KRY01 - MNG

Model 2019

Kryptografie

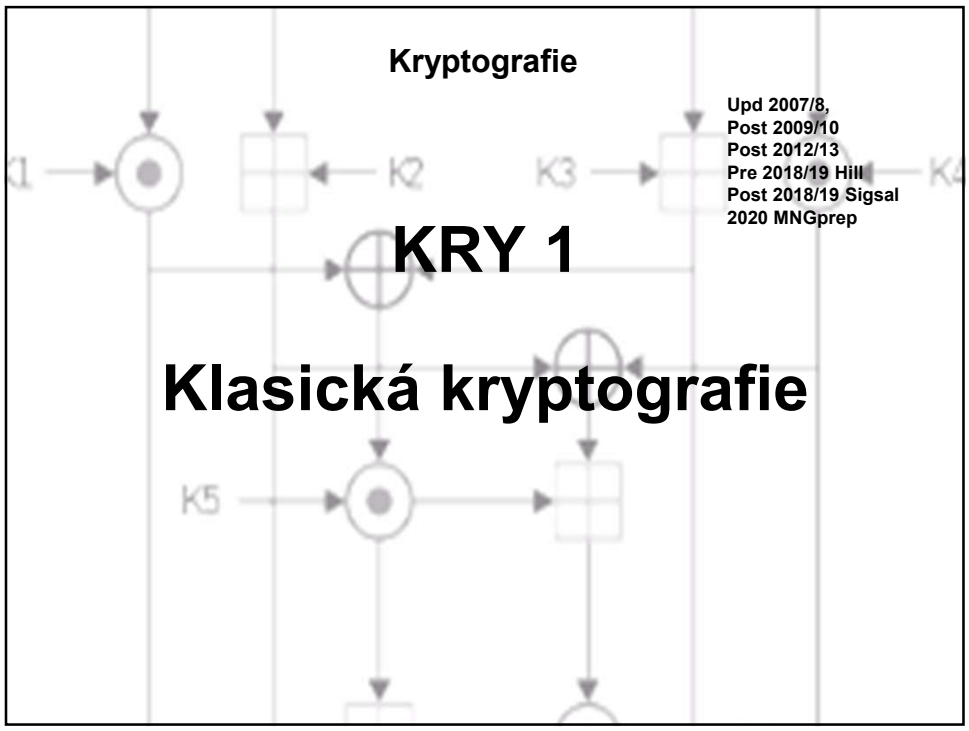
Část 1 Klasická kryptografie

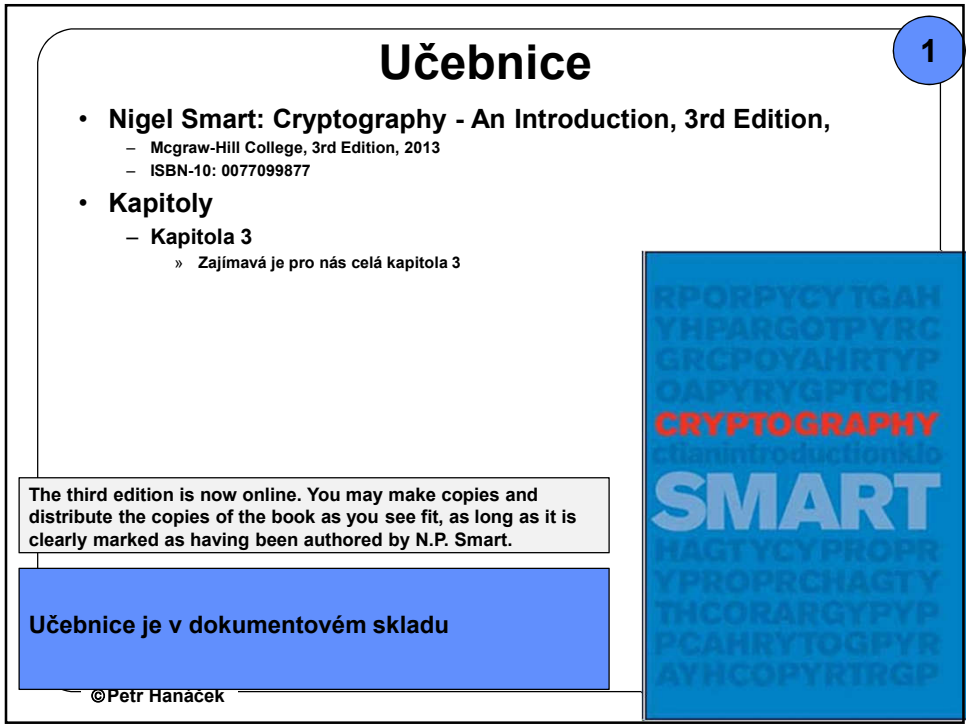
Post 19/20

Souhrnné materiály

Ver 0.1

KRY


Kryptografie
Upd 2007/8,
Post 2009/10
Post 2012/13
Pre 2018/19 Hill
Post 2018/19 Sigsal
2020 MNGprep
KRY 1
Klasická kryptografie

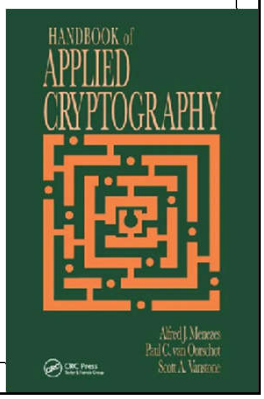

Učebnice
1
Nigel Smart: Cryptography - An Introduction, 3rd Edition,
Mcgraw-Hill College, 3rd Edition, 2013
ISBN-10: 0077099877
Kapitoly
Kapitola 3
Zajímavá je pro nás celá kapitola 3
CRYPTOGRAPHY
SMART
The third edition is now online. You may make copies and distribute the copies of the book as you see fit, as long as it is clearly marked as having been authored by N.P. Smart.
Učebnice je v dokumentovém skladu
©Petr Hanáček

Učebnice

1

- Menezes, Van Oorschot, Vanstone: Handbook of Applied Cryptography, CRC Press, Hardcover, 816 pages, CRC Press, 1997.
- Kapitoly
  - Kapitola 7.3
    - » Zajímavá je pro nás celá kapitola 7.3

http://www.cacr.math.uwaterloo.ca/hac/

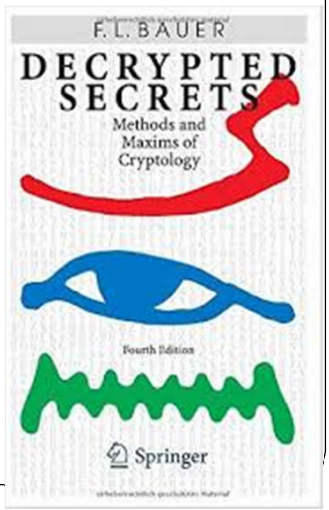


©Petr Hanáček

Doporučená četba

1

- Decrypted Secrets
- [Bauer] Friedrich L. Bauer: Decrypted Secrets - Methods and Maxims of Cryptology, ISBN-10 3-540-24502-2 Springer Berlin Heidelberg New York



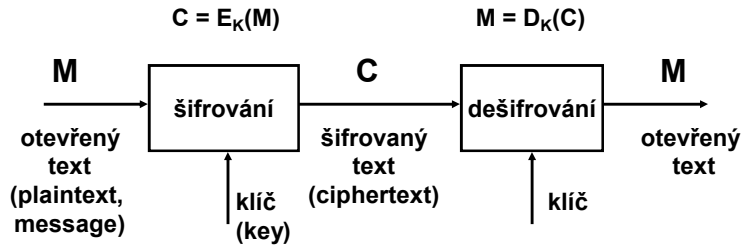
©Petr Hanáček

Kryptografie

- **Účel kryptografie**
- **Klasická kryptografie x Moderní kryptografie**
  - **Historie**
  - **Základní rozdíly**

 CLACRYPT Slide 5

Kryptografie

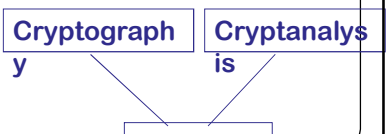


- **podle klíčů**
  - **symetrické X asymetrické**
  - **tajný klíč X veřejný klíč, soukromý klíč**

 CLACRYPT Slide 6

Pojmy

- **Kryptografie - cryptography**
  - Transformace otevřeného textu na šifrovaný text a (obvykle) naopak
- **Kryptoanalýza - cryptanalysis (codebreaking)**
  - Transformace šifrovaného textu na otevřený bez znalosti klíče
- **Kryptologie - cryptology**
  - Kryptografie a kryptoanalýza
- **Šifra, šifrovací algoritmus - cipher**
  - Algoritmus na transformaci otevřeného textu na šifrovaný text a (obvykle) naopak
- **Otevřený text**
  - Srozumitelný text, zpráva, plaintext, message
- **Šifrovaný text**
  - Kryptogram, ciphertext, cryptogram
- **Šifrování - encryption**
- **Dešifrování - decryption**



©Petr Hanáček CLACRYPT Slide 7

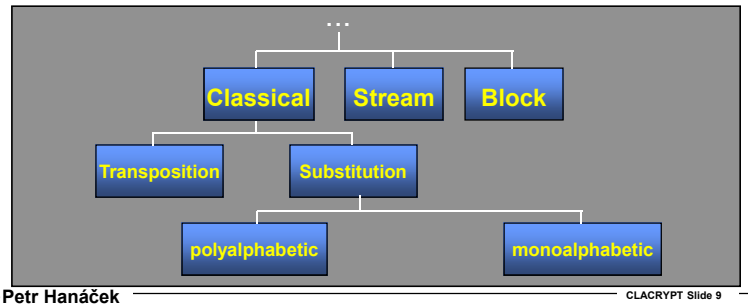
Útoky

- **Ciphertext only attack**
  - Útočník zná pouze šifrovaný text, snaží se zjistit klíč nebo otevřený text
- **Known plaintext attack**
  - Útočník zná šifrovaný text a odpovídající otevřený text, snaží se zjistit klíč
- **Chosen plaintext attack**
  - Útočník zná šifrovaný text a odpovídající otevřený text, který si mohl zvolit, snaží se zjistit klíč

©Petr Hanáček CLACRYPT Slide 8

Typy klasických šifer

- **Steganografické postupy**
  - ukryjí přenášený text uvnitř jiného textu
- **Substituční šifry**
  - nahrazují jednotlivé znaky/symboly textu jinými znaky
- **Transpoziční šifry**
  - mění pořadí znaků v textu (typicky pomocí nějakého geometrického obrazce, např. matice)



 CLACRYPT Slide 9

Steganografie

 CLACRYPT Slide 10

Steganografie

- Stega = hidden
- Graph = writing

- Steganografie je způsob jak ukrýt informaci uvnitř něčeho jiného
- Příklady:
  - Ukrytí informace uvnitř textového souboru
  - Vodoznak (watermark) uvnitř obrazu

 CLACRYPT Slide 11

Steganografie

Dear George, 3rd March
Greetings to all at Oxford. Many thanks for your letter and for the Summer examination package. All Entry Forms and Fees Forms should be ready for final dispatch to the Syndicate by Friday 20th or at the very least, I'm told, by the 21st. Admin has improved here, though there's room for improvement still; just give us all two or three more years and we'll really show you! Please don't let these wretched 16+ proposals destroy your basic O and A pattern. Certainly this sort of change, if implemented immediately, would bring chaos.

- Například první písmeno každého slova, nejméně významný bit každého pixelu
- Princip „Security through obscurity"
- Nepoužitelná jakmile se prozradí metoda

 CLACRYPT Slide 12

Steganografie

Dear George, 3rd March
Greetings to all at Oxford. Many thanks for your letter and for the Summer examination package. All Entry Forms and Fees Forms should be ready for final dispatch to the Syndicate by Friday 20th or at the very least, I'm told, by the 21st. Admin has improved here, though there's room for improvement still; just give us all two or three more years and we'll really show you! Please don't let these wretched 16+ proposals destroy your basic O and A pattern. Certainly this sort of change, if implemented immediately, would bring chaos.

 CLACRYPT Slide 13

Moderní použití steganografie

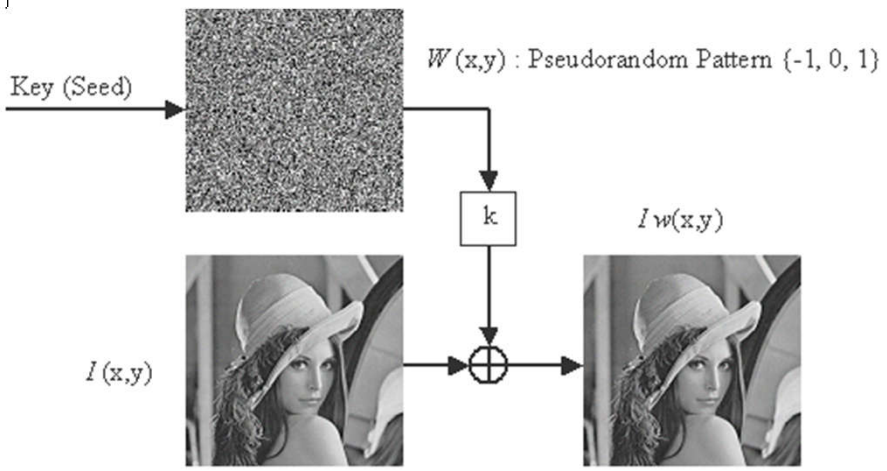
- Zpráva se skládá z posloupnosti bitů
- Některé bity jsou bity zprávy, jiné jsou „vata“
- Pomocí klíče oprávněný příjemce zjistí, které bity jsou „vata“ a mají se vypustit

Nevýhoda steganografie:

- Šifrovaný text je delší (často významně) než otevřený text

 CLACRYPT Slide 14

Vodoznaky (Watermarking)



$$I_w(x,y) = I(x,y) + k * W(x,y)$$

©Petr Hanáček CLACRYPT Slide 15

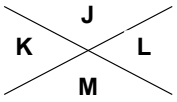
Substituční šifry

©Petr Hanáček CLACRYPT Slide 16

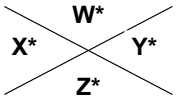
Freemason cipher

- Ukázka, že abeceda se nemusí skládat pouze ze znaků ale obecně z jakýchkoli symbolů
- Nemá klíč
- Tabulka:

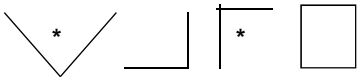
A	B	C
D	E	F
G	H	I



N*	O*	P*
Q*	R*	S*
T*	U*	V*



Příklad:



©Petr Hanáček CLACRYPT Slide 17

CAESAR

©Petr Hanáček CLACRYPT Slide 18

Caesarova šifra

- První známé algoritmické šifrování
- Julius Caesar šifroval své zprávy tak, že nahradil každé písmeno třetím následujícím písmenem v abecedě
- "caesar" se zašifruje jako "FDHVDU"
- Slabiny
  - každý, kdo zná algoritmus, může zprávu dešifrovat
- Možné vylepšení
  - odesílatel a příjemce mají domluven klíč (číslo od 1 do 25), který znamená o kolik písmen se posouvá
  - šifra je stále slabá - stavový prostor klíčů je příliš malý

a → D	w → Z
b → E	x → A
c → F	y → B
d → G	z → C

Caesarova šifra - útoky

- Útok silou – pouhých 26 možností
- Frekvenční analýza

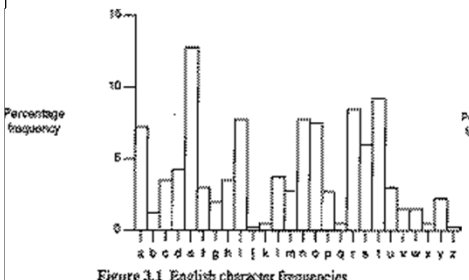


Figure 3.1 English character frequencies

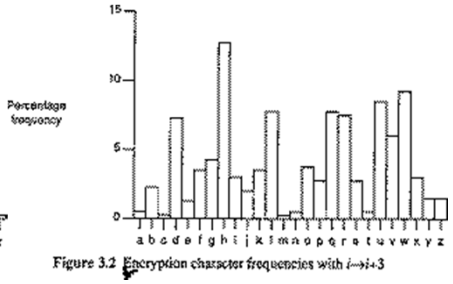


Figure 3.2 Encryption character frequencies with $i \rightarrow i+3$

Kerckhoffův princip

- A. Kerckhoffs byl holandský kryptolog v 19. století
- Bezpečnost musí záviset pouze na utajení klíče
  - Je třeba předpokládat, že útočník zná všechny podrobnosti o použitém algoritmu
- Ergo, *Security by obscurity doesn't work!*

Monoalfabetické substituční šifry

Monoalfabetické substituční šifry

- Substituční šifry nahrazují jednotlivá písmena textu pomocí klíče, kterým je permutace všech 26 písmen (mixed alphabet).

Example: The key is a permutation:

```
abcdefghijklmnopqrstuvwxyz
PDUIRMFOHSBNCGVKTJWEYAQXZL

Encryption:
 Plaintext:  monoalphabeticsubstitution
 Ciphertext: CVGVBNKOPDREHUWYDWEHEYEHVG
```

- Počet všech možných klíčů je $26! = 4*10^{26}$
- Nerozluštitelné během celého prvního tisíciletí našeho letopočtu

©Petr Hanáček CLACRYPT Slide 23

Zapamatovatelný klíč

- Speciální případ monoalfabetické šifry
- Klíčem je slovo nebo několik slov (keyphrase)
- Použití klíče:
  - write key (with repeated letters deleted)
  - then write all remaining letters in columns underneath
  - then read off by columns to get ciphertext equivalents

```
STARW
BCDEF
GHIJK
LMNOP
QUVXY
Z

Plain:  ABCDEFGHIJKLMNOPQRSTUVWXYZ
Cipher: SBGLQZTCHMUADINVREJOXWFKPY

Plaintext:  I KNOW ONLY THAT I KNOW NOTHING
Ciphertext: H UINF NIAP OCSO H UINF INOCHIT
```

©Petr Hanáček CLACRYPT Slide 24

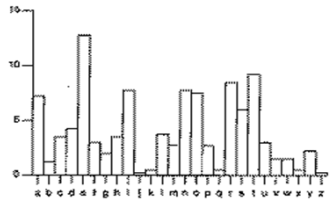
Anatomie jazyka: Frekvence

Frekvence písmen

e	12.31%	l	4.03%	b	1.62%
t	9.59	d	3.65	g	1.61
a	8.05	c	3.20	v	0.93
o	7.94	u	3.10	k	0.52
n	7.19	p	2.29	q	0.20
i	7.18	f	2.28	x	0.20
s	6.59	m	2.25	j	0.10
r	6.03	w	2.03	z	0.09
h	5.14	y	1.88		

Frekvence slov

the	6.421%	that	1.244%
of	4.028%	is	1.034%
and	3.150%	i	0.945%
to	2.367%	it	0.930%
a	2.091%	for	0.770%
in	1.778%	as	0.764%



Frekvence je invariantní vzhledem k monoalfabetické substituci

©Petr Hanáček — CLACRYPT Slide 25

Gadsby

Chapter XXIX.

Gadsby was walking back from a visit down in Branton Hill's manufacturing district on a Saturday night. A busy day's traffic had had its noisy run; and with not many folks in sight, His Honor got along without having to stop to grasp a hand, or talk; for a Mayor out of City Hall is a shining mark for any politician. And so, coming to Broadway, a booming bass drum and sounds of singing, told of a small Salvation Army unit carrying on amidst Broadway's night shopping crowds. Gadsby, walking toward that group, saw a young girl, back towards him, just finishing a long, soulful oration, saying:

"... and I can say this to you, for I know what I am talking about; for I was brought up in a pool of liquor!"

As that army group was starting to march on, with this girl turning towards Gadsby, His Honor had to grasp, astonishingly:

"Why! Mary Antor!"

"Oh! If it isn't Mayor Gadsby! I don't run across you much, nowadays. How is Lady Gadsby holding up during this awful war?"

From the novel Gadsby by Ernest Vincent Wright

©Petr Hanáček — CLACRYPT Slide 26

Užitečné vlastnosti pro angličtinu

- Nejčastější písmena
  - Etaoinshrdlu
- Kliky
  - » {e} {t} {aoin} {srh} {ld}
- Nejčastější digramy
  - TH, HE, IN, ER, AN, RE, ED, ON, ES, ST, EN, AT, TO, NT, HA, ND, OU, EA, NG, AS, OR, TI, IS, ET, IT, AR, TE, SE, HI, OF
- Nejčastější trigramy
  - THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR, DTH

 CLACRYPT Slide 27

Příklad – kryptoanalýza

- Použijeme ciphertext-only attack
- Šifrovaný text je:

BTLDXFETMDGLGMVMYFQEMQAPMVBZQMXZQEGZVXFTL
XGUWFVXBFWDYUXUQFQXUBGQZBMYMBBFHQXFPXGU
VHISUBXZVCMGQVXGUBFAUITUMCUTVXGZVIFFCXTMBUV
BTLDXFETMDGLPTFWZXVZQZXZMYMQAYZWZXUAHVUIL
XGUUELDXZMQVVFWUPFHTXGFHVMQALUMTVMEFXFXGU
XKUQXZUXGBUQXHTLKGUTUZXDYMLUAMBTHBZMYTFYU
ZQXGUFHXBFWUFPIFXGKFTYAKMTVBFWDYUXUAZQ
QZQUXUUQVZJXLXGTUUXGUIFFCBFOUTVXGFVUMVDUBXV
FPXGUGZVXFTLKGZBGKUTUWFVXVZEQZPZBMQXXFXGU
AUOUYFDWUQXFPXGUVHISUBX

 CLACRYPT Slide 28

Příklad – kryptoanalýza

- **Statistika písmen:**

letter	prob	letter	prob	letter	prob
A	.023	J	.003	S	.005
B	.054	K	.015	T	.054
C	.010	L	.030	U	.120
D	.026	M	.061	V	.066
E	.018	N	.000	W	.023
F	.090	O	.005	X	.118
G	.066	P	.020	Y	.028
H	.023	Q	.059	Z	.064
I	.018	R	.000		

- **Nejčastější digramy: XG (16), GU (11), XF (8), QX (7), VX (7), BF (6), UX (6), ZQ (6)**
- **Nejčastější trigramy: XGU (10), BFW, FPX, FXG, GZV, LDX, LXG, MQA, PXG, UBX, UQX, UXU, VXG – ostatní se vyskytují třikrát**

©Petr Hanáček — CLACRYPT Slide 29

Příklad – kryptoanalýza

- **Několik předpokladů:**
  - **U a X jsou nejčastější písmena. Předpokládejme, že jde o písmena E a T** — U=>E, X=>T
  - **Nejčastější digram je XG. To znamená, že X = T a pak G = H. THE je nejčastější trigram v angličtině, takže lze odvodit, že U = E.** — G=>H
  - **XF je poměrně častý digram. Víme, že X = T. Pak XF může být TO nebo TI. O je v angličtině o něco častější než I, takže předpokládejme, že F = O.** — F=>O
- **X = T, G = H, U = E, F = O**

BTLDtoETMDhLhMVMYoQEMQAPMVBZQMtZQEhZVtoTL
theWoVtBoWDYeteQoQteBhQZBMYMBBoHQtoPthe
VHISeBtZVCMhQVtheBoAeITeMCeTVthZVIoCtTMBeV
BTLDtoETMDhLPToWZtVZQZtZMYMQAYZWZteAHVeIL
theeELDtZMQVVoWePoHTthoHVMQALeMTVMEototHe
tKeQtZethBeQtHTLKheTeZtDYMLeAMBTHBZMYToYe
ZQtheoHtBoWeFPIothKoTYAKMTVBoWDYeteAZQ
QZQeteeQVZJtLthTeetheIooCBFOeTVthFVeMVDeBtV
oPthehZVtoTLKhZBhKeTeWoVtVZEQZPZBMQttothe
AeOeYoDWeQtoPtheVHISeBt

©Petr Hanáček — CLACRYPT Slide 30

Příklad – kryptoanalýza

U=>E
X=>T
G=>H
F=>O

- Nyní analyzujme QX a UQX. QX může být AT, NT nebo IT. Pokud se však podíváme na trigram UQX (U=E), pak pravděpodobně Q = N.
- MQA je častý trigram. Vzhledem k tomu, že Q = N, můžeme říct, že MQA = AND. Tedy M = A a A = D.
- Nyní víme, že Q = N, M = A, A = D, X = T, G = H, U = E, F = O.
- Budeme pokračovat stejným způsobem dále.

 CLACRYPT Slide 31

Příklad – kryptoanalýza

- Výsledná zpráva je:

cryptography has a long and fascinating history
the most complete nontechnical account of the
subject is kahns the codebreakers this book traces
cryptography from its initial and limited use by
the egyptians some four thousand years ago to the
twentieth century where it played a crucial role
in the outcome of both world wars completed in
nineteen sixty three the book covers those aspects
of the history which were most significant to the
development of the subject

(Z knihy Handbook of Applied Cryptography, A. Menezes, P. van Oorschot, S. Vanstone)

 CLACRYPT Slide 32

Anatomie jazyka: Patterns

- **Invariance :**
  - **"Pro všechny *monoalfabetické jednoduché substituční šifry je vzorek opakování jednotlivých znaků v textu invariantní vzhledem k této substituci*."**
- **Notace vzorku**
  - **NRGRN -> 12321, NRGRNOR -> 1232142, ...**
  - **ale 123245678 == 121 !!!**
- **Realizace vzorku je často jedinečná**
  - **1221 má asi 250 realizací, ale jenom málo z nich je z vojenského prostředí (assault, attack, battalion, barrack, zeppelin, shipping, missile, commodore) a pouze několik z diplomatického (affair, ambassador,...)**
  - **Opakování slov (např. pro zdůraznění) má vynikající vzorek: SC48SC48 pro spojenecký konvoj**
- **Metoda pravděpodobného slova**
  - **Pokud pravděpodobné slovo je "division", hledáme všechny výskyty vzorku 12131**
- **Test: co je 1234135426 (v kontextu WW2) ?**

 CLACRYPT Slide 33

Zvýšení odolnosti monoalfabetické substituce

- **„Uříznutí vrcholů" tabulky frekvencí**
  - ***Homofonní šifry* - převádějí jeden znak textu na několik jiných znaků textu**
- **Šifrování skupin znaků (např. dvojic znaků)**
  - **Symbol abecedy se skládá z více znaků**
  - ***Polygramové šifry* - nahrazují skupinu znaků jinou skupinou znaků**
- **Každý znak zprávy se zašifruje jinou transformací**
  - ***Polyalfabetické šifry* - různé znaky zprávy jsou převáděny různým způsobem**

 CLACRYPT Slide 34

KRY

Příklad homofonní šifry

- **Zobrazuje písmena na číslice 00, ..., 99**

Plaintext	Ciphertext
A	17 19 34 41 56 60 67 83
I	08 22 53 65 88 90
L	03 44 76
N	02 09 15 27 32 40 59
O	01 11 23 28 42 54 70 80
P	33 91
T	05 10 20 29 45 58 64 78 99

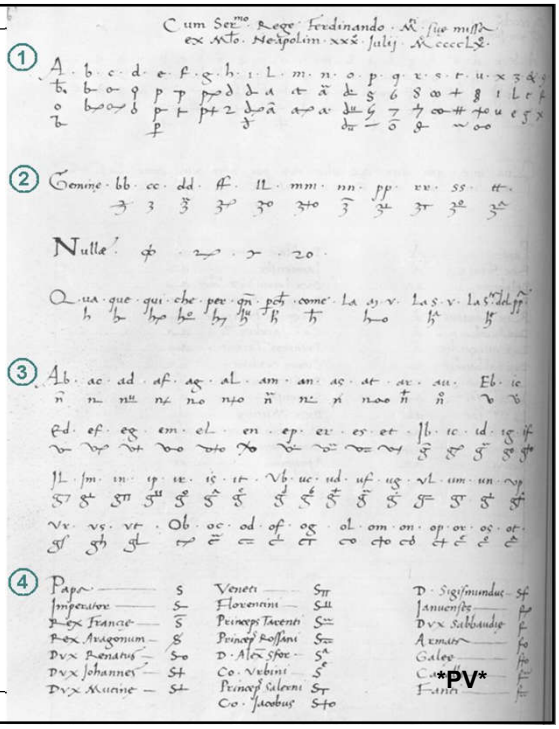
- **Zpráva:** P L A I N P I L O T
- **Šifrovaný text:** 91 44 56 65 59 33 08 76 28 78

©Petr Hanáček — CLACRYPT Slide 35

Nomenklátor 16.století

Nomenklátory z konce patnáctého století obsahují již mimo klasické substituční homofonní záměny (na obrázku v odstavci označeném 1) znaky pro zdvojená písmena (odstavec 2) a slabiky (odstavec 3) a dále kódy pro nejfrekventovanější slova a jména (odstavec 4) a klamače (druhý řádek odstavce 2)

(Klamače – nevýznamové skupiny písmen, které měly ztížit kryptoanalýzu zašifrovaných textů).

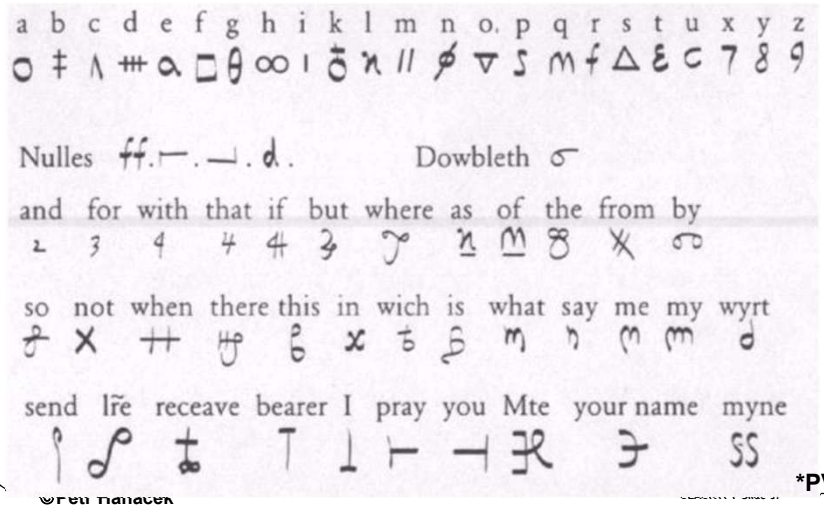


©Petr Hanáček

KRY

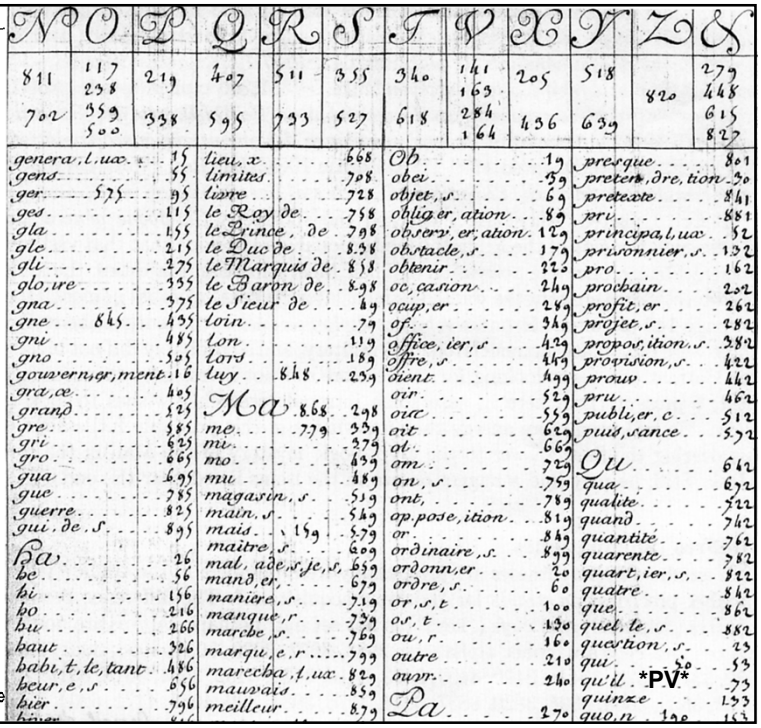
Nomenklátor

- **Přepis osobního nomenklátoru skotské královny Marie Stuartovny (kolem roku 1600)**



PV

Přerod nomenklátoru v kódovou knihu



PV

Kódová kniha - Codebook

- Monoalfabetická polygramová (často homofonní) šifra
- Ukázka kódové knihy (písmeno K)
  - at 5003
  - attack 1701
  - begins 7803
  - the 3243

- plaintext: The attack begins at ...
- ciphertext: 3243 1701 7803 5003 ...

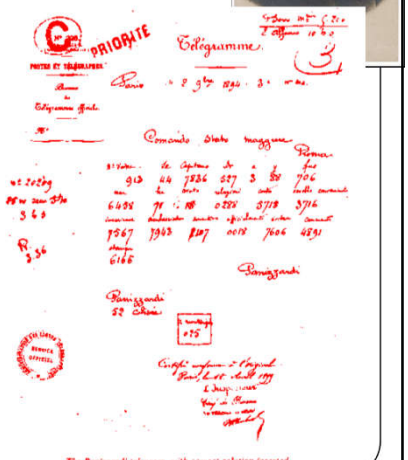
- Např. Dreyfusův telegram, Zimmermannův telegram

 CLACRYPT Slide 39

Dreyfusův telegram

- 1. 11. 1894 francouzské noviny uvedly, že francouzský důstojník kapitán Alfred Dreyfus, předával Německu tajné informace
- 2. 11. Panizzardi, italský vojenská ataše v Paříži, posílá do Říma šifrovaný telegram:
- "Commando stato maggiore Roma 913 44 7836 527 3 88 706 6458 71 18 0288 5715 3716 7567 7943 2107 0018 7606 4891 6165

  Panizzardi"



The Panizzardi telegram, with correct solution inserted

 CLACRYPT Slide 40

- **Francouzi odposlechli telegram**
  - **Bylo třeba zjistit, jaká kódová kniha byla použita**
  - **Podle skupin číslic to vypadalo na komerční tzv. Baravelliho kód**
  - **Dešifrování dalo:**

913 44 7836 527 3 88 706 6458 71 ...
us le rimprovera nar i te ren pensato sara ...

  - **což jsou nesmysly**
- **Další hypotéza byla, že v telegramu je Dreyfusovo jméno, které je podle Baravelliho kódové knihy**
  - **Dreyfus = 227 1 98 306**
  - **Ve zprávě ale tato posloupnost není**
  - **Zato v ní je**
    - » **527 3 88 706**
  - **Což může znamenat, že Panizzardi zašifroval nejlevější číslici skupiny nějakou jinou šifrou**

 CLACRYPT Slide 41

- **10. 11. byla nalezena i Panizzardiho šifra**
- **Jde o substituci:**

First plaincode digit 0 1 2 3 4 5 6 7 8 9
First ciphertext digit 1 3 5 7 9 0 2 4 6 8

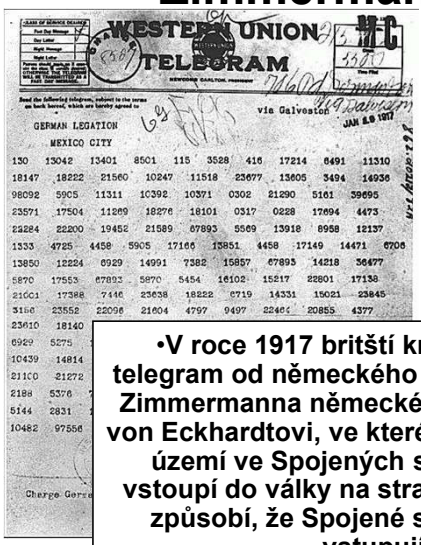
- **A zpráva je:**

74 1336 227 1 98 306 5858 31 08 7588 ...
Se Capitano Dr e y fus non ha avuto relazione...

- **Což znamená "Pokud kapitán Dreyfus s vámi není v žádném vztahu, bylo by moudré, aby to velvyslanec oficiálně popřel, aby se zabránilo spekulacím novinářů."**

 CLACRYPT Slide 42

Zimmermannův telegram


WESTERN UNION TELEGRAM

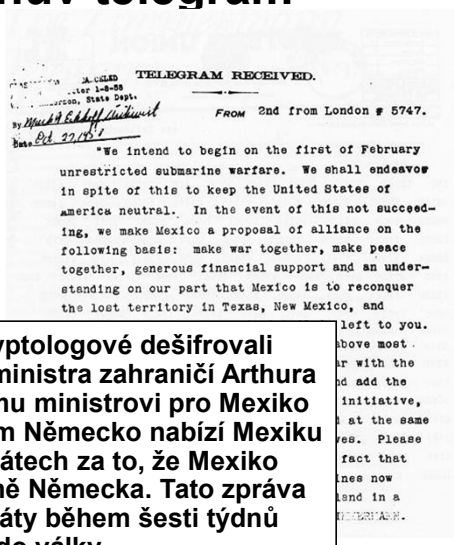
via Galveston

JAN 16 1917

GERMAN LEGATION

MEXICO CITY

130 13042 13401 8501 115 3528 416 17214 6491 11310
18147 18222 21560 10247 11518 23677 13605 3494 14936
98092 5905 11311 10392 10371 0302 21290 5161 39695
23571 17504 11269 18276 18101 0317 0228 17694 4473
22284 22200 19452 21589 67893 5569 13918 8958 12137
1333 4725 4458 5905 17166 13851 4458 17149 14471 6706
13850 12224 6929 14991 7382 15857 67893 14218 36477
5870 17553 67893 5870 5454 16102 15217 22801 17138
21001 17388 7446 23638 18222 6719 14331 15021 23845
3156 23552 22096 21604 4797 9497 22464 20855 4377
23610 18140
6929 5275
10439 14814
21100 21272
2188 5376
5144 2831
10482 97556


TELEGRAM RECEIVED.

FROM 2nd from London # 5747.

"We intend to begin on the first of February unrestricted submarine warfare. We shall endeavor in spite of this to keep the United States of America neutral. In the event of this not succeeding, we make Mexico a proposal of alliance on the following basis: make war together, make peace together, generous financial support and an understanding on our part that Mexico is to reconquer the lost territory in Texas, New Mexico, and

left to you.
above most
with the
add the
initiative,
at the same
Please
fact that
lines now
land in a

•V roce 1917 britští kryptologové dešifrovali telegram od německého ministra zahraničí Arthura Zimmermanna německému ministrovi pro Mexiko von Eckhardtovi, ve kterém Německo nabízí Mexiku území ve Spojených státech za to, že Mexiko vstoupí do války na straně Německa. Tato zpráva způsobí, že Spojené státy během šesti týdnů vstupují do války.

Playfair - digrafická substituce

- **Objevena v r. 1854 Charlesem Weatstonem**
- **25 znaků (abeceda vyjma „j“), počínaje heslem, je vepsáno do čtverce 5x5, který je chápán jako torus**
  - **P A L M E T O N R S**
  - **R S T O N D F G B C**
  - **B C D F G or K Q U H I**
  - **H I K Q U X Y Z V W**
  - **V W X Y Z L M E P A**
  - **Pokud se digram otevřeného textu nachází v jednom řádku (sloupci) je nahrazen písmeny od něj vpravo (dolů)**
    - » **am -> LE, dl -> KT**
  - **Jinak je první písmeno nahrazeno písmenem ve stejném řádku ale ve sloupci druhého znaku a naopak**
    - » **ag -> EC, ho -> QR**
  - **Pokud je v textu dvojice stejných znaků, je jeden znak vynechán**

Bezpečnost Playfair

- Bezpečnost je výrazně větší než u monoalfabetické šifry
- Jelikož máme 26 x 26 = 676 digramů
- Potřebovali bychom pro analýzu frekvenční tabulku s 676 položkami (oproti 26 u monoalfabetické substituce)
- Tedy je třeba mnohem větší množství textu pro analýzu
- Široce používaná po mnoho let (např. armáda USA a GB ve WWI)
- Bohužel je ji však možno rozluštit při délce zprávy několik set znaků
- Stále obsahuje příliš mnoho informací o struktuře zprávy

 CLACRYPT Slide 45

Identifikace Playfairu

- Text, zašifrovaný šifrou Playfair má jisté charakteristiky, které lze použít pro identifikaci této šifry
- Jelikož jde o substituční šifru, vzácné souhlásky (pro angličtinu) j, k, q, x a z se vyskytují častěji než v otevřeném textu a digramy s těmito souhláskami se objevují také častěji
- Ve zprávě je vždy sudý počet písmen
- V digramech se neobjevují dvojice stejných písmen jako SS, EE, MM, . . .

 CLACRYPT Slide 46

Vlastnosti šifry Playfair

- **Playfair má některé vlastnosti, které lze využít pro rozluštění textu**
  - **Opakování a frekvence digramů je obecně stejná jako u jejich ekvivalentů v otevřeném textu**
  - **Žádné písmeno v otevřeném textu nemůže být zašifrováno samo na sebe, tj. "a" nemůže být zašifrováno jako "a"**
  - **Dva reverzované digramy v otevřeném textu (např. ER a RE) budou vždy zašifrovány jako dva reverzované digramy v zašifrovaném textu**
  - **Každé písmeno v otevřeném textu může být zašifrováno jenom jako jedno z pěti písmen – jedno, které je bezprostředně pod ním ve stejném sloupci a čtyři, které jsou ve stejném řádku**
  - **Výsledkem je, že 3 nebo 4 nejčastější písmena se nacházejí ve stejném řádku jako "e"**
  - **A mnoho dalších…**
- **Příklad rozluštění Playfairu — cca 20 slajdů**

 CLACRYPT Slide 47

Hillova šifra - polygrafická substituce

- **Objevena v r. 1929 Lesterem Hillem**
- **První polygrafická šifra použitelná s více než třemi znaky zároveň**

Let d=2 , M= m1 m2, C= c1, c2 where:

$C1 = (k_{11}m1 + k_{12}m_2) \bmod n$

$C2 = (k_{21}m1 + k_{22}m_2) \bmod n$

Where K = $\begin{bmatrix} K11 & k12 \\ K21 & K22 \end{bmatrix}$

That is : $\begin{bmatrix} C1 \\ C2 \end{bmatrix} = \begin{bmatrix} K11 & k12 \\ K21 & K22 \end{bmatrix} * \begin{bmatrix} m1 \\ m2 \end{bmatrix} \bmod n$

$Ek(M) = K*M$

$Dk(C) = K^{-1} *C \bmod n$

$= K^{-1} K M \bmod n$

$= M$

Where $K.K^{-1} \bmod n = I$ (Identical matrix)

 lide 48

Hillova šifra - příklad

$\sum$ =A...Z

M = EG

K

3	2
3	5

K^{-1}

15	20
17	9

Mod 26 =

I

1	0
0	1

M = E G = 4 6

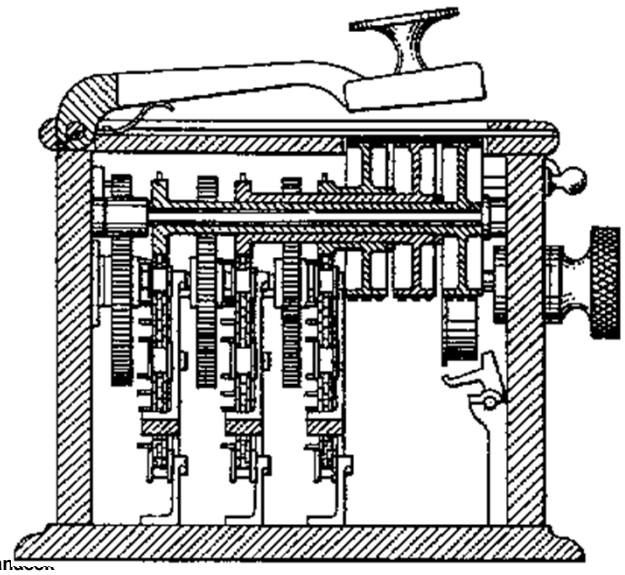
C1	=	3	2	*	4	mod 26 =	24	=	Y
C2		3	5		6		16		Q

To decipher:-

m1	=	15	20	*	24	mod 26 =	4	=	E
m2		17	9		16		6		G

©Petr Hanáček CLACRYPT Slide 49

Hillova šifra – HW řešení



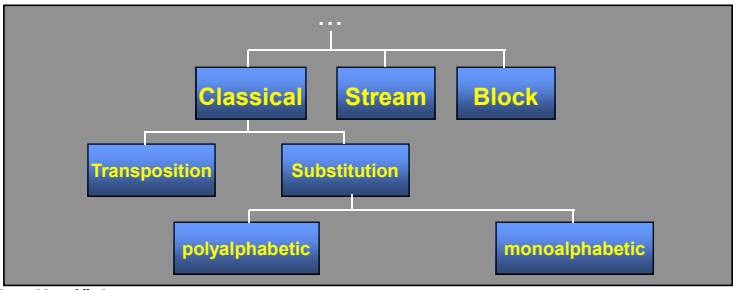
©Petr Hanáček Slide 50

Polyalfabetické substituční šifry

 CLACRYPT Slide 51

Polyalfabetická substituce

- Použití více různých substitucí
- Každý znak je zašifrován jinou substituční funkcí
- Zploštění frekvenční charakteristiky jazyka
- Frekvenční analýzu ani jiné statistické metody nelze použít



 CLACRYPT Slide 52

První pokusy (bez klíče)

- **Leon Battista Alberti**
  - **Albertiho disk - první polyalfabetický šifrový systém**
  - **Doporučoval posunout abecedy po třech nebo čtyřech slovech**
- **Johannes Trithemius (tabula recta)**
  - **Kniha Polygraphia (1518)**
  - **Tabula recta**

1 ABCDEFGHIJKLMNOPQRSTUVWXYZ
2 BCDEFGHIJKLMNOPQRSTUVWXYZA
3 CDEFGHIJKLMNOPQRSTUVWXYZAB
4 DEFGHIJKLMNOPQRSTUVWXYZABC
5 EFGHIJKLMNOPQRSTUVWXYZABCD
6 FGHIJKLMNOPQRSTUVWXYZABCDE
7 GHIJKLMNOPQRSTUVWXYZABCDEF
8 HIJKLMNOPQRSTUVWXYZABCDEFG
9 IJKLMNOPQRSTUVWXYZABCDEFGH
10 JKLMNOPQRSTUVWXYZABCDEFGHI
11 KLMNOPQRSTUVWXYZABCDEFGHIJ
12 LMNOPQRSTUVWXYZABCDEFGHIJK
. . .
24 XYZABCDEFGHIJKLMNOPQRSTUVW
25 YZABCDEFGHIJKLMNOPQRSTUVWX
26 ZABCDEFGHIJKLMNOPQRSTUVWXY

Vigenerova šifra

- **Blaise de Vigenère, ~1550**
  - **Působil ve službách vévody Navarrského**
  - **S kryptografií přišel do styku jako diplomat ve Vatikánu**
- **Nerozlomitelná cca 300 let**
- **Používá Caesarova principu**
  - **s rozdílnými posuvy pro jednotlivé znaky, aby se zakryla frekvence znaků**
  - **znaky klíče definují posuv pro jednotlivá písmena**
  - **klíč je periodicky opakován, aby obsáhl celou délku šifrovaného textu**
- **Příklad:**

```
Otevřený text:   vigenerescipher
Klíč:            keykeykeykeykey
Šifrovaný text:  FMEORCBIQMMNRIP
```

  - **a=0, b=1, c=2, ... z=25 mod 26**
- **Vigenerova tabulka**

Vigenerova tabulka

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

Vigenere - autoklíč

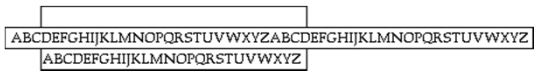
- **Mechanismus, kdy je domluven pouze počáteční klíč (často jen jedno písmeno), který je pak modifikován**
  - **Samotnou zprávou (autoklíč otevřeného textu)**
  - **Zašifrovanou zprávou (autoklíč zašifrovaného textu)**
- **Autoklíč otevřeného textu**
  - **Domluvený klíč bude D, otevřený text ALBATROS:**
  - **Autoklíč = D+otevřený text**
  - **Klíč: DALBATRO**
  - **Otevřený text: ALBATROS**
  - **Šifrový text: DLMBTKFG**

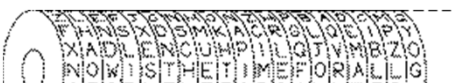
Jiné varianty

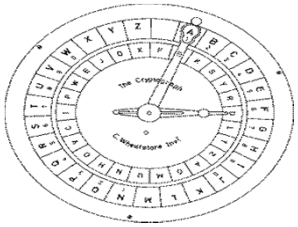
	šifrování	dešifrování
• Vigenére:	M + K = C	C – K = M
• Beaufort:	K – M = C	K – C = M
• Beaufort':	M – K = C	C + K = M
• Dneska:	M ⊕ K = C	C ⊕ K = M

Šifrovací strojky

- Saint-Cyrovo pravítko
- Jeffersonův cylindr
  - Vytvořen v r. 1790, skládal se z 36 disků, každý s náhodnou abecedou, pořadí disků bylo klíčem, v jedné řadě se nastavila zpráva, v jiné řadě se přečetla zašifrovaná zpráva
- Wheatstonův disk
  - Původně vynalezen Wadsworthem v r.1817, pak vyvinut Wheatstonem v r. 1860, skládal se ze dvou soustředných kruhů, které generovaly polyalfabetickou šifru







Vigenerova šifra (útok)

- **Rozluštěna Charlesem Babbagem, ale postup byl utajován**
- **Nezávisle rozluštěna Friedrichem Kasiskim, 1863.**

- **1. Nalezni délku klíče *k***
  - **pro krátký klíč zkus 1, 2, 3, …, nebo**
  - **vytvoř tabulku všech vzdáleností stejných znaků v zašifrovaném textu**
  - **gcd nejčastějších vzdáleností je délka klíče**
- **2. Nalezni písmena klíče jedno po druhém**
  - **rozděl zprávu na *k* menších zpráv, znichž každá obsahuje znaky, šifrované stejným písmenem klíče**
  - **řeš šifru jako *k* zpráv, zašifrovaných Caesarovou šifrou**

 CLACRYPT Slide 60

Původní Babbageův útok

- **Využití opakování pro uhodnutí délky klíče:**

  Posloupnost XFO se nachází na pozicích 65, 71, 122, 176.

  Vzdálenosti jsou = (71 – 65) = 6 = 3 * 2

  (122 – 65) = 57 = 3 * 19

  (176 – 122) = 54 = 3 * 18

  Klíč má pravděpodobně délku 3 znaky.

 CLACRYPT Slide 61

Příklad – kryptoanalýza Vigenère

- **Vigenèrova šifra je polyalfabetická, takže analýza je obtížnější.**

- **Zašifrovaný text je:**

MRGFNIATXZQVFFNUXFFYBTCETYXIIXGZKACJLRGKQYEIX
OYYAUAPXYIJLHPRGVTSFPAYNNYURZOPHXWYXLFRNUTZBR
FKAHFWFZESYUWZMOLLBSBZBJHFPLXKHVIVMZTZHUIWAET
IUEDFGLXDIEXIYJIUXPNNEIXABVCINTVCIEZYYDAZGZIW
TYXJIKTRZLMFFKALGZNVKZXIIMXUUNAPGVXFUSMISKHVY
VOCRVXRIWTYXZOIRFNUXZNXLDUDPZGVHVOWMOYJERLAUG
LVTUXTHRBUQZTYTXORNKBASFFXGHQVDSHUYJSYHDYUWYX
YYKHVTUCDACAHXSEVGJIEFZGLXRSBXSYKOEPPNYAKTUAC
EFYILFWEAHCIAUALLZNXMVCKLRRHGFNXMOYUESKPM

 CLACRYPT Slide 62

Příklad – kryptoanalýza Vigenère

- Jak bylo řečeno, Vigenèrova šifra má klíč o délce *m*
- Prvním krokem je zjištění délky klíče
- Lze použít dva postupy - Kasiskiho test a index koincidence

 CLACRYPT Slide 63

Příklad – kryptoanalýza Vigenère

- Kasiskiho test byl vytvořen v roce 1863 pruským důstojníkem Friedrichem Kasiskim.
- Metoda je založena na pozorování, že dva identické úseky otevřeného textu budou zašifrovány na stejný zašifrovaný text tehdy, pokud jsou vzdáleny δ pozic od sebe ($\delta \equiv 0 \pmod{m}$)
- Našim cílem je nalézt několik identických úseků textu, každý o délce nejméně 3, a zaznamenat si vzdálenost mezi jejich počátky. Délka klíče *m* dělí všechny vzdálenosti $\delta_1, \delta_2, \ldots, \delta_n$. Pak je tedy *m* největší společný dělitel všech vzdáleností δ_i
- V předloženém zašifrovaném textu se trigram TYX vyskytuje třikrát. Počáteční pozice jsou 25, 181 a 235. Vzdálenost mezi první a druhou je 156, mezi první a třetí je 210. GCD těchto čísel je 6, takže můžeme předpokládat, že délka klíče je také 6

 CLACRYPT Slide 64

Příklad – kryptoanalýza Vigenère

- Nyní použijeme pro stejný účel index koincidence
- Index koincidence je definován takto:
  - Nechť $x = x_1 x_2 \ldots x_n$ je řetězec *n* znaků. Index koincidence řetězce *x*, označený $I_c(x)$ je definován jako pravděpodobnost, že dva náhodné prvky *x* jsou identické.
- Označíme frekvence písmen *A*, *B*, *C*, …, *Z* v řetězci *x* symboly $f_1, f_2, f_3, \ldots, f_{25}$.
- Dva prvky v *x* můžeme vybrat $\binom{n}{2}$ způsoby
- Existuje $\binom{f_i}{2}$ způsobů jak vybrat stejný prvek
- Pak získáme vzorec pro průměrný index koncidence (někdy značený Phi(x)) :

$$I_c(x) = \frac{\sum_{i=0}^{25}\binom{f_i}{2}}{\binom{n}{2}} = \frac{\sum_{i=0}^{25} f_i(f_i-1)}{n(n-1)}$$

 CLACRYPT Slide 65

Příklad – kryptoanalýza Vigenère

- Index koincidence řetězce v angličtině je přibližně 0.065

$$I_c(x) \approx \sum_{i=0}^{25} p_i^2 = 0.065$$

  - Kde p_i je frekvence písmene a_i. (pravděpodobnost, že na stejném místě bude totéž písmeno je tedy p^2)
- Totéž platí, pokud *x* je zašifrovaný text, vytvořený monoalfabetickou šifrou
- Nyní můžeme zapsat zašifrovaný text následujícím způsobem:

$$\begin{aligned} c_1 &= c_1 c_{m+1} c_{2m+1} \ldots \\ c_2 &= c_2 c_{m+2} c_{m+2} \ldots \\ &\ldots \\ c_m &= c_m c_{2m} c_{3m} \ldots \end{aligned}$$

- Pokud jsou $c_1, c_2, \ldots, c_m$ vytvořeny tak, že tímto *m* je délka klíče, pak každé $I_c(c_i)$ by mělo být přibližně 0.065

 CLACRYPT Slide 66

Příklad – kryptoanalýza Vigenère

- Na druhou stranu, pokud *m* není délka klíče, pak řetězce c_i vypadají náhodněji. Zcela náhodný řetězec by měl

$$I_c \approx 26\left(\frac{1}{26}\right)^2 = \frac{1}{26} = 0.038$$

- Následující tabulka obsahuje I_c pro různé hodnoty *m*:

m	I_c
1	0.043
2	0.052; 0.051
3	0.05; 0.059; 0.045
4	0.049; 0.053; 0.052; 0.051
5	0.034; 0.05; 0.048; 0.038; 0.045
6	0.063; 0.07; 0.083; 0.062; 0.071; 0.048
7	0.033; 0.041; 0.038; 0.046; 0.041; 0.04; 0.047

- Tato metoda také ukazuje, že *m* = 6

©Petr Hanáček — CLACRYPT Slide 67

Příklad – kryptoanalýza Vigenère

- Pro zjištění samotného hesla použijeme metodu podobnou indexu koincidence
- Každý podřetězec c_i byl vytvořen pomocí monoalfabetické šifry. Pokud posuv abecedy nazveme g, pak lze vytvořit následující vzorec

$$M_g = \sum_{i=0}^{25} \frac{p_i f_{i+g}}{n}$$

  kde f_1, f_2, …, f_i jsou frekvence písmen *A*, *B*, …, *Z* v podřetězci c_i, a *n* je délka podřetězce
- Pokud je *g* správná hodnota posuvu, pak, M_g bude zhruba rovno 0.065
- Nyní vybereme nejpravděpodobnější hodnotu M_g pro každý podřetězec

©Petr Hanáček — CLACRYPT Slide 68

Příklad – kryptoanalýza Vigenère

i	M_g
1	0.062; 0.042; 0.033; 0.035; 0.041; 0.039; 0.030; 0.040; 0.036; 0.039; 0.026; 0.040; 0.043; 0.046; 0.038; 0.046; 0.032; 0.033; 0.042; 0.043; 0.037; 0.029; 0.047; 0.035; 0.032; 0.036
2	0.033; 0.037; 0.035; 0.035; 0.046; 0.042; 0.048; 0.040; 0.032; 0.028; 0.043; 0.040; 0.038; 0.046; 0.037; 0.026; 0.042; 0.065; 0.037; 0.033; 0.041; 0.044; 0.029; 0.036; 0.038; 0.034
3	0.038; 0.030; 0.038; 0.029; 0.043; 0.041; 0.052; 0.034; 0.041; 0.041; 0.036; 0.033; 0.040; 0.040; 0.028; 0.050; 0.031; 0.025; 0.036; 0.073; 0.039; 0.035; 0.034; 0.044; 0.033; 0.038
4	0.040; 0.043; 0.034; 0.047; 0.038; 0.031; 0.042; 0.064; 0.037; 0.027; 0.030; 0.042; 0.036; 0.036; 0.038; 0.039; 0.043; 0.041; 0.040; 0.034; 0.044; 0.042; 0.040; 0.033; 0.027; 0.034
5	0.030; 0.037; 0.034; 0.030; 0.046; 0.047; 0.041; 0.036; 0.035; 0.043; 0.047; 0.035; 0.038; 0.035; 0.033; 0.036; 0.049; 0.034; 0.027; 0.044; 0.065; 0.037; 0.026; 0.044; 0.045; 0.028
6	0.031; 0.039; 0.041; 0.041; 0.038; 0.044; 0.044; 0.034; 0.030; 0.037; 0.039; 0.036; 0.035; 0.039; 0.034; 0.034; 0.042; 0.059; 0.043; 0.029; 0.036; 0.043; 0.037; 0.033; 0.039; 0.035

- **Našli jsme klíč, je to ARTHUR**

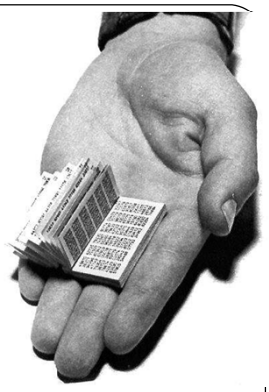
Příklad – kryptoanalýza Vigenère

- **Rozluštěná zpráva (s přidanými mezerami) je:**

many traces we found of him in the boggirt island where he had hid his savage ally a huge drivingwheel and a shaft halffilled with rubbish showed the position of an abandoned mine beside it were the crumbling remains of the cottages of the miners driven away no doubt by the foul reek of the surrounding swamp in one of these a staple and chain with a quantity of gnawed bones showed where the animal had been confined a skeleton with a tangle of brown hair adhering to it lay among the debris.

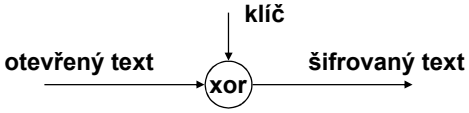
(Z knihy Pes Baskervillský, Arthur Conan Doyle)

Vernamova šifra
One Time Pad



Vernamova šifra

- **Polyalfabetická substituce bez opakování klíče**
- **Objevena Gilbertem S. Vernamem z AT&T v roce 1917 pro šifrování telegrafních zpráv**



- **Šifra je nerozluštitelná, pokud:**
  - **Klíč má stejnou délku jako všechny šifrované zprávy**
  - **Klíč se nikdy nepoužije znovu**
  - **Klíč je náhodně zvolen (opravdu náhodně)**
- **Claud Elwood Shannon prokázal, že tento systém je *absolutně bezpečný šifrový systém***

ETCRRM

- **Horká linka mezi USA a SSSR**
- **Vytvořena po kubánské krizi 30.8. 1963**
- **Byla použita Vernamova šifra pomocí zařízení ETCRRM-II (Electronic Teleprinter Cryptographic Regenerative Repeater Mixer II)**
- **Šifrovaný text se odečítal od klíčové pásky, klíčová páska byla automaticky po použití ničena**

PV

 CLACRYPT Slide 73

DEINSTAR

- **Varianta Vernamovy šifry používaná během studené války BND (Bundesnachrichtendienst)**
- **Text se tabulkou převedl na jedno až dvoumístné číselné symboly a k těm se přičetl klíč**

	0	1	2	3	4	7	8	9
	D	E	I	N	S	T	A	R
5	B	C	F	G	H	K	L	M
6	O	P	Q/J	U	V	W	X/Y	Z
								PV

Otevřený text (o): Děkuji za pozornost.
Převod do mezinárodní abecedy: **DEKUJI ZA POZORNOST**
Převod podle tabulky (O): 0 1 57 63 62 2 69 8 61 60 69 60 9 3 60 4 7
Rozpis do pětimístných skupin: 01576 36226 98616 06960 93604 7
Heslo (K): 75409 78210 87302 10834 52019 3
Vigenére (O + K = Š): **76975 04436 75918 16794 45613 0**

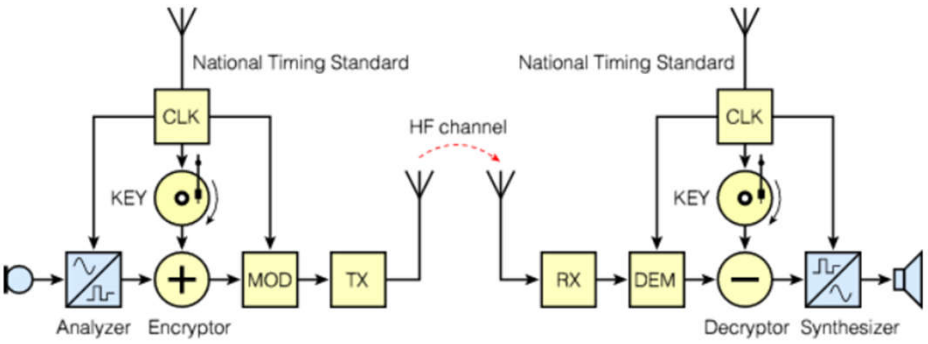
 CLACRYPT Slide 74

SIGSALY

- Bezpečný radiotelefon, 1943-1946
- SIGSALY - Secure Digital Voice Communications
- První využití:
  - Zakódování hlasu pomocí kompandovaného PCM
  - Šifrované telefonie
  - Komprese šířky pásma hlasu
  - Technologie rozprostřeného spektra
  - Spolehlivého přenosu pomocí Frequency Shift Keying (FSK) a FDM (Frequency Division Multiplex)
- Váha asi 50 tun, spotřeba 30 kW
- Jednorázový klíč na gramofonové desce
- Monitorováno Němci, ale nerozluštěno
- Asi 12 instalací, např. Pentagon, Oxford Street v Londýně, na jedné z lodí generála Douglase MacArthura
- Deklasifikováno v roce 1976
- http://en.wikipedia.org/wiki/SIGSALY

©Petr Hanáček CLACRYPT Slide 75

SIGSALY



©Petr Hanáček CLACRYPT Slide 76

Source: https://www.cryptomuseum.com/crypto/usa/sigsaly/index.htm

SIGSALY

©Petr Hanáček

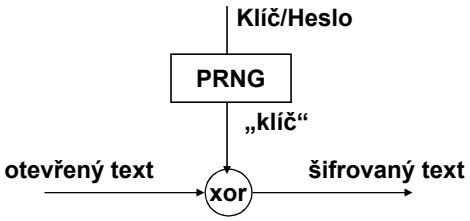
Co NENÍ Vernamova šifra

- Šifra Autokey
- Navržená Vigenèrem
- Klíč je tak dlouhý jako zpráva – nejdřív se použije krátké heslo, pak místo hesla nastupuje posunutý otevřený text
- Příklad:
  - Heslo je „deceptive“
  - `key:` `deceptivewearediscoveredsav`
  - `plaintext:` `wearediscoveredsaveyourself`
  - `ciphertext:` `ZICVTWQNGKZEIIGASXSTSLVVWLA`
- Klíč nesplňuje požadavek náhodnosti

©Petr Hanáček CLACRYPT Slide 78

Co NENÍ Vernamova šifra

- Nekonečný klíč se vyrobí z konečného klíče/hesla (např. o délce 128 bitů) pomocí generátoru pseudonáhodné posloupnosti (PRNG)
- Tento princip se často nazývá „proudová šifra“
- Klíč nesplňuje požadavek dostatečné délky pro Vernamovu šifru



©Petr Hanáček CLACRYPT Slide 79

Transpoziční šifry

©Petr Hanáček CLACRYPT Slide 80

Ne moc dobrá transpoziční šifra…

Aoccdrnig to rscheearch at an Elingsh uinervtisy,
it deosn't mttaer in waht oredr the ltteers in a wrod are,
olny taht the frist and lsat ltteres are at the rghit pcleas.
The rset can be a toatl mses and you can sitll raed it
wouthit a porbelm. Tihs is bcuseae we do not raed
ervey lteter by ilstef, but the wrod as a wlohe.

©Petr Hanáček CLACRYPT Slide 81

Scytale

- **Šifra Scytale**
  - **Stará transpoziční šifra používaná v Řecku**
  - **Pomocí proužku papíru na tyči**
  - **Zpráva se zapisuje na omotaný proužek papíru po řádcích**
  - **Na proužku jsou zdánlivě náhodné znaky**

- **Není příliš bezpečná – klíčem je pouze průměr tyče**

©Petr Hanáček CLACRYPT Slide 82

KRY

Rail Fence

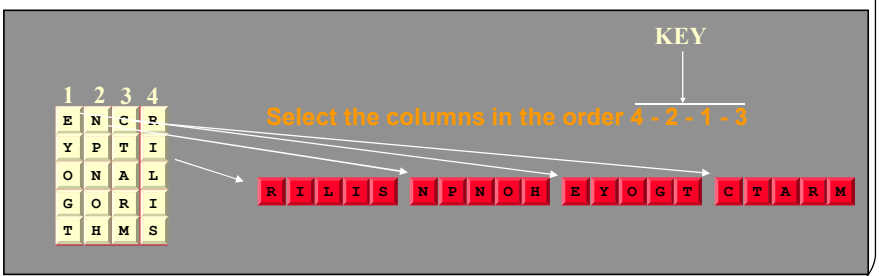
•

T	**5**	T												R							
H	**3**			H									C		E						E
E	**1**					I						E				T				G	
K	**4**						S				S						M				A
E	**2**							I		A								E		S	
Y	**6**								S										S		

- **Railfence: TRHCEEIETGSSMAIAEASS**
- **Redfence (klíčovaná): IETGIAESHCEESSMATRSS**

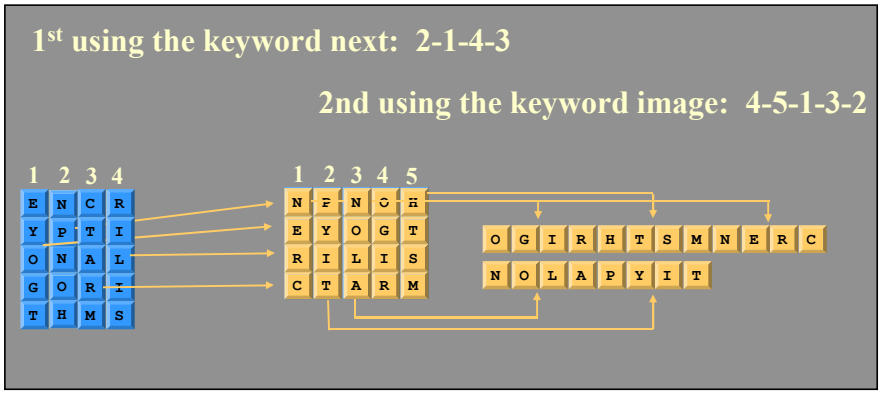
Sloupcová transpozice
Columnar Transposition

- **Zpráva se zapíše po řádcích a šifrovaný text se vytvoří tak, že se čte po sloupcích v daném pořadí**
- **Např. zpráva je “encryption algorithms”, matice je 5x4 a klíč je 4 - 2 - 1 - 3**



Dvojitá sloupcová transpozice
Double Columnar Transposition

- Pozn: Pro snadnější zapamatování klíče je možné mu přiřadit slovo. Klíč je pak pořadí písmen slova v abecedě (next = 2143).
- Příklad:



©Petr Hanáček CLACRYPT Slide 85

Složené šifry
Product ciphers

©Petr Hanáček CLACRYPT Slide 86

Složené šifry - Product Ciphers

- Jedna šifra (substituční nebo transpoziční) se nejevila dostatečně bezpečná
- Nabízí se zašifrovat zprávu postupně několika šiframi, ale
  - Dvě substituce za sebou tvoří jenom jednu (složitější) substituci
  - Dvě transpozice za sebou tvoří jenom jednu (složitější) transpozici
  - Ale substituce následovaná transpozicí vytvoří novou, bezpečnější šifru
- Složené šifry se obvykle skládají z kombinací substitucí a transpozicí
- Pro ruční realizaci dost komplikovaná, ale používala se
- Pro mechanický stroj také příliš komplikovaná
- Princip se ale používá v moderní kryptografii

 CLACRYPT Slide 87

KONEC

Credits:

Obrázky označené *PV* byly převzaty z prezentace Pavla Vondrušky

 CLACRYPT Slide 88

CHAPTER 3

Historical Ciphers

Chapter Goals

- To explain a number of historical ciphers, such as the Caesar cipher, substitution cipher.
- To show how these historical ciphers can be broken because they do not hide the underlying statistics of the plaintext.
- To introduce the concepts of substitution and permutation as basic cipher components.
- To introduce a number of attack techniques, such as chosen plaintext attacks.

1. Introduction

An encryption algorithm, or cipher, is a means of transforming plaintext into ciphertext under the control of a secret key. This process is called encryption or encipherment. We write

$$c = e_k(m),$$

where

- m is the plaintext,
- e is the cipher function,
- k is the secret key,
- c is the ciphertext.

The reverse process is called decryption or decipherment, and we write

$$m = d_k(c).$$

Note, that the encryption and decryption algorithms e, d are public, the secrecy of m given c depends totally on the secrecy of k.

The above process requires that each party needs access to the secret key. This needs to be known to both sides, but needs to be kept secret. Encryption algorithms which have this property are called *symmetric cryptosystems* or secret key cryptosystems. There is a form of cryptography which uses two different types of key, one is publicly available and used for encryption whilst the other is private and used for decryption. These latter types of cryptosystems are called *asymmetric cryptosystems* or *public key cryptosystems*, to which we shall return in a later chapter.

Usually in cryptography the communicating parties are denoted by A and B. However, often one uses the more user-friendly names of Alice and Bob. But you should not assume that the parties are necessarily human, we could be describing a communication being carried out between two autonomous machines. The eavesdropper, bad girl, adversary or attacker is usually given the name Eve.

In this chapter we shall present some historical ciphers which were used in the pre-computer age to encrypt data. We shall show that these ciphers are easy to break as soon as one understands the statistics of the underlying language, in our case English. In Chapter 5 we shall study this relationship between how easy the cipher is to break and the statistical distribution of the underlying plaintext.

TABLE 1. English letter frequencies

Letter	Percentage	Letter	Percentage
A	8.2	N	6.7
B	1.5	O	7.5
C	2.8	P	1.9
D	4.2	Q	0.1
E	12.7	R	6.0
F	2.2	S	6.3
G	2.0	T	9.0
H	6.1	U	2.8
I	7.0	V	1.0
J	0.1	W	2.4
K	0.8	X	0.1
L	4.0	Y	2.0
M	2.4	Z	0.1

FIGURE 1. English letter frequencies

The distribution of English letter frequencies is described in Table 1, or graphically in Fig. 1. As one can see the most common letters are E and T. It often helps to know second order statistics about the underlying language, such as which are the most common sequences of two or three letters, called bigrams and trigrams. The most common bigrams in English are given by Table 2, with the associated approximate percentages. The most common trigrams are, in decreasing order,

THE, ING, AND, HER, ERE, ENT, THA, NTH, WAS, ETH, FOR.

Armed with this information about English we are now able to examine and break a number of historical ciphers.

TABLE 2. English bigram frequencies

Bigram	Percentage	Bigram	Percentage
TH	3.15	HE	2.51
AN	1.72	IN	1.69
ER	1.54	RE	1.48
ES	1.45	ON	1.45
EA	1.31	TI	1.28
AT	1.24	ST	1.21
EN	1.20	ND	1.18

2. Shift Cipher

We first present one of the earliest ciphers, called the shift cipher. Encryption is performed by replacing each letter by the letter a certain number of places on in the alphabet. So for example if the key was three, then the plaintext A would be replaced by the ciphertext D, the letter B would be replaced by E and so on. The plaintext word HELLO would be encrypted as the ciphertext KHOOR. When this cipher is used with the key three, it is often called the Caesar cipher, although in many books the name Caesar cipher is sometimes given to the shift cipher with any key. Strictly this is not correct since we only have evidence that Julius Caesar used the cipher with the key three.

There is a more mathematical explanation of the shift cipher which will be instructive for future discussions. First we need to identify each letter of the alphabet with a number. It is usual to identify the letter A with the number 0, the letter B with number 1, the letter C with the number 2 and so on until we identify the letter Z with the number 25. After we convert our plaintext message into a sequence of numbers, the ciphertext in the shift cipher is obtained by adding to each number the secret key k modulo 26, where the key is a number in the range 0 to 25. In this way we can interpret the shift cipher as a *stream cipher*, with key stream given by the repeating sequence

$$k, k, k, k, k, k, \ldots$$

This key stream is not very random, which results in it being easy to break the shift cipher. A naive way of breaking the shift cipher is to simply try each of the possible keys in turn, until the correct one is found. There are only 26 possible keys so the time for this exhaustive key search is very small, particularly if it is easy to recognize the underlying plaintext when it is decrypted.

We shall show how to break the shift cipher by using the statistics of the underlying language. Whilst this is not strictly necessary for breaking this cipher, later we shall see a cipher that is made up of a number of shift ciphers applied in turn and then the following statistical technique will be useful. Using a statistical technique on the shift cipher is also instructive as to how statistics of the underlying plaintext can arise in the resulting ciphertext.

Take the following example ciphertext, which since it is public knowledge we represent in blue.

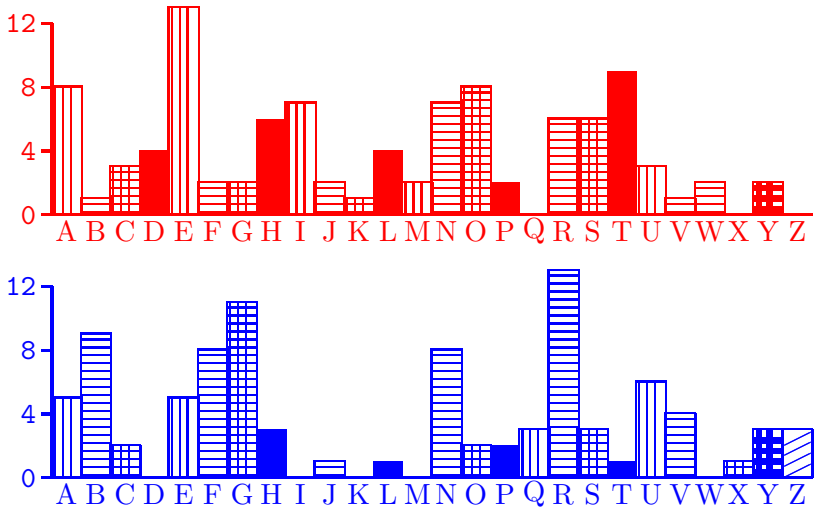
GB OR, BE ABG GB OR: GUNG VF GUR DHRFGVBA:
JURGURE 'GVF ABOYRE VA GUR ZVAQ GB FHSSRE
GUR FYVATF NAQ NEEBJF BS BHGENTRBHF SBEGHAR,
BE GB GNXR NEZF NTNVAFG N FRN BS GEBHOYRF,
NAQ OL BCCBFVAT RAQ GURZ? GB QVR: GB FYRRC;
AB ZBER; NAQ OL N FYRRC GB FNL JR RAQ
GUR URNEG-NPUR NAQ GUR GUBHFNAQ ANGHENY FUBPXF
GUNG SYRFU VF URVE GB, 'GVF N PBAFHZZNGVBA
QRIBHGYL GB OR JVFU'Q. GB QVR, GB FYRRC;
GB FYRRC: CREPUNAPR GB QERNZ: NL, GURER'F GUR EHO;
SBE VA GUNG FYRRC BS QRNGU JUNG QERNZF ZNL PBZR
JURA JR UNIR FUHSSYRQ BSS GUVF ZBEGNY PBVY,
ZHFG TVIR HF CNHFR: GURER'F GUR ERFCRPG
GUNG ZNXRF PNYNZVGL BS FB YBAT YVSR;

One technique of breaking the previous sample ciphertext is to notice that the ciphertext still retains details about the word lengths of the underlying plaintext. For example the ciphertext letter N appears as a single letter word. Since the only single letter words in English are A and I we can conclude that the key is either 13, since N is thirteen letters on from A in the alphabet, or the key is equal to 5, since N is five letters on from I in the alphabet. Hence, the moral here is to always remove word breaks from the underlying plaintext before encrypting using the shift

cipher. But even if we ignore this information about the words we can still break this cipher using frequency analysis.

We compute the frequencies of the letters in the ciphertext and compare them with the frequencies obtained from English which we saw in Fig. 1. We present the two bar graphs one above each other in Fig. 2 so you can see that one graph looks almost like a shift of the other graph. The statistics obtained from the sample ciphertext are given in blue, whilst the statistics obtained from the underlying plaintext language are given in red. Note, we do not compute the red statistics from the actual plaintext since we do not know this yet, we only make use of the knowledge of the underlying language.

FIGURE 2. Comparison of plaintext and ciphertext frequencies for the shift cipher example



By comparing the two bar graphs in Fig. 2 we can see by how much we think the blue graph has been shifted compared with the red graph. By examining where we think the plaintext letter E may have been shifted, one can hazard a guess that it is shifted by one of

$$2, 9, 13 \text{ or } 23.$$

Then by trying to deduce by how much the plaintext letter A has been shifted we can guess that it has been shifted by one of

$$1, 6, 13 \text{ or } 17.$$

The only shift value which is consistent appears to be the value 13, and we conclude that this is the most likely key value. We can now decrypt the ciphertext, using this key. This reveals, that the underlying plaintext is:

To be, or not to be: that is the question:
Whether 'tis nobler in the mind to suffer
The slings and arrows of outrageous fortune,
Or to take arms against a sea of troubles,
And by opposing end them? To die: to sleep;
No more; and by a sleep to say we end
The heart-ache and the thousand natural shocks
That flesh is heir to, 'tis a consummation
Devoutly to be wish'd. To die, to sleep;
To sleep: perchance to dream: ay, there's the rub;
For in that sleep of death what dreams may come

When we have shuffled off this mortal coil,
Must give us pause: there's the respect
That makes calamity of so long life;

The above text is obviously taken from *Hamlet* by William Shakespeare.

3. Substitution Cipher

The main problem with the shift cipher is that the number of keys is too small, we only have 26 possible keys. To increase the number of keys a *substitution cipher* was invented. To write down a key for the substitution cipher we first write down the alphabet, and then a permutation of the alphabet directly below it. This mapping gives the substitution we make between the plaintext and the ciphertext

Plaintext alphabet	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Ciphertext alphabet	GOYDSIPELUAVCRJWXZNHBQFTMK

Encryption involves replacing each letter in the top row by its value in the bottom row. Decryption involves first looking for the letter in the bottom row and then seeing which letter in the top row maps to it. Hence, the plaintext word HELLO would encrypt to the ciphertext ESVVJ if we used the substitution given above.

The number of possible keys is equal to the total number of permutations on 26 letters, namely the size of the group S_{26}, which is

$$26! \approx 4.03 \cdot 10^{26} \approx 2^{88}.$$

Since, as a rule of thumb, it is feasible to only run a computer on a problem which takes under 2^{80} steps we can deduce that this large key space is far too large to enable a brute force search even using a modern computer. Still we can break substitution ciphers using statistics of the underlying plaintext language, just as we did for the shift cipher.

Whilst the shift cipher can be considered as a stream cipher since the ciphertext is obtained from the plaintext by combining it with a keystream, the substitution cipher operates much more like a modern block cipher, with a block length of one English letter. A ciphertext block is obtained from a plaintext block by applying some (admittedly simple) key dependent algorithm.

Substitution ciphers are the types of ciphers commonly encountered in puzzle books, they have an interesting history and have occurred in literature. See for example the Sherlock Holmes story *The Adventure of the Dancing Men* by Arthur Conan–Doyle. The plot of this story rests on a substitution cipher where the ciphertext characters are taken from an alphabet of 'stick men' in various positions. The method of breaking the cipher as described by Holmes to Watson in this story is precisely the method we shall adopt below.

We give a detailed example, which we make slightly easier by keeping in the ciphertext details about the underlying word spacing used in the plaintext. This is only for ease of exposition, the techniques we describe can still be used if we ignore these word spacings, although more care and thought is required.

Consider the ciphertext

XSO MJIWXVL JODIVA STW VAO VY OZJVCO'W LTJDOWX KVAKOAXJTXIVAW VY SIDS XOKSAVLVDQ IAGZWXJQ. KVUCZXOJW, KVUUZAIKTXIVAW TAG UIKJVOLOKXJVAIKW TJO HOLL JOCJOWOAXOG, TLVADWIGO GIDIXTL UOGIT, KVUCZXOJ DTUOW TAG OLOKXJVAIK KVUUOJKO. TW HOLL TW SVWXIAD UTAQ JOWOTJKS TAG CJVGZKX GONOLVCUOAX KOAXJOW VY UTPVJ DLVMTL KVUCTAIOW, XSO JODIVA STW T JTCIGLQ DJVHIAD AZUMOJ VY IAAVNTXINO AOH KVUCTAIOW. XSO KVUCZXOJ WKIOAKO GOCTJXUOAX STW KLVWO JOLTXIVAWSICW HIXS UTAQ VY XSOWO VJDTAIWTXIVAW NIT KVLLTMVJTXINO CJVPOKXW, WXTYY WOKVAGUOAXW TAG NIWIXIAD IAGZWXJITL WXTYY. IX STW JOKOAXLQ IAXJVGZKOG

WONOJTL UOKSTAIWUW YVJ GONOLVCIAD TAG WZCCVJXIAD OAXJOCJOAOZJITL WXZGOAXW TAG WXTYY, TAG TIUW XV CLTQ T WIDAIYIKTAX JVLO IA XSO GONOLVCUOAX VY SIDS-XOKSAVLVDQ IAGZWXJQ IA XSO JODIVA.

XSO GOCTJXUOAX STW T LTJDO CJVDJTUUO VY JOWOTJKS WZCCVJXOG MQ IAGZWXJQ, XSO OZJVCOTA ZAIVA, TAG ZE DVNOJAUOAX JOWOTJKS OWXTMLIW-SUOAXW TAG CZMLIK KVJCVJTXIVAW. T EOQ OLOUOAX VY XSIW IW XSO WXJ-VAD LIAEW XSTX XSO GOCTJXUOAX STW HIXS XSO KVUCZXOJ, KVUUZAIKTXIVAW, UIKJVOLOKXJVAIKW TAG UOGIT IAGZWXJIOW IA XSO MJIWXVL JODIVA . XSO TKT-GOUIK JOWOTJKS CJVDJTUUO IW VJDTAIWOG IAXV WONOA DJVZCW, LTADZTDOW TAG TJKSIXOKXZJO, GIDIXTL UOGIT, UVMILO TAG HOTJTMLO KVUCZXIAD, UTK-SIAO LOTJAIAD, RZTAXZU KVUCZXIAD, WQWXOU NOJIYIKTXIVA, TAG KJQCXVD-JTCSQ TAG IAYVJUTXIVA WOKZJIXQ.

We can compute the following frequencies for single letters in the above ciphertext:

Letter	Freq	Letter	Freq	Letter	Freq
A	8.6995	B	0.0000	C	3.0493
D	3.1390	E	0.2690	F	0.0000
G	3.6771	H	0.6278	I	7.8923
J	7.0852	K	4.6636	L	3.5874
M	0.8968	N	1.0762	O	11.479
P	0.1793	Q	1.3452	R	0.0896
S	3.5874	T	8.0717	U	4.1255
V	7.2645	W	6.6367	X	8.0717
Y	1.6143	Z	2.7802		

In addition we determine that the most common bigrams in this piece of ciphertext are

TA, AX, IA, VA, WX, XS, AG, OA, JO, JV,

whilst the most common trigrams are

OAX, TAG, IVA, XSO, KVU, TXI, UOA, AXS.

Since the ciphertext letter O occurs with the greatest frequency, namely 11.479, we can guess that the ciphertext letter O corresponds to the plaintext letter E. We now look at what this means for two of the common trigrams found in the ciphertext

- The ciphertext trigram OAX corresponds to E * *.
- The ciphertext trigram XSO corresponds to * * E.

We examine similar common similar trigrams in English, which start or end with the letter E. We find that three common ones are given by ENT, ETH and THE. Since the two trigrams we wish to match have one starting with the same letter as the other finishes with, we can conclude that it is highly likely that we have the correspondence

- X = T,
- S = H,
- A = N.

Even after this small piece of analysis we find that it is much easier to understand what the underlying plaintext should be. If we focus on the first two sentences of the ciphertext we are trying to break, and we change the letters which we think we have found the correct mappings for, then we obtain:

THE MJIWTVL JEDIVN HTW VNE VY EZJVCE'W LTJDEWT
KVNKENTJTTIV NW VY HIDH TEKHNVLVDQ INGZWTJQ.
KVUCZTEJW, KVUUZNIKTTIVNW TNG UIKJVELEKTJVNIKW

TJE HELL JECJEWENTEG, TLVNDWIGE GIDITTL UEGIT,
KVUCZTEJ DTUEW TNG ELEKTJVNIK KVUUEJKE.

Recall, this was after the four substitutions

$$O = E,\ X = T,\ S = H,\ A = N.$$

We now cheat and use the fact that we have retained the word sizes in the ciphertext. We see that since the letter T occurs as a single ciphertext letter we must have

$$T = I \text{ or } T = A.$$

The ciphertext letter T occurs with a probability of 8.0717, which is the highest probability left, hence we are far more likely to have

$$T = A.$$

We have already considered the most popular trigram in the ciphertext so turning our attention to the next most popular trigram we see that it is equal to TAG which we suspect corresponds to the plaintext AN*. Therefore it is highly likely that G = D, since AND is a popular trigram in English.

Our partially decrypted ciphertext is now equal to

THE MJIWTVL JEDIVN HAW VNE VY EZJVCE'W LAJDEWT
KVNKENTJATIV NW VY HIDH TEKHNVLVDQ INDZWTJQ.
KVUCZTEJW, KVUUZNIKATIVNW AND UIKJVELEKTJVNIKW
AJE HELL JECJEWENTED, ALVNDWIDE DIDITAL UEDIA,
KVUCZTEJ DAUEW AND ELEKTJVNIK KVUUEJKE.

This was after the six substitutions

$$O = E,\ X = T,\ S = H,$$
$$A = N,\ T = A,\ G = D.$$

We now look at two-letter words which occur in the ciphertext:

- IX
  This corresponds to the plaintext *T. Therefore the ciphertext letter I must be one of the plaintext letters A or I, since the only two-letter words in English ending in T are AT and IT. We already have worked out what the plaintext character A corresponds to, hence we must have I = I.
- XV
  This corresponds to the plaintext T*. Hence, we must have V = O.
- VY
  This corresponds to the plaintext O*. Hence, the ciphertext letter Y must correspond to one of F, N or R. We already know the ciphertext letter corresponding to N. In the ciphertext the probability of Y occurring is 1.6, but in English we expect F to occur with probability 2.2 and R to occur with probability 6.0. Hence, it is more likely that Y = F.
- IW
  This corresponds to the plaintext I*. Therefore, the plaintext character W must be one of F, N, S and T. We already have F, N, T, hence W = S.

All these deductions leave the partial ciphertext as

THE MJISTOL JEDION HAS ONE OF EZJOCE'S LAJDEST
KONKENTJATIONS OF HIDH TEKHNOLODQ INDZSTJQ.
KOUCZTEJS, KOUUZNIKATIONS AND UIKJOELEKTJONIKS AJE
HELL JECJESENTED, ALONDSIDE DIDITAL UEDIA,
KOUCZTEJ DAUES AND ELEKTJONIK KOUUEJKE.

This was after the ten substitutions

$$O = E,\ X = T,\ S = H,\ A = N,\ T = A,$$
$$G = D,\ I = I,\ V = O,\ Y = F,\ W = S.$$

Even with half the ciphertext letters determined it is now quite easy to understand the underlying plaintext, taken from the website of the University of Bristol Computer Science Department. We leave it to the reader to determine the final substitutions and recover the plaintext completely.

4. Vigenère Cipher

The problem with the shift cipher and the substitution cipher was that each plaintext letter always encrypted to the same ciphertext letter. Hence underlying statistics of the language could be used to break the cipher. For example it was easy to determine which ciphertext letter corresponded to the plaintext letter E. From the early 1800s onwards, cipher designers tried to break this link between the plaintext and ciphertext.

The substitution cipher we used above was a mono-alphabetic substitution cipher, in that only one alphabet substitution was used to encrypt the whole alphabet. One way to solve our problem is to take a number of substitution alphabets and then encrypt each letter with a different alphabet. Such a system is called a polyalphabetic substitution cipher.

For example we could take

Plaintext alphabet	ABCDEFGHIJKLMNOPQRSTUVWXYZ
Ciphertext alphabet one	TMKGOYDSIPELUAVCRJWXZNHBQF
Ciphertext alphabet two	DCBAHGFEMLKJIZYXWVUTSRQPON

Then the plaintext letters in an odd position we encrypt using the first ciphertext alphabet, whilst the plaintext letters in even positions we encrypt using the second alphabet. For example the plaintext word HELLO, using the above alphabets would encrypt to SHLJV. Notice that the two occurrences of L in the plaintext encrypt to two different ciphertext characters. Thus we have made it harder to use the underlying statistics of the language. If one now does a naive frequency analysis we no longer get a common ciphertext letter corresponding to the plaintext letter E.

We essentially are encrypting the message two letters at a time, hence we have a block cipher with block length two English characters. In real life one may wish to use around five rather than just two alphabets and the resulting key becomes very large indeed. With five alphabets the total key space is

$$(26!)^5 \approx 2^{441},$$

but the user only needs to remember the key which is a sequence of

$$26 \cdot 5 = 130$$

letters. However, just to make life hard for the attacker, the number of alphabets in use should also be hidden from his view and form part of the key. But for the average user in the early 1800s this was far too unwieldy a system, since the key was too hard to remember.

Despite its shortcomings the most famous cipher during the 19th-century was based on precisely this principle. The *Vigenère cipher*, invented in 1533 by Giovan Batista Belaso, was a variant on the above theme, but the key was easy to remember. When looked at in one way the Vigenère cipher is a polyalphabetic block cipher, but when looked at in another, it is a stream cipher which is a natural generalization of the shift cipher.

The description of the Vigenère cipher as a block cipher takes the description of the polyalphabetic cipher above but restricts the possible plaintext alphabets to one of the 26 possible cyclic shifts of the standard alphabet. Suppose five alphabets were used, this reduces the key space down to

$$26^5 \approx 2^{23}$$

and the size of the key to be remembered as a sequence of five numbers between 0 and 25.

However, the description of the Vigenère cipher as a stream cipher is much more natural. Just like the shift cipher, the Vigenère cipher again identifies letters with the numbers $0, \ldots, 25$. The secret key is a short sequence of letters (e.g. a word) which is repeated again and again to form

a keystream. Encryption involves adding the plaintext letter to a key letter. Thus if the key is SESAME, encryption works as follows,

```
THISISATESTMESSAGE
SESAMESESAMESESAME
LLASUWSXWSFQWWKASI
```

Again we notice that A will encrypt to a different letter depending on where it appears in the message.

But the Vigenère cipher is still easy to break using the underlying statistics of English. Once we have found the length of the keyword, breaking the ciphertext is the same as breaking the shift cipher a number of times.

As an example, suppose the ciphertext is given by

UTPDHUG NYH USVKCG MVCE FXL KQIB. WX RKU GI TZN, RLS BBHZLXMSNP KDKS; CEB IH HKEW IBA, YYM SBR PFR SBS, JV UPL O UVADGR HRRWXF. JV ZTVOOV YH ZCQU Y UKWGEB, PL UQFB P FOUKCG, TBF RQ VHCF R KPG, OU KFT ZCQU MAW QKKW ZGSY, FP PGM QKFTK UQFB DER EZRN, MCYE, MG UCTFSVA, WP KFT ZCQU MAW KQIJS. LCOV NTHDNV JPNUJVB IH GGV RWX ONKCGTHKFL XG VKD, ZJM VG CCI MVGD JPNUJ, RLS EWVKJT ASGUCS MVGD; DDK VG NYH PWUV CCHIIY RD DBQN RWTH PFRWBBI VTTK VCGNTGSF FL IAWU XJDUS, HFP VHCF, RR LAWEY QDFS RVMEES FZB CHH JRTT MVGZP UBZN FD ATIIYRTK WP KFT HIVJCI; TBF BLDPWPX RWTH ULAW TG VYCHX KQLJS US DCGCW OPPUPR, VG KFDNUJK GI JIKKC PL KGCJ IAOV KFTR GJFSAW KTZLZES WG RWXWT VWTL WP XPXGG, CJ FPOS VYC BTZCUW XG ZGJQ PMHTRAIBJG WMGFG. JZQ DPB JVYGM ZCLEWXR: CEB IAOV NYH JIKKC TGCWXF UHF JZK.

WX VCU LD YITKFTK WPKCGVCWIQT PWVY QEBFKKQ, QNH NZTTW IRFL IAS VFRPE ODJRXGSPTC EKWPTGEES, GMCG
TTVVPLTFFJ; YCW WV NYH TZYRWH LOKU MU AWO, KFPM VG BLTP VQN RD DSGG AWKWUKKPL KGCJ, XY OPP KPG ONZTT ICUJCHLSF KFT DBQNJTWUG. DYN MVCK ZT MFWCW HTWF FD JL, OPU YAE CH LQ! PGR UF, YH MWPP RXF CDJCGOSF, XMS UZGJQ JL, SXVPN HBG!

There is a way of finding the length of the keyword, which is repeated to form the keystream, called the *Kasiski test*. First we need to look for repeated sequences of characters. Recall that English has a large repetition of certain bigrams or trigrams and over a long enough string of text these are likely to match up to the same two or three letters in the key every so often. By examining the distance between two repeated sequences we can guess the length of the keyword. Each of these distances should be a multiple of the keyword, hence taking the greatest common divisor of all distances between the repeated sequences should give a good guess as to the keyword length.

Let us examine the above ciphertext and look for the bigram WX. The gaps between some of the occurrences of this bigram are 9, 21, 66 and 30, some of which may have occurred by chance, whilst some may reveal information about the length of the keyword. We now take the relevant greatest common divisors to find,

$$\begin{aligned}\gcd(30,66) &= 6,\\ \gcd(3,9) = \gcd(9,66) &= \gcd(9,30) = \gcd(21,66) = 3.\end{aligned}$$

We are unlikely to have a keyword of length three so we conclude that the gaps of 9 and 21 occurred purely by chance. Hence, our best guess for the keyword is that it is of length 6.

Now we take every sixth letter and look at the statistics just as we did for a shift cipher to deduce the first letter of the keyword. We can now see the advantage of using the histograms to break the shift cipher earlier. If we used the naive method and tried each of the 26 keys in turn we

could still not detect which key is correct, since every sixth letter of an English sentence does not produce an English sentence. Using our earlier histogram based method is more efficient in this case.

FIGURE 3. Comparison of plaintext and ciphertext frequencies for every sixth letter of the Vigenère example, starting with the first letter

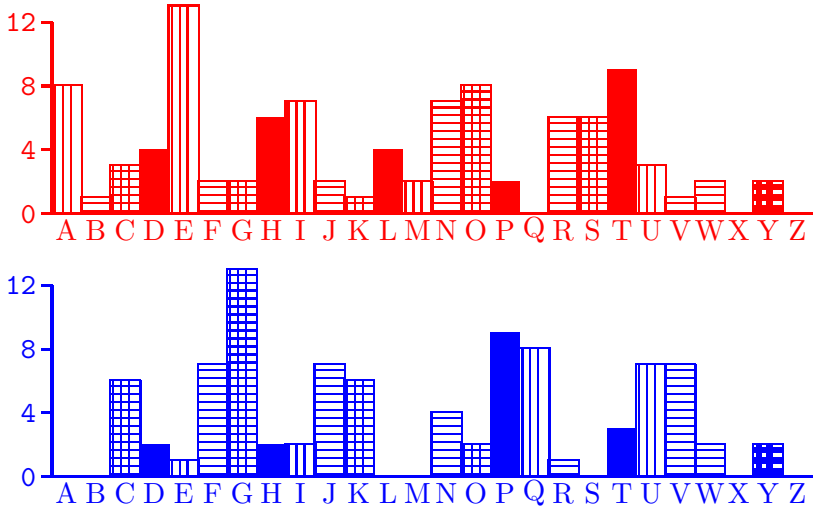
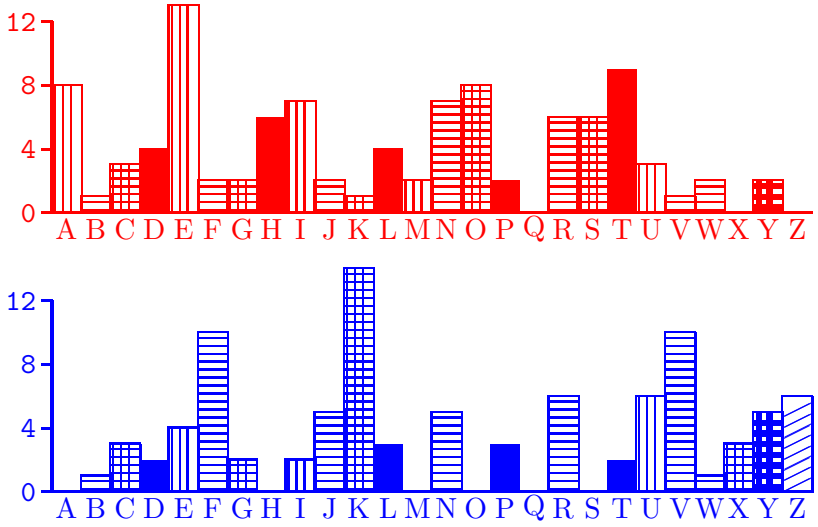


FIGURE 4. Comparison of plaintext and ciphertext frequencies for every sixth letter of the Vigenère example, starting with the second letter



The relevant bar charts for every sixth letter starting with the first are given in Fig. 3. We look for the possible locations of the three peaks corresponding to the plaintext letters A, E and T. We see that this sequence seems to be shifted by two positions in the blue graph compared with the red graph. Hence we can conclude that the first letter of the keyword is C, since C corresponds to a shift of two.

We perform a similar step for every sixth letter, starting with the second one. The resulting bar graphs are given in Fig. 4. Using the same technique we find that the blue graph appears to

have been shifted along by 17 spaces, which corresponds to the second letter of the keyword being equal to R.

Continuing in a similar way for the remaining four letters of the keyword we find the keyword is

CRYPTO.

The underlying plaintext is then found to be:

Scrooge was better than his word. He did it all, and infinitely more; and to Tiny Tim, who did not die, he was a second father. He became as good a friend, as good a master, and as good a man, as the good old city knew, or any other good old city, town, or borough, in the good old world. Some people laughed to see the alteration in him, but he let them laugh, and little heeded them; for he was wise enough to know that nothing ever happened on this globe, for good, at which some people did not have their fill of laughter in the outset; and knowing that such as these would be blind anyway, he thought it quite as well that they should wrinkle up their eyes in grins, as have the malady in less attractive forms. His own heart laughed: and that was quite enough for him.

He had no further intercourse with Spirits, but lived upon the Total Abstinence Principle, ever afterwards; and it was always said of him, that he knew how to keep Christmas well, if any man alive possessed the knowledge. May that be truly said of us, and all of us! And so, as Tiny Tim observed, God bless Us, Every One!

The above text is taken from *A Christmas Carol* by Charles Dickens.

5. A Permutation Cipher

The ideas behind substitution type ciphers forms part of the design of modern symmetric systems. For example later we shall see that both DES and Rijndael make use of a component called an S-Box, which is simply a substitution. The other component that is used in modern symmetric ciphers is based on permutations.

Permutation ciphers have been around for a number of centuries. Here we shall describe the simplest, which is particularly easy to break. We first fix a permutation group S_n and a permutation

$$\sigma \in S_n.$$

It is the value of σ which will be the secret key. As an example suppose we take

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 4 & 1 & 3 & 5 \end{pmatrix} = (1243) \in S_5.$$

Now take some plaintext, say

Once upon a time there was a little girl called snow white.

We break the text into chunks of 5 letters

onceu ponat imeth erewa salit tlegi rlcal ledsn owwhi te.

We first pad the message, with some random letters, so that we have a multiple of five letters in each chunk.

onceu ponat imeth erewa salit tlegi rlcal ledsn owwhi teahb.

Then we take each five-letter chunk in turn and swap the letters around according to our secret permutation σ. With our example we obtain

coenu npaot eitmh eewra lsiat etgli crall dlsdn wohwi atheb.

We then remove the spaces, so as to hide the value of n, producing the ciphertext

coenunpaoteitmheewralsiatetglicralldlsdnwohwiatheb.

However, breaking a permutation cipher is easy with a chosen plaintext attack, assuming the group of permutations used (i.e. the value of n) is reasonably small. To attack this cipher we mount a chosen plaintext attack, and ask one of the parties to encrypt the message

abcdefghijklmnopqrstuvwxyz,

to obtain the ciphertext

cadbehfigjmknlorpsqtwuxvyz.

We can then deduce that the permutation looks something like

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & \dots \\ 2 & 4 & 1 & 3 & 5 & 7 & 9 & 6 & 8 & 10 & 12 & 14 & 11 & 13 & 15 & \dots \end{pmatrix}.$$

We see that the sequence repeats (modulo 5) after every five steps and so the value of n is probably equal to five. We can recover the key by simply taking the first five columns of the above permutation.

Chapter Summary

- Many early ciphers can be broken because they do not successfully hide the underlying statistics of the language.
- Important principles behind early ciphers are those of substitution and permutation.
- Ciphers can either work on blocks of characters via some keyed algorithm or simply consist of adding some keystream to each plaintext character.
- Ciphers which aimed to get around these early problems often turned out to be weaker than expected, either due to some design flaw or due to bad key management practices adopted by operators.

Further Reading

The best book on the history of ciphers is that by Kahn. Kahn's book is a weighty tome so those wishing a more rapid introduction should consult the book by Singh. The book by Churchhouse also gives an overview of a number of historical ciphers.

R. Churchhouse. *Codes and Ciphers. Julius Caesar, the Enigma and the Internet.* Cambridge University Press, 2001.

D. Kahn. *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet.* Scribner, 1996.

S. Singh. *The Codebook: The Evolution of Secrecy from Mary, Queen of Scots to Quantum Cryptography.* Doubleday, 2000.