Cryptography

BMEVITMAV52

Information and Network Security

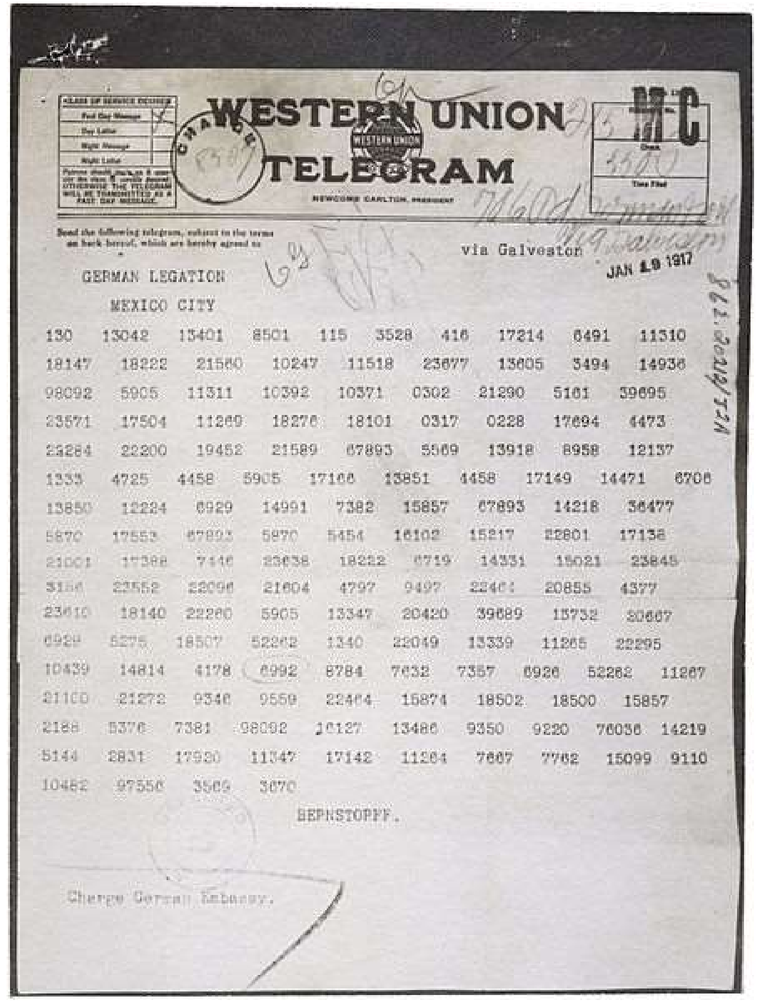
feher.gabor@tmit.bme.hu

Cryptography

- Cryptography
  - Greek word: secret writing
    - The encrypted message is visible, but the meaning is unknown

- Basic notations
  - Plaintext (P)
  - Ciphertext (C)
  - Key (K)

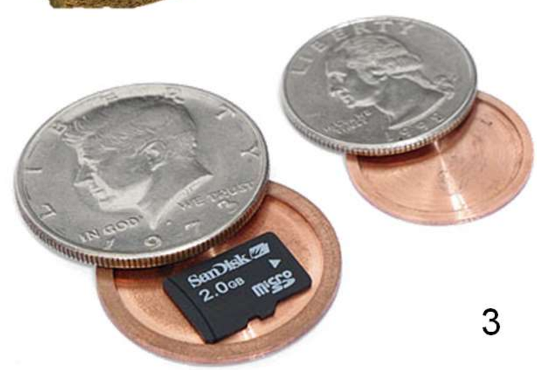
  - Encryption: C = E(P, K)
  - Decryption: P = D(C, K)



Steganography

- Steganography
  - Greek word: covert writing
    - The “encrypted” message is invisible

- First steganography
  - Based on Herodotus notes (~400 BC)
    - Demeratus: Message under the clay tables, information about military movements
    - Histiaeus: Message under the hair of a servant
  - Pliny the Elder (roman empire)
    - Invisible ink (milk of the thithymallus plant), message between the rows



Cryptanalysis

- To get the key
  - Known plaintext
    - Some parts of the text (P_1) and its secret form (C_1) is known by the attacker
    - E.g.: ZIP archives with known files
  - Ciphertext only attack
    - Only the secret form of the message is known.
    - Usually this is the case
  - Brute force – dictionary – smart force attacks
    - Testing the keys
    - Rainbow tables
  - Side channel attack
    - Attack on the implementation (not on theory)
    - E.g. RSA attacks
- To get the message without the key
- Modify message without the key

Security, obscurity, design

- Security by obscurity
  - The encryption method is not publicly known. It is a secret of the inventors
  - May contain design errors
  - May result severe errors when the method is discovered by others

- Security by design
  - The encryption method is well known by the public. (OPEN) The key s the only secret
  - The method is investigated by many cryptanalyst

- Kerckhoffs' principle and Shannon's maxim
  - The enemy knows the system (but not the key)

History of cryptography
Classic encryption

- ~ 2500 BC
  - Mystic hieroglyphs in Egypt. Now really a secret, but importance
- ~ 600 BC
  - Simple, *monoalphabetic substitution ciphers*
    - E.g.: Atbash cipher: Hebrew cipher inverting the ABC
      - ABCDEFGHIJKLMNOPQRSTUVWXYZ
      - ZYXWVUTSRQPONMLKJIHGFEDCBA
- ~400 BC
  - Greek: Born of *Steganography*
    - Herodotus writing about bald servants, clay table
  - Using *transposition (around 700 BC ?)*
    - Ciphering in Sparta militia: Scytale
      - Writing to a paper that is wrapped around a stick. The diameter of the stick is important

Classic encryption (cont.)

- ~ 200 BC
  - Ploybius (greek) table. Writing the ABC into a 5x5 –table. Indicate the position of the letters
  - Signaling over a public channel

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

- ~ 50 BC
  - Roman: *Ceasar cipher. Shift 3 cipher*
    - ABCDEFGHIJKLMNOPQRSTUVWXYZ
    - DEFGHIJKLMNOPQRSTUVWXYZABC
- ~ 400 AD
  - Indian: Secret communication (Kama sutra)
    - Mostly steganography

History of cryptography
Medieval encryption

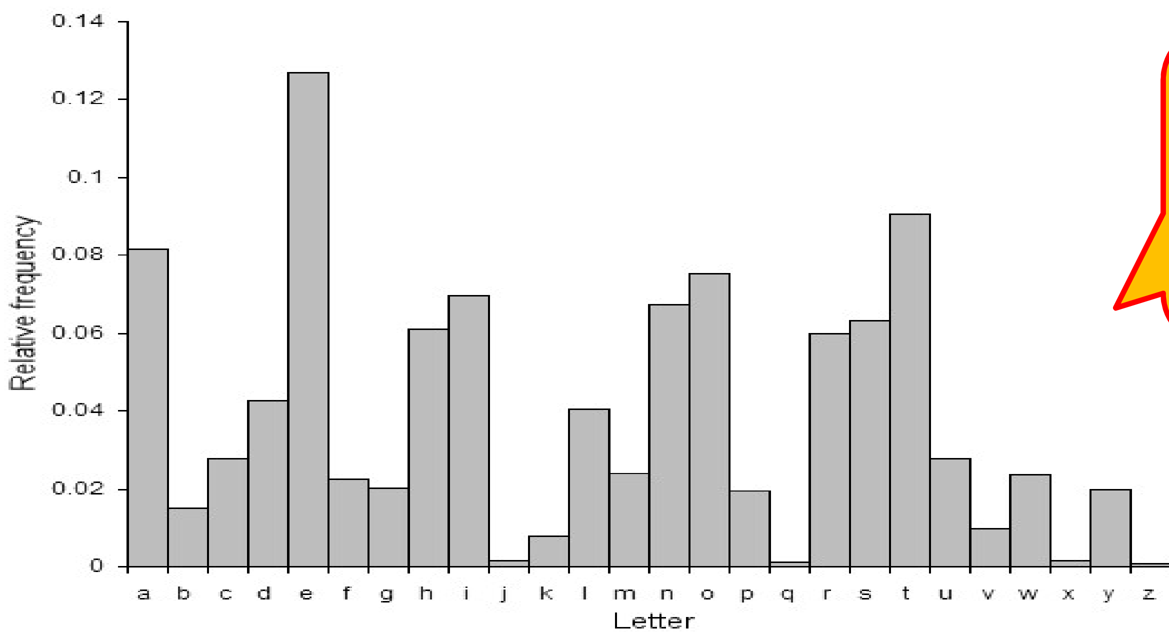
- ~ 800
  - Al-Kindi (Iraq): Breaking the monoalphabetic ciphers using frequency analysis. At the same time reference to the *polyalphabetic ciphers*.

- ~ 1350
  - Taj ad-Din Ali ibn ad-Duraihim ben Muhammad ath-Tha 'alibi al-Mausili (Egypt): ciphering using multiple substitutions (only reference, the literature is lost)

- 1466
  - Leon Battista Alberti (Italian artist and scientist): the “inventor” of the polyalphabetic encryption
    - Father of Western Cryptology

Frequency Analysis

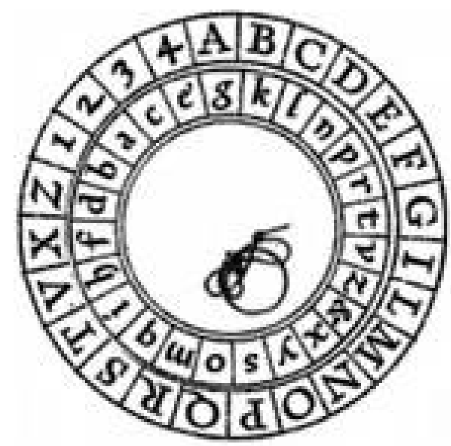
- The frequency of the letters depends on the language
  - The substitution key can be discovered



The letters in the English alphabet, and their frequency

Polyalphabetic ciphering

- Using different monoalphabetic cipher for each letter
- 1466 Alberti: Alberti ciphering disc
  - Two discs: Stable outer disk (plaintext) and a moving inner one (ciphertext). On the outer discs there are numbers also. It may refer to codebooks
  - **Substitution can be changed at any time!**
  - Encryption method 1:
    - The alphabet is denoted by capital letter in the ciphertext. Changing the alphabet at will.
  - Encryption method 2:
    - There is a change in the alphabet at decoding numbers. The first letter is the starting word of the new alphabet



Polyalphabetic ciphering (cont.)

- Johannes Trithemius (German)
  - 1499: Steganographia
    - Book about steganography
  - 1518: Polygraphia:
    - the first printed book about cryptography
  - Tabula recta
    - Using substitutions

Letter to be encrypt

Key

	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

Tabula Recta

- Key: F
- Plain text: G

- Cipher text: L

	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

Polyalphabetic ciphering (cont.)

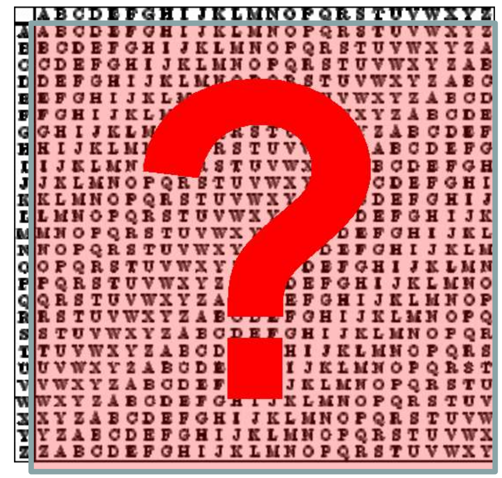
- 1553 Giovan Battista Bellaso (Italian)
  - Using the „tabula recta” and a key to get the right substitution
    - Plaintext: `ATTACKATDAWN`
    - Key: `LEMONLEMONLE`
    - Ciphertext: `LXFOPVEFRNHR`

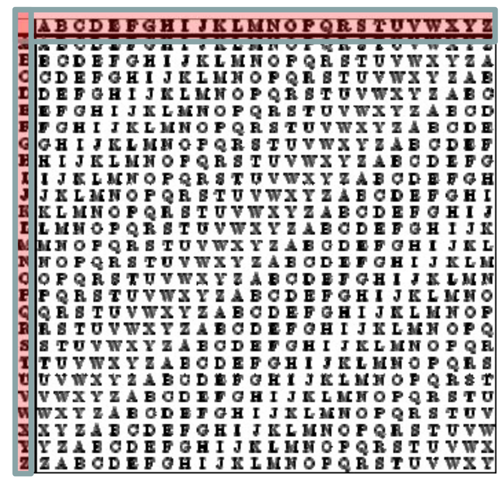
Called Vigenere cipher

- 1586 Blaise de Vigenère (French)
  - „Autokey” cipher
    - After a short secret work (key) the plaintext is used to be the key
    - Plaintext: `ATTACKATDAWN…`
    - Key: `QUEENLYATTACKATDAWN....`
    - Ciphertext: `QNXEPVYTWTWP...`

Mixed cipher alphabets

- Creating an alternative ‘tabula recta’:
  - Each letter appears only once in columns and rows
    - We have all the letters
- Could be a permutation of rows in tablua recta, but could be random also
- It is easier to create the table based on a keyword, than distribute it with the message





Mixed cipher alphabets

- A keyword can mix the substitution alphabet
  - A trivial form is to have the keyword first and then the remaining letters behind
    - E.g.: CRYPTOGRAPHIC
      CRYPTOGAHIBDEFJKLMNQSUVWXZ
  - However besides the keyword, there could be an other “keyword” to modify the remaining alphabet
    - Examples: Matrices, matrices using numbers, paths in matrices, n-th letters based on the key, …

Examples for mixed alphabets

Keyword— ARTILLERY

Keyword mixed sequence in matrix:

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

The keyword gives the size of the matrix. The missing letters follow the keyword. Read out the ABC in columns

Resulting sequence:

ABKURCMVTDNWIFOXLGPZEHQYJS

Keyword- CALIFORNIA

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

Similar to the previous, but the reading columns are numbered

Resulting sequence:

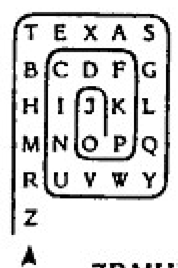
ADQZCBPYFHUIGTLESNMXOJVRKW

Keyword— TEXAS

In by rows:

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

Out spirally:



ZRMHBTEXASGLQYWVUNICDFKPOJ

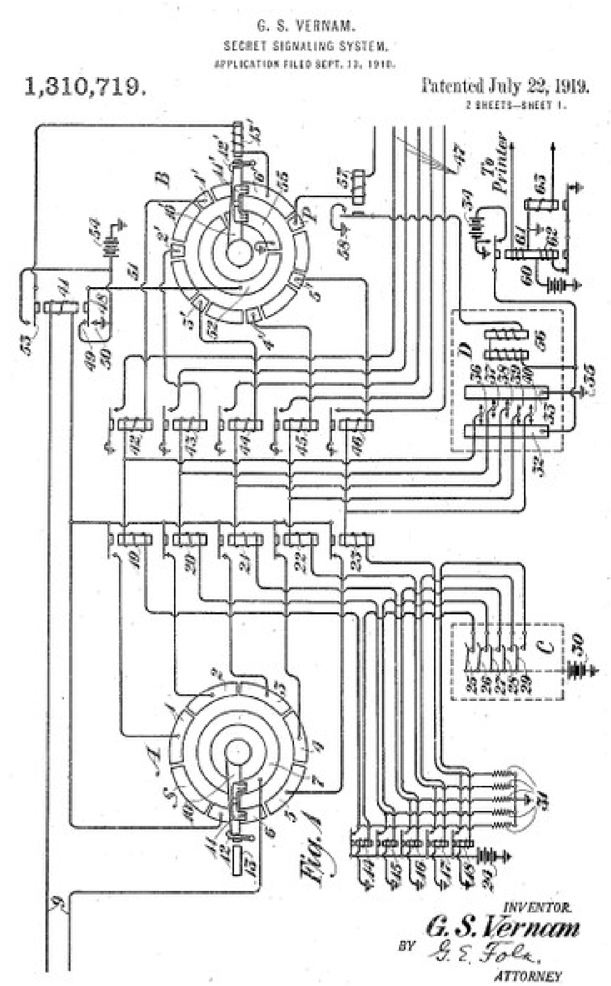
Breaking the polyalphabetic encryption

- The number of substitution ABCs could be limited
  - Tabula recta: the number of letters
- Frequency analysis is difficult however
  - When the key used multiple times: First, we should know the length of the keyword. Second, we can use frequency analysis
  - Babbage's (English) method (1854)
    - Shifting the ciphertext
  - Kasiski (Prussian) method(1863)
    - Bigram detection

To find out the key length

History of cryptography
WW I.

- One time pad (OTP)
  - 1917 Gilbert Stanford Vernam (AT&T) invention
    - Encryption for teletypewriters (TTY). Keypaper and the message
    - Using relays
    - plain: A = "++−−−“
    - key: B = "+−−++“
    - secret: G = "−+−++”
  - 1920 Joseph Oswald Mauborgne (American)
    - Key should be random!
  - Using pads for storing the key

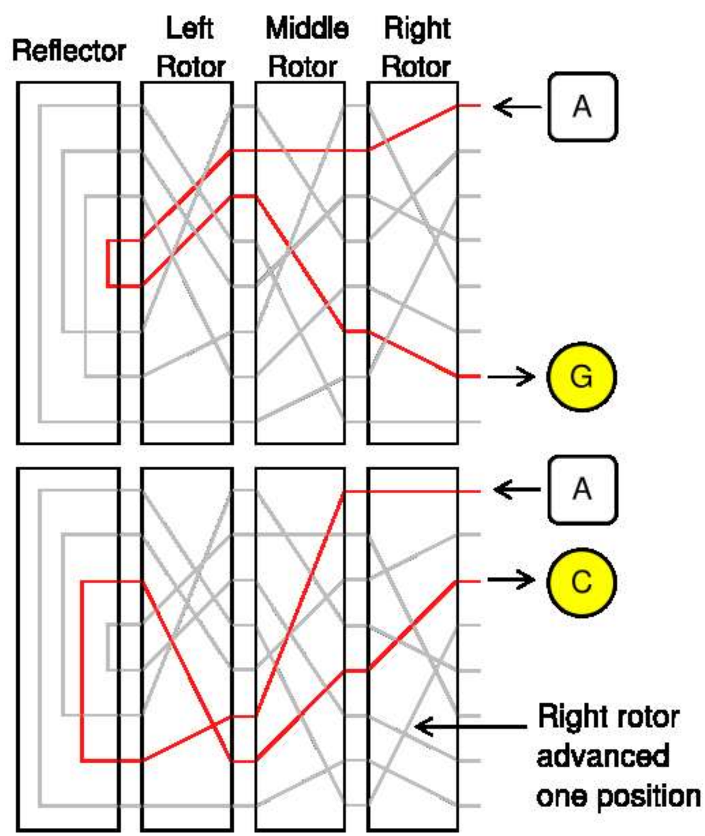


History of cryptography
WW II.

- Mechanical and electromechanical ciphering machines
  - German:
    - Enigma (1920) - rotor machine
  - Japan:
    - Purple – stepping switch
  - English:
    - TypeX – rotor machine
  - USA:
    - SIGABA – rotor machine

Rotor

- Complex polyalphabetic substitution
  - The rotor does the substitution + steps
  - Multiple rotors
    - The key length are multiplied
    - 3 rotor with 26 letters:
      - 26^3 -> 17576 variations
  - A new alphabet for each letter due to stepping



History of cryptography
Modern cryptography

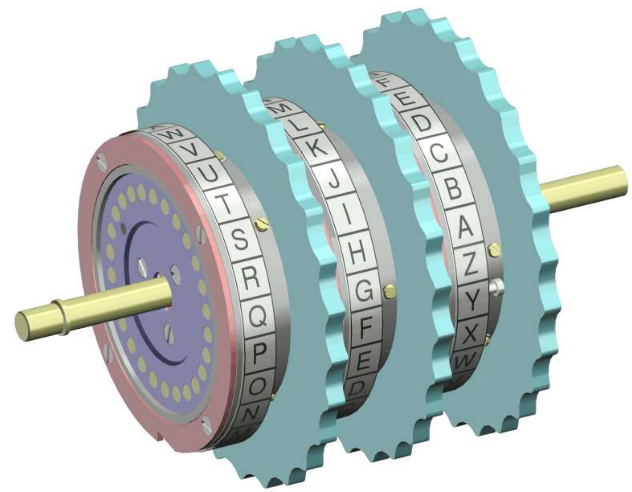
- Claude Shannon: Communication Theory of Secrecy Systems (1949)
    - The basics of cryptography and cryptanalysis
  - No more alphabet, just bits and bytes
- 1975: DES – Data Encryption Standard: Block cipher
  - Horst Feistel
- 1976: Diffie-Hellman key exchange: key management
  - Bailey Whitfield Diffie and Martin Edward Hellman
- 1977: RSA: asymetric block cipher
  - Ron Rivest, Adi Shamir and Leonard Max Adleman
  - 1973 Clifford Cocks (UK) basically the same invention
- 1987: RC4: stream cipher
  - Ron Rivest
- 1991: DSA – Digital Signature Algorithm
  - David W. Kravitz
- 1998: AES – Advanced Encryption Standard: block cipher
  - Joan Daemen and Vincent Rijmen

ENIGMA

- Encryption based on rotors
  - Discs with ABC rings
  - Place for 3-4 discs, but the disc set is greater
- Additional wires for switching (plugboard)
  - Arbitrary letters can be exchanged. Change after or before disc encryption
  - From 1939: 10 change (max 13)

- Key and strength:
  - Selection of the discs (~2^7), Disc positions (~2^9), Initial rotation of the discs and ABCs (~2^{14}), Plugboard connections (~2^{47}) -> **2^{77} key length**
  - Settings were changed day by day

- There were flaws in the design + there were too much confirmation messages
  - ENIGMA was broken



Cryptanalysis

- Breaking the cipher
- Generally working on substitution ciphers
  - Monoalphabetic
  - Polyalphabetic

Practice: Monoalphabetic ciphers

- Try these tools
  - www.counton.org/explorer/codebreaking/
  - cryptoclub.math.uic.edu/menu/tools.htm

1. XAOOR QRBOS

SRJEOTD TSTWD
FXA XMFIX XMLAB'D EJMSA FR FXA ETOTUV

NTB RJF MC FXA JCIXTBFAS HTILQTFABD RN FXA JCNTDXMRCTHOA ACS RN FXA QADFABC DYMBTO TBW RN FXA ETOTUV OMAD T DWTOO JCBAETBSAS VAOORQ DJC. RBHMFMCE FXMD TF T SMDFTCIA RN BRJEXOV CMCAFV-FQR WMOOMRC WMOAD MD TC JFFABOV MCDMECMNMITCF OMFFOA HOJA EBAAC YOTCAF QXRDA TYA-SADIACSAS OMNA NRBWD TBA DR TWTZMCEOV YBMWMFMKA FXTF FXAV DFMOO FXMCL SMEMFTO QTFIXAD TBA T YBAFFV CATF MSAT. FXMD YOTCAF XTD — RB BTFXAB XTS — T YBRHOAW, QXMIX QTD FXMD: WRDF RN FXA YARYOA RC MF QABA JCXTYYV NRB YBAFFV WJIX RN FXA FMWA. WTCV DROJFMRCD QABA DJEEADFAS NRB FXMD YBRHOAW, HJF WRDF RN FXADA QABA OTBEAOV IRCIABCAS QMFX FXA WRKAWACFD RN DWTOO EBAAC YMAIAD RN YTYAB, QXMIX MD RSS HAITJDA RC FXA QXROA MF QTDC'F FXA DWTOO EBAAC YMAIAD RN YTYAB FXTF QABA JCXTYYV. TCS DR FXA YBRHOAW BAWTMCAS; ORFD RN FXA YARYOA QABA WATC, TCS WRDF RN FXAW QABA WMDABTHOA, AKAC FXA RCAD QMFX SMEMFTO QTFIXAD. WTCV QABA MCIBATDMCEOV RN FXA RYMCMRC FXTF FXAV'S TOO WTSA T HME WMDFTLA MC IRWMCE SRQC NBRW FXA FBAAD MC FXA NMBDF YOTIA. TCS DRWA DTMS FXTF AKAC FXA FBAAD XTS HAAC T HTS WRKA, TCS FXTF CR RCA DXRJOS AKAB XTKA OANF FXA RIATCD. TCS FXAC, RCA FXJBDSTV, CATBOV FQR FXRJDTCS VATBD TNFAB RCA WTC XTS HAAC CTMOAS FR T FBAA NRB DTVMCE XRQ EBATF MF QRJOS HA FR HA CMIA FR YARYOA NRB T IXTCEA, RCA EMBO DMFFMCE RC XAB RQC MC T DWTOO ITNA MC BMILWTCDQRBFX DJSSACOV BATOMZAS QXTF MF QTD FXTF XTS HAAC ERMCE QBRCE TOO FXMD FMWA, TCS DXA NMCTOOV LCAQ XRQ FXA QRBOS IRJOS HA WTSA T ERRS TCS XTYYV YOTIA. FXMD FMWA MF QTD BMEXF, MF QRJOS QRBL, TCS CR RCA QRJOS XTKA FR EAF CTMOAS FR TCVFXMCE. DTSOV, XRQAKAB, HANRBA DXA IRJOS EAF FR T YXRCA FR FAOO TCVRCA THRJF MF, T FABBMHOV DFJYMS ITFTDFBRYXA RIIJBBAS, TCS FXA MSAT QTD ORDF NRBAKAB. FXMD MD CRF XAB DFRBV.

Breaking polyalphabetic ciphers

- Polyaplhabetic ciphers change the substitution based on the Key
  - First, the length of the password should be known
    - Inspecting coincidences
    - Kasiski analysis
  - Second, using frequency analysis based on groups using the same part of the Key

Example

- Ciphertext

```
VKMHG QFVMO IJOII OHNSN IZXSS CSZEA WWEXU
LIOZB AGEKQ UHRDH IKHWE OBNSQ RVIES LISYK
BIOVF IEWEO BQXIE UUIXK EKTUH NSZIB SWJIZ
BSKFK YWSXS EIDSQ INTBD RKOZD QELUM AAAEV
MIDMD GKJXR UKTUH TSBGI EQRVF XBAYG UBTCS
XTBDR SLYKW AFHMM TYCKU JHBWV TUHRQ XYHWM
IJBXS LSXUB BAYDI OFLPO XBULU OZAHE JOBDT
ATOUT GLPKO FHNSO KBHMW XKTWX SX
```

(www.murky.org)

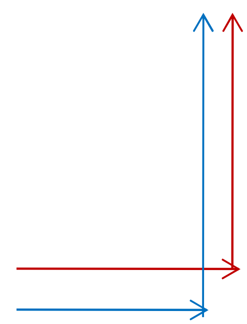
Example (cont.)

- In this example we know the cipher
- Beaufort cipher (Sir Francis Beaufort)
  - Similar to Vigenere cipher, but
    - (1st appr.) We use a slightly modified substitution table (left figure)
    - (2nd appr.) We use the Vigenere table a different way (right figure)
  - The method of the encryption is the same as the method of decryption

Cleartext

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

Ciphertext



Discovering the key length

- Method of coincidences
  - We shift the text by a given amount of characters and count the matches by position. There are matches. In the case of the most frequent characters there are more matches if we use the same substitution, so the shift is done by exactly or the multiple of the key length.

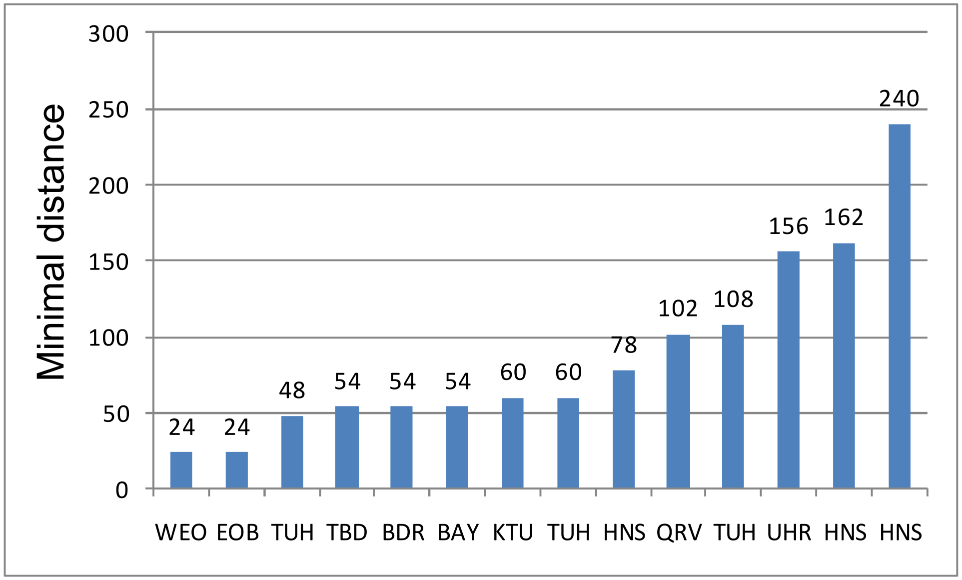
Original:	VK**M**HGQFVM**O**IJOI**I**OHN**S**NI**Z**XSSSCSZEA...	
Shift 1:	KMHGQFVMOIJOIIOHNSNIZXSSCSZEAW...	8
Shift 2:	MHGQFVMOIJOIIOHNSNIZXSSCSZEAWW...	12
Shift 3:	HGQFVMOIJOIIOHNSNIZXSSCSZEAWWE...	11
Shift 4:	GQFVMOIJOIIOHNSNIZXSSCSZEAWWEX...	13
Shift 5:	QFVMOIJOIIOHNSNIZXSSCSZEAWWEXU...	9
Shift 6:	FV**M**OIJOII**O**HNSN**I**ZXS**S**CS**Z**EAWWEXUL...	**25**
Shift 7:	VMOIJOIIOHNSNIZXSSCSZEAWWEXULI...	11

Discovering the key length (cont.)

- Kasiski analysis
  - We are searching for small letter groups (min 3 letters). The distance of two similar group is the number of characters between the first letters.
  - There are common small letter groups in the languages (E.g.: English: *ing*, *the* in the plaintext)
  - The same letter group means that most probably the key part, used to encrypt the plaintext there was the same. Here the distance is exactly or the multiple of the key length. The key length is the greatest common divisor of the lengths.
  - Errors should be ignored.

Discovering the key length (cont.)

VKMHG QFVMO IJOII O**HNS**N
IZXSS CSZEA WWEXU LIOZB
AGEKQ UHRDH IKHWE OBNSQ
RVIES LISYK BIOVF IEWEO
BQXIE UUIXK EKTU**H NS**ZIB
SWJIZ BSKFK YWSXS EIDSQ
INTBD RKOZD QELUM AAAEV
MIDMD GKJXR UKTUH TSBGI
EQRVF XBAYG UBTCS XTBDR
SLYKW AFHMM TYCKU JHBWV
TUHRQ XYHWM IJBXS LSXUB
BAYDI OFLPO XBULU OZAHE
JOBDT ATOUT GLPKO F**HNS**O
KBHMW XKTWX SX



Greatest common divisor: 6

Frequency analysis based on key letters

- Monoaplhabetic substitution based on the actual letter of the key
  - Creating groups that use the same substitution – Their position in the key is the same

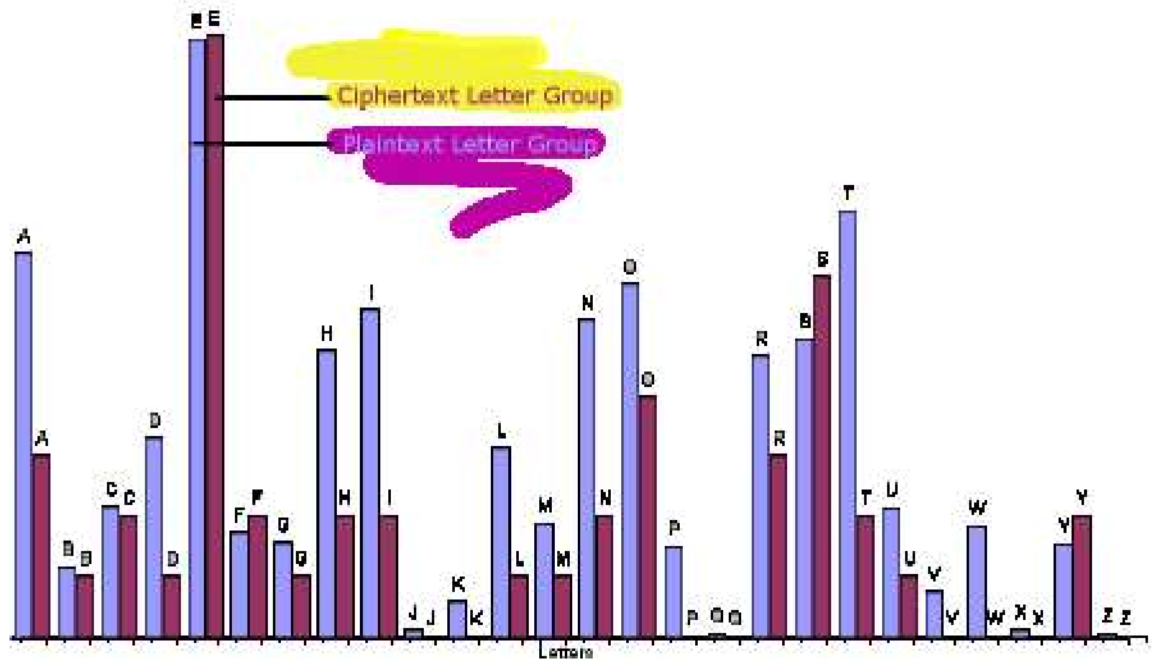
```
0: VFOSSWIEDERIOEEESJFSIKLEDUSRYSSFCWQISYPUJTPSWS
1: KVINCWOKHOVSVOUKZIKENOUVGKBVGXLHKVXJXDOOOOKOXX
2: MMIISEZQIBIYFBUTIZYITZMMKTGFUTYMUTYBUIXZBUOKK
3: HOOZZXBUKNEKIQIUBBWDBDAIJUIXBBKMJUHXBOBADTFBT
4: GIHXEUAHHSSBEXXHSSSSDQADXHEBTDWTHHWSBFUHTGHHW
5: QJNSALGRWQLIWIKNWKXQREAMRTQACRAYBRMLALLEALNMX
```

```
VKMHGQ
FVMOIJ
OIIOHN
SNIZXS
SCSZEA
WWEXUL
IOZBAG
EKQUHR
```

Frequency analysis (cont.)

- Investigation of the first letter of the key
  - Using frequency analysis

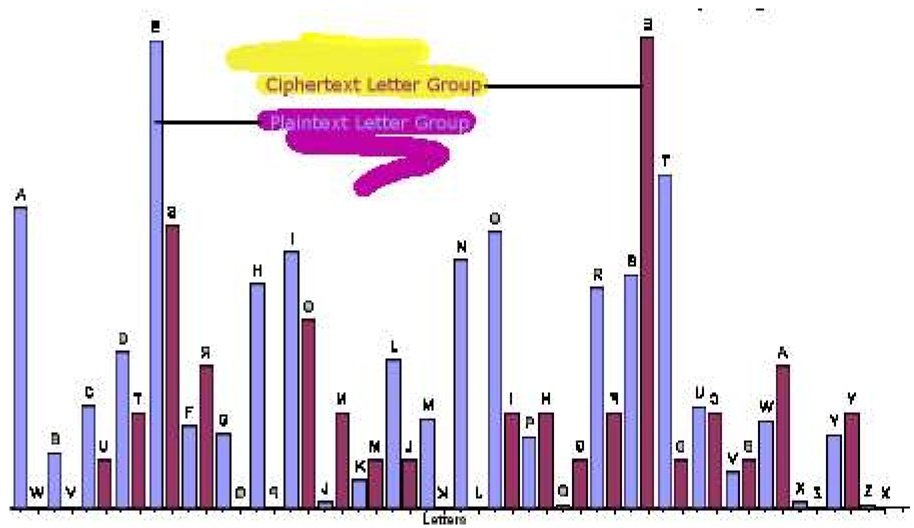
Here we think that E = E. Looks like it is not Beaufort. However, we know that this is Beaufort, so this is a wrong substitution. We should use an other pairing!



Frequency analysis (cont.)

- We should try all the possible substitutions and try to find the one, where the difference of the frequencies are minimal.
  - E.g.: Using square sum for the difference measurement

Password: **W**OMBLE



Solution

```
BEAUF ORTAN DVIGE NEREB ECOME MUCHE ASIER
TOANA LYSEW HENTH EREIS ALOTO FTEXT TOWOR
KWITH THISA LLOWS USTOU SETHE REPEA TINGN
ATURE OFTHE KEYTO OBTAI NMANY VALUA BLEST
ATIST ICSON CETHE LENGT HOFTH EKEYI SASCE
RTAIN EDORP ERHAP SGUES SEDAT THENG ROUPS
OFLET TERSA KEYLE NGTHA PARTC ANBEA NALYS
EDASI FTHEY WEREA CAESA RCIPH ER
```

- "Beaufort and Vigenere become much easier to analyse when there is a lot of text to work with. This allows us to use the repeating nature of the key to obtain many valuable statistics. Once the length of the key is ascertained or perhaps guessed at, then groups of letters a key length apart can be analysed as if they were a Caesar cipher"

Practice: Polyalphabetic ciphers

- Try these tools
  - pages.central.edu/emp/lintont/classes/spring01/cryptography/java/kasiski.html

Ccgr mtgn m xwbi, gzije ieg p pvlxde smfa aug pavqh wc e iapdasi btee llw favshx. Jzifehif hlr oift ayh, ilr dmltxi uxvy osje m vss vvvmfg opcpo, fg inedccci vf xze hmzaetw gslxir wie Dmltxi Fth Eahans Lcdh.

Bfi eodrwck, Yaxllq Vss Vvvmfg Tscs efciv hqv adxuwv af els rshdh yo fs jxwvl lwr svochzgxzed eg xx ush teqr ollvdi kizgs ilrq'h keqr spgu gxzed.

"Xvpx'f s kgop mrte," uwv eoflsg wnah. Ko flsn tnuowd m rwri oswcef jcg Pvlxde Dir Gmqary Hasr is gsow ta lsg kesrvmaxvtv.

Breaking autokey ciphers

- The plaintext is in the key

- Example, using Vigenere cipher

```
Plaintext:     MEETATTHEFOUNTAIN
Key:           KILTMEETATTHEFOUN
Ciphertext:    WMPMMXXAEYHBRYOCA
```

Breaking autokey ciphers (cont.)

- Searching for most common words (E.g.: THE)

```
– ciphertext: WMP MMX XAE YHB RYO CA
– key:        THE THE THE THE THE ..
– plaintext:  DFL TFT ETA FAX YRK ..

– ciphertext: W MPM MXX AEY HBR YOC A
– key:        . THE THE THE THE THE .
– plaintext:  . TII TQT HXU OUN FHY .

– ciphertext: WM PMM XXA EYH BRY OCA
– key:        .. THE THE THE THE THE
– Plaintext:  .. WFI EQW LRD IKU VVW
```

- The fragments we get in the plaintext are more or less probable fragments of the language
  - (most probable) FAX OUN ETA … … FTF DFL EQW (least probable)

Breaking autokey ciphers (cont.)

- Investigating the length of the key (Hopefully not so long)
  - The most probable FAX is not working here, try the second OUN

  - Key length: 4:

```
cipher:     WMPMMXXAEYHBRYOCA
key:        ......ETA.THE.OUN
plain:      ......THE.OUN.AIN
```

Here we have fragments, that seems to be OK according their probability

  - Key length: 5:

```
cipher:     WMPMMXXAEYHBRYOCA
key:        .....EQW..THE..OU
plain:      .....THE..OUN..OG
```

  - Key length: 6:

```
cipher:     WMPMMXXAEYHBRYOCA
key:        ....TQT...THE...O
plain:      ....THE...OUN...M
```

Breaking autokey ciphers (cont.)

- Searching for possible key/plain text

```
cipher:     WMPMMXXAEYHBRYOCA
key:        ..LTM.ETA.THE.OUN
plain:      ..ETA.THE.OUN.AIN
```

- Discovering the plaintext
  - As the plaintext is in the key, we can check it immediately

```
plain: M.ETA.THE.OUN.AIN
plain: MEETATTHEFOUNTAIN
```

Breaking autokey ciphers II.

- Using frequency analysis
  - We can use it on autokey ciphers through the discovered plaintext

- Example:

```
VFPJUDEEVUHCUWRNGSZNKARFFNVXILDPFNVXI?ANLBDHYUBYV
GYAIXDSMXKFBPITVXDUYNWWTTPIZVUITXOYBXQENNTXMJQKHM
FBTJZBHBFLHZYKOLFOJFQISQQJHNPCYKDKYAWQYFIIHMDSFFE
RJGSDFJQZJWTWNFG?FNSSDYQRUXKSFVKVSUZCRFZIKFUEKVIE
ZFFLPIZYHTSBTRYJELFSDUNQMYVHW?VXKCRFCAQZHCPENQSGP
EXZUFXQLYVZUAEIVGLYNEIIFKXQJZWLPLVYWBTNURIALZAGVK
NTDMTQHEKYCOZYTEFGNZUYTXOSQLAATPIIAVALTZXROPKZSNX
QJWJWWJJRGEFGAOIRXLLGDLBBFDRP
```

(Vorlath blog)

- Here we use an unknown cipher

Breaking autokey ciphers II. (cont.)

Key for the susbtitution:
KRYPTOS

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

Breaking autokey ciphers II. (cont.)

- To discover the plaintext, we should know
  - The substitution table
  - The length of the keyword
    - There is no repeating key, so we cannot use the previous methods

- We should try all the possibilities
  - Not for the key itself

Breaking autokey ciphers II. (cont.)

- Known alphabet and key length
  - We create groups using the length of the key
  - We create alphabet long (English: 26) subgroups. The first letter is the suspected key. The second is the part of the ciphertext decrypted with the key. The third is the ciphertext part decrypted with the previously found letter (autokey use the plaintext). And so on…
  - We try to find the most probable subgroup using the frequency of the letters in a language.
    - We give number to the letters according their frequency
      - E.g.: English: `ABCDEFGHIJKLMNOPQRSTUVWXYZ`
        `84779657812768862889655360`

Breaking autokey ciphers II. (cont.)

- In the example the key length is 8
  - 8 groups
- In the figure there is the 26 cases for the first group
  - 1., 9., 17., 25., … letters of the ciphertext decrypted

```
AIASSSYEGENOHTTGANNNWIOUTEHHONSAGYGTIXVSVESNTCIZ 350
BHBOAOPDHDQTIPOFBMQMXHSQODIGSMASHRHPJWWOWDAMOBOX 297
CGCTBTTCICUPJYSECLULZGANSCJFALBOIKIYLVXTXCBLSASW 303
DFDPCPOBJBVYLRADDJVJKFBMABLEBJCTJZJRMUZPZBCJASAV 247
EEEYDYSALAWRMKBCEIWIRECLBAMDCIDPLXLKNQKYKADIBOBU 307
FDFRERASMSXKNZCBFHXHYDDJCSNCDHEYMWMZQNRRRSEHCTCQ 301
GCGKFKBONOZZQXDAGGZGPCEIDOQBEGFRNVNXUMYKYOFGDPDN 266
HBHZGZCTQTKXUWESHFKFTBFHETUAFFGKQUQWVLPZPTGFEYEM 270
IAIXHXDPUPRWVVFOIEREOAGGFPVSGEHZUQUVWJTXTPHEFRFL 300
JSJWIWEYVYYVWUGTJDYDSSHFGYWOHDIXVNVUXIOWOYIDGKGJ 279
KVKGZGUMSMFFAEWQKKFKJVXYWMAUXKZHSCSEBOIGIMZKWJWS 237
LOLVJVFRWRPUXQHPLCPCAOIEHRXTICJWWMWQZHSVSRJCHZHI 282
MTMULUGKXKTQZNIYMBTBBTJDIKZPJBLVXLXNKGAUAKLBIXIH 251
NPNQMQHZZZONKMJRNAOACPLCJZKYLAMUZJZMRFBQBZMAJWJG 224
OLOBTBKGEGLAFSYIOULUULPWYGFJPUTCETESGKQBQGTUYEYR 300
PNPDYDXICIICDBKLPWIWNNRZKIDMRWYECSCBEYMDMIYWKGKP 294
QYQNNNIXKXSMRLLKQSSSDYMBLXRRMSNQKIKLYECNCXNSLVLF 287
RURFKFVLALGEBDXNRZGZLUZRXLBQZZKGABADCTJFJLKZXIXO 238
SJSAOARFFFMSGOPHSQMQVJTVPFGITQOBFPFOHZUAUFOQPDPK 283
TMTCPCZHDHJBEARJTVJVQMYXRHELYVPDDODAFRNCNHPVRFRY 301
URUMQMJWRWALYJMZUOAOERNAMWYKNOQNRHRJPDDMDWQOMUME 288
VKVLULLVYVBJPINXVTBTFKQSNVPZQTUMYGYITCELEVUTNQND 283
WZWJVJMUPUCITHQWWPCPGZUOQUTXUPVLPFPHOBFJFUVPQNQC 252
XXXIWINQTQDHOGUVXYDYHXVTUQOWVYWJTETGSAGIGQWYUMUB 274
YQYEREWJBJHDCCZMYXHXMQKKZJCNKXRFBABCDPLELJRXZHZT 238
ZWZHXHQNONEGSFVUZRERIWWPVNSVWRXIODOFASHHHNXRVLVA 296
```

Breaking autokey ciphers II. (cont.)

- Investigating all the groups

```
1 AIASSSYEGENOHTTGANNNWIOUTEHHONSAGYGTIXVSVESNTCTZ 350
2 BTLITIUANLFNERERNLXGASURHWOECLWSEERYNPENENEUYOO  348
3 SWLBHBSREDOWRADUUODLBTLIEHKEAYATXIESUOSONDITFNX  330
4 CAYLALETTXRAENUUNCOEOHDERENXTWSMTGEETIERTEGEODX  353
5 ISIETEDHITMSDSNNKAEYUEIDEROAIWHEHHSVENCTYGHSUSY  350
6 STNHPTTSCHAGAMDDNTSKTYTOSEWCOTISITFESTOHSRTFRWI  356
7 SOVOOHHMFETANIETOILNTSSUOXSTNHSSRDINSFNSEEMOSEZ  354
8 ATIWSEEAIIITDTROWOAOHHBTMWTLOILATEFMIIDEVEIRESU  358
```

- The first letters are the key. The rest is the plaintext.

Statistical tests

- Statistical test may help during the cryptanalysis

- Index of coincidence
  - Comparing two random English letters (latin) we have a chance for the match as 1/26 = 0.0385
  - Comparing two English letter from a written text, we get 0.0667 probability for the matching
  - We can use this difference in the tests

Index of coincidence

- φ_r: In an alphabet, containing c letters, and in a totally random N length text. The expected results for the number of matches is
  - $1/c \cdot N(N-1)$
- φ_p: In the case of a text from a written language, the expected results of the matches is
  - $IC \cdot N(N-1)$, where IC is language specific
- φ_o: The empirical number of matches:
  - $\sum n_i(n_i-1)$

Index of coincidence

- Friedman test
- ΔIC: the ratio of the empirical and expected

  $\varphi o/\varphi r = \frac{1}{1/c}\sum_{i=1}^{c}\frac{n_i(n_i-1)}{N(N-1)}$
- Language specific

English	1.73	Italian	1.94
French	2.02	Portuguese	1.94
German	2.05	Russia	1.76
		Spanish	1.94

Index of coincidence

- Example ciphertext (using Vigenere cipher):

```
QPWKA LVRXC QZIKG RBPFA EOMFL  JMSDZ VDHXC XJYEB IMTRQ WNMEA
IZRVK CVKVL XNEIC FZPZC ZZHKM  LVZVZ IZRRQ WDKEC HOSNY XXLSP
MYKVQ XJTDC IOMEE XDQVS RXLRL  KZHOV
```

- We group the ciphertext based on the investigated key length and test whether they are from an English text
  - Results (key length:ΔIC):
    1:1.12, 2:1.19, 3:1.05, 4:1.17, **5:1.82**, 6:0.99, 7:1.00, 8:1.05, 9:1.16, **10:2.07**

ΔIC example 2.

Letters:	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
f:	3	3	0	7	2	1	1	4	0	0	1	0	0	0	4	1	6	3	0	4	1	0	5	1	0	3
f-1:	2	2		6	1			3							3		5	2		3			4			2
f(f-1):	6	6		42	2			12							12		30	6		12			20			6

$$
\begin{aligned}
\phi o &= \Sigma f\,(f-1) \\
&= 6+6+42+2+12+12+30+6+12+20+6 \\
&= 154
\end{aligned}
$$

$$
\begin{aligned}
\phi p &= .0667\,N\,(N-1) \\
&= .0667 \times 50 \times 49 \\
&= 163
\end{aligned}
$$

$$
\begin{aligned}
\phi r &= .0385\,N\,(N-1) \\
&= .0385 \times 50 \times 49 \\
&= 94
\end{aligned}
$$

$$
\begin{aligned}
\Delta IC &= \phi o / \phi r \\
&= 154/94 \\
&= 1.64
\end{aligned}
$$

References

- US ARMY Cryptography manual
  - http://www.umich.edu/~umich/fm-34-40-2/
- Codes
  - http://www.secretcodebreaker.com/codes.html
  - http://25yearsofprogramming.com/fun/ciphers.htm