dots, p_n) = H(p_1, \dots, p_n, 0) \tag{2}$$

und entsprechend auch, wenn die Null an einer anderer Position steht.

Schließlich sollten wir wir nicht aus den Augen verlieren, daß wir es in der Kryptographie üblicherweise mit mehreren, voneinander abhängigen Nachrichtenquellen zu tun haben, typischerweise einer Quelle von Klartext, einer Schlüsselquelle und einer Quelle von Chiffretext.

Wenn wir ein Klartextalphabet $A = \{x_1, \dots, x_n\}$ und ein Chiffretextalphabet $B = \{y_1, \dots, y_m\}$ haben mit Wahrscheinlichkeiten p_i für x_i und q_j für y_j, so wissen wir, daß der Buchstabe y_j insgesamt mit Wahrscheinlichlichkeit q_j auftritt. Falls wir allerdings bereits wissen, daß es sich hier um die Verschlüsselung des Klartextbuchstabens x_i handelt, ändern sich möglicherweise die Wahrscheinlichkeiten der verschiedenen y_j – auch wenn dies bei einem idealen Kryptosystem nicht der Fall sein sollte.

Wir bezeichnen die neuen Wahrscheinlichkeiten mit $q_j(x_i)$ und die Information

$$H_{x_i}(q_1, \dots, q_m) \underset{\text{def}}{=} H\big(q_1(x_i), \dots, q_m(x_i)\big)$$

als *bedingte Information* unter der Voraussetzung des Ereignisses (Klartextbuchstabens) x_i. Die bedingte Information unter der Voraussetzung, daß wir den Klartextbuchstaben kennen, ist der Mittelwert über alle H_{x_i}, d.h.

$$H_A(q_1,\dots,q_m) \underset{\text{def}}{=} \sum_{i=1}^{n} p_i H_{x_i}(q_1,\dots,q_m)\,.$$

Wenn wir nun Klartext und Chiffretext oder sonst zwei Quellen gemeinsam betrachten, können wir dies formal beschreiben durch das Alphabet $A \times B$, dessen „Buchstabe" (x_i, y_j) bedeuten soll, daß die erste Quelle x_i und die zweite y_j aussendet. Die Wahrscheinlichkeit hierfür sei p_{ij}.

Für ein vernünftiges Maß der Information muß dann gelten

$$H(\dots,p_{ij},\dots) = H(p_1,\dots,p_n) + H_A(q_1,\dots,q_m)\,, \tag{3}$$

denn die Gesamtinformation ändert sich natürlich nicht, wenn wir sie aufteilen in die Information, die uns die erste Quelle liefert, und die Information, die uns die zweite liefert, *nachdem* wir die der ersten kennen.

Durch die Bedingungen (1) – (3) ist die Funktion H im wesentlichen festgelegt:

Satz: Ist H stetig in allen seinen Argumenten und erfüllt die Bedingungen (1) – (3), so gibt es eine reelle Zahl $a > 1$ derart, daß gilt

$$H(p_1,\dots,p_n) = -\sum_{i=1}^{n} p_i \log_a p_i\,.$$

Insbesondere ist H also bis auf einen positiven Faktor eindeutig bestimmt.

Beweis: Wir betrachten zunächst die Funktion

$$L(n) \underset{\text{def}}{=} H\left(\frac{1}{n},\dots,\frac{1}{n}\right)\,.$$

Nach (2) und (1) ist

$$\begin{aligned} L(n) = H\left(\frac{1}{n},\dots,\frac{1}{n}\right) &= H\left(\frac{1}{n},\dots,\frac{1}{n},0\right) \\ &\le H\left(\frac{1}{n+1},\dots,\frac{1}{n+1}\right) = L(n+1)\,, \end{aligned}$$

die Funktion ist also monoton steigend.

Als nächstes betrachten wir q voneinander unabhängige Quellen A_1 bis A_q, die jeweils r verschiedene Buchstaben mit gleicher Wahrscheinlichkeit $1/r$ liefern. Das Produkt $A_1 \times \cdots \times A_q$ enthält dann r^q Tupel, die allesamt mit derselben Wahrscheinlichkeit $1/r^q$ auftreten; die Gesamtinformation ist also

$$H\left(\frac{1}{r^q},\dots,\frac{1}{r^q}\right) = L(r^q)\,.$$

Alternativ läßt sich dies auch induktiv nach (3) berechnen, indem wir zunächst die Information der ersten $q-1$ Buchstaben betrachten, d.h.

$$H\left(\frac{1}{r^{q-1}},\dots,\frac{1}{r^{q-1}}\right) = L(r^{q-1})\,,$$

und dann die bedingte Information der letzten Quelle betrachten unter der Voraussetzung, daß wir die $q-1$ Buchstaben der anderen bereits kennen.

Da die A_j als voneinander unabhängig vorausgesetzt wurden, liefern uns die ersten $q-1$ Quellen keinerlei Information über die q-te: Das ist in der Stochastik die Definition der Unabhängigkeit. Somit ist die bedingte Entropie $H_{A_1\times\cdots\times A_{q-1}}$ gleich der gewöhnlichen, d.h. nach (3) gilt

$$L(r^q) = L(r^{q-1}) + H\left(\frac{1}{r},\dots,\frac{1}{r}\right) = L(r^{q-1}) + L(r)\,.$$

Durch Induktion nach q folgt:

$$L(r^q) = qL(r) \quad \text{für alle } r,q \in \mathbb{N}\,. \tag{4}$$

Nun betrachten wir natürliche Zahlen r, s, p, q mit $r^q \le s^p \le s^{q+1}$; dann ist

$$q\log r \le p\log s \le (q+1)\log r \qquad \text{oder} \qquad \frac{q}{p} \le \frac{\log s}{\log r} \le \frac{q+1}{p}\,.$$

Wegen der Monotonie von L folgt auch

$L(r^q) \le L(s^p) \le L(r^{q+1})$, also nach (4) $qL(r) \le pL(s) \le (q+1)L(r)$.

Division durch $pL(r)$ macht daraus

$$\frac{q}{p} \leq \frac{L(s)}{L(r)} \leq \frac{q+1}{p}\,,$$

$L(s)/L(r)$ liegt daher im selben Intervall wie $\log s/\log r$, so daß

$$\left|\frac{L(s)}{L(r)} - \frac{\log s}{\log r}\right| \leq \frac{1}{p}$$

sein muß. Da p beliebig groß gewählt werden kann, folgt

$$\frac{L(s)}{L(r)} = \frac{\log s}{\log r} \qquad \text{oder} \qquad L(s) = \frac{L(r)}{\log r} \cdot \log s\,.$$

Damit ist also $L(s)$ proportional zu einem Logarithmus, wobei die Proportionalitätskonstante $L(r)/\log r$ wegen der Monotonie sowohl von L als auch des Logarithmus positiv sein muß. Mithin gibt es eine reelle Zahl $a > 1$ mit $L(n) = \log_a n$ für alle $n \in \mathbb{N}$.

Im speziellen Fall von Quellen, die alle Buchstaben mit gleicher Wahrscheinlichkeit ausgeben, ist der Satz damit bewiesen. Für allgemeine Quellen genügt es, wenn wir uns auf den Fall beschränken, daß alle Wahrscheinlichkeiten p_i rational sind; denn da H in allen seines Argumenten stetig sein soll, ist es durch seine rationalen Werte eindeutig bestimmt.

Wir betrachten also ein Alphabet $A = \{x_1, \ldots, x_n\}$ aus n Buchstaben, deren i-ter die Wahrscheinlichkeit $p_i = g_i/g$ habe mit $g_i \in \mathbb{N}$ und $\sum g_i = g$. Weiter betrachten wir ein Alphabet B aus g Buchstaben $y_1, \ldots, y_g$ mit Wahrscheinlichkeiten $q_1, \ldots, q_g$. Dessen Buchstaben verteilen wir auf n disjunkte Teilmengen $B_i \subseteq B$ derart, daß B_i aus g_i Buchstaben besteht. Die Ereignisse aus B sollen wie folgt von denen aus A abhängen: Wenn A den Buchstaben x_i liefert, liefert B mit jeweils gleicher Wahrscheinlichkeit $1/g_i$ einen der Buchstaben aus der Teilmenge B_i. Durch mehrfache Anwendung von (2) folgt

$$H_{x_i}(q_1, \ldots, q_g) = H\left(\frac{1}{g_i}, \ldots, \frac{1}{g_i}\right) = L(g_i) = \log_a g_i$$

und

$$\begin{aligned} H_A(q_1, \ldots, q_g) &= \sum_{i=1}^{n} p_i \log_a g_i = \sum_{i=1}^{n} p_i \log_a(gp_i) \\ &= \sum_{i=1}^{n} p_i \log_a p_i + \sum_{i=1}^{n} p_i \log_a g \\ &= \sum_{i=1}^{n} p_i \log_a p_i + \log_a g\,. \end{aligned}$$

Nun betrachten wir das Produkt $A \times B$. Der Buchstabe (x_i, y_j) daraus kann nur auftreten, wenn y_j in B_i liegt; nur g der nq Buchstaben aus $A \times B$ können also mit einer von Null verschiedenen Wahrscheinlichkeit auftreten. Da x_i die Wahrscheinlichkeit g_i/g hat und beim Auftreten von x_i der Buchstabe $y_j \in B_i$ mit Wahrscheinlichkeit $1/g_i$ gewählt wird, hat jedes der g möglichen Paare (x_i, y_j) dieselbe Wahrscheinlichkeit $1/g$. Also ist die Entropie der Quelle $A \times B$ gleich $L(g)$, und wegen (3) folgt

$$\begin{aligned} H(p_1, \ldots, p_n) &= L(g) - H_A(q_1, \ldots, q_g) \\ &= \log_a g - \left(\sum_{i=1}^{n} p_i \log_a p_i + \log_a g\right) \\ &= -\sum_{i=1}^{n} p_i \log_a p_i\,, \end{aligned}$$

wie behauptet. ■

Damit sind wir allerdings noch nicht ganz fertig: Zwar wissen wir nun, daß jede Funktion, die den Bedingungen (1)–(3) genügt, die angegebene Form haben muß, wir wissen aber noch nicht, ob es wirklich solche Funktionen gibt. Mit anderen Worten: Wir müssen noch nachprüfen, ob

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_a p_i$$

wirklich die Bedingungen (1) – (3) erfüllt. (2) und (3) sind klar, problematisch dagegen ist die Bedingung (1). Das folgende Lemma zeigt, daß auch sie erfüllt ist:

Lemma: Für m Zahlen $p_1, \dots, p_m \in [0,1]$ mit $\sum_{i=1}^{m} p_i = 1$ gilt stets

$$0 \le -\sum_{i=1}^{m} p_i \log_a p_i \le \log m\,;$$

dabei steht rechts genau dann ein Gleichheitszeichen, wenn alle p_i gleich $1/m$ sind.

Der *Beweis* erfolgt durch vollständige Induktion nach m.

Der Fall $m = 1$ ist trivial, denn dann muß $p_1 = 1$ sein, und die Behauptung gilt, da sowohl $\log_a p_1$ also auch $\log_a m$ verschwinden.

Für $m > 1$ können wir uns sofort auf den Fall $m = 1$ berufen, falls eine der Zahlen p_i gleich eins sein sollte. Interessant ist daher nur der Fall, daß kein p_i den Wert eins hat; insbesondere ist $p_m \ne 1$.

Wir betrachten die Summe $-\sum_{i=1}^{m} p_i \log_a p_i$ also unter der Nebenbedingung $p_m = q < 1$. Dann ist

$$\sum_{i=1}^{m-1} p_i = 1 - q\,,$$

und für $p_i^* = \dfrac{p_i}{1-q}$ ist $\sum_{i=1}^{m-1} p_i^* = 1$.

Daher können wir die Induktionsannahme auf die p_i^* anwenden und erhalten

$$0 \le -\sum_{i=1}^{m-1} p_i^* \log_a p_i^* \le \log_a(m-1)$$

mit Gleichheitszeichen rechts genau dann, wenn alle $p_i^* = \frac{1}{m-1}$ sind. Dies müssen wir mit $\sum_{i=1}^{m} p_i \log_a p_i$ in Verbindung bringen. Setzen wir

also die Definition von p_i^* ein:

$$\begin{aligned}
\sum_{i=1}^{m-1} p_i^* \log_a p_i^* &= \frac{1}{1-q} \sum_{i=1}^{m-1} p_i\big((\log_a p_i - \log_a(1-q)\big) \\
&= \frac{1}{1-q} \sum_{i=1}^{m-1} p_i \log_a p_i - \frac{1}{1-q} \sum_{i=1}^{m-1} p_i \log_a(1-q) \\
&= \frac{1}{1-q} \sum_{i=1}^{m-1} p_i \log_a p_i - \frac{1}{1-q}(1-q)\log_a(1-q) \\
&= \frac{1}{1-q} \sum_{i=1}^{m-1} p_i \log_a p_i - \log_a(1-q)\,.
\end{aligned}$$

Somit ist die hier interessierende Summe

$$-\sum_{i=1}^{n} p_i \log p_i = -(1-q) \sum_{i=1}^{m-1} p_i^* \log_a p_i^* - (1-q)\log_a(1-q) - p_m \log_a p_m\,.$$

Die negative Summe über die ersten $m-1$ der p_i^* wird nach Induktionsannahme genau dann maximal, wenn alle $p_i^* = \frac{1}{m-1}$ sind. Außerdem ist $p_m = q$, der gesamte Ausdruck kann also höchstens gleich

$$(1-q)\log_a(m-1) - (1-q)\log_a(1-q) - q\log_a q$$

sein. Diese Funktion müssen wir in Abhängigkeit von q maximieren. Ihre Ableitung nach q ist

$$\begin{aligned}
&-\log_a(m-1) + \log_a(1-q) + \frac{1-q}{1-q} - \log_a q - \frac{q}{q} \\
= &-\log_a(m-1) + \log_a(1-q) - \log_a q = \log_a \frac{1-q}{(m-1)q}\,,
\end{aligned}$$

und dies wird genau dann gleich null, wenn der Bruch eins ist, d.h. für

$$1 - q = (m-1)q \qquad \text{oder} \qquad q = p_m = \frac{1}{m}\,.$$

Für $i < m$ erhalten wir

$$p_i = (1-q)p_i^* = \left(1 - \frac{1}{m}\right) \cdot \frac{1}{m-1} = \frac{m-1}{m} \cdot \frac{1}{m-1} = \frac{1}{m}\,,$$

womit das Lemma bewiesen ist. ■

Damit haben wir für die Definition des Informationsgehalts nur noch die Freiheit, die Basis a des Logarithmus festzulegen; die traditionelle Wahl ist $a = 2$.

Definition: Die Entropie einer Quelle A mit einem m-buchstabigen Alphabet und Wahrscheinlichkeit p_i für das Auftreten des i-ten Buchstaben ist

$$H(A) = -\sum_{i=1}^{m} p_i \log_2 p_i \,.$$

Diese Entropie wollen wir nun berechnen für eine Quelle, die deutschen Klartext liefert; das Alphabet enthalte, wie in der klassischen Kryptographie üblich, nur die 26 Buchstaben von A bis Z. Nach obiger Definition erhalten wir mit den beim Auszählen von *Dr. Katzenbergers Badereise* ermittelten Buchstabenhäufigkeiten

$$H(A) = -\sum_{i=1}^{26} p_i \log_2 p_i \approx 4{,}04088 < \log_2 26 \approx 4{,}70044 \,.$$

Die Entropie ist also nicht sehr viel kleiner als die einer Quelle, die alle Buchstaben mit gleicher Wahrscheinlichkeit liefert.

Nun ist allerdings eine Quelle, die zusammenhangslos Buchstaben ausstößt, keine sonderlich gute Annäherung an die deutsche Sprache – selbst wenn die Buchstabenhäufigkeiten stimmen. Wie wir bei der Kryptanalyse von allgemeinen monoalphabetischen Substitutionen gesehen haben, ergeben bereits die Kontakthäufigkeiten ein deutlich klareres Bild der Sprache als die bloßen Buchstabenhäufigkeiten; wir sollten daher eine Entropie definieren und berechnen, die auch unsere Kenntnis der Kontakthäufigkeiten wiederspiegelt. Dazu dürfen wir allerdings nicht vom bisher verwendeten Alphabet A mit 26 Buchstaben ausgehen, sondern von einem Alphabet, dessen Elemente ganze Buchstabenfolgen einer festen Länge N sind, d.h. also von A^N.

Wir bezeichnen die Wahrscheinlichkeit dafür, daß der (gewöhnliche) Buchstabe y auf den Buchstaben x folgt mit $p(x|y)$ und die Wahrscheinlichkeit für das Auftreten von x mit $p(x)$. Die Buchstabenfolge $x_1x_2\ldots x_N$ aus unserem neuen Alphabet hat dann, wenn wir nur mit Buchstaben- und Kontakthäufigkeiten argumentieren, die Wahrscheinlichkeit

$$p(x_1\ldots x_N) = p(x_1)p(x_2|x_1)\cdots p(x_N|x_{N-1})\,.$$

Die Entropie pro „Buchstabe“ aus A^N ist somit

$$H_N = -\sum_{\mathbf{x}\in A^N} p(\mathbf{x})\log_2 p(\mathbf{x})\,,$$

wobei H_1 gerade die oben berechnete Entropie für das Alphabet A ist.

Versuchen wir, H_N für $N \ge 2$ aus H_{N-1} zu berechnen! Dazu schreiben wir

$$\mathbf{x} = \mathbf{y}x_N \quad \text{mit} \quad \mathbf{y} \in A^{N-1}\,;$$

dann ist

$$p(\mathbf{x}) = p(x_1\ldots x_N) = p(\mathbf{y})p(x_N|x_{N-1})$$

und

$$\log_2 p(\mathbf{x}) = \log_2 p(\mathbf{y}) + \log_2 p(x_N|x_{N-1})\,,$$

also

$$p(\mathbf{x})\log_2 p(\mathbf{x}) = p(\mathbf{y})p(x_N|x_{N-1})\bigl(\log_2 p(\mathbf{y}) + \log_2 p(x_N|x_{N-1})\bigr)$$

und

$$\begin{aligned} H_N &= -\sum_{\mathbf{x}\in A^N} p(\mathbf{x})\log_2 p(\mathbf{x}) \\ &= -\sum_{\mathbf{x}\in A^N} p(\mathbf{y})\log_2 p(\mathbf{y})\cdot p(x_N|x_{N-1}) \\ &\quad -\sum_{\mathbf{x}\in A^N} p(\mathbf{x})\log_2 p(x_N|x_{N-1})\,. \end{aligned}$$

Dabei ist

$$\begin{aligned} &-\sum_{\mathbf{x}\in A^N} p(\mathbf{y})\log_2 p(\mathbf{y})\cdot p(x_N|x_{N-1}) \\ &= -\sum_{\mathbf{y}\in A^N} p(\mathbf{y})\log_2 p(\mathbf{y}) \sum_{x_N\in A} p(x_N|x_{N-1})\,. \end{aligned}$$

Für jedes $x_{N-1} \in A$ und damit auch für jedes $\mathbf{y} \in A^{N-1}$ ist

$$\sum_{x_N \in A} p(x_N|x_{N-1}) = 1 ,$$

denn irgendein Buchstabe muß schließlich auf x_{N-1} folgen. Also ist

$$-\sum_{\mathbf{x} \in A^N} p(\mathbf{y}) \log_2 p(\mathbf{y}) \cdot p(x_N|x_{N-1}) = H_{N-1}$$

und

$$H_N = H_{N-1} + \Delta_N \quad \text{mit} \quad \Delta_N = -\sum_{\mathbf{x} \in A^N} p(\mathbf{x}) \log_2 p(x_N|x_{N-1}) .$$

Speziell für $N = 2$ erhalten wir

$$\Delta_2 = -\sum_{x_1x_2 \in A^2} p(x_1x_2) \log_2 p(x_2|x_1) ,$$

was sich leicht durch Auszählen von Buchstabenpaaren bestimmen läßt. Für $N \geq 3$ ist nach Definition

$$p(x_1 \dots x_N) = p(x_1)p(x_2|x_1) \cdots p(x_N|x_{N-1}) ,$$

und

$$\begin{aligned} &\sum_{x_1 \in A} p(x_1)p(x_2|x_1) \cdots p(x_N|x_{N-1}) \\ &= \sum_{x_1 \in A} p(x_1x_2)p(x_3|x_2) \cdots p(x_N|x_{N-1}) \\ &= p(x_2)p(x_3|x_2) \cdots p(x_N|x_{N-1}) , \end{aligned}$$

denn $\sum_{x_1 \in A} p(x_1x_2) = p(x_2)$, da auf jeden Fall *irgendein* Buchstabe $x_1 \in A$ vor x_2 stehen muß.

Somit ist $\Delta_N = \Delta_{N-1}$, d.h. $\Delta_N = \Delta_2$ für alle $N \geq 2$ und wir erhalten

Lemma: $H_N = H_1 + (N-1)H_\infty$, wobei H_1 die Entropie ist und

$$H_\infty = -\sum_{x_1x_2 \in A^2} p(x_1x_2) \log_2 p(x_2|x_1) .$$

■

Asymptotisch, für $N \to \infty$, ist also der Informationsgehalt pro (gewöhnlichem) Buchstabe gleich H_∞.

Auf der Basis von JEAN PAULS Roman *Dr. Katzenbergers Badereise* erhält man

$$H_2 \approx 7{,}43853 \text{ Bit} .$$

Da wir

$$H_1 \approx 4{,}04088 \text{ Bit}$$

bereits kennen, liefert das Lemma

$$H_\infty \approx 3{,}39765 \text{ Bit/Buchstabe} .$$

Genauso lassen sich H_1, H_2 und H_∞ auch für andere Arten von Daten bestimmen: Betrachten wir etwa Dr. Katzenbergers Badereise als ASCII-Text, also mit Groß- und Kleinschreibung, Leerzeichen und Satzzeichen, so ergeben sich die Werte

$$H_1 \approx 4{,}423089, \quad H_2 \approx 7{,}797414 \quad \text{und} \quad H_\infty \approx 3{,}374325 ,$$

beim gleichen Text im Format von MS WORD ist

$$H_1 \approx 5{,}077718, \quad H_2 \approx 8{,}248848 \quad \text{und} \quad H_\infty \approx 3{,}171130 ,$$

je formatierter der Text ist, desto kleiner wird also H_∞.

Große Werte von H_∞ erhält man für komprimierte Dateien; Auszählen einer `.zip`-Datei etwa ergab

$$H_1 \approx 7{,}998676, \quad H_2 \approx 15{,}925112 \quad \text{und} \quad H_\infty \approx 7{,}926437 .$$

Programmdateien führen auf ähnliche Werte wie Texte: Für eine Programmdatei unter Windows ergab sich

$$H_1 \approx 5{,}278473, \quad H_2 \approx 8{,}791534 \quad \text{und} \quad H_\infty \approx 3{,}513061 ,$$

für eine unter UNIX

$$H_1 \approx 6{,}550956, \quad H_2 \approx 11{,}462319 \quad \text{und} \quad H_\infty \approx 4{,}911363 .$$

Schließlich kann man auch Trigrammhäufigkeiten und so weiter im statistischen Modell berücksichtigen; hierfür habe ich nur die Entropie H_n

pro n-Gramm und, als ein n-tel davon, die pro Buchstabe ausgerechnet, ausgezählt auf der Basis von 26 Buchstaben ohne Leerzeichen *usw.:*

n	H_n	H_n/n
1	4,04088	4,04088
2	7,43853	3,71926
3	10,37095	3,45698
4	12,81915	3,20479

Man schätzt, daß bei voller Ausnutzung aller statistischer Strukturen der Sprache schließlich noch etwa 30% der $\log_2 26$ Bit/Buchstabe notwendig sind. Zwar ist eine Sprache mit all ihrer Orthographie und Grammatik viel zu komplex, als daß man eine solche Zahl exakt berechnen könnte, man kann sie aber doch experimentell schätzen, indem man muttersprachlichen Lesern Lückentexte vorlegt und schaut, bis zu welchem Prozentsatz lesbarer Buchstaben sie das Original noch rekonstruieren können.

§2: Kryptanalyse durch den Bayesschen Gegner

a) Die Grundidee

Gegeben seien ein Alphabet A (über das wir im Augenblick keine weitere Annahmen machen) und ein Kryptosystem

$$\left\{T_s \colon A \to A \;\middle|\; s \in S\right\}.$$

Dabei sind die Abbildungen $T_s \colon A \to A$ Verschlüsselungsverfahren, die von einem Schlüssel $s \in S$ abhängen.

Bei der Einschätzung der Sicherheit dieses Systems gehen wir, aus den im vorigen Kapitel genannten Gründen, davon aus, daß der Gegner das Kryptosystem kennt; wir beurteilen die als nur auf der Grundlage des Schlüssels.

Ein Gegner muß zur Entschlüsselung einer Nachricht also „nur“ noch den Schlüssel finden; ein gutes Kryptosystem versucht, ihm dieses so schwer wie möglich zu machen.

Wie eingangs erwähnt, betrachten wir in diesem Paragraphen einen „idealisierten“ Gegner, der über unbegrenzte Rechenmöglichkeiten

verfügt, so daß er jeden endlichen Algorithmus, unabhängig von der Zahl der notwendigen Rechenschritte, in kurzer Zeit durchführen kann; diesen Gegner bezeichnet man in der Kryptologie als den BAYESschen Gegner.

Seine Strategie ist die folgende: Er bestimmt bei vorliegendem Chiffretext c für jeden Schlüssel $s \in S$ die Entschlüsselung $T_s^{-1}(c)$ und berechnet deren Wahrscheinlichkeit vor dem Hintergrund seines Wissens über die Natur der Klartexte. Er entscheidet sich für den Schlüssel, für den diese Wahrscheinlichkeit maximal ist.

b) Einführendes Beispiel

Um zu sehen, wie das funktioniert, beobachten wir den BAYESschen Gegner bei der Entschlüsselung einer CAESAR-Chiffre – bei komplizierteren Chiffren macht sich zu schnell bemerkbar, daß er deutlich schneller rechnen kann als wir.

Angenommen, er empfängt das schon im letzten Kapitel betrachtete Kryptogramm QJULIILPPRUJHQJUDXH Q .

Unter der Annahme, daß es sich hier um eine CAESAR-Chiffre handelt, ist es nicht nur für den BAYESschen Gegner, sondern auch für uns leicht möglich, alle 26 möglichen Klartexte sowie deren jeweiligen Wahrscheinlichkeiten zu berechnen. Zu jedem dieser Klartexte berechnen wir dann dessen Wahrscheinlichkeit in der deutschen Sprache.

Um zu sehen, wie viele Chiffretextbuchstaben zur Entschlüsselung notwendig sind, betrachten wir nicht das gesamte Kryptogramm auf einmal, sondern wir betrachten in der i-ten Zeile der folgenden Tabelle nur die ersten i Buchstaben. Für diese ist jeweils die wahrscheinlichstse Entschlüsselung angegeben, zunächst auf der Basis der Buchstabenwahrscheinlichkeiten, dann auf Basis der Kontaktwahrscheinlichkeiten. Die angegebenen Wahrscheinlichkeiten sind der Übersichtlichkeit halber nicht die absoluten Klartextwahrscheinlichkeiten der jeweiligen Entschlüsselungen, sondern die relativen: Wir dividieren also durch die Summe der Wahrscheinlichkeiten der 26 möglichen Klartexte. Das Ergebnis ist in Prozent angegeben:

E	18,498	E	18,498
RE	25,379	ER	54,251
NAT	25,832	ERK	40,153
NATE	71,640	NATE	77,103
ANGRI	50,664	ANGRI	58,662
NATEV S	58,718	ANGRI F	95,260
NATEV SS	78,445	ANGRI FF	99,281
NATEV SSV	33,482	ANGRI FFI	99,311
ANGRI FFIM	42,358	ANGRI FFIM	99,974
ANGRI FFIMM	31,116	ANGRI FFIMM	99,997
FSLWN KKNRR T	35,311	ANGRI FFIMM O	100,000
ANGRI FFIMM OR	43,708	ANGRI FFIMM OR	100,000
ANGRI FFIMM ORG	33,840	ANGRI FFIMM ORG	100,000
ANGRI FFIMM ORGE	65,818	ANGRI FFIMM ORGE	100,000
ANGRI FFIMM ORGEN	84,217	ANGRI FFIMM ORGEN	100,000
ANGRI FFIMM ORGEN G	73,487	ANGRI FFIMM ORGEN G	100,000
ANGRI FFIMM ORGEN GR	51,676	ANGRI FFIMM ORGEN GR	100,000
NATEV SSVZZ BETRA TEN	62,256	ANGRI FFIMM ORGEN GRA	100,000
NATEV SSVZZ BETRA TENH	66,592	ANGRI FFIM MORGENG RAU	100,000
ANGRI FFIMM ORGEN GRAUE	57,125	ANGRI FFIMM ORGEN GRAUE	100,000
ANGRI FFIMM ORGEN GRAUE N	70,387	ANGRI FFIMM ORGEN GRAUE N	100,000

Mit Buchstabenhäufigkeiten allein schwankt der BAYESsche Gegner also noch ziemlich lange zwischen zwei Schlüsseln, beachtet er auch Kontakthäufigkeiten, hat er schon nach fünf Buchstaben die richtige Entschlüsselnung und ab dem sechsten Buchstaben weiß er das mit einer Wahrscheinlichkeit von über 95%.

Als nächstes Kryptogramm betrachten wir PWDUYFSFQDXJ . Hier ergeben sich folgende wahrscheinlichste Entschlüsselungen:

E	18,498	E	18,498
EL	20,611	EL	27,369
ELS	36,252	ELS	61,022
ZGNE	20,524	MTAR	27,712
ZGNEI	51,027	MTARV	32,544
NUBSW D	25,059	DKRIM T	46,246
ELSJN UH	42,422	KRYPT AN	63,235

ELSJN UHU	44,039	KRYPT ANA	72,856
DKRIM TGTE	81,297	KRYPT ANAL	53,404
DKRIM TGTER	85,742	DKRIM TGTER	99,705
DKRIM TGTER L	87,118	DKRIM TGTER L	98,089
ZGNEI PCPAN HT	75,437	KRYPT ANALY SE	99,797

Hier gelingt die Entschlüsselung also nur mittels Kontakthäufigkeiten, und auch da nur dank des guten Doktors STRYKIUS, der dafür sorgt, daß das Digramm „RY“ in *Dr. Katzenbergers Badereise* mit positiver Wahrscheinlichkeit auftritt: Bei vielen anderen Referenztexten tritt diese Buchstabenkombination nie auf, so daß die Wahrscheinlichkeit eines jeden Worts, daß sie enthält auf null gesetzt würde.

Man muß freilich beachten, daß der ideale BAYESsche Gegner nicht mit der Auszählung eines einzelnen Texts arbeitet, sondern *die* korrekte Wahrscheinlichkeitsverteilung kennt – wie immer diese auch definiert sein mag.

Es ist kein Zufall, das der BAYESsche Gegner einbuchstabige Nachrichten immer als „E“ entschlüsselt: Für einbuchstabige Texte gibt es nach beiden Wahrscheinlichkeitsmodellen keine bessere Alternative. Entsprechend einfach werden auch Wörter erkannt, die mit „E“ beginnen und viele häufige Buchstaben und Buchstabenkombinationen enthalten wie beispielsweise im Kryptogramm JSIQNHMPJNY , wo beide Modelle sofort den richtigen Schlüssel finden:

E	18,498	E	18,498
EN	43,447	EN	66,315
END	49,997	END	96,907
ENDL	60,346	ENDL	93,935
ENDLI	89,043	ENDLI	99,306
ENDLI C	89,309	ENDLI C	99,995
ENDLI CH	96,787	ENDLI CH	100,000
ENDLI CHK	92,988	ENDLI CHK	100,000
ENDLI CHKE	98,564	ENDLI CHKE	100,000
ENDLI CHKEI	99,529	ENDLI CHKEI	100,000
ENDLI CHKEI T	99,939	ENDLI CHKEI T	100,000

Auch nicht zu exotische Substantive mit Artikel sind meist problemlos:
IJWYNXHM führt zu

E	18,498	E	18,498
DE	22,057	DE	38,733
DER	38,490	DER	87,654
DERT	51,481	DERT	96,107
DERTI	75,471	DERTI	99,393
DERTI S	88,877	DERTI S	99,717
DERTI SC	87,934	DERTI SC	99,999
DERTI SCH	90,031	DERTI SCH	100,000

und PUQEFGQTXQ zu

E	18,498	E	18,498
IN	23,056	CH	40,135
DIE	41,643	DIE	43,465
DIES	64,018	DIES	70,969
DIEST	69,681	DIEST	90,466
DIEST U	53,199	DIEST U	76,378
DIEST UE	83,884	DIEST UE	91,854
DIEST UEH	87,958	DIEST UEH	97,642
DIEST UEHL	96,798	DIEST UEHL	99,523
DIEST UEHLE	99,372	DIEST UEHLE	99,978

Nur der sächliche Artikel ist etwas problematischer: EBTIBVT wird entschlüsselt als

E	18,498	E	18,498
HE	29,769	HE	28,458
EBT	20,051	DAS	47,768
EBTI	34,481	DASH	32,720
HEWLE	45,218	DASHA	75,568
DASHA U	47,297	DASHA U	98,212
DASHA US	57,475	DASHA US	99,609

Hier gibt es also Schwierigkeiten mit den Buchstabenhäufigkeiten allein, da der Text aus eher nicht so häufigen Buchstaben zusammengesetzt ist.

Dieses Problem tritt auch bei anderen Texten auf wie etwa HAQQNAA mit

E	18,498	E	18,498
LE	20,611	UN	25,577
LEU	29,391	UND	69,721
LEUU	24,785	UNDD	88,852
LEUUR	44,254	UNDDA	97,663
LEUUR E	68,174	UNDDA N	99,279
LEUUR EE	81,555	UNDDA NN	99,890

und selbst bei bei so zentralen deutschen Wörtern wieYRVGX HYGHE mit

E	18,498	E	18,498
LE	20,611	UN	25,577
HAE	33,485	LEI	39,068
LEIT	54,851	LEIT	71,632
UNRCT	43,649	LEITK	41,025
UNRCT D	51,689	VOSDU E	56,410
UNRCT DU	59,272	LEITK UL	71,093
UNRCT DUC	47,724	LEITK ULT	98,296
UNRCT DUCD	48,308	LEITK ULTU	91,777
UNRCT DUCDA	50,558	LEITK ULTUR	97,113

oder FNHZNTRA mit

E	18,498	E	18,498
EM	15,918	EM	21,504
LTN	19,493	CKE	37,812
AICU	20,640	SAUM	70,182
AICUI	28,589	SAUMA	97,483
AICUI O	29,941	SAUMA G	98,658
SAUMA GE	52,561	SAUMA GE	99,995
SAUMA GEN	78,116	SAUMA GEN	100,000

Zusammenfassend können wir festhalten, daß dem BAYESschen Gegner zumindest mit Kontakthäufigkeiten bereits wenige Buchstaben zur korrekten Entschlüsselung reichen. Beachtet man noch, daß sich weder ein

BAYESsche Gegner noch ein realer Kryptanalytiker auf Kontakthäufigkeiten beschränken muß, sondern auch Sprachkenntnisse einsetzt, die sich auf Trigramme, Worte und so weiter beziehen, kommen sie mit noch weniger Buchstaben aus, als obige Beispiele nahelegen. Während des ersten Weltkriegs aktive Militärkryptographen gehen davon aus, daß man nur knapp zwei Buchstaben pro Alphabet (das heißt also pro CAESAR-Chiffre) benötigt, um eine VIGENÈRE-Chiffre zu brechen.

Als Kuriositär am Rande sei noch erwähnt, daß der BAYESsche Gegner auch ohne Chiffretext bereits wahrscheinlichste „Entschlüsselungen“ findet: Falls er nur mit Buchstabenhäufigkeiten arbeitet, ist das natürlich eine Folge von lauter „E“s, schon bei Kontakthäufigkeiten ergeben sich aber interessantetere Ergebnisse, die hier bis zur Nachrichtenlänge acht aufgeführt sind:

n	Text	Wahrscheinlichkeit
1	e	0.184945672750
2	en	0,042051665485
3	ich	0,011662006378
4	ende	0,003537155688
5	eiche	0,000825895113
6	endich	0,000201549003
7	endende	0,000067649431
8	eichende	0,000015795555

c) Der allgemeine Ansatz

Der BAYESsche Gegner braucht als erstes Wissen oder zumindest Annahmen über den Inhalt der Nachrichten: In den obigen Beispielen handelte es sich um deutschen Klartext, verschlüsselt auf der Grundlage von 26 Buchstaben ohne Zwischenräume und Satzzeichen. Das ist natürlich nicht mehr der typische Fall für heutige Kryptanalyse. Dort geht es eher um Dokumente von Textverarbeitungs-, Tabellenkalkulations- oder Datenbankprogrammen oder um ausführbare Programme für ein gegebenes Betriebssystem oder ähnliches.

In jedem dieser Fälle verschafft er sich als erstes durch Auszählen eine Wahrscheinlichkeitsverteilung für die möglichen Klar„texte“ der

Länge N und ordnet dadurch jedem solchen Klartext W eine

Klartextwahrscheinlichkeit $p_K(W)$

zu; diese kann beispielsweise so wie im letzten Paragraphen über Buchstabenhäufigkeiten oder über Kontakthäufigkeiten definiert sein, aber auch viele andere Ansätze sind möglich.

Außerdem ordnet er jedem Schlüssel s aus dem Schlüsselraum S eine

Schlüsselwahrscheinlichkeit $p_S(s)$

zu. Bei einem gut gemanagten Kryptosystem sollten alle Schlüssel mit gleicher Wahrscheinlichkeit auftreten, so daß

$$p_S(s) = \frac{1}{\#S} \quad \text{für alle } s \in S$$

ist, aber in der Praxis kann der Gegner oft große Vorteile aus der Tatsache ziehen, daß dies nicht der Fall ist. (Wie zufällig sind Ihre Paßwörter?)

Falls nun ein Chiffretext $C \in A^N$ aufgefangen wird, muß der BAYESsche Gegner berechnen, wie wahrscheinlich jeder Schlüssel $s \in S$ vor dem Hintergrund dieses Chiffretexts ist. Dazu müssen bedingte Wahrscheinlichkeiten berechnet werden, also muß man zunächst die Wahrscheinlichkeit des Chiffretexts $C \in A^N$ kennen.

C entsteht aus einem Klartext $W \in A^N$ durch Verschlüsselung mit einem Schlüssel $s \in S$; da der Schlüssel vom Klartext unabhängig sein sollte, ist die Wahrscheinlichkeit für das Zusammentreffen von Klartext W und Schlüssel s gleich

$$p_K(W) \cdot p_S(s) .$$

Die Wahrscheinlichkeit, einen bestimmten Chiffretext C zu empfangen, ist also *a priori* gleich

$$p_{Ch}(C) = \sum_{\substack{(W,s) \in A^N \times S \\ T_s(W) = C}} p_K(W) \cdot p_S(s) .$$

Die Wahrscheinlichkeit, daß C auftritt und mit Schlüssel $s \in S$ verschlüsselt wurde, ist entsprechend gleich

$$p_{Ch,S}(C, s) = \sum_{\substack{W \in A^N \\ T_s(W) = C}} p_K(W) \cdot p_S(s) ,$$

wobei die Summe hier im allgemeinen wegen der Injektivität von T_s nur einen Summanden hat. (T_s muß nicht injektiv auf A^N sein; es muß nur injektiv auf der Teilmenge der tatsächlich vorkommenden Nachrichten sein.)

Die eigentlich interessante Wahrscheinlichkeit, die Wahrscheinlichkeit, daß ein gegebener Chiffretext C durch Verschlüsselung mit $s \in S$ entstanden ist, berechnet sich nun als

$$p_{S|Ch}(s|C) = \frac{p_{Ch,S}(C,s)}{p_{Ch}(C)} \, .$$

Der BAYESsche Gegner entscheidet sich bei seinem Entschlüsselungsversuch für einen Schlüssel, der diese bedingte Wahrscheinlichkeit maximal macht. Um einen solchen Schlüssel zu bestimmen, muß er $p_{S|Ch}(s|C)$ möglicherweise für alle Schlüssel $s \in S$ berechnen, jedoch ist dies angesichts seiner unbegrenzten Rechenfähigkeit kein Problem.

Solange es sich nur um eine Nachricht dreht, wird der BAYESsche Gegner nicht in erster Linie am Schlüssel interessiert sein, sondern am wahrscheinlichsten Klartext. Hier ist entsprechend die Wahrscheinlichkeit für das Zusammentreffen von Klartext W und Chiffretext C gleich

$$p_{Ch,K}(C,W) = \sum_{\substack{s \in S \\ T_s(W)=C}} p_K(W) \cdot p_S(s)$$

und die bedingte Wahrscheinlichkeit für Klartext W, nachdem Chiffretext C aufgefangen wurde, ist

$$p_{K|Ch}(W|C) = \frac{p_{Ch,K}(C,W)}{p_{Ch}(C)} \, .$$

Definition: Eine BAYESsche Entscheidungsfunktion ist eine Familie von Abbildungen

$$\delta_N : A^N \to A^N$$

mit der Eigenschaft, daß für jeden Chiffretext $C \in A^N$ gilt:

$$p_{K|Ch}(\delta(C)|C) = \max_{W \in A^N} p_{K|Ch}(W|C) \, .$$

Weniger formal ausgedrückt: Für jeden Chiffretext $C \in A^N$ ist $\delta(C)$ ein Klartext mit höchstmöglicher Wahrscheinlichkeit – idealerweise *der* Klartext mit höchstmöglicher Wahrscheinlichkeit, allerdings könnte es auch mehrere Klartexte mit gleicher Wahrscheinlichkeit geben, so daß $\delta(C)$ im allgemeinen durch die obige Definition nicht eindeutig bestimmt ist. Trotzdem ist klar, daß man nichts besseres tun kann, als mit einer solchen BAYESschen Entscheidungsfunktion zu arbeiten; wenn dies zu nichts führt, ist ein Kryptosystem sicher.

d) Perfekte Sicherheit

Am allersichersten ist ein Kryptosystem, wenn der BAYESsche Gegner aus dem aufgefangenen Chiffretext C *überhaupt keine* Information gewinnen kann, die über seine ursprünglich gewählte Wahrscheinlichkeitsfunktion p_K hinausgeht, wenn also die bedingten Wahrscheinlichkeiten, die er in Kenntnis von C errechnet, gleich den ursprünglichen Wahrscheinlichkeiten sind:

Definition: Ein Kryptosystem $\{T_s \mid s \in S\}$ über dem Alphabet A hat *perfekte Sicherheit* für Nachrichten der Länge N, wenn für jeden Chiffretext $C \in A^N$ und jeden Klartext $W \in A^N$ gilt:

$$p_{K|Ch}(W|C) = p_K(W) \, .$$

Wir kennen bereits ein System mit perfekter Sicherheit, den *one time pad* aus Kapitel eins. Leider ist er für große Kommunikationsnetzwerke, etwa im Bankenbereich, nicht praktikabel, da jeder Teilnehmer mit jedem anderen riesige Schlüssel austauschen müßte. Auch ist die Erzeugung *wirklich* zufälliger Buchstaben- oder Zahlenfolgen nicht ganz einfach; die typischen Zufallsgeneratoren der gängigen Betriebssysteme und Compiler sind kryptographisch extrem schwach und können für solche Zwecke nicht verwendet werden.

Ideal wäre deshalb perfekte Sicherheit mit deutlich kleineren Schlüssellängen, aber dies ist nach einem Satz von SHANNON leider nur sehr eingeschränkt möglich:

Satz: Falls ein Kryptosystem $\{T_s \mid s \in S\}$ perfekte Sicherheit für Nachrichten der Länge N hat, ist die Anzahl der Schlüssel $s \in S$ mindestens gleich der Anzahl der möglichen (d.h. mit Wahrscheinlichkeit $p_K(W) > 0$ auftretenden) Klartexte der Länge N.

Beweis: K^+ sei die Menge aller Klartexte aus A^N, die mit positiver Wahrscheinlichkeit auftreten, und C^+ sei die Menge der mit positiver Wahrscheinlichkeit auftretenden Chiffretexte. Für jedes feste $W \in K^+$ und jedes $C \in C^+$ muß es dann einen Schlüsses $s \in S$ geben, so daß $T_s(W) = C$ ist; denn sonst wäre $p_{K|Ch}(W|C) = 0$, aber $p_K(W) > 0$; der Gegner könnte also aus dem Auftreten von C Information gewinnen. Damit muß es mindestens so viele Schlüssel geben, wie es Chiffretexte in C^+ gibt. Für jeden Schlüssel $s \in S$ ist aber T_s eine injektive Abbildung von K^+ nach C^+, d.h. es muß erst recht so viele Schlüssel geben, wie es mögliche Klartexte gibt.

■

e) Die Mehrdeutigkeit eines Schlüssels

In der Praxis ist es nicht unbedingt notwendig, daß der Gegner aus einem aufgefangenen Chiffretext *überhaupt keine* Information ziehen kann; es reicht, wenn er *nicht genügend* Information bekommt.

Als nächstes wollen wir daher abschätzen, wieviel Information der (besonders gefährliche) BAYESsche Gegner aus einem Chiffretext ziehen kann. Besonders wichtig ist der Schutz des Schlüssels s, denn sobald dieser bekannt ist, können bis zum nächsten Schlüsselwechsel alle Nachrichten entschlüsselt werden.

Die Information, die der Gegner zur Rekonstruktion des Schlüssels gewinnen muß, läßt sich nach den Vorarbeiten des letzten Paragraphen leicht quantifizieren: Es ist die Entropie

$$H_S = -\sum_{s\in S} p_S(s) \log_2 p_S(s)\,.$$

Das Lemma über die Maximalität der Entropie bei Gleichverteilung zeigt also noch einmal die eigentlich auch so selbstverständliche Tatsache, daß ein Kryptosystem umso sicherer ist, je mehr die Wahrscheinlichkeitsverteilung der Schlüssel einer Gleichverteilung entspricht.

Der BAYESsche Gegner kennt die Entropie

$$H_{Ch} = -\sum_{C\,\in\,A^N} p_{Ch}(C) \log_2 p_{Ch}(C)$$

der Menge aller Chiffretexte der Länge N; nachdem er eine Nachricht C aufgefangen hat, hat er im Mittel diese Information gewonnen.

Um daraus Informationen über den Schlüssel zu erhalten, berechnet er zunächst die Entropie der Menge aller Paare (C, s):

$$\begin{aligned}
H_{Ch,S} &= -\sum_{(C,s)\,\in\,A^N\times S} p_{Ch,S}(C,s) \log_2 p_{Ch,S}(C,s)\\
&= -\sum_{(C,s)\,\in\,A^N\times S} p_{Ch,S}(C,s)\bigl(\log_2 p_{Ch}(C) + \log_2 p_{S|Ch}(s|C)\bigr)\\
&= -\sum_{(C,s)\,\in\,A^N\times S} p_{Ch,S}(C,s) \log_2 p_{Ch}(C)\\
&\qquad -\sum_{(C,s)\,\in\,A^N\times S} p_{Ch,S}(C,s) \log_2 p_{S|Ch}(s|C)\\
&= -\sum_{C\,\in\,A^N} p_{Ch}(C) \log_2 p_{Ch}(C)\\
&\qquad -\sum_{(C,s)\,\in\,A^N\times S} p_{Ch,S}(C,s) \log_2 p_{S|Ch}(s|C)\\
&= H_{Ch} - \sum_{(C,s)\,\in\,A^N\times S} p_{Ch,S}(C,s) \log_2 p_{S|Ch}(s|C)\,.
\end{aligned}$$

Die letzte Summe in der letzten Zeile bezeichnen wir als *bedingte Entropie* oder *Mehrdeutigkeit*

$$H_{S|Ch} \underset{\text{def}}{=} -\sum_{(C,s)\,\in\,A^N\times S} p_{Ch,S}(C,s) \log_2 p_{S|Ch}(s|C)$$

des Schlüssels bei vorliegendem Chiffretext; diese Information muß sich der Gegner verschaffen, um den Schlüssel zu bestimmen. Sobald $H_{S|Ch} = 0$ ist, kennt er den Schlüssel, denn da in der obigen Summe alle Summanden kleiner oder gleich null sind, kann $H_{S|Ch}$ nur dann gleich null sein, wenn jeder einzelne Summand verschwindet, wenn also für jeden Schlüssel $s \in S$ entweder $p(C,s) = 0$ ist, womit der Schlüssel mit Sicherheit ausgeschlossen ist, oder $\log_2 p(s|C) = 0$, d.h. $p(s|C) = 1$, womit der Schlüssel s mit Sicherheit richtig ist.

Um die Information abzuschätzen, die der Gegner maximal aus einem Chiffretext gewinnen kann, müssen wir also $H_{S|Ch}$ berechnen. Für ein beliebiges Kryptosystem und beliebige Wahrscheinlichkeitsverteilung der Klartexte können wir offensichtlich nichts konkreteres hinschreiben als die definierende Summe. Die für die gesamte Informationstheorie grundlegende Idee von CLAUDE SHANNON war es, spezifischere Aussagen nicht über ein spezielles, sondern über das *durchschnittliche* Kryptosystem zu machen; diese Aussagen sollten dann *im wesentlichen* für die meisten Kryptosysteme gelten.

f) Randomisierung

Um von Durchschnitten reden zu können, müssen wir zunächst einige Annahmen machen über das Verhalten der Klartexte. Wir betrachten wieder Klartexte über einem Alphabet A aus n Buchstaben; für $x \in A$ sei $p(x)$ die Wahrscheinlichkeit, mit der der Buchstabe x auftritt, und für $(x,y) \in A^2$ sei $p(x|y)$ die bedingte Wahrscheinlichkeit dafür, daß er auf ein y folgt.

Wir hatten bislang immer angenommen, daß etwa die Buchstabenhäufigkeiten, die wir durch Auszählen eines hinreichend langen Texts erhalten, in hinreichend guter Näherung mit denen übereinstimmen, die wir in einem gegebenen Klartext vorfinden. Dies müssen wir für die folgenden Überlegungen etwas präziser fassen:

Definition: Für einen gegebenen Klartext $W = w_1 \dots w_N \in A^N$ sei

$$m_x(W) = \#\{i \le N \mid w_i = x\}$$

und

$$m_{xy}(W) = \#\{i \le N-1 \mid w_i = x \text{ und } w_{i+1} = y\}\,.$$

$m_x(W)$ und $m_{xy}(W)$ zählen also, wie oft der Buchstabe x *bzw.* das Buchstabenpaar xy in W vorkommen. Idealerweise sollte für einen hinreichend langen Text W und zwei Buchstaben $x,y \in A$ gelten

$$p(x) \approx \frac{m_x(W)}{N} \quad \text{und} \quad p(x|y) \approx \frac{m_{yx}(W)}{m_y(W)}\,.$$

Exakt kann das natürlich schon aus zahlentheoretischen Gründen nicht gelten, aber wir wollen verlangen, daß die Abweichung für große N mit sehr kleiner Wahrscheinlichkeit größer als eine vorgebbare Schranke δ ist:

Definition: Eine Quelle für Klartexte heißt *ergodisch,* wenn es zu zwei beliebig vorgebbaren reellen Zahlen $\delta, \varepsilon > 0$ stets eine natürliche Zahl N_0 gibt, so daß für alle $N \ge N_0$ und alle $x \in A$ für Texte $W \in A^N$ gilt

$$p\left(\left|\frac{m_x(W)}{N} - p(x)\right| > \delta\right) < \varepsilon\,.$$

Unsere Ansätze zur Kryptanalyse klassischer Systeme beruhten somit stets darauf, daß wir annahne, daß die deutsche Sprache eine ergodische Quelle für Klartexte ist; der Erfolg bei unseren Entschlüsselungsversuchen spricht dafür, daß diese Annahme nicht garzu falsch sein sollte. In der Tat zeigt man in der Stochastik, daß jede Quelle, bei der es zu jedem Buchstabenpaar (x,y) einen Text gibt, in dem y irgendwo hinter x steht, ergodisch ist. Da wohl niemand bezweifelt, daß es solche Texte zu jedem Buchstabenpaar gibt (für (q, y) etwa **q**uer durch Z**y**pern), kann so die Ergodizität auch bewiesen werden.

Wir wollen im folgenden Aussagen machen über die *durchschnittliche* ergodische Quelle. Es ist klar, daß wir keine Chance haben, alle ergodischen Quellen zu bestimmen und dann einen Mittelwert zu bilden; der folgende Satz von SHANNON zeigt aber, daß wir durch Zusammenfassen der Buchstaben zu hinreichend langen Wörtern erreichen können, daß die Wahrscheinlichkeitsverteilung sehr einfach ist: Die Wörter zerfallen in zwei Klassen S und O, so daß Wörter aus S (wie *selten*) praktisch nie vorkommen, während die restlichen Wörter (aus O wie *oft*) alle praktisch dieselbe Wahrscheinlichkeit haben. Aus technischen Gründen arbeiten wir nicht mit Wahrscheinlichkeiten, sondern mit der Entropie:

Die buchstabenweise Entropie der Quelle ist, wenn wir wie in §1, *3)* von Kontakthäufigkeiten ausgehen, gleich

$$H = -\sum_{x\in A}\sum_{y\in A} p(x)p(y|x)\log_2 p(y|x)\,,$$

und wir fordern, daß sich diese Entropie gleichmäßig auf die $W \in O$ verteilen soll:

Satz: Für eine ergodische Quelle gibt es zu zwei beliebig vorgebbaren reellen Zahlen $\varepsilon, \eta > 0$ stets ein $N_0 \in \mathbb{N}$, so daß A^N für $N \geq N_0$ als Vereinigung zweier Teilmengen O und S geschrieben werden kann mit den Eigenschaften

1.) $p(W \in S) < \varepsilon$

2.) $\left|\frac{-\log_2 p(W)}{N} - H\right| < \eta$ für alle $W \in O$.

Zum *Beweis* wählen wir zunächst willkürlich eine natürliche Zahl N und eine reelle Zahl $\delta > 0$ und setzen dann

$$O \underset{\text{def}}{=} \left\{ W \in A^N \;\middle|\; \begin{array}{l} p(W) > 0 \text{ und für alle } x, y \in A \text{ ist} \\ \left|m_{xy}(W) - Np(x)p(y|x)\right| < N\delta \end{array} \right\}$$

und $S = A^N \setminus O$. Für $W \in O$ können wir dann $m_{xy}(W)$ schreiben als

$$m_{xy}(W) = Np(x)p(y|x) + N\delta_{xy} \quad \text{mit} \quad \left|\delta_{xy}\right| < \delta$$

und erhalten damit für $W = w_1 \dots w_N$

$$\begin{aligned} p(W) &= p(w_1)\prod_{i=2}^{N} p(w_i|w_{i-1}) \\ &= p(w_1)\prod_{x\in A}\prod_{y\in A} p(y|x)^{m_{xy}(W)} \\ &= p(w_1)\prod_{x\in A}\prod_{y\in A} p(y|x)^{Np(x)p(y|x)+N\delta_{xy}}\,. \end{aligned}$$

Logarithmieren führt auf

$$\begin{aligned} -\log_2 p(w) &= -\log_2 p(w_1) - N\sum_{x\in A}\sum_{y\in A} p(x)p(y|x)\log_2 p(y|x) \\ &\qquad - N\sum_{x\in A}\sum_{y\in A}\delta_{xy}\log_2 p(y|x) \\ &= -\log p(w_1) + NH - N\sum_{x\in A}\sum_{y\in A}\delta_{xy}\log_2 p(y|x) \end{aligned}$$

und somit zur Ungleichung

$$\begin{aligned} \left|\frac{-\log_2 p(W)}{N} - H\right| &= \left|\frac{-\log_2 p(W)}{N} - \sum_{x\in A}\sum_{y\in A}\delta_{xy}\log_2 p(y|x)\right| \\ &< \frac{-\log_2 p(W)}{N} + \delta\sum_{x\in A}\sum_{y\in A} -\log_2 p(y|x)\,. \end{aligned}$$

Wählt man N hinreichend groß und δ hinreichend klein, so läßt sich dieser Ausdruck offensichtlich kleiner als jedes vorgegebene $\eta > 0$ machen, so daß die Eigenschaft 2.) erfüllt ist.

Für die Eigenschaft 1.) müssen wir nachrechnen, mit welcher Wahrscheinlichkeit eine Nachricht $W \in A^N$ in S liegt. Sie liegt genau dann in S, wenn sie nicht in O liegt, für mindestens ein Buchstabenpaar (x, y) muß also

$$\left|m_{xy}(W) - Np(x)p(y|x)\right| \geq N\delta$$

sein. Die Wahrscheinlichkeit hierfür ist sicherlich nicht größer als die Summe über alle Paare (x, y) für die entsprechenden Wahrscheinlichkeiten, d.h.

$$p(W \in S) \leq \sum_{x\in A}\sum_{y\in A} p\left(\left|m_{xy}(W) - Np(x)p(y|x)\right| \geq N\delta\right)\,.$$

Wegen der Ergodizität der Quelle können wir für jedes $\widetilde{\varepsilon} > 0$ ein $N_1 \in \mathbb{N}$ finden, so daß für $N \geq N_1$ gilt

$$p\left(\left|\frac{m_x(W)}{N} - p(x)\right| > \frac{\delta}{2}\right) < \widetilde{\varepsilon}$$

oder, was dasselbe ist,

$$p\Big(|m_x(W) - Np(x)| > \frac{\delta}{2}N\Big) < \widetilde{\varepsilon}\,.$$

Für große N wird, wieder wegen der Ergodizität der Quelle, auch die Häufigkeit $m_x(W)$ groß; falls wir N so groß wählen, daß wir die Wahrscheinlichkeit des Ereignisses $m_x(W) < N_1$ vernachlässigen können, ist also auch

$$p\left(\left|\frac{m_{xy}(W)}{m_x(W)} - p(y|x)\right| > \frac{\delta}{2}\right) < \widetilde{\varepsilon}\,.$$

Die komplementäre Wahrscheinlichkeit dafür, daß

$$|m_x(W) - Np(x)| < \frac{\delta}{2}N$$

bzw.

$$\big|m_{xy}(W) - m_x(W)p(y|x)\big| < \frac{\delta}{2}m_x(W) \le \frac{\delta}{2}N$$

ist, beträgt jeweils mindestens $1-\widetilde{\varepsilon}$; die Wahrscheinlichkeit dafür, daß beides eintritt, ist also mindestens

$$(1-\widetilde{\varepsilon})^2 > 1 - 2\widetilde{\varepsilon}\,.$$

Falls die erste der beiden Ungleichungen erfüllt ist, gilt erst recht

$$|m_x(W)p(y|x) - Np(x)p(y|x)| < \frac{\delta}{2}Np(y|x) \le \frac{\delta}{2}N$$

und damit nach der Dreiecksungleichung zusammen mit der zweiten Ungleichung

$$\big|m_{xy}(W) - Np(x)p(y|x)\big| < \frac{\delta}{2}N + \frac{\delta}{2}N = N\delta\,.$$

Diese Ungleichung ist somit mindestens mit Wahrscheinlichkeit $1-2\widetilde{\varepsilon}$ erfüllt; die komplementäre Wahrscheinlichkeit ist

$$p\Big(\big|m_{xy}(W) - Np(x)p(y|x)\big| > N\delta\Big) < 2\widetilde{\varepsilon}\,.$$

Dies können wir in die oben abgeleitete Formel

$$p(W \in S) \le \sum_{x\in A}\sum_{y\in A} p\left(\big|m_{xy}(W) - Np(x)p(y|x)\big| \ge N\delta\right)$$

einsetzen und erhalten, da es n^2 Buchstabenpaare gibt,

$$p(W \in S) < 2n^2\widetilde{\varepsilon}\,.$$

Da n eine Konstante ist, müssen wir nun nur $\widetilde{\varepsilon}$ hinreichend klein wählen und erhalten dann, daß diese Wahrscheinlichkeit höchstens gleich ε ist, wie behauptet.

■

Die Mengen O und S aus dem gerade bewiesenen Satz lassen sich für eine gegebene Quelle natürlich nur schwer konstruieren, denn im allgemeinen muß N doch schon sehr groß gewählt werden, damit dieser Satz gilt. Wir können allerdings trotzdem Aussagen über die Größe von O und S machen: Für kleine Werte von ε und η machen wir keinen großen Fehler, wenn wir davon ausgehen, daß alle Nachrichten aus O mit *exakt* derselben Wahrscheinlichkeit vorkommen und daß Nachrichten aus S *nie* vorkommen. Dann ist für $W \in A^N$ also

$$p(W) = \begin{cases} 1/\#O & \text{falls } W \in O \\ 0 & \text{falls } W \in S \end{cases}$$

und die Entropie der Quelle ist

$$-\sum_{W\in A^N} p(W)\log_2 p(W) = \sum_{W\in O}\frac{1}{\#O}\log_2\#O = \log_2\#O\,.$$

Die Entropie pro Buchstabe des Alphabets A ist damit gleich

$$\frac{\log_2\#O}{N}\,.$$

Diese Entropie pro Buchstabe können wir berechnen; für die Quellen aus §1 haben wir dies dort bereits getan: Geht man nur von den Buchstabenhäufigkeiten aus, ist sie gleich 4,04088 Bit, und bei Berücksichtigung der Kontakthäufigkeiten gleich

$$\frac{4{,}04088 + 3{,}39765(N-1)}{N}\ \text{Bit}\,.$$

Also ist im ersten Fall

$$\#O \approx 2^{4{,}04088N}$$

und im zweiten

$$\#O \approx 2^{4{,}04088+3{,}39765(N-1)}\,.$$

In A^N gibt es insgesamt

$$26^N \approx 2^{4,7044N}$$

Nachrichten; das Verhältnis $\#O$ zu $\#A^N$ läßt sich daher schreiben als

$$\frac{\#O}{\#A^N} = 2^{-R_N} ,$$

wobei im ersten Fall

$$R_N \approx (4{,}7044 - 4{,}04088)N = 0{,}66352N$$

ist und im zweiten

$$R_N \approx 4{,}7044N - 4{,}04088 + 3{,}39765(N-1) = 1{,}30675N - 0{,}64323 .$$

Damit können wir uns daran machen, die Mehrdeutigkeit eines Schlüssels für diese (und viele andere) Quellen abzuschätzen: Wir ersetzen die jeweilige Quelle durch ein „durchschnittliche“ ergodische Quelle derselben Entropie und rechnen mit dieser in der Hoffnung, daß wir dadurch keinen allzu großen Fehler machen. Außerdem nehmen wir an, daß N hinreichend groß sei, so daß wir den obigen Satz anwenden können.

Nach Definition ist die Mehrdeutigkeit des Schlüssels nach Empfang einer Chiffretext-Nachricht C der Länge N gleich

$$H_{S|Ch} \underset{\text{def}}{=} - \sum_{(C,s)\,\in\, A^N \times S} p_{Ch,S}(C,s) \log_2 p_{S|Ch}(s|C)$$

Wenn wir davon ausgehen, daß jeder Schlüssel mit derselben Wahrscheinlichkeit auftritt, ist für jeden Schlüssel $s \in S$

$$p_S(s) = \frac{1}{\#S} .$$

und für einen Klartext $W \in A^N$ ist

$$p_K(W) = \begin{cases} 1/\#O & \text{falls } W \in O \\ 0 & \text{sonst} \end{cases} ;$$

eine Nachricht $C \in A^N$ kann also nur dann als Chiffretext auftreten, wenn es (mindestens) einen Schlüssel $s \in S$ gibt, so daß der zugehörige Klartext $W = T_s^{-1}(C)$ in O liegt. Wenn wir die Anzahl aller solcher Schlüssel mit $N_S(C)$ bezeichnen, ist also die Wahrscheinlichkeit des Chiffretexts C gleich

$$p_{Ch}(C) = \frac{N_S(C)}{\#S} \cdot \frac{1}{\#O} = \frac{N_S(C)}{\#S \cdot \#O} ;$$

die Wahrscheinlichkeit für Chiffretext C verschlüsselt mit $s \in S$ ist

$$p_{Ch,S}(C,s) = \begin{cases} \frac{1}{\#O \cdot \#S} & \text{falls } T_s^{-1}(C) \in O \\ 0 & \text{sonst} \end{cases} ,$$

und damit ist die bedingte Wahrscheinlichkeit für den Schlüssel s nach Empfang des Chiffretexts C

$$p_{S|Ch}(s|C) = \frac{p_{Ch,S}(C,s)}{p_{Ch}(C)} = \frac{1}{N_S(C)} .$$

Somit ist

$$\begin{aligned} H_{S|Ch} &= - \sum_{(C,s)\,\in\, A^N \times S} p_{Ch,S}(C,s) \log_2 p_{S|Ch}(s|C) \\ &= - \sum_{C \in A^N} \sum_{\substack{s \in S \\ T_s^{-1}(C) \in O}} \frac{1}{\#O \cdot \#S} (-\log_2 N_S(C)) \\ &= \sum_{C \in A^N} \frac{N_S(C)}{\#O \cdot \#S} \log_2 N_S(C) . \end{aligned}$$

Diese Summe können wir nicht weiter ausrechnen, da wir die Zahlen $N_S(C)$ nicht kennen. Nun haben wir aber angenommen, daß wir ein *durchschnittliches* Kryptosystem haben, d.h. wir interessieren uns für den Mittelwert von $H_{S|Ch}$ über *alle* Kryptosysteme, und darüber können wir Aussagen machen:

Für einen zufällig gewählen Text $C \in A^N$ und einen zufällig gewählten Schlüssel $s \in S$ ist $T_s^{-1}(C) \in O$ mit Wahrscheinlichkeit

$$\frac{\#O}{\#A^N} = \frac{\#O}{n^N} = 2^{-R_N} ,$$

wobei die reelle Zahl R_N wie oben so gewählt wird, daß die rechtsstehende Gleichung gilt. Entsprechend ist die Wahrscheinlichkeit dafür, daß $T_s^{-1}(C)$ nicht in O liegt, gleich

$$1 - 2^{-R_N} .$$

Die Wahrscheinlichkeit dafür, daß $N_S(C)$ gleich einer festen Zahl m ist, können wir interpretieren als die Wahrscheinlichkeit dafür, daß für m Schlüssel s gilt $T_s^{-1}(C) \in O$, während für die restlichen $\#S - m$ Schlüssel gilt $T_s^{-1}(C) \notin O$. Falls die m Schlüssel vorgegeben sind, ist diese Wahrscheinlichkeit gleich

$$\left(2^{-R_N}\right)^m \left(1 - 2^{-R_N}\right)^{\#S-m} ,$$

und da es $\binom{\#S}{m}$ Möglichkeiten gibt, aus $\#S$ Schlüsseln m auszuwählen, ist die Wahrscheinlichkeit dafür, daß $N_S(C) = m$ ist, gleich

$$\binom{\#S}{m} \left(2^{-R_N}\right)^m \left(1 - 2^{-R_N}\right)^{\#S-m} ,$$

die Anzahl der Chiffretexte C mit $N_S(C) = m$ ist also gleich

$$\binom{\#S}{m} \left(2^{-R_N}\right)^m \left(1 - 2^{-R_N}\right)^{\#S-m} \#A^N .$$

Der durchschnittliche Wert von $H_{S|Ch}$ berechnet sich somit zu

$$\begin{aligned}
&\sum_{m=0}^{\#S} \binom{\#S}{m} \left(2^{-R_N}\right)^m \left(1 - 2^{-R_N}\right)^{\#S-m} \#A^N \frac{m}{\#O\#S} \log_2 m \\
&= \frac{\#A^N}{\#O} \frac{1}{\#S} \sum_{m=0}^{\#S} \binom{\#S}{m} \left(2^{-R_N}\right)^m \left(1 - 2^{-R_N}\right)^{\#S-m} m \log_2 m \\
&= \frac{2^{R_N}}{\#S} \sum_{m=0}^{\#S} \binom{\#S}{m} \left(2^{-R_N}\right)^m \left(1 - 2^{-R_N}\right)^{\#S-m} m \log_2 m .
\end{aligned}$$

Falls wir $\#S = 26$ setzen und von den beiden oben betrachteten Wahrscheinlichkeitsmodellen ausgehen, wenn wir also ein *zufälliges* Kryptosystem betrachten, das dieselben Parameter hat wie das System der 26 CAESAR-Substitutionen, läßt sich dies leicht berechnen; das Ergebnis ist für die beiden oben betrachteten Wahrscheinlichkeitsmodelle in den beiden folgenden Abbildungen dargestellt. Wie man sieht, ist die Mehrdeutigkeit des Schlüssels ab etwa zehn bis zwölf *bzw.* sechs bis acht Buchstaben Chiffretext praktisch gleich null, was auch ungefähr unseren Experimenten mit der CAESAR-Chiffre entspricht.

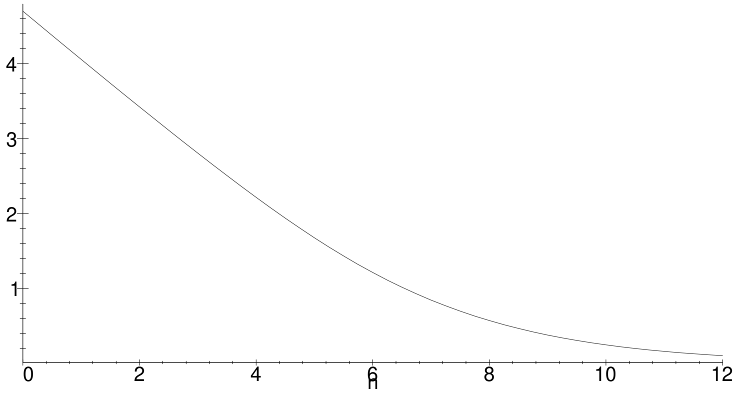

Abb. 1: Mehrdeutigkeit des Schlüssels via Buchstabenhäufigkeiten

Wie man den Kurven ansieht, nimmt der Mehrdeutigkeit des Schlüssels anfänglich etwa linear ab mit n. Wer die Binomialverteilung kennt, sieht auch leicht, warum dies so ist: Falls der Erwartungswert $2^{-R_N}\#S$ von $N_S(C)$ hinreichend groß ist, sind alle einigermaßen wahrscheinliche Werte in seiner näheren Umgebung, man kann daher ohne allzu großen Fehler $N_S(C)$ durch diesen Wert ersetzen und erhält

$$\begin{aligned}
H_{S|Ch} &= \sum_{C \in A^N} \frac{N_S(C)}{\#O \cdot \#S} \log_2 N_S(C) \\
&\approx \#A^N \cdot \frac{2^{-R_N}\#S}{\#O \cdot \#S}(-R_N + \log_2 \#S) \\
&= \log_2 \#S - R_N .
\end{aligned}$$

Anfänglich gewinnt man also aus N Buchstaben Chiffretext etwa R_N Bit des Schlüssels. Falls R_N wie in unseren Beispielen eine lineare Funktion von N ist, nimmt die Mehrdeutigkeit des Schlüssels also anfänglich linear ab.

Diese Rechnungen zeigen einmal mehr, daß ein Gegner vor allem von der Redundanz R_N der Nachrichtenquelle profitieren kann; je kleiner diese ist, desto weniger Information kann er gewinnen. Durch vorherige Datenkomprimierung läßt sich daher die Sicherheit der meisten Kryptosysteme deutlich erhöhen.