Kryptografie a počítačová bezpečnost

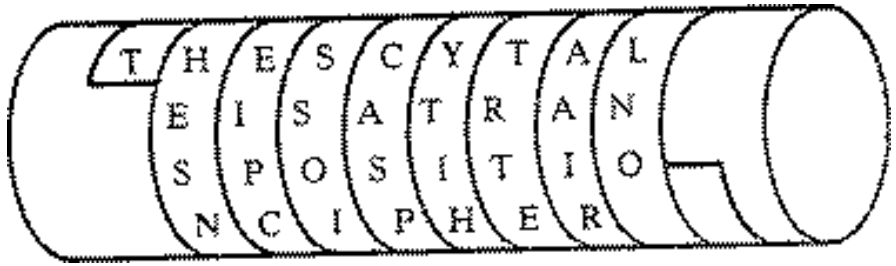
Historická kryptografie

Literatura

- https://www.cs.vsb.cz/ochodkova/courses/kpb/cryptography-and-network-security_-principles-and-practice-7th-global-edition.pdf kapitoly 3.1., 3.2
- https://cacr.uwaterloo.ca/hac/ kapitoly 1.1 – 1.5, 7.3

Kdy „vznikla“ kryptografie? (1)

- Egypt, Mezopotámie
- Hebrejci – Bible (Starý zákon) obsahuje úseky šifrované hebrejskou šifrou **atbaš** (500 př.n.l.)
  - první znak *alef* se zamění za poslední *tav*,
  - druhé *bet* se zamění za poslední předposlední *šin*
  - http://cs.wikipedia.org/wiki/Hebrejsk%C3%A1_abeceda
- Řekové – 500 let př.n.l. Scytale



- Kniha kódů a šifer - Simon Singh (https://www.dokoran.cz/index.php?Kniha_kodu_a_sifer_4_vydani&p=book&id=1279)

Kdy „vznikla“ kryptografie? (2)

- **Caesar** (100-44 př.n.l.)

  C = L FDPH L VDZ L FRQTXHUHG
  M = I CAME I SAW I CONQUERED

- Klíč = 3, posun v abecedě
- Posun v abecedě o 3 znaky:

  M = ABCDEFGHIJKLMNOPQRSTUVWXYZ
  C = DEFGHIJKLMNOPQRSTUVWXYZABC

Základní kryptografické principy

- V minulosti se uplatňovaly dva základní principy konstrukce šifrovacích algoritmů (které používaly sdílený tajný klíč):
  - **Substituce** (záměna) - nahrazení znaků otevřeného textu jinými znaky.
    - Monoalfabetické šifry
    - Polyalfabetické šifry
  - **Transpozice** (permutace) - nemění znaky otevřeného textu, ale mění jejich pořadí.
- V současnosti se čistě substituční nebo permutační algoritmy nepoužívají.
- Používají se algoritmy (symetrické), které kombinují obě techniky, říká se jim **složené šifry** (product ciphers).

Substituční algoritmy

- Substituční algoritmy
  - Monoalfabetické substituční algoritmy (monoalfabetická substituce, jednoduchá záměna)
    - Každý znak otevřeného textu je v šifrovém textu nahrazen stejně (stejným znakem), tedy všechny výskyty jsou nahrazeny stejně, např. znak A bude vždy nahrazen znakem X
  - Polyalfabetické substituční algoritmy (polyalfabetická substituce, složitá záměna)
    - Pro znak otevřeného textu může být v textu šifrovém použito znaků více, tedy „více abeced“, více substitucí. Např. jednou bude otevřené A nahrazeno šifrovým X, jindy třeba šifrovým B apod.
- https://en.wikipedia.org/wiki/Substitution_cipher

Monoalfabetické substituční algoritmy (1)

- **Shift Cipher** (zobecnění Caesarovy šifry)
  - Abeceda (OA i ŠA): (AAbM) anglická bez mezery ‘A’=0, …, ‘Z’=25, počet znaků n = 26
  - M=C=K=Z_{26}
    - Nechť Z je množina celých čísel, pak Z_{26} je množina celých čísel modulo 26, Z_{26} = {0, 1, …, 25}
  - zpráva m se šifruje po blocích (délka bloku je 1 znak)
  - klíčem k je n (=1) znaků, k∈{0, 1, …, 25},
  - $c = e_k(m)=(m+k) \bmod 26$
  - $m = d_k(c)=(c-k) \bmod 26$
  - 26 možných klíčů (resp. 25, protože posun o 0 pozic nemá význam)
  - Caesarova šifra je Shift šifra s k=3
    - $c = e_k(m)=(m+3) \bmod 26$
    - $m = d_k(c)=(c-3) \bmod 26$

Monoalfabetická substituce (2)

- **Obecný algoritmus** obecná permutace 26 znaků, např.

```
OA: ABCDEFGHIJKLMNOPQRSTUVWXYZ
ŠA: DKVQFIBJWPESCXHTMYAUOLRGZN
```

Př.:

```
M = IFWEW ISHTO REPLA CELET TERS
C = WIRFR WAJUH YFTSD VFSFU UFYA
```

- Klíčem je tedy libovolná permutace OA. Klíčů je 26! $\approx$ 4 x 10^{26}

- https://en.wikipedia.org/wiki/Substitution_cipher

Monoalfabetická substituce (3)

- Další modifikace – abeceda je určená klíčem (heslem) a mechanismem doplnění zbývajících znaků abecedy
- Výhoda – nemusí se předávat celá permutovaná abeceda ale jen „heslo“ k jejímu vytvoření

```
K = JULIUSCAESAR
M = ABCDEFGHIJKLMNOPQRSTUVWXYZ
C = JULISCAERTVWXYZBDFGHKMNOPQ
```

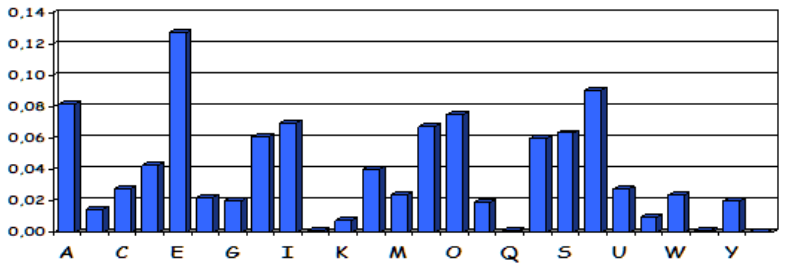
Frekvenční analýza

- Útok ze známého šifrového textu (known ciphertext only attack)
  - Existují různé typy kryptoanalytických útoků podle toho co má útočník k dispozici
- Popsáno Abu Al-Kindim v „Manuscript on Deciphering Cryptographic Message“, 9 stol. n.l.
- Založena na analýze vlastností přirozeného jazyka, na jeho statistických vlastnostech, na frekvencích výskytu jednotlivých hlásek (znaků), dvojic (bigramů), trojic (trigramů) znaků apod.
- Vhodné pro delší texty (a nevhodné pro specifické texty, např.: „from Zanzibar to Zambia to Zaire, ozone zones make zebras run zany zigzags“)
- https://en.wikipedia.org/wiki/Frequency_analysis

Některé tabulky četností

Písmeno	Angl.	Franc.	Něm.	Češ.	Slov.
A	7,96	7,68	5,52	8,99	9,49
B	1,60	0,80	1,56	1,86	1,90
C	2,84	3,32	2,94	3,04	3,45
D	4,01	3,60	4,91	4,14	4,09
E	12,86	17,76	19,18	10,13	9,16
F	2,62	1,06	1,96	0,33	0,31
G	1,99	1,10	3,60	0,48	0,40
H	5,39	0,64	5,02	2,06	2,35
I	7,77	7,23	8,21	6,92	6,81
J	0,16	0,19	0,16	2,10	2,12
K	0,41	0,00	1,33	3,44	3,80
L	3,51	5,89	3,48	4,20	4,56
M	2,43	2,72	1,69	2,99	2,97
N	7,51	7,61	10,20	6,64	6,34
O	6,62	5,34	2,14	8,39	9,34
P	1,81	3,24	0,54	3,54	2,87
Q	0,17	1,34	0,01	0,00	0,00
R	6,83	6,81	7,01	5,33	5,12
S	6,62	8,23	7,07	5,74	5,94
T	9,72	7,30	5,86	4,98	5,06
U	2,48	6,05	4,22	3,94	3,70
V	1,15	1,27	0,84	4,50	4,85
W	1,80	0,00	1,38	0,06	0,06
X	0,17	0,54	0,00	0,04	0,03
Y	1,52	0,21	0,00	2,72	2,57
Z	0,05	0,07	1,17	3,44	2,72

Angl.	Franc.	Něm.	Čeština
TH: 3,30	ES: 3,05	EN: 4,43	PR: 1,98
HE: 2,70	EL: 2,46	ER: 3,75	NI: 1,94
IN: 2,02	EM: 2,42	CH: 2,80	ST: 1,81
ER: 1,91	DE: 2,15	EI: 2,42	NA: 1,68
RE: 1,69	RE: 2,09	DE: 2,33	NE: 1,61
AN: 1,67	NT: 1,97	ND: 2,08	EN: 1,55
ES: 1,49	ON: 1,64	IN: 1,97	RA: 1,35
EN: 1,46	ER: 1,63	GE: 1,96	OV: 1,32
ON: 1,34	TE: 1,63	IE: 1,88	TE: 1,30
AT: 1,27	SE: 1,55	TE: 1,76	AN: 1,25



Angl.	Franc.	Něm.	Čeština	Slov.
E: 12,86	E: 17,76	E: 19,18	E: 10,13	A: 9,49
T: 9,72	S: 8,23	N: 10,20	A: 8,99	O: 9,34
A: 7,96	A: 7,68	I: 8,21	O: 8,39	E: 9,16
I: 7,77	N: 7,61	S: 7,07	I: 6,92	I: 6,81
N: 7,51	T: 7,30	R: 7,01	N: 6,64	N: 6,34
R: 6,83	I: 7,23	T: 5,86	S: 5,74	S: 5,94
Σ: 52,65	Σ: 55,81	Σ: 57,53	Σ: 46,81	Σ: 47,08

Začátek	Konec
P: 12,50	E: 16,67
S: 9,72	I: 13,96
V: 9,19	A: 10,94
Z: 8,95	O: 8,93
N: 7,64	U: 7,94
O: 5,56	Y: 7,03
Σ: 53,56	Σ: 65,47
souhl.: 84,51	souhl.: 34,53
samohl.: 15,49	samohl: 65,47

Relativní četnost (vliv volby korpusu)

- Angličtina

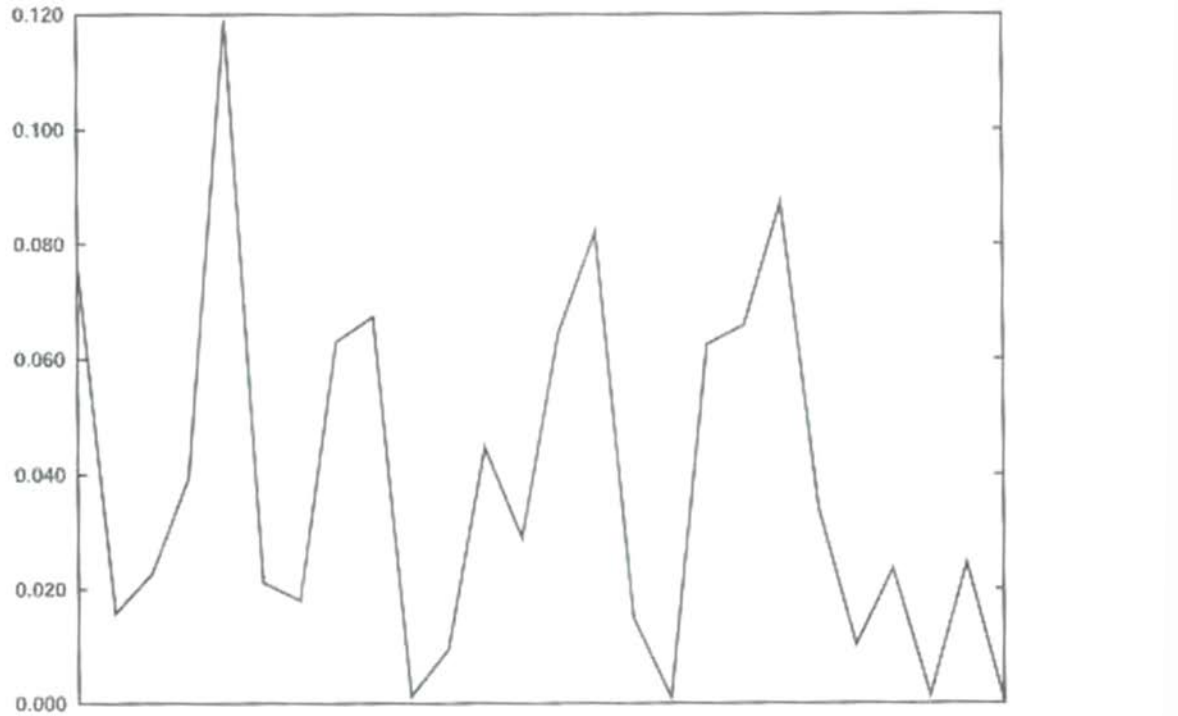


Figure 1-1 Frequency distribution table for Shakespeare's complete works [3]. The letters are shown left to right, *A* through *Z*, with the *y*-value being the frequency of that character occurring in *The Complete Works of William Shakespeare* [3].

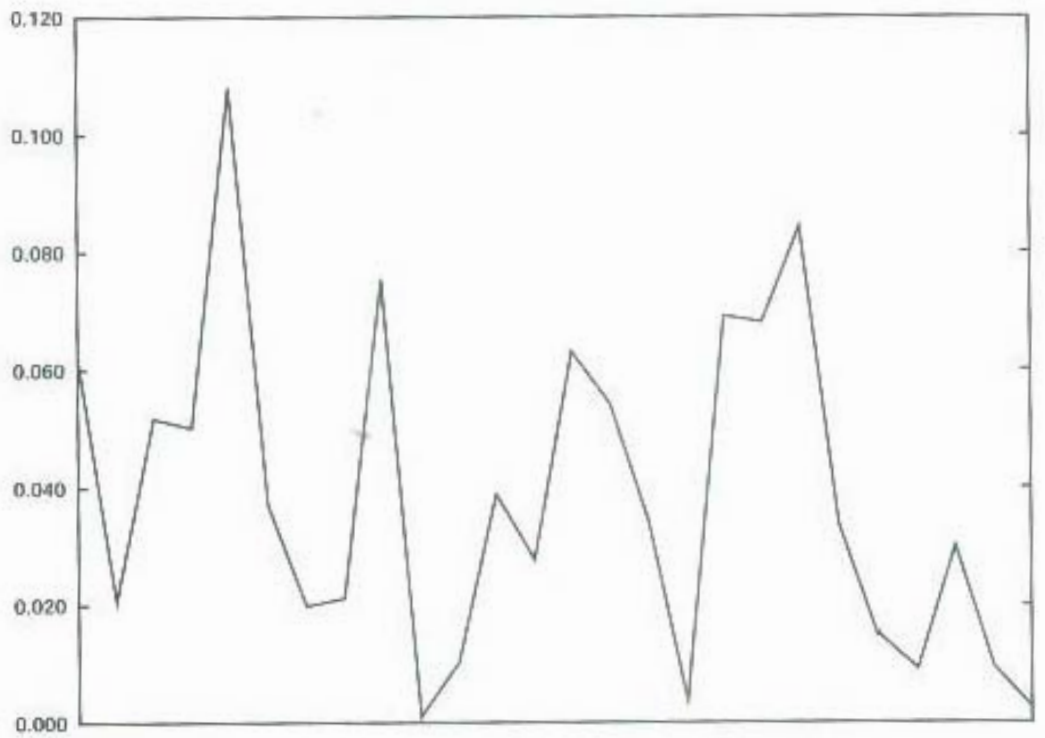


Figure 1-2 Frequency distribution table for "vanilla" Linux 2.6.15.1 source code (including only alphabetic characters). The total size is approximately 205 megabytes.

Frekvenční analýza

- Založena na vlastnostech jazyka
  - Pořadí hlásek v češtině:
    E,O,A,I,N,S,T,R,V,U,L,Z,D,K,P,M,C,Y,H,J,B,G,F,X,W,Q
  - Pořadí hlásek v češtině na začátku slov:
  - **P,S,V,Z,N,T,O,J,K,D,A,B,M,R,U,C,I,H,E,L,F,G,W,Y,Q,X**
  - Pořadí hlásek v češtině na konci slov:
    E,I,A,O,U,Y,M,T,H,V,L,K,S,Z,D,N,R,C,J,B,P,G,F,W,X,Q

  - **Pozor - mezera 16 – 20%**

Frekvenční analýza

- **Bigramy** ST, PR, SK, CH, DN, TR
  - Zvláštnosti souhláskových bigramů v češtině
  - **ST:** S a T má přibližně stejnou frekvenci
    - existuje i bigram TS
    - Je součástí velkého počtu souhláskových trigramů (STR, STN, STL, STV)
    - vyskytuje se uprostřed i na konci slova
  - **PR:** P má asi poloviční frekvenci než R
    - Obrácený bigram RP se téměř nevyskytuje (chrpa)
    - Zpravidla nelze rozšířit „dozadu“ na souhláskový trigram (PRV)
    - Lze rozšířit dopředu na samohláskový trigram (SPR, ZPR, ...)
    - Zpravidla stojí na počátku slov
  - **CH**
    - H má jen o něco menší frekvenci než C (u kratších textů nemusí platit)
    - Bývá zpravidla na konci slov spolu se samohláskami Y, I, A, E (YCH, ICH, ACH, ECH)
    - Většinou platí: předchází-li CH souhláska, pak je po něm samohláska a naopak (OBCHOD, NECHŤ)
- **Trigramy**
    - PRO, UNI, OST, STA, ANI, OVA, YCH, STI, PRI, PRE, OJE, REN, IST, **STR (nejběžnější** souhláskový trigram!), EHO, TER, RED, ICH, ...

Monoalfabetická substituce - Polygramové šifry

- Polygramové šifry (polygramové šifrovací algoritmy)
  - znak OT se šifruje do skupiny znaků ŠT
  - nebo skupina znaků → na skupinu znaků
- Polyboisův čtverec (https://en.wikipedia.org/wiki/Polybius_square)

CTVEREC →ACDDEAAEDBAEAC

	A	B	C	D	E
A	A	B	C	D	E
B	F	G	H	I	K
C	L	M	N	O	P
D	Q	R	S	T	U
E	V	W	X	Y	Z

Monoalfabetická substituce - Polygramové šifry

- Bigram na bigram – Playfair, písmena z každého z bigramů se mohou ve čtverci vyskytnout ve třech pozicích: na stejném řádku, na stejném sloupci nebo na jiném řádku a sloupci:
    - Pokud obě písmena z bigramu leží na stejném řádku, nahradí se písmeny ležícími napravo od nich. Pokud je jedno z písmen poslední v řádku, nahradí se prvním ze stejného řádku.
    - Pokud obě písmena leží ve stejném sloupci, nahradí se písmeny ležícími pod nimi. Pokud je jedno z písmen poslední ve sloupci, nahradí se prvním z téhož sloupce.
    - Pokud obě písmena leží na jiném řádku a v jiném sloupci, je každé z nich nahrazeno písmenem ležícím na průsečíku řádku daného písmena a sloupce druhého písmena.
  - Příklad viz https://en.wikipedia.org/wiki/Playfair_cipher

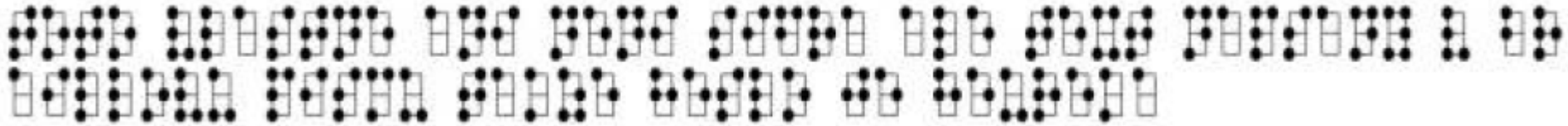
Monoalfabetická substituce – různé abecedy

- Není nutné nahrazovat písmena opět písmeny. Můžeme použít libovolné znaky, tj. **otevřená a šifrová abeceda mohou být různé**
- Př.: OT je v angličtině bez mezery.
- ŠT:

```
5 3 ‡ ‡ † 3 0 5 ) ) 6 * ; 4 8 2 6 ) 4 ‡ . ) 4 ‡ ) ; 8 0 6 *
; 4 8 † 8 ¶ 6 0 ) ) 8 5 ; 1 ‡ ( ; : ‡ * 8 † 8 3 ( 8 8 ) 5 *
† ; 4 6 ( ; 8 8 * 9 6 * ? ; 8 ) * ‡ ( ; 4 8 5 ) ; 5 * † 2 :
* ‡ ( ; 4 9 5 6 * 2 ( 5 * - 4 ) 8 ¶ 8 * ; 4 0 6 9 2 8 5 ) ;
) 6 † 8 ) 4 ‡ ‡ ; 1 ( ‡ 9 ; 4 8 0 8 1 ; 8 : 8 ‡ 1 ; 4 8 † 8
5 ; 4 ) 4 8 5 † 5 2 8 8 0 6 * 8 1 ( ‡ 9 ; 4 8 ; ( 8 8 ; 4 (
‡ ? 3 4 ; 4 8 ) 4 ‡ ; 1 6 1 ; : 1 8 8 ; ‡ ? ;
```

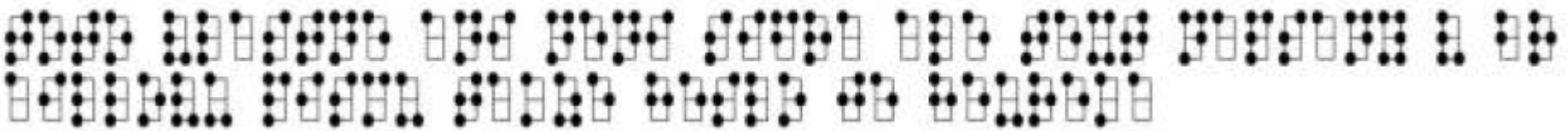
- Řešení v povídce Zlatý brouk (The Gold-Bug), Edgar Allan Poe, např. http://en.wikipedia.org/wiki/The_Gold-Bug nebo http://www.musilek.eu/michal/pdf/Poe.pdf

Monoalfabetická substituce – různé abecedy

- 0 B1 2 345 ox6 r7Ma 890 juFy k1 L234 k enG5 I6ex e7 Gxx 8 9a s 01rZ h2 Gpk k x345 e 6y nkj7 k8 T9v dT a0 1G2v s 3e B4 56 789 I0h krT F 123 vuo4 5x6 xn7 c89n uy ZFs 01 s 2w F3 d456 uje 789 oM0 1 r 2L3s 4L5 6u78 G9 0T1h Gj deF v2Qv Qe 3d4h p 5n6M d7 y8y 9M0 1s23 j4c v wwe u 5x Zw e678 M Q901 jn2 3I j4 5x6
- 

Monoalfabetická substituce – různé abecedy

- 0 B1 2 345 ox6 r7Ma 890 juFy k1 L234 k enG5 I6ex e7 Gxx 8 9a s 01rZ h2 Gpk k x345 e 6y nkj7 k8 T9v dT a0 1G2v s 3e B4 56 789 I0h krT F 123 vuo4 5x6 xn7 c89n uy ZFs 01 s 2w F3 d456 uje 789 oM0 1 r 2L3s 4L5 6u78 G9 0T1h Gj deF v2Qv Qe 3d4h p 5n6M d7 y8y 9M0 1s23 j4c v wwe u 5x Zw e678 M Q901 jn2 3I j4 5x6
  - Text byl převeden do šifrového textu pomocí morseovky. K zápisu morseovky bylo použito toto kódování : tečka = libovolné písmeno, čárka = libovolná číslice



  - Jedná se o kódování, kde pro kód písmena je použito Braillovo slepecké písmo.

Polyalfabetická substituce

- Pro každé písmeno otevřené abecedy se používá více substitucí, tedy více šifrových abeced.
- To, která šifrová abeceda se použije, závisí na klíči.
- Lépe skrývá frekvenční závislosti, poskytuje lepší frekvenční distribuci znaků šifrového textu.

Vigenèrova šifra

- Blaise de Vigenère – Vigenèrova šifra
  - le chiffre indéchiffrable
  - publikoval v „*Traicté des Chiffres*“ r. 1585 polyalfabetickou šifru, která odolala až do r. 1854, reps. 1863 (zlomili Ch. Babbage resp. F. Kasiski),
  - Kryptoanalýza, známá jako Kasiského test, vychází z předpokladu, že klíč je kratší než otevřený text a musí tak být používán opakovaně.

Vigenèrova šifra

```
M = THISP ROCES SCANA LSOBE EXPRE SSED
K = CIPHE RCIPH ERCIP HERCI PHERC IPHE
C = VPXZT IQKTZ WTCVP SWFDM TETIG AHLH

  C -> CDEFGHIJKLMNOPQRSTUVWXYZAB
  I -> IJKLMNOPQRSTUVWXYZABCDEFGH
  P -> PQRSTUVWXYZABCDEFGHIJKLMNO
  H -> HIJKLMNOPQRSTUVWXYZABCDEFG
  E -> EFGHIJKLMNOPQRSTUVWXYZABCD
  R -> RSTUVWXYZABCDEFGHIJKLMNOPQ

'T' klíč 'C' mapuje se na 'V'
'H' klíč 'I' mapuje se na 'P'
'I, klíč 'P' mapuje se na 'X' atd.
```

Vigenèrova tabulka

```
    ABCDEFGHIJKLMNOPQRSTUVWXYZ

A | ABCDEFGHIJKLMNOPQRSTUVWXYZ |
B | BCDEFGHIJKLMNOPQRSTUVWXYZA |
C | CDEFGHIJKLMNOPQRSTUVWXYZAB |
D | DEFGHIJKLMNOPQRSTUVWXYZABC |
E | EFGHIJKLMNOPQRSTUVWXYZABCD |

F | FGHIJKLMNOPQRSTUVWXYZABCDE |

G | GHIJKLMNOPQRSTUVWXYZABCDEF |

H | HIJKLMNOPQRSTUVWXYZABCDEFG |
I | IJKLMNOPQRSTUVWXYZABCDEFGH |
J | JKLMNOPQRSTUVWXYZABCDEFGHI |
K | KLMNOPQRSTUVWXYZABCDEFGHIJ |
L | LMNOPQRSTUVWXYZABCDEFGHIJK |
M | MNOPQRSTUVWXYZABCDEFGHIJKL |
N | NOPQRSTUVWXYZABCDEFGHIJKLM |
O | OPQRSTUVWXYZABCDEFGHIJKLMN | …
```

Polyalfabetická substituce

- $M=C=K=(Z_{26})^n$, zpráva m se šifruje po blocích o n znacích,
- klíčem je řetěz n znaků, $k=(k_1,k_2,\ldots,k_n)$,
- $e_k(m_1,m_2,\ldots,m_n)=(m_1+k_1) \bmod 26,\ldots,(m_n+k_n) \bmod 26,$
- $d_k(c_1,c_2,\ldots,c_n)=(c_1-k_1) \bmod 26,\ldots,(c_n-k_n) \bmod 26.$
- Počet všech různých klíčů je 26^n,

M = PRIKL ADPOL YALFA BETIC KESIF RY,
K = TOTOJ EKLIC TOTOJ EKLIC TOTOJ EK,
c_1=P + T mod 26, c_2=R + O mod 26...,
tj.c_1=15+19 mod 26, c_2=17+14 mod 26...

Autokey Cipher

- Vigenère navrhnul „the **autokey** cipher“ (chtěl najít způsob jak vytvořit klíč stejně dlouhý jako OT)
- Klíčové slovo DECEPTIVE

M = WEAREDISCOVEREDSAVEYOURSELF

K = **DECEPTIVE**WEAREDISCOVEREDSAV

C = ZICVTWQNGKZEIIGASXSTSLVVWLA