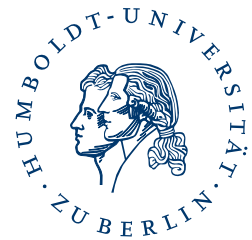
Einführung in die Kryptologie

Johannes Köbler



Institut für Informatik
Humboldt-Universität zu Berlin

SS 2022

Kerckhoffs' Prinzip

- Die Erfolgsaussichten eines Angriffs gegen ein Kryptosystem hängen sehr stark von der Ausgangslage des Angreifers ab
- Prinzipiell sollte man weder die Fähigkeiten des Gegners noch die Unvorsichtigkeit der Anwender von Kryptosystemen unterschätzen
- Bereits vor mehr als einem Jahrhundert postulierte Kerckhoffs, dass die Frage der Sicherheit nicht von irgendwelchen obskuren Annahmen über den Wissensstand des Gegners abhängig gemacht werden sollte

Goldene Regel für Kryptosystem-Designer

Unterschätze niemals den Kryptoanalytiker. Gehe insbesondere immer von der Annahme aus, dass dem Gegner das angewandte System bekannt ist.

- Tatsächlich sind die Prinzipien fast aller heute im Einsatz befindlichen Kryptosysteme bekannt
- Nur so kann einer Vielzahl von Kryptoanalytikern die Suche nach Schwachstellen ermöglicht werden

Klassifikation von Angriffen gegen Kryptosysteme

Die folgende Liste enthält eine Auswahl von Angriffsszenarien mit zunehmender Gefährlichkeit

Angriff bei bekanntem Kryptotext (ciphertext-only attack)

- Der Gegner fängt Kryptotexte ab und versucht, Rückschlüsse auf den zugehörigen Klartext oder den benutzten Schlüssel zu ziehen

Angriff bei bekanntem Klartext (known-plaintext attack)

- Der Gegner ist im Besitz von einigen zusammengehörigen Klartext-Kryptotext-Paaren

Angriff bei frei wählbarem Klartext (chosen-plaintext attack)

- Der Gegner ist (zumindest vorübergehend) in der Lage, sich zu Klartexten seiner Wahl die zugehörigen Kryptotexte zu besorgen
- Kann hierbei die Wahl der Klartexte in Abhängigkeit von zuvor erhaltenen Verschlüsselungsergebnissen getroffen werden, so spricht man von einem Angriff bei adaptiv wählbarem Klartext

Klassifikation von Angriffen gegen Kryptosysteme

Angriff bei frei wählbarem Kryptotext (chosen-ciphertext attack)

- Vor der Beobachtung des zu entschlüsselnden Kryptotextes konnte sich der Gegner zu Kryptotexten seiner Wahl die zugehörigen Klartexte besorgen, ohne dabei jedoch in den Besitz des Dechiffrierschlüssels zu kommen (Mitternachtsattacke)
- Das dabei erworbene Wissen steht ihm nun bei der Durchführung seines Angriffs zur Verfügung
- Auch in diesem Fall können sich die Erfolgsaussichten des Gegners erhöhen, wenn ein Angriff bei adaptiv wählbarem Kryptotext (adaptive chosen-ciphertext attack) möglich ist, also der Kryptotext in Abhängigkeit von den zuvor erzielten Entschlüsselungsergebnissen wählbar ist

Angriff bei frei bzw. adaptiv wählbarem Text (chosen-text attack)

- Sowohl Klartexte als auch Kryptotexte sind frei bzw. adaptiv wählbar

Klassifikation von Angriffen gegen Kryptosysteme

- Ohne Frage ist ein Kryptosystem, das bereits bei einem Angriff mit bekanntem Kryptotext Schwächen zeigt, kaum brauchbar
- Tatsächlich müssen aber an ein praxistaugliches Kryptosystem noch weit höhere Anforderungen gestellt werden
- Denn häufig unterlaufen den Anwendern sogenannte Chiffrierfehler, die den Gegner oft in eine deutlich bessere Ausgangsposition versetzen
- So ermöglicht beispielsweise das Auftreten stereotyper Klartext-Formulierungen einen Angriff bei bekanntem Klartext, sofern der Gegner diese Formulierungen kennt bzw. errät
- Begünstigt durch solche Unvorsichtigkeiten, die im praktischen Einsatz häufig vorkommen, können sich bereits winzige Konstruktionsschwächen leicht zu einer ernsthaften Bedrohung auswachsen
- Wie die Geschichte der Kryptografie eindrucksvoll belegt, sind es häufig die Anwender selbst, die – im unerschütterlichen Glauben an die kryptografische Stärke ihres Verfahrens – den Erfolg eines Angriffs ermöglichen

Klassifikation von Angriffen gegen Kryptosysteme

- Zusammenfassend lässt sich also festhalten, dass die Gefährlichkeit von Angriffen, denen ein Kryptosystem im praktischen Einsatz ausgesetzt ist, kaum zu überschätzen ist
- Andererseits kann selbst das beste Kryptosystem keinen Schutz vor einer unbefugten Dechiffrierung bieten, wenn es dem Gegner etwa gelingt, in den Besitz des geheimen Schlüssels zu kommen – sei es aus Unachtsamkeit der Anwender oder infolge von Manipulationsversuchen von Seiten des Gegners (Social Engineering bzw. Social Hacking)
- Auch Implementierungsangriffe nutzen nicht direkt Schwachstellen des Kryptoverfahrens aus, sondern zielen vielmehr darauf ab, durch physikalische Messungen wie bspw. des Stromverbrauchs oder der Laufzeit von Berechnungen (sog. Seitenkanalangriffe) Informationen über den unbekannten Schlüssel zu gewinnen

Vollständige Schlüsselsuche

Manche der bisher betrachteten Chiffrierverfahren haben einen so kleinen Schlüsselraum, dass ohne großen Aufwand eine vollständige Schlüsselsuche (auch Brute-Force Angriff genannt) möglich ist

Beispiel

- Es sei bekannt, dass das Kryptotextstück $y = saxp$ mit einer additiven Chiffre erzeugt wurde ($K = A = B = A_{lat}$)
- Entschlüsseln wir y probeweise mit allen möglichen Schlüsselwerten, so erhalten wir folgende Zeichenketten:

k	A	B	C	D	E	F	G	H	I	J	K	L	M
$D(k, y)$	SAXP	RZWO	QYVN	PXUM	OWTL	NVSK	MURJ	LTQI	KSPH	JROG	IQNF	HPME	GOLD

	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
	FNKC	EMJB	DLIA	CKHZ	BJGY	AIFX	ZHEW	YGDV	XFCU	WEBT	VDAS	UCZR	TBYQ

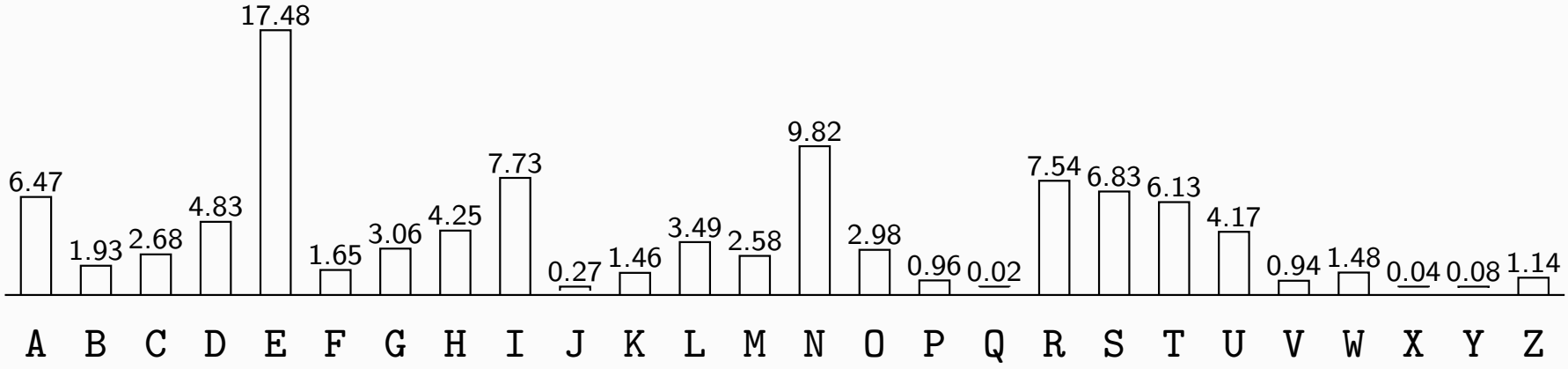
- Unter diesen springen vor allem die beiden Klartextkandidaten $x =$ GOLD (Schlüsselwert $k = M$) und $x =$ WEBT ($k = W$) ins Auge ◁

Vollständige Schlüsselsuche

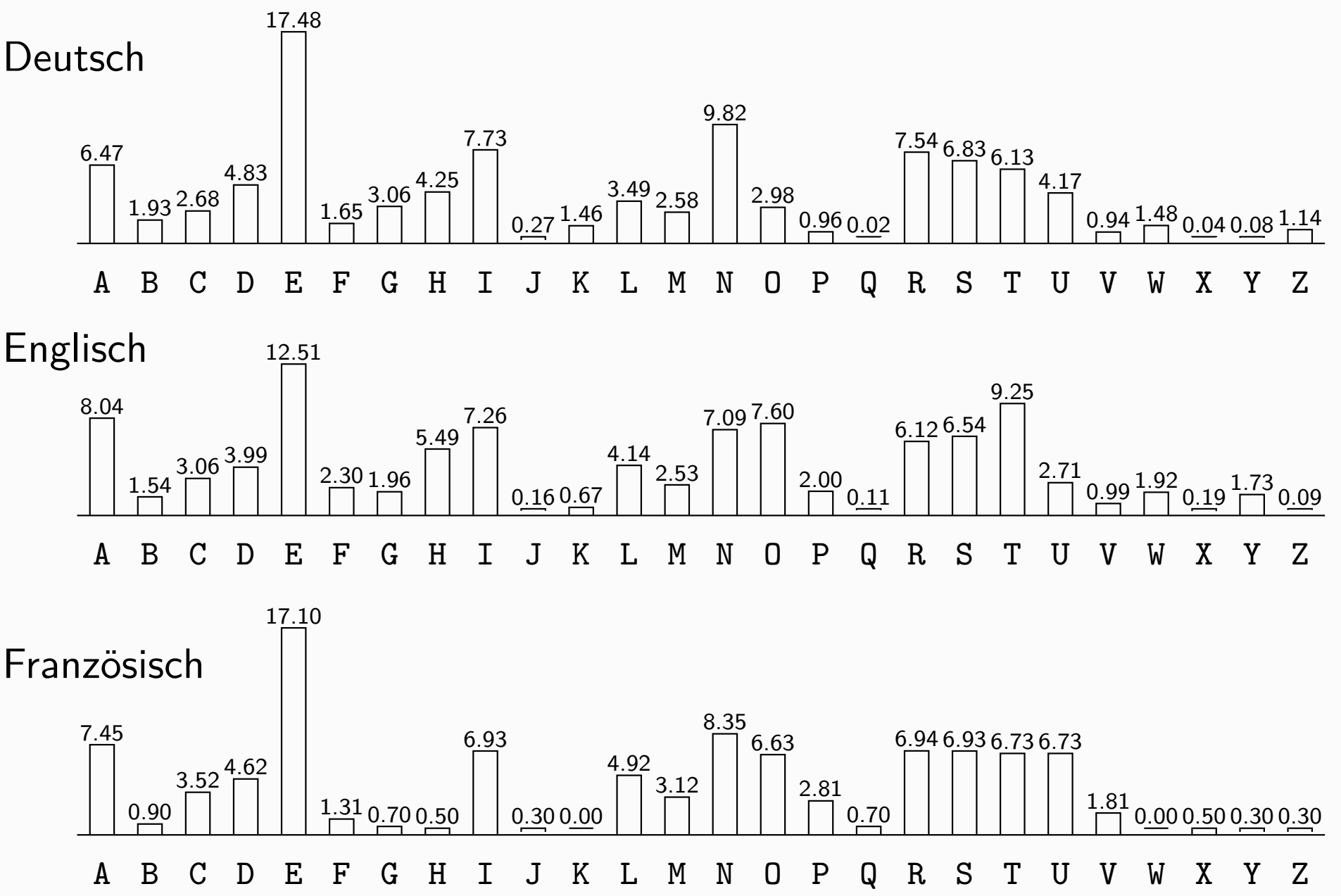
- Ist $s = |K|$ die Größe des Schlüsselraums, so kann der Gegner bei bekanntem Kryptotext y die Suche nach dem zugehörigen Klartext x auf eine Menge von maximal s Texten $x_1, \ldots, x_s$ beschränken
- Daneben hat der Gegner meist ein *a priori* Wissen über den Klartext
- Weiß er zum Beispiel, dass er in deutscher Sprache verfasst ist, kann er einen Großteil der Texte x_i ausschließen
- Mit jedem Text x_i, der nicht als Klartext infrage kommt, kann auch mindestens ein Schlüssel ausgeschlossen werden
- Sind noch mehrere Schlüsselwerte möglich, so kann weiteres Kryptotextmaterial Klarheit bringen
- Manchmal hilft auch eine Inspektion der verbliebenen Schlüsselwerte weiter, etwa wenn der Schlüssel nicht rein zufällig erzeugt wurde, sondern aus einem einprägsamen Schlüsselwort ableitbar ist
- Auch wenn der Gegner die Klartextsprache nicht kennt, kann eine Häufigkeitsanalyse erfolgreich sein

Häufigkeitsanalyse

- Mit zunehmender Länge gleichen sich die Häufigkeitsverteilungen der Buchstaben in natürlichsprachigen Texten einer „Grenzverteilung“ an, die in erster Linie von der benutzten Sprache und nur in geringem Umfang von der Art des Textes abhängt
- Selbst zwischen unterschiedlichen Sprachen gibt es oft Gemeinsamkeiten
- So kommt in fast allen europäischen Sprachen der Buchstabe `E` sehr häufig vor, während `X`, `Y` und `Z` nur selten auftreten
- Die Ungleichmäßigkeit der Buchstabenhäufigkeiten ist darauf zurückzuführen, dass natürliche Sprachen relativ viel Redundanz enthalten



Häufigkeitsverteilung von Einzelbuchstaben (in %)



Häufigkeitsanalyse von einfachen Substitutionen

- Durch eine Häufigkeitsanalyse können insbesondere einfache Substitutionen g leicht gebrochen werden
- Der Grund dafür ist, dass die einzelnen Zeichen a in der Klartextsprache meist mit unterschiedlichen Wahrscheinlichkeiten $p(a)$ auftreten

	Deutsch	Englisch	Französisch
sehr häufig	E	E	E
häufig	N | I R S | A T	T | A O I N | S R H	A S | T I | R N | U
durchschnittlich	D H U | L G O | C M	L D | C U M F	L O | D M C P
selten	B F W K Z | P V	P G W Y B | V K	V | H G F B Q | J
sehr selten	J Y X Q	X J Q Z	X | Z Y K | W

- Berechnet man die relativen Häufigkeiten h der Zeichen im Kryptotext, so gilt $p(a) \approx h(g(a))$ (vorausgesetzt der Klartext ist genügend lang)
- Eine nach dieser Methode durchgeführte Kryptoanalyse wird in der Erzählung „Der Goldkäfer“ von Edgar Allan Poe geschildert

Häufigkeitsanalyse bei der additiven Chiffre

- Ein typischer deutscher Text besteht zu 62% aus den sieben Zeichen `E`, `N`, `I`, `R`, `S`, `A`, `T` (das sind nicht einmal 27% aller Zeichen)
- Bei additiven Chiffren lässt sich der Schlüssel k meist schon bestimmen, indem man die Differenz zwischen dem häufigsten Buchstaben im Kryptotext und dem häufigsten Buchstaben der Klartextsprache bildet
- Bei affinen Chiffren reicht es dagegen, die beiden häufigsten Buchstaben zu bestimmen, um den Schlüssel $k = (b, c)$ zu ermitteln

Häufigkeitsanalyse bei der affinen Chiffre

Beispiel

- Es sei bekannt, dass sich hinter dem Kryptotext

  $y =$ `laoea ehoap hwvae ixobg jcbho thlob lokhe ixope vbcix ockix qoppo boapo mohqc euogk opeho jhkpl eappj seobe ixoap opmcu`

  ein mit einer affinen Chiffre verschlüsselter deutscher Klartext verbirgt
- Die Einzelnen Zeichen y_i treten in y mit folg. Häufigkeiten $H_y(y_i)$ auf

y_i	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
$H_y(y_i)$	7	6	5	0	10	0	2	8	5	3	4	4	2	0	19	11	2	0	1	1	2	2	1	5	0	0

- Daher vermuten wir, dass das häufigste in y vorkommende Zeichen o für das Klartextzeichen `E` und das zweithäufigste Zeichen p für `N` steht
- Dies führt auf folgendes Gleichungssystem für den Schlüssel $k = (b, c)$:

$$\begin{aligned} b \cdot \mathtt{E} + c &= o \\ b \cdot \mathtt{N} + c &= p \end{aligned}$$

Häufigkeitsanalyse bei der affinen Chiffre

Beispiel

- Dies führt auf folgendes Gleichungssystem für den Schlüssel $k = (b, c)$:

$$b \cdot \mathtt{E} + c = o$$
$$b \cdot \mathtt{N} + c = p$$

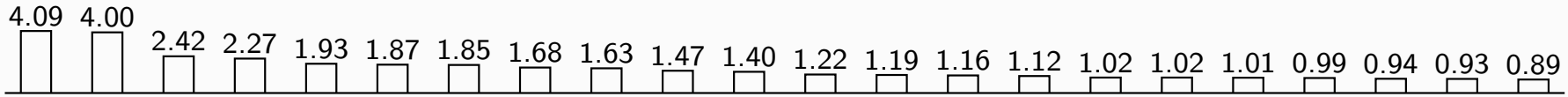
- Subtrahieren wir die erste von der zweiten Gleichung, so erhalten wir die Kongruenz $9 \cdot b \equiv_{26} 1$, woraus sich $b = 3$ und damit $c = 2$ ergibt
- Tatsächlich stimmen für $k = (3, 2)$ die Häufigkeiten $H_y(y_i)$ und die gerundeten erwarteten Häufigkeiten $H_{100}(D(k, y_i))$ von $D(k, y_i)$ in einem typischen deutschen Text der Länge 100 nicht nur für $y_i \in \{o, p)$ sehr gut überein, sondern auch für alle übrigen Zeichen:

y_i	o	p	e	h	a	b	c	x	i	l	k	j	u	m	g	v	q	s	t	w	r	f	n	z	y	d
$H_y(y_i)$	19	11	10	8	7	6	5	5	5	4	4	3	2	2	2	2	2	1	1	1	0	0	0	0	0	0
$H_{100}(D(k, y_i))$	17	10	7	6	8	8	6	4	3	5	4	3	3	3	1	1	1	3	0	0	2	2	1	1	0	0
$D(k, y_i)$	E	N	S	T	I	R	A	H	C	D	U	L	G	M	K	P	W	O	X	Y	F	B	V	Z	Q	J

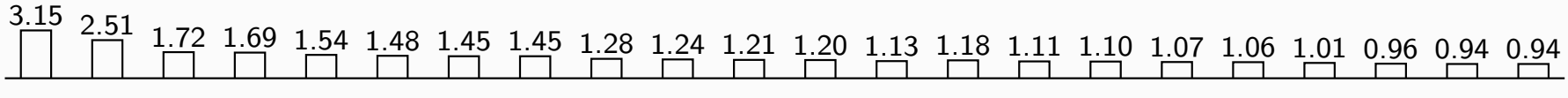
◁

Kryptoanalyse von Blocktranspositionen

- Blocktranspositionen lassen sich mit Hilfe von Bigrammhäufigkeiten brechen, die manchmal auch als Kontakthäufigkeiten bezeichnet werden
- Die häufigsten Bigramme im Deutschen und im Englischen (in %):



ER EN CH DE EI ND TE IN IE GE ES NE UN ST RE HE AN BE SE NG DI SC



TH HE AN IN ER RE ON ES TI AT ST EN OR ND TO NT ED IS AR OU OF TE

- Ist die Blocklänge ℓ bekannt, so schreibt man den Kryptotext zeilenweise in eine Matrix $S = (s_{ij}) = (S_1 \ldots S_\ell)$ mit ℓ Spalten $S_1, \ldots, S_\ell$
- Da jede Zeile dieser Matrix aus dem zugehörigen Klartextblock mit derselben Permutation π erzeugt wurde, müssen die Spalten S_j jetzt nur noch in die „richtige“ Reihenfolge gebracht werden
- Die Nachfolgespalte S_k von S_j (bzw. die Vorgängerspalte S_j von S_k) kann anhand der Werte von $\hat{p}(S_j, S_k) = \sum_i p(s_{ij}, s_{ik})$ bestimmt werden

Kryptoanalyse von Blocktranspositionen

Beispiel (Häufigkeitsanalyse von Bigrammen)

- Für den mit einer Blocktransposition (mit $\ell = 5$) erzeugten Kryptotext

  $$y = \texttt{ihehr bwean rneii nrkeu elnzk rxtae vlotr engie}$$

  erhalten wir eine Matrix S mit den folgenden fünf Spalten:

S_1	S_2	S_3	S_4	S_5
I	H	E	H	R
B	W	E	A	N
R	N	E	I	I
N	R	K	E	U
E	L	N	Z	K
R	X	T	A	E
V	L	O	T	R
E	N	G	I	E

- Um die richtige Vorgänger- oder Nachfolgerspalte von S_1 zu finden, bestimmen wir für jede potentielle Spalte S_j, $j = 2, \ldots, 5$, wieviele der Bigramme $s_{ij}s_{i1}$ (bzw. $s_{i1}s_{ij}$) zu den 20 häufigsten gehören:

Kryptoanalyse von Blocktranspositionen

Beispiel (Häufigkeitsanalyse von Bigrammen)

- Um die richtige Vorgänger- oder Nachfolgerspalte von S_1 zu finden, bestimmen wir für jede potentielle Spalte S_j, $j = 2, \ldots, 5$, wieviele der Bigramme $s_{ij}s_{i1}$ (bzw. $s_{i1}s_{ij}$) zu den 20 häufigsten gehören:

				↓		↓		
S_2	S_3	S_4	S_5	S_1	S_2	S_3	S_4	S_5
H	**E**	H	R	**I**	H	**E**	H	R
W	E	A	N	**B**	W	**E**	A	N
N	**E**	I	I	**R**	N	**E**	I	I
R	K	**E**	**U**	**N**	R	K	**E**	U
L	**N**	Z	K	**E**	L	**N**	Z	K
X	T	A	**E**	**R**	X	T	A	**E**
L	O	T	R	V	L	O	T	R
N	**G**	**I**	E	**E**	**N**	G	**I**	E
1	4	2	2		1	4	2	1

- Da die Paare (S_3, S_1) und (S_1, S_3) jeweils vier häufige Bigramme bilden, nehmen wir an, dass im Klartext S_1 auf S_3 oder S_3 auf S_1 folgen muss

Kryptoanalyse von Blocktranspositionen

Beispiel (Häufigkeitsanalyse von Bigrammen)

- Wir entscheiden uns für die zweite Möglichkeit und betrachten nun die Spaltenpaare (S_j, S_1) und (S_3, S_j), $j = 2, 4, 5$:

				↓			↓
S_2	S_4	S_5	S_1	S_3	S_2	S_4	S_5
H	H	R	I	**E**	H	H	**R**
W	A	N	B	**E**	W	A	**N**
N	I	I	R	**E**	**N**	**I**	**I**
R	**E**	**U**	**N**	K	R	E	U
L	Z	K	E	N	L	Z	K
X	A	**E**	**R**	**T**	X	A	**E**
L	T	R	V	O	L	T	R
N	**I**	E	**E**	**G**	N	I	**E**
1	2	2			1	1	5

- Da der Wert von $\hat{p}(S_3, S_5)$ hoch ist, nehmen wir an, dass auf S_3 die Spalte S_5 folgt

Kryptoanalyse von Blocktranspositionen

Beispiel (Häufigkeitsanalyse von Bigrammen)

- Im nächsten Schritt erhalten wir daher die folgende Tabelle:

		↓	↓		↓	↓
S_2	S_4	S_1	S_3	S_5	S_2	S_4
H	H	I	E	R	H	H
W	A	B	E	N	W	A
N	I	R	E	I	N	I
R	E	N	K	U	R	E
L	Z	E	N	K	L	Z
X	A	R	T	E	X	A
L	T	V	O	R	L	T
N	I	E	G	E	N	I
1	2				2	1

- Diese lässt die Spaltenanordnung S_4, S_1, S_3, S_5, S_2 vermuten, welche tatsächlich auf den gesuchten Klartext führt:

$$x = \texttt{HIER HABEN WIR EINEN KURZEN KLARTEXT VORLIEGEN}$$

◁

Kryptoanalyse von polygrafischen Chiffren

- Blocksysteme mit kleiner Blocklänge ℓ (z.B. bigrafische Systeme) lassen sich wie einfache Substitutionen durch Häufigkeitsanalysen brechen
- Wird bei Hill-Chiffren ℓ sehr groß gewählt, ist eine solche statistische Analyse nicht mehr möglich
- Das Hill-System kann dann zwar einem Kryptotextangriff widerstehen, jedoch nicht einem Angriff mit gewähltem Klartext

Beispiel (Angriff mit gewähltem Klartext auf die Hill-Chiffre)

- Bei einem GK-Angriff verschafft sich der Gegner den Kryptotext $g(x)$ zu den ℓ Blöcken $x \in \{100\ldots0, \ldots, 0\ldots001\}$, um die Schlüsselmatrix zu erhalten (oBdA. sei $A = \{0, \ldots, m-1\}$):

$$\begin{aligned} g(100\ldots0) &= k_{11}\, k_{12} \ldots k_{1\ell} \\ g(010\ldots0) &= k_{21}\, k_{22} \ldots k_{2\ell} \\ &\vdots \\ g(0\ldots001) &= k_{\ell 1}\, k_{\ell 2} \ldots k_{l\ell} \end{aligned}$$

◁

Auch einem Angriff mit bekanntem Klartext kann die Hill-Chiffre kaum widerstehen

Beispiel (Angriff mit bekanntem Klartext auf die Hill-Chiffre)

- Ist bei einem BK-Angriff eine ausreichende Menge von Klartext-Kryptotextpaaren (x_i, y_i), $i = 1, \ldots, \mu$ bekannt, so kann das Hill-System wie folgt gebrochen werden
- Wir suchen nach l Blöcken x_i, $i \in I$, so dass für die aus diesen Zeilen gebildete Matrix X die Bedingung $\mathrm{ggT}(\det(X), m) = 1$ erfüllt
- Wegen $x_i k = y_i$ für alle $i \in I$ lässt sich dann die Schlüsselmatrix k zu $k = YX^{-1}$ bestimmen, wobei Y die aus den Blöcken y_i mit $i \in I$ gebildete Matrix ist

Kryptoanalyse von polyalphabetischen Chiffren

Angriff auf die Vigenère-Chiffre bei bekannter Periode

- Die Vigenère-Chiffre galt bis ins 19. Jahrhundert als sicher
- Da ihr Schlüsselstrom jedoch periodisch ist, lässt sie sich mit statistischen Methoden leicht brechen, zumindest wenn der Kryptotext im Verhältnis zur Periode (Länge des Schlüsselwortes) genügend lang ist
- Sofern die Periode d bekannt ist, kann man den Kryptotext zeilenweise in eine d-spaltige Matrix schreiben
- Verfahrensbedingt wurden dann die einzelnen Spalten $y_1, \ldots, y_d$ durch eine additive Chiffre verschlüsselt
- Sie können daher einzeln durch eine Häufigkeitsanalyse gebrochen werden
- Hierbei liefert jede Spalte y_i den entsprechenden Buchstaben k_i des Schlüsselwortes

Kryptoanalyse von polyalphabetischen Chiffren

Bestimmung der Periode bei der Vigenère-Chiffre

- Zur Bestimmung von d kann man den Kasiski-Test oder den Koinzidenzindex benutzen
- Die früheste generelle Methode zur Bestimmung der Periode bei der Vigenère-Chiffre stammt von Friedrich W. Kasiski (1860)
- Kommt ein Wort w an zwei verschiedenen Stellen im Kryptotext vor, so kann es sein, dass die gleiche Klartextsequenz zweimal auf die gleiche Weise, d. h. mit der gleichen Schlüsselsequenz, verschlüsselt wurde
- In diesem Fall ist die Entfernung δ der beiden Vorkommen von w ein Vielfaches der Periode d
- Werden mehrere solche Wortpaare im Kryptotext mit verschiedenen Entfernungen δ_i gefunden, so liegt die Vermutung nahe, dass d gemeinsamer Teiler aller (oder zumindest vieler) δ_i ist
- Dadurch wird die Anzahl der noch infrage kommenden Werte für d stark einschränkt

Beispiel (Kasiski-Test)

- Angenommen, der Klartext $x = \texttt{DERERSTEUNDLETZTEVERS}\ldots$ wurde mit einer Vigenère-Chiffre und Schlüssel $k = \textit{KAS}$ zu einem Kryptotext $y = \textit{nejorkdemxddotrdenork}\ldots$ verschlüsselt:

$$\begin{array}{rll} & \texttt{DERERSTEUNDLETZTEVERS}\ldots & \text{(Klartext } x\text{)} \\ + & \underline{\textit{KASKASKASKASKASKASKAS}\ldots} & \text{(Schlüsselstrom } \hat{k}\text{)} \\ & \textit{nejorkdemxddotrdenork}\ldots & \text{(Kryptotext } y\text{)} \end{array}$$

- Da die Sequenzen *ork* bzw. *de* im Kryptotext

$$y = \textit{nejorkdemxddotrdenork}$$

mit den Entfernungen $\delta_1 = 15$ und $\delta_2 = 9$ vorkommen, liegt die Vermutung nahe, dass die Periode $d = \text{ggT}(9, 15) = 3$ ist ◁

Kryptoanalyse von polyalphabetischen Chiffren

Koinzidenzindex-Untersuchungen

- Zur Bestimmung der Periode d gibt es neben heuristischen Methoden auch folgenden statistischen Ansatz, der erstmals von William Frederick Friedman im Jahr 1920 beschrieben wurde
- Er basiert auf der Beobachtung, dass eine längere Periode eine zunehmende Glättung der Buchstabenhäufigkeiten im Kryptotext bewirkt

Definition

- Der Koinzidenzindex (engl. Index of Coincidence) eines Textes y der Länge n über dem Alphabet $\mathcal{B}$ ist definiert als

$$IC(y) = \frac{1}{n \cdot (n-1)} \cdot \sum_{a \in \mathcal{B}} H_y(a) \cdot (H_y(a) - 1)$$

- Hierbei ist $H_y(a)$ die absolute Häufigkeit des Buchstabens a im Text y

Kryptoanalyse von polyalphabetischen Chiffren

Koinzidenzindex-Untersuchungen

- $IC(y)$ gibt also die Wahrscheinlichkeit an, mit der man im Text y an zwei zufällig gewählten Positionen den gleichen Buchstaben vorfindet
- Er ist umso größer, je ungleichmäßiger die Häufigkeiten $H_y(a)$ sind
- Um die Periode d einer Vigenère-Chiffre zu bestimmen, schreibt man den Kryptotext y für $d = 1, 2, 3, \ldots$ in eine Matrix mit d Spalten und berechnet für jede Spalte y_i den Koinzidenzindex $IC(y_i)$
- Für genügend lange Kryptotexte ist dasjenige d, welches das maximale arithmetische Mittel der Spaltenindizes $IC(y_i)$ liefert, mit hoher Wahrscheinlichkeit die gesuchte Periode
- Besteht eine Spalte nur aus Kryptozeichen, die mit demselben Schlüsselbuchstaben k erzeugt wurden, so stimmt der Koinzidenzindex dieser Spalte mit dem Koinzidenzindex des zugehörigen Klartextes überein
- Wurden dagegen die Kryptozeichen einer Spalte mit unterschiedlichen Schlüsselbuchstaben generiert, so bewirkt dies eine Glättung der Häufigkeitsverteilung, weshalb der Spaltenindex kleiner ausfällt

Koinzidenzindex-Untersuchungen

- Ist die Einzelbuchstabenverteilung $p : A \to [0, 1]$ der Klartextsprache bekannt, so lässt sich der Suchraum für d wie folgt einschränken
- Sei Y die Zufallsvariable, die den mittels einer Vigenère-Chiffre mit einem zufälligen Schlüsselwort der Länge d aus einem zufälligen Klartext der Länge n generierten Kryptotext beschreibt
- Dann hängt der erwartete Koinzidenzindex $E(IC(Y))$ von Y neben der Klartextsprache nur von d und n ab, weshalb er auch mit $\mathrm{E}_{d,n}(IC)$ bezeichnet wird
- Im Fall $d = 1$ gilt $IC(Y) = IC(X)$, wobei X die zufällige Wahl des Klartextes beschreibt
- Zudem können wir bei sehr langen Texten von den gegenseitigen Abhängigkeiten der Zeichen im Text absehen und erhalten

$$\mathrm{E}_{1,\infty}(IC) = \sum_{a \in A} p(a)^2$$

- Dieser Wert wird auch als Koinzidenzindex der Klartextsprache bezeichnet

Koinzidenzindex-Untersuchungen

Definition

Der Koinzidenzindex IC_L einer Sprache mit Buchstabenverteilung p ist

$$IC_L = \sum_{a \in A} p(a)^2$$

IC_L ist zudem ein Maß für die Rauheit der Verteilung p

Definition (Rauheitsgrad; Measure of Roughness)

Der Rauheitsgrad MR_L einer Sprache L mit Buchstabenverteilung p ist

$$MR_L = \sum_{a \in A} (p(a) - 1/m)^2 = \sum_{a \in A} p(a)^2 - 1/m = IC_L - 1/m,$$

wobei $m = |A|$ ist

Beispiel

Für die englische Sprache ($m = 26$) gilt beispielsweise $IC_{\text{Englisch}} \approx 0.0687$ und $MR_{\text{Englisch}} \approx 0.0302$ ◁

Koinzidenzindex-Untersuchungen

- Sobald die Periode d die Klartextlänge n übersteigt, ist der Kryptotext bei zufälliger Wahl des Schlüsselwortes ebenfalls rein zufällig
- Dies führt auf einen erwarteten Koinzidenzindex von

$$\mathrm{E}_{d,n}(IC) = \sum_{a \in A} |A|^{-2} = |A|^{-1} = m^{-1}, \quad \text{falls } d \geq n \geq 2$$

- Allgemein gilt für hinreichend großes n,

$$\mathrm{E}_{d,n}(IC) = \frac{n-d}{d \cdot (n-1)} \cdot IC_L + \frac{n \cdot (d-1)}{d \cdot (n-1)} \cdot m^{-1}, \quad \text{falls } 1 \leq d \leq n,$$

- Von den $\binom{n}{2}$ möglichen Positionspaaren liegen nämlich
  - $d \cdot \binom{n/d}{2} = n(n-d)/2d$ Paare in einer Spalte, was einem Anteil von $(n-d)/d(n-1)$ entspricht, und
  - $\binom{d}{2}(n/d)^2 = n^2(d-1)/2d$ Paare in zwei verschiedenen Spalten, was einem Anteil von $n(d-1)/d(n-1)$ entspricht

Koinzidenzindex-Untersuchungen

Beispiel (Schätzung der Periode d bei einer Vigenère-Chiffre)

d	1	2	3	4	5	6	8	10	100
$\mathrm{E}_{d,100}(IC)$	69	54	48	46	44	43	42	41	39

- Die Tabelle zeigt den Erwartungswert $\mathrm{E}_{d,n}(IC)$ des Koinzidenzindexes (in Promille) eines Kryptotextes für einen zufälligen englischen Klartext der Länge $n = 100$ in Abhängigkeit von der Periode d
- Berechnet sich der Koinzidenzindex eines Vigenère-Kryptotextes für einen englischen Klartext der Länge 100 zu $0,045$, so liegt die Vermutung nahe, dass das verwendete Schlüsselwort die Länge vier oder fünf hat

◁

Koinzidenzindex-Untersuchungen

- Der Koinzidenzindex kann auch Hinweise dafür liefern, mit welchem Kryptoverfahren ein vorliegender Kryptotext erzeugt wurde
- Bei Transpositionschiffren sowie bei einfachen Substitutionen bleibt nämlich der Koinzidenzindex im Gegensatz zu polyalphabetischen und polygrafischen Verfahren erhalten
- Erstere lassen sich von letzteren zudem dadurch unterscheiden, dass bei ihnen sogar die Buchstabenhäufigkeiten unverändert bleiben

Koinzidenzindex-Untersuchungen

- Sind die Periode d sowie die Einzelzeichenverteilung der Klartextsprache bekannt, kann man das Schlüsselwort auch wie folgt bestimmen
- Man schreibt den Kryptotext y in eine Matrix mit d Spalten und berechnet für jedes Zeichen $a \in A$ die relative Häufigkeit $h_i(a)$ in jeder Spalte y_i
- Da y_i aus dem Klartext durch Addition von k_i entstanden ist, stimmt die Verteilung

  $$h_i(a + k), a \in A$$

  für $k = k_i$ am besten mit der Klartextverteilung $p(a)$, $a \in A$, überein
- Da

  $$\alpha_i(k) := \sum_{a \in A} p(a) h_i(a + k)$$

  ein Maß für die Ähnlichkeit der beiden Verteilungen $p(a)$ und $h_i(a + k)$ ist (siehe Übungen), nimmt $\alpha_i(k)$ wahrscheinlich für $k = k_i$ einen maximalen Wert an

Koinzidenzindex-Untersuchungen

Beispiel

- Durch eine Vigenère-Chiffre mit Periode $d = 4$ wurde der Kryptotext

$$y = \texttt{huds kuae zgxr avtf pgws wgws zhtp pbil lrtz pzhw loij vfic vbth}$$
$$\texttt{lugi lgpr khwm yhti uaxr bhtw ucgx ospw aoch imcs yhwq hwcf yocg}$$
$$\texttt{ogtz lbil swbf lohx zwsi zvds atgs thwi ssux lmts mhwi kspx ogwi}$$
$$\texttt{hrpf lsam usuv vail lhgi lhwv vivl avtw ocij ptic mstx vii}$$

  aus einem englischem Klartext der Länge 203 erzeugt

- Schreiben wir y in 4 Spalten y_i der Länge $|y_i| = 51$ für $i = 1, 2, 3$ und $|y_4| = 50$, so ergeben sich folgende Werte für $\alpha_i(k)$ (in Promille):

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$\alpha_1(k)$	36	31	31	45	38	26	42	**73**	44	26	36	47	30	32	36	29	28	39	48	42	42	39	42	42	35	31
$\alpha_2(k)$	44	41	40	51	41	31	37	43	34	28	36	26	28	43	**68**	45	35	27	42	43	40	35	30	24	31	45
$\alpha_3(k)$	47	41	48	37	49	40	35	30	48	32	25	42	31	26	43	**76**	37	31	39	45	35	34	37	26	30	25
$\alpha_4(k)$	38	40	27	41	**65**	47	28	34	39	33	35	36	30	30	48	44	35	42	47	38	39	34	27	38	36	37

- Da $\alpha_1(k)$ für $k = a_7 = H$, $\alpha_2(k)$ für $k = a_{14} = O$, $\alpha_3(k)$ für $k = a_{15} = P$ und $\alpha_4(k)$ für $k = a_4 = E$ einen maximalen Wert annimmt, erhalten wir das Schlüsselwort *HOPE*

Koinzidenzindex-Untersuchungen

- Zur Bestimmung des Schlüsselwortes kann man auch die Methode des gegenseitigen Koinzidenzindexes verwenden
- Dabei ist die verwendete Klartextsprache (und somit deren Häufigkeitsverteilung) irrelevant, da die Spalten – wie der Name schon sagt – gegenseitig in Relation gesetzt werden

Definition

Der gegenseitge Koinzidenzindex von zwei Texten y und y' mit den Längen n und n' über dem Alphabet $\mathcal{B}$ ist definiert als

$$IC(y, y') = \frac{1}{n \cdot n'} \cdot \sum_{a \in \mathcal{B}} H_y(a) \cdot H_{y'}(a)$$

- $IC(y, y')$ ist also die Wahrscheinlichkeit, dass bei zufälliger Wahl je einer Position in y und einer in y' das gleiche Zeichen vorgefunden wird
- $IC(y, y')$ ist umso größer, je besser die Häufigkeitsverteilungen von y und y' (d. h. H_y und $H_{y'}$) übereinstimmen

Koinzidenzindex-Untersuchungen

- Sei nun y ein Kryptotext, der mit einem Schlüsselwort bekannter Länge d erzeugt wurde, und seien y_i $(i = 1, \ldots, d)$ die zugehörigen Spalten
- Dann gibt der gegenseitige Koinzidenzindex der Spalten $y_i + \delta$ und y_j (für $1 \leq i < j \leq d$ und $0 \leq \delta \leq 25$) die Wahrscheinlichkeit an, dass man bei zufälliger Wahl einer Position in $y_i + \delta$ und in y_j dasselbe Zeichen vorfindet
- Da die Buchstabenverteilungen von $y_i - k_i$ und von $y_j - k_j$ der der Klartextsprache entsprechen, haben $y_i + \delta$ und y_j für $\delta = k_j - k_i$ eine ähnliche Verteilung
- Mit großer Wahrscheinlichkeit nimmt also $IC(y_i + \delta, y_j)$ für $\delta = \delta_{ij} = k_j - k_i$ einen relativ großen Wert an, während für $\delta \neq \delta_{ij}$ mit kleinen Werten zu rechnen ist

Koinzidenzindex-Untersuchungen

Beispiel

- Betrachten wir den Kryptotext aus vorigem Beispiel, so ergeben sich für $IC(y_i + \delta, y_j)$ die folgenden Werte (in Promille):

δ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$IC(y_1+\delta, y_2)$	40	31	25	38	25	21	46	**74**	50	33	31	44	43	34	31	28	24	31	44	45	37	48	64	44	25	31
$IC(y_1+\delta, y_3)$	26	47	25	21	47	32	18	49	**91**	42	27	51	45	31	29	32	23	29	27	39	45	46	39	58	44	24
$IC(y_1+\delta, y_4)$	38	40	29	31	35	24	32	58	42	32	44	50	43	39	31	20	34	36	30	40	45	24	42	**78**	47	22
$IC(y_2+\delta, y_3)$	50	**85**	49	21	28	35	24	34	46	25	24	27	59	50	50	53	51	24	22	26	43	36	35	32	24	34
$IC(y_2+\delta, y_4)$	46	53	40	37	51	42	29	23	24	32	40	55	38	31	32	45	**67**	49	25	27	29	29	34	37	38	35
$IC(y_3+\delta, y_4)$	49	36	38	60	36	25	34	19	29	42	41	33	54	27	36	**78**	47	25	29	33	27	28	47	32	27	54

- Also ist (mit großer Wahrscheinlichkeit)

$$\delta_{1\,2} = 7,\ \delta_{1\,3} = 8,\ \delta_{1\,4} = 23,\ \delta_{2\,3} = 1,\ \delta_{2\,4} = 16,\ \delta_{3\,4} = 15$$

- Wir können nun alle Spalten relativ zur ersten Spalte so verschieben, dass der ganze Text eine einheitliche Verschiebung δ hat, also die zweite Spalte um -7, die dritte um -8 und die vierte um -23

Koinzidenzindex-Untersuchungen

Beispiel

- Wir können nun alle Spalten relativ zur ersten Spalte so verschieben, dass der ganze Text eine einheitliche Verschiebung δ hat, also die zweite Spalte -7, die dritte um -8 und die vierte um -23
- Für die Bestimmung von δ, muss man nur den häufigsten Buchstaben in dem auf diese Weise erzeugten Text bestimmen
  (oder eine vollständige Suche durchführen)
- Der häufigste Buchstabe ist `L` $(16,3\%)$
- Also ist $\delta = \texttt{L} - \texttt{E} = H = 7$ und das Schlüsselwort lautet *HOPE*
  $(H+7 = O,\ H+8 = P,\ H+23 = E)$ ◁

Analyse der Lauftextverschlüsselung

- Zum Brechen einer Stromchiffre mit Klartextschlüsselstrom kann man wie folgt vorgehen
- Man geht zunächst davon aus, dass jedes Kryptotextzeichen durch Summation eines Klartext- und Schlüsselstromzeichs mit jeweils mittlerer bis hoher Wahrscheinlichkeit entstanden ist
- Dies sind etwa im Englischen die Zeichen `E`, `T`, `A`, `O`, `I`, `N`, `S`, `R`, `H`
- Zu einem Teilwort w des Kryptotextes bestimmt man dann alle Wortpaare (w_1, w_2) mit $w_1 + w_2 = w$ und $w_1, w_2 \in \{\mathtt{E}, \mathtt{T}, \mathtt{A}, \mathtt{O}, \mathtt{I}, \mathtt{N}, \mathtt{S}, \mathtt{R}, \mathtt{H}\}^*$
- In der Regel ergeben sich nur sehr wenige sinnvolle Paare, aus denen durch Kontextbetrachtungen und Erweitern von w nach links und rechts der Kryptotext entschlüsselt werden kann
- Wird die Analyse durch ein Computerprogramm durchgeführt, kann an die Stelle der Kontextbetrachtungen auch die Häufigkeitsverteilung von n-Grammen der Sprache treten
- Das Programm wählt dann solche Wortpaare (w_1, w_2), die eine hohe Wahrscheinlichkeit haben

Analyse der Lauftextverschlüsselung

Beispiel

- Gegeben ist der Kryptotext *moqkthcblmwxf*...
- Wir beginnen die Untersuchung mit einer Wortlänge von vier Buchstaben, also $w = moqk$
- Der erste Buchstabe *m* kann nur auf eine der folgenden Arten zustande gekommen sein:

	ABCDE...I...T...Z	(Klartextzeichen)
+	*MLKJI...E...T...N*	(Schlüsselzeichen)
=	*mmmmm...m...m...m*	(Kryptotextzeichen)

- Es ergeben sich folgende wahrscheinliche Paare für die Zeichen von w:

m:	*o*:	*q*:	*k*:
(E,*I*)	(A,*O*)	(I,*I*)	(R,*T*)
(I,*E*)	(H,*H*)		(S,*S*)
(T,*T*)	(O,*A*)		(T,*R*)

Analyse der Lauftextverschlüsselung

Beispiel (Fortsetzung)

- Es ergeben sich folgende wahrscheinliche Paare für die Zeichen von w:

m:	(E,I)	o:	(A,O)	q:	(I,I)	k:	(R,T)
	(I,E)		(H,H)				(S,S)
	(T,T)		(O,A)				(T,R)

- Diese führen auf folgende $3 \cdot 3 \cdot 1 \cdot 3 = 27$ Wortpaare (w_1, w_2):

w_1	EAIR	EAIS	EAIT	EHIR	...	THIS	...	TOIT
w_2	*IOIT*	*IOIS*	*IOIR*	*IHIT*	...	*THIS*	...	*TAIR*

- Als sinnvoll stellt sich aber nur die Wahl $w_1 = w_2 =$ THIS heraus ◁

Analyse von Autokey Chiffren

Autokey Chiffren mit Kryptotextschlüsselstrom

- Diese Systeme bieten so gut wie keinen Schutz, da sie ohne Kenntnis des Schlüsselwortes sehr leicht entschlüsselt werden können (zumindest falls die Länge des Schlüsselwortes im Verhältnis zur Länge des Kryptotextes relativ kurz ist)
- Man subtrahiert dazu den Kryptotext y für $\delta = 1, 2, \ldots$ von dem um δ Positionen verschobenen Kryptotext bis sinnvoller Klartext erscheint:

	dumsqmozkfn...	(Kryptotext y)
$-$	*DUMSQMO...*	(„Kryptotextschlüsselstrom“)
$=$	....NSCHUTZ...	(Klartext x)

Analyse von Autokey Chiffren

Autokey Chiffren mit Klartextschlüsselstrom

- Neben der oben beschriebenen Analyse der Lauftextverschlüsselung kann das Brechen der *Autokey*-Systeme mit Klartextschlüsselstrom auch analog zur Kasiski-Methode erfolgen
- Sei d die Länge des Schlüsselwortes $k_0 \dots k_{d-1}$
- Falls im Klartext die gleiche Buchstabenfolge $x_i \dots x_{i+l-1}$ im Abstand $2d$ auftritt (beispielsweise $d = 3$ und $l = 2$),

$$
\begin{array}{rll}
 & \downarrow\ \downarrow \quad\ \downarrow\ \downarrow & \\
 & x_0\ x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ x_7\ x_8\ x_9\ x_{10}\ x_{11}\ x_{12}\ x_{13}\ x_{14}\ \dots & \text{Klartext } x \\
+ & k_0\ k_1\ k_2\ x_0\ x_1\ x_2\ x_3\ x_4\ x_5\ x_6\ \ x_7\ \ x_8\ \ x_9\ \ x_{10}\ x_{11}\ \dots & \text{Schlüsselstrom } kx \\
\hline
 & y_0\ y_1\ y_2\ y_3\ y_4\ y_5\ y_6\ y_7\ y_8\ y_9\ y_{10}\ y_{11}\ y_{12}\ y_{13}\ y_{14}\ \dots & \text{Kryptotext } y
\end{array}
$$

so tritt im Kryptotext die gleiche Buchstabenfolge im Abstand d auf und d kann auf diese Art unter Umständen leicht bestimmt werden

Analyse von Autokey Chiffren

Autokey Chiffren mit Klartextschlüsselstrom

- Ist die Schlüsselwortlänge d bekannt, so können die Buchstaben $k_1 \dots k_d$ des Schlüsselwortes der Reihe nach bestimmt werden
- Da nämlich durch y und k_i die Klartextzeichen an den Positionen $i, d+i, 2d+i, \dots$ eindeutig festgelegt sind, kann jedes einzelne k_i unabhängig von den anderen Schlüsselwortbuchstaben durch eine statistische Analyse bestimmt werden
- Tatsächlich lässt sich die Kryptoanalyse der Autokey Chiffre mit Klartextschlüsselstrom bei Kenntnis von d auf die Kryptoanalyse einer Vigenère-Chiffre mit Periode d reduzieren (siehe Übungen)