Friedmann und Kasiski gegen Vigenere

Obwohl die Häufigkeitsanalyse bei der Vigenere-Verschlüsselung zunächst sinnlos erscheint, ist es Kasiski und auch Friedmann gelungen, mit zwei unterschiedlichen Verfahren Vigenere-verschlüsselte Texte zu „knacken".
Die Grundidee dieser beiden Verfahren ist folgende: Ist die Länge *n* des Schlüsselwortes bekannt, dann kann man die Vigenere-Verschlüsselung auf *n* monoalphabetische Verschlüsselungen zurückführen, die jede für sich mit der Häufigkeitsanalyse „geknackt" werden kann.
Denn der *erste* Geheimtextbuchstabe, der *n+1.*, der *2n+1.*, der *3n+1.* .... stehen unter demselben Schlüsselbuchstaben und werden somit mit derselben Caesar-Verschiebung codiert. Dasselbe gilt für den *2.*, den *n+2.*, den *2n+2.* Geheimtextbuchstaben usw.
Die Frage reduziert sich somit auf das Finden der Schlüsselwortlänge.
Im 19. Jh wurde der **Kasiski-Test** entwickelt, dessen Grundidee folgende ist:
Gleiche Geheimbuchstabenfolgen (mind. 3 Buchstaben) sind höchstwahrscheinlich gleiche Klartextfolgen, die an gleicher Stelle unter dem Schlüsselwort stehen. Der Zeichenabstand zwischen diesen Wiederholungen ist dann ein Vielfaches der Schlüssellänge.
Beispiel:

stt woyej lllkisef tfmekc fatr ek gy mazeef oy dwx yaune lskftwzp dsy eedkqof jceasll, mto dak dtasxe ss lnvkcef kydw lcayzp nsis jwslnvkx, dwx pr fonhl clr nopl kvlelkc, ady pr at oej rlgw clr, vgcuwhpr fgnh ra oefqpn, ogd maz the mpsunlh, kuwllk pr ra oee ynhdads cuxmwt, yiunes aye waxvlais amydej jpm raqadr lbwx oak clr nopl kvlelkc ae gyfstr wsxpn woyfsis nmx oak kceamyik ayd kktnw lzlyky

(Text aus: Albrecht Beutelspacher, Heike Neumann, Thomas Schwarzpaul: „Kryptographie in Theorie und Praxis, vieweg-Verlag, 2005)

Dieser Text ist Vigenere-verschlüsselt. Die farbigen Zeichenfolgen sind höchstwahrscheinlich auch gleicheKlartextfolgen, die an gleicher Stelle zwischen dem Schlüsselwort stehen. Zwischen den ersten beiden Folgen „clr" beträgt der Zeichenabstand 28, so dass das Schlüsselwort die Länge 2, 4, 7, 14 oder 28 haben kann. Der Abstand zwischen der zweiten und dritten Zeichenfolge „clr" beträgt 104. Als Schüssellängen kommen jetzt nur noch 2 und 4 in Frage, denn 28, 14 und 7 sind keine Teiler von 104. So kann man nach und nach die Schlüssellänge eingrenzen, die hier in der Tat 4 beträgt.
Der **Friedmann-Test** ist weniger intuitiv aber besser zu berechnen und zu implementieren.
Wenn man berechnen will, wie groß die Wahrscheinlichkeit ist, in einem Text zufällig zwei gleiche Buchstaben zu ziehen, dann ist das (bei der Wahrscheinlichkeit p_i für das Auftreten des *i*-ten Buchstabens) $\sum_{i=1}^{26} p_i^2$. In einem deutschen Text kommen die Buchstaben unterschiedlich häufig vor und es gilt: $\sum_{i=1}^{26} p_i^2 = 0{,}0762$. In einem Text, in dem

jeder Buchstabe etwa gleich häufig auftritt gilt: $\sum_{i=1}^{26} p_i^2 = \sum_{i=1}^{26} \left(\frac{1}{26}\right)^2 = 0{,}0385$. Bei monoalphabetisch verschlüsselten Texten ist die Wahrscheinlichkeit zweimal denselben Buchstaben zu ziehen also 0,0762, bei polyalphabetisch verschlüsselten Texten im besten Fall 0,0385.

Jetzt berechnet man für den vorliegenden Geheimtext diesen sogenannten Friedmannschen Koinzidenzindex. Wie groß ist hier die Wahrscheinlichkeit zwei mal denselben Buchstaben zu ziehen? Ist der Geheimtext n Zeichen lang und sei n_i die absolute Häufigkeit des auftretend des i-ten Buchstabens, dann gibt es $\frac{n_i(n_i-1)}{2}$ Paare gleicher Buchstaben. Also ist die Wahrscheinlichkeit, dass man genau den i-ten Buchstaben zwei mal zieht: $\frac{\frac{n_i(n_i-1)}{2}}{\frac{n(n-1)}{2}}$. Summiert man über die 26 Buchstaben, so ergibt sich als Wahrscheinlichkeit, dass man irgendeinen Buchstaben im vorliegenden Geheimtext zwei mal zieht der Friedmannsche Koinzidenzindex des Geheimtextes:

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

Hat man den Friedmannschen Koinzidenzindex des Geheimtextes berechnet, dann kann man mit Hilfe der Friedmann-Formel:

$$l = \frac{0{,}0377n}{(n-1)\cdot I - 0{,}0385n + 0{,}0762}$$

die Schlüssellänge l, bzw. eine gute Annäherung an die Schlüssellänge berechnen.
Vielleicht noch ein paar Worte, woher diese Formel plötzlich kommt.
Der Koinzidenzindex gibt die Wahrscheinlichkeit an, aus dem vorliegenden Txt zwei mal denselben Buchstaben zu ziehen. Wenn wir jetzt aber wüssten, dass die Schlüssellänge l ist, dann ist die Wahrscheinlichkeit 0,0762, wenn die Buchstaben an derselben Stelle unter dem Schlüsselwort stehen. Nun gibt es jeweils $\frac{n}{l}$ Buchstaben, die an derselben Stelle unter dem Schlüsselwort stehen. Das heißt, die Anzahl von gleichen Buchstabenpaaren, die an derselben Stelle unter dem Schlüsselwort stehen ist: $l \cdot \frac{\frac{n}{l}\cdot\left(\frac{n}{l}-1\right)}{2} = \frac{n\cdot\left(\frac{n}{l}-1\right)}{2}$. Die Anzahl von gleichen Buchstabenpaaren, die an unterschiedlichen Stellen unter dem Schlüsselwort stehen ist aber

$\frac{n\left(n-\frac{n}{l}\right)}{2}$. Die gesamt zu erwartenden Anzahl von gleichen Buchstabenpaaren ist damit:

$0{,}0762 \cdot \frac{n \cdot \left(\frac{n}{l} - 1\right)}{2} + 0{,}0385 \cdot \frac{n\left(n - \frac{n}{l}\right)}{2}$ oder, wenn es um die Wahrscheinlichkeit gleicher Buchstabenpaare im gesamten Text geht:

$I = \frac{0{,}0762 \cdot \frac{n \cdot \left(\frac{n}{l} - 1\right)}{2} + 0{,}0385 \cdot \frac{n\left(n - \frac{n}{l}\right)}{2}}{\frac{n(n-1)}{2}}$. Stellt man diese Formel geschickt um und löst sie nach l auf, kommen wir zu der oben genannten.

Wir haben also eine Formel kennen gelernt, mit der man die Schlüssellänge berechnen kann. Dies funktioniert nicht immer ganz exakt und manchmal ergeben sich auch gar keine natürlichen Zahlen, wie im Text oben, der nach dem Friedmann-Test eine Schlüssellänge von 4,3 hat. In Kombination mit dem Kasiski-Test kann man aber gute Ergebnisse erzielen.