Kryptologie: Kryptographie und Kryptoanalyse

Kryptologie ist die Wissenschaft, die sich mit dem Ver- und Entschlüsseln von Informationen befasst.

Beispiel Internetkommunikation: Versiegeln (Integrität der Nachricht)
Signieren (Authentizität)
Verschlüsseln (Geheimhaltung)

Steganografie: Verdeckte Nachrichtenübermittlung in Texten, Bildern, Tönen

Techniken der Kryptographie

Prinzipien:

Diffusion: Der Zusammenhang zwischen Klartextzeichen und Geheimtextzeichen wird durch Substitutionen verwischt

Konfusion: Die Zeichen des Klartextes werden im Geheimtext durch Transposition verteilt.

Transposition:

Anagramm: Bsp: Lied - Leid, Palindrom: Leben, Nebel; Reliefpfeiler,... ;
Schüttelreim: Wir mögen einen schwarzen Wein und weniger das Warzenschwein):

Skytale

Transposition im Quadrat : 'Horizontal rein und vertikal raus' (Konfusion)

Monoalphabetische Codierung:

Codierungen durch positionsunabhängige Substitutionen

Codierung durch Geheimzeichen

a b c d e f g h i⓪ j① k② l③ m④ n⑤

ιυγενδφορυμΙνφορματικ – ΙΥΓΕΝΔΦΟΡΥΜΙΝΦΟΡΜΑΤΙΚ – JugendforumInformatik

Grundbegriffe: (z.B. bei Transpostion)

Klartext: plaintext
Verschlüsselungsfunktion (Algorithmus); i.a. nicht geheim!
Schlüssel: key - geheim!
Geheimtext: Ciphertext

Der gesamte Klartext wird mit einer 'festen' Permutation des Alphabets codiert,
d.h. 26! also ca. 403 Quadrillionen Möglichkeiten. Trotzdem nicht sicher!

Verschiebung (Caesar):

ciphertextzeichennummer c ; plaintextzeichennummer p; Schlüsselkonstante s
$c = p + s \mod 26$
Realisierung mit Chiffrierscheiben

Verschiedene Permutationsalgorithmen:

Beispiele: Schlüsselwort und restliches Alphabet angehängt
A B C D E F G H I J K L M N O P Q R S T U V X Y Z
F O R U M A BCDEG H ...

Multiplikative Chiffren C = p*s mod 26, s Primzahl

involutorische (wie XOR bei Bit-'Addition')

Bigrammsubstitution (Sonderfall einer polygraphischen Codierung):
Zeichenpaare des Klartextes werden durch Zeichenpaare im Geheimtext codiert.

Homophone Chiffren: Verschleierung der Häufigkeiten
Häufigen Klartextzeichen werden verschiedene Geheimtextzeichen zugeordnet, die in zufälliger! Reihenfolge eingesetzt werden.

Techniken der Kryptoanalyse

Geheimtextangriff (ciphertext-only attack)
Klartextangriff (known-plaintext attack)
Angriff mit ausgewähltem Klartext (chosen-plaintext attack)
Probleme / Chancen: lange Texte; wiederholte Texte; bekannte Themen, Begriffe, Anlässe;

Kriterien von C. E. Shannon:

Kryptoanalytische Sicherheit; Schlüssellänge; (De-)Chiffrieraufwand; Aufblähung des Geheimtextes; Verschleppung von Chiffrierfehlern;
Wie stellt man sich den 'optimalen' Kryptoanalytiker vor: unbegrenzt clever, ausdauernd, heimtückisch, mit großen Ressourcen (z.B. Rechenleistung), mit Kenntnis von Verschlüsselungsfunktion, Klar- und Geheimtext - aber ohne Kenntnis des Schlüssels!

Kryptoanalyse:

Jede Sprache hat eine eigene Anatomie in Form der Häufigkeitsgebirge für Buchstaben, Buchstabenpaare, oder Trigramme u.s.w.
Bei der monoalphabetischen Codierung bleiben charakteristische Häufigkeitsverteilungen erhalten.

Perfektes Chiffriersystem: One time pad ; System von Vernam

Jedes Zeichen des Klartextes der Länge n wird mit einem Zeichen des Schlüssels der Länge n codiert. (meist durch 'Addition' mod 26; bzw. bitweise binäre Addition)
Der Schlüssel wird genau einmal verwendet:

Fortlaufende Textstelle aus einem Buch
Berechenbare Zufallsfolge (der übermittele Schlüssel ist dann kleiner, aber nicht mehr sicher!)
(Automat mit gegebenen Zuständen produziert 'zufällige' 0,1 Folgen)

Ein neueres Verfahren wie z. B. die asymmetrische Codierung beim RSA-Verfahren ist ebenfalls sehr sicher und auch interessant.
Zuerst betrachten wir das klassische Verfahren der polyalphabetischen Codierung.

Polyalphabetische Codierung

Polyalphabetische Codierungen sind positionsabhängige Substitutionen, d.h. es wird mit wechselnder 'Taktik' codiert.
Das bekannteste Verfahren zur polyalphabetischen Codierung ist die Vigenére-Chiffrierung.
Für jedes Zeichen wird ein 'anderer' Schlüssel (z.B. Verschiebechiffren) verwendet.
Dadurch kann die Häufigkeitsverteilung der Zeichen einer Sprache verdeckt werden.
Beispiel:
Notiert man unter den Text:
'D i e s i s t e i n e p o l y a l p h a b e t i s c h e C o d i e r u n g' das Schlüsselwort '
J u g e n d f o r u m ….. so entsteht daraus:

M c k w v v y s z h q …..

Kryptoanalyse:

Wie sicher ist nun diese Verschlüsselung? Spätestens seit Anfang dieses Jahrhunderts war klar, dass diese Verschlüsselungstechnik angreifbar ist.
Der Knackpunkt bei der Analyse ist die Schlüsselwortlänge.
Wenn es gelingt die Schlüsselwortlänge zu finden, dann kann der durch die Schlüsselwortlänge neu strukturierte Text wieder mit Hilfe der statistischen Analyse von monoalphabetischen Codierungen untersucht werden.

Ermittlung der Schlüsselwortlänge:

KASISKI-Test (Kasiski war preußischer Infanteriemajor im 19 Jhd.)
Grundgedanke: Ermittlung der Schlüsselwortlänge - bis auf ein Vielfaches - durch 'Parallelstellensuche'.
Klartext-Buchstaben die mit den gleichen Schlüsselwort-Buchstaben verschlüsselt werden, ergeben den gleichen Geheimtext-Buchstaben. Wiederholen sich nun Folgen von Buchstaben des Klartextes zufälligerweise in dem 'Takt des Schlüsselwortes - also nur abhängig von der Schlüsselwortlänge - , dann wiederholen sich auch die zugehörigen Geheimtextbuchstaben.
Wenn man im Geheimtext solche Wiederholungssequenzen der Länge größer 2 im Abstand **d** findet, kann man vermuten, dass die Schlüsselwortlänge ein Teiler von **d** ist.

FRIEDMAN-Test (Entwicklung durch Colonel W. Friedman Anfang dieses! Jahrhunderts)

Ziel: Ermittlung der ungefähren Schlüsselwortlänge.
Grundgedanken des Verfahrens:
Entnehme aus dem vorgegebenen Text zwei zufällig ausgewählte, - nicht unbedingt nebeneinanderliegende - Buchstaben.
Mit welcher Wahrscheinlichkeit besteht das gebildete Paar aus gleichen Buchstaben?
Im gegebenen Text mit n Buchstaben sind n_1 Buchstaben a, n_2 Buchstaben b, ... und n_{26} Buchstaben z.
Dann ist die Anzahl der möglichen Paare 'aa' gleich $n_1 (n_1-1)/2$; entsprechendes gilt für die anderen Buchstaben. Insgesamt ergibt sich der Koinzidenzindex K
(vgl. Kappa bei Friedman 1922, index of coincidence; IC)

$$\text{probability(aa oder bb oder...zz)} = K = \frac{\sum_{i=1}^{26} n_i \cdot (n_i - 1)}{n(n-1)}$$

K ist also das Verhältnis der Anzahl der Paare aus gleichen Buchstaben und der Anzahl aller möglichen Paare.

Dieses Verhältnis kann gemäß der Wahrscheinlichkeitsverteilung der Buchstaben im Text einer festgewählten Sprache allgemein berechnet werden.
Sei die Wahrscheinlichkeit für a gleich p_1, für b gleich p_2 , dann ergibt sich:
Die Wahrscheinlichkeit dafür, dass zwei willkürlich herausgegriffene Buchstaben gleich sind ist $K = \sum_{i=1}^{26} p_i^2$. Dieser Wert ist für verschiedene Sprachen jeweils charakteristisch:
Deutsch K = 0.0762 ; Englisch: K = 0.0658 ; Französisch K = 0.0778

Bei einem zufällig gewürfelten Buchstabensalat ergibt sich K mit $\sum_{i=1}^{26} \frac{1}{26^2} = 0.0385$

d.h. je unregelmäßiger ein Text ist, desto kleiner ist K, mit Minimum 0.0385
Eine monoalphabetische Codierung ändert nichts an K ! Dies ermöglicht einen Test auf den Codierungstyp.

Vermutet man bei einem Text eine polyalphabetische Codierung mit der Schlüsselwortlänge p , so kann der Text zeilenweise in p Spalten geschrieben werden. Die Zeichen einer Spalte i sind dann durch monoalphabetische Codierung (z.B. durch Caesar-Chiffre) mit dem i-ten Schlüsselwortbuchstaben entstanden. Die Chance in einer Spalte ein paar aus gleichen Buchstaben zu treffen ist somit im deutschen Text ca. 0.0762.
Betrachtet man Paare von Buchstaben aus verschiedenen Spalten, so wird ein Paar nur zufällig! aus gleichen Buchstaben bestehen, also mit der Wahrscheinlichkeit von ca. 0.0385
(Es sei denn man muss/kann bei einem langen Schlüsselwort dessen statistische Daten berücksichtigen)

Bisher können wir den Koinzidenzwert für einen gegebenen Text mit der Formel

$$K = \frac{\sum_{i=1}^{26} n_i \cdot (n_i - 1)}{n(n-1)}$$ berechnen.

Man könnte nun den Text für verschiedene Schlüsselwortlängen p in Spalten zeilenweise notieren und jeweils den Koinzidenzwert K der Spalten bestimmen. Diejenige Einteilung die den größten Wert für K liefert ergibt das gesuchte p.

Aus der Suche nach einem formelmäßigen Zusammenhang zwischen K und p berechnen wir den Koinzidenzindex des Textes aus n Zeichen durch eine Fallunterscheidung auf eine neue Art und Weise.
In jeder Spalte stehen n/p Buchstaben. (Für große n ohne Beachtung von Rundungsfehlern)
Um einen Buchstaben zu wählen gibt es n Möglichkeiten und $\frac{n}{p} - 1$ Möglichkeiten einen zweiten Buchstaben in derselben Spalte zu wählen. Es können also $n \cdot (\frac{n}{p} - 1) / 2$ Paare von Buchstaben einer Spalte gebildet werden. Da es gerade $n - \frac{n}{p}$ Buchstaben außerhalb einer festen Spalte sind insgesamt $n \cdot (n - \frac{n}{p}) / 2$ Paare von Buchstaben aus verschiedenen Spalten zu bilden.

Für die Anzahl der Buchstabenpaare aus *gleichen* Buchstaben gilt daher:

$$A = n \cdot (\frac{n}{p} - 1) / 2 \bullet 0.0762 + n \cdot (n - \frac{n}{p}) / 2 \bullet 0.0385 = \frac{n(n-p)}{2p} \cdot 0.0762 + \frac{n^2(p-1)}{2p} \cdot 0.0385$$

Die Wahrscheinlichkeit P dafür , ein Paar aus gleichen Buchstaben zu treffen ist somit gleich $\frac{A}{n(n-1)/2}$, also gilt $P = \frac{n-p}{p(n-1)} \cdot 0.0762 + \frac{n(p-1)}{p(n-1)} \cdot 0.0385$

$$= \frac{1}{p(n-1)} [0.0377\, n + p(0.0385\, n - 0.0762)]$$

Der schon auf allgemeine Art berechnete Koinzidenzindex K eines Textes entspricht aber dem soeben bestimmten Wert P.

Die Formel $K = \frac{1}{p(n-1)} [0.0377\, n + p(0.0385\, n - 0.0762)]$ führt durch Umformen und Auflösen nach der Schlüsselwortlänge p auf das Ergebnis:

$p = \frac{0.0377}{(n-1)K - 0.0385\, n + 0.0762}$, wobei n die Anzahl der Buchstaben des Textes ist und K wie oben beschrieben berechnet wird mit $K = \frac{\sum_{i=1}^{26} n_i \cdot (n_i - 1)}{n(n-1)}$.

Bei einem konkreten Text brauchen also nur die Häufigkeiten n_i der einzelnen Buchstaben bestimmt werden, um K und somit die Schlüsselwortlänge p zu bestimmen.

Je kleiner K ist, um so größer muss die Schlüsselwortlänge p sein. Es zeigt sich also, dass der Koinzidenzindex mit der Schlüsselwortlänge korreliert ist und aus dem Wert von K ein Wert für p ermittelt werden kann.