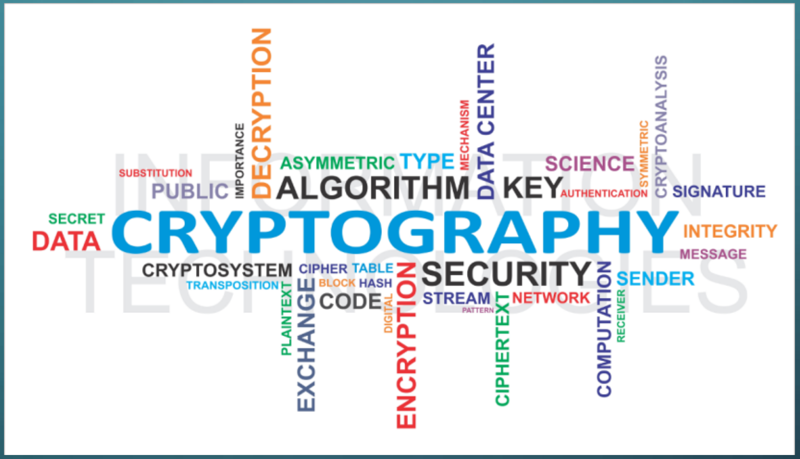
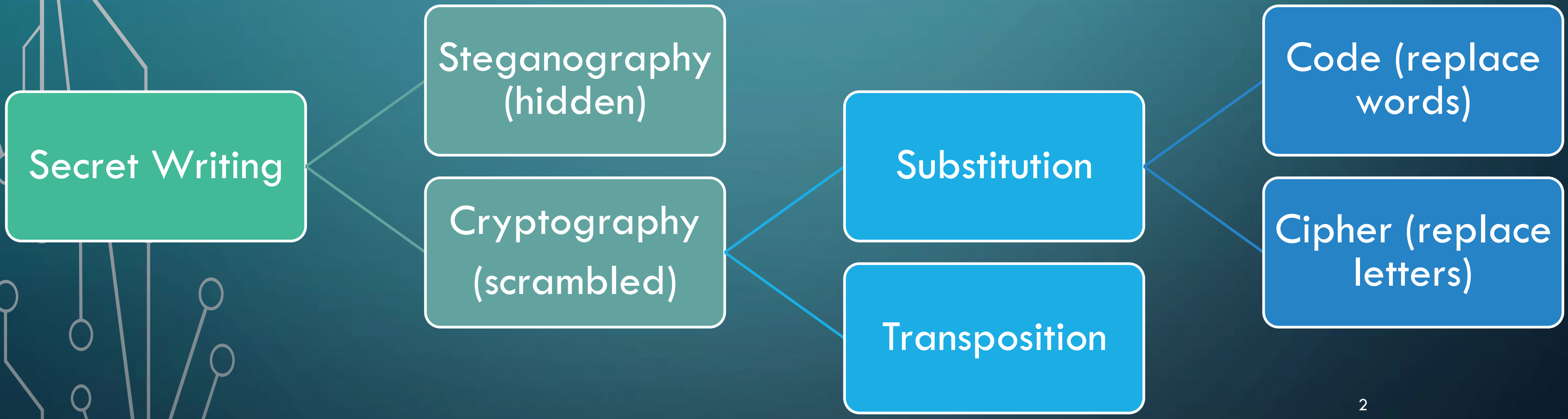
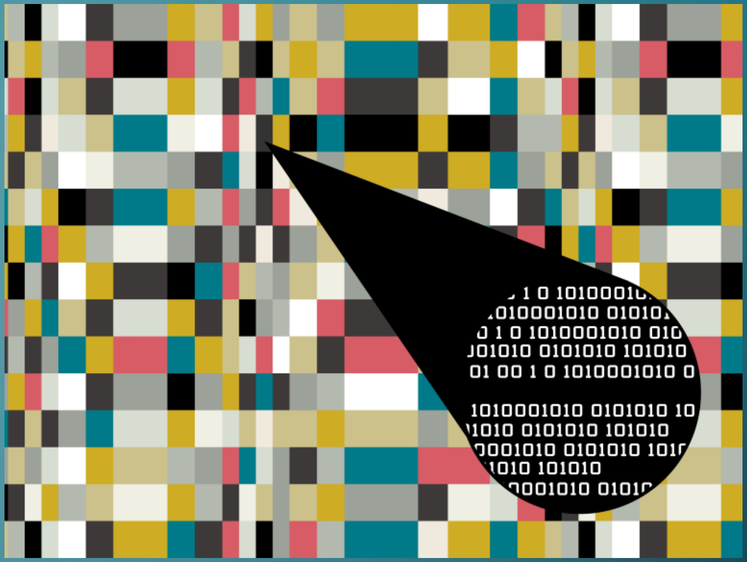
<BASIC AND HISTORICAL>



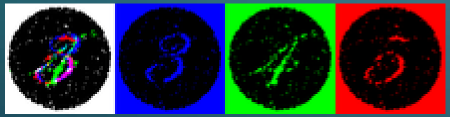
The science of secret writing and its main branches.





STEGANOGRAPHY

THE SAME IMAGE VIEWED BY WHITE, BLUE, GREEN AND RED LIGHTS REVEALS DIFFERENT HIDDEN NUMBERS.



Sources: The Code Book / Singh
https://en.wikipedia.org/wiki/Steganography

STEGANOGRAPHY

- In Greek
  - Steganos = covered
  - Graphein = to write
- Steganography is about hiding messages
- Historically, secret messages were often hidden (or memorized)
- Today, steganography is used primarily to protect digital rights
  - "watermarking" copyright notices
  - "fingerprinting" a serial ID

HISTORY OF STEGANOGRAPHY (PHYSICALLY HIDING)

- Runners were memorizing messages
  - Sometimes killed after delivering the message

  - This is Sparta!! (300-film)

HISTORY OF STEGANOGRAPHY (PHYSICALLY HIDING)

- Demaratus tells Athens of Persia's attack plans
  - Writes the secret message directly on the wooden backing of a wax tablet before applying its beeswax surface.

Wax tablet and a Roman stylus

Writing with stylus and folding wax tablet. painter, Douris, ca 500 BC

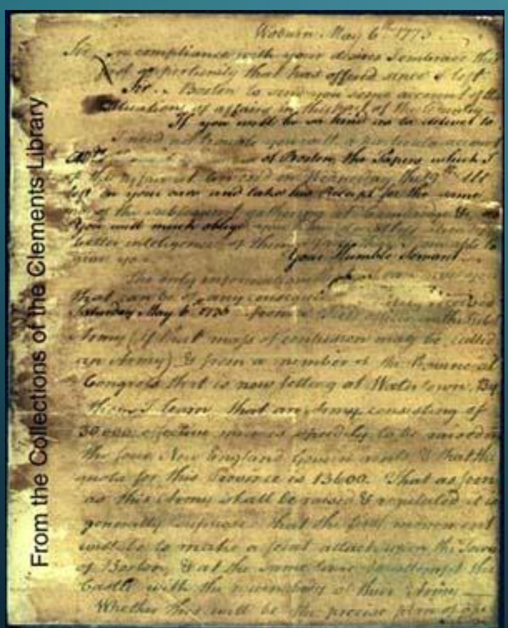
HISTORY OF STEGANOGRAPHY (PHYSICALLY HIDING)

- Greek Histaiaeus encouraged Aristagoras of Miletus to revolt against the Persian King.
  - Writes message on the shaved head of the messenger, and sends him after his hair grew, with the instruction, "When thou art come to Miletus, bid Aristagoras shave thy head, and look thereon."

- Chinese silk balls
  - Message is written on silk, turned into wax-covered ball that was swallowed by the messenger

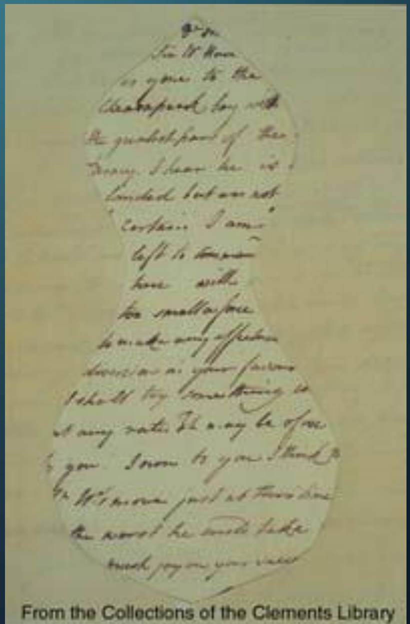
HISTORY OF STEGANOGRAPHY (CONT.)

- Invisible Ink
  - Certain organic fluids (milk, fruit juice) are transparent when dried but the deposit can be charred and is then visible
  - Romans used to write between the lines
  - A mixture of alum and vinegar may be used to write on hardboiled eggs, so that can only be read once shell is broken
  - Pin punctures: Small pin punctures on selected letters are ordinarily not visible unless the paper is held up in front of a light



From the Collections of the Clements Library

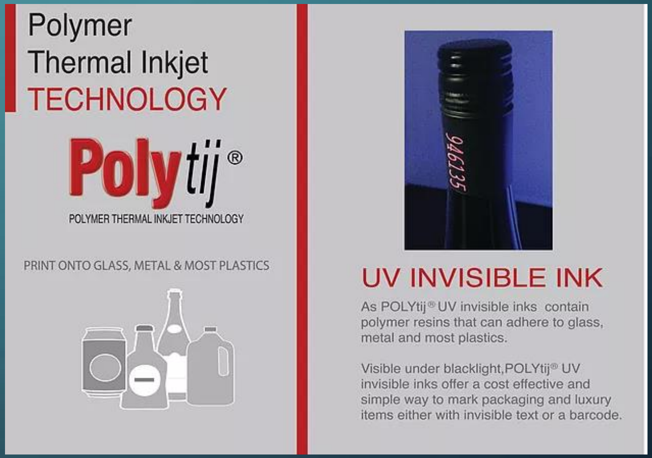
From the Collections of the Clements Library



From the Collections of the Clements Library

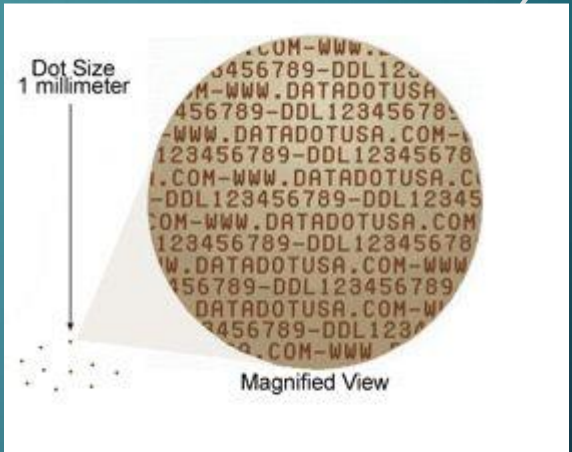
HISTORY OF STEGANOGRAPHY (CONT.)

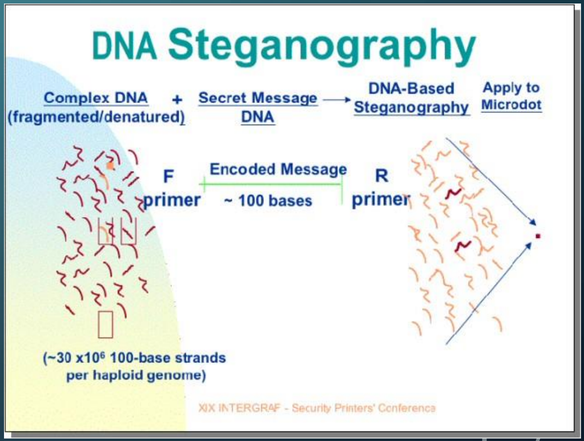
- Invisible Ink (today)
  - Invisible ink-jet technology, Ink that is too small for human eye (Univ of Buffalo, 2000)
  - Invisible inks that are only visible under UV light



HISTORY OF STEGANOGRAPHY (CONT.)

- Microdots
  - is text or an image substantially reduced in size onto a small disc to prevent detection by unintended recipients.
  - WW2 Germany - documents shrunk to the size of a dot, and embedded within innocent letters
  - DNA microdot, embedding synthetically formed DNA sequence (secret) into a normal DNA strand, then posting as microdot
  - Microdots with barcode-like information

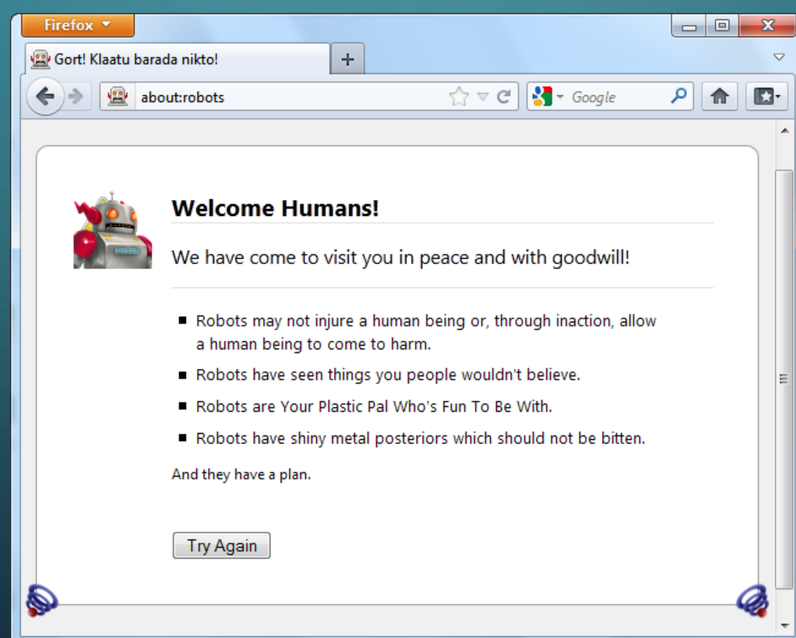




HISTORY OF STEGANOGRAPHY (CONT.)

- Easter eggs 

  - Programmers embed in software
    - See http://www.eeggs.com
  - Claims that Beatles embedded secret messages in their music http://www.bbc.com/culture/story/20141003-the-hidden-messages-in-songs



HIDING A MESSAGE WITHIN A TEXT

- An actual message from a German spy during world war I

"Apparently, neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affect pretext for embargo on by products, ejecting suets and vegetable oils."

The codebreakers/Khan

HIDING A MESSAGE WITHIN A TEXT

- An actual message from a German spy during world war I
  - read second letter in each word

"Apparently, neutral's protest is thoroughly discounted and ignored. Isman hard hit. Blockade issue affect pretext for embargo on by products, ejecting suets and vegetable oils."

"Pershing Sails from NY June 1" The codebreakers/Khan

Whoever the sender was, his ingenuity was expanded in vain, since Pershing actually sailed May 28.

HIDING A MESSAGE WITHIN A TEXT (MORE)

- Shift some words by one point/pixel.
  - Shifted words (or their first letters) make the sentence
- Use different fonts
  - Letter by letter or word by word (Francis Bacon Cipher)
- Lexical steganography uses the redundancy of the English language
  - “I feel well” and “I feel fine” seem the same, but one may be used to encode “SOS”
- Chaffing and winnowing
  - Riddle text with extra parts that the receiver will know how to remove (e.g., those that don’t “authenticate”)

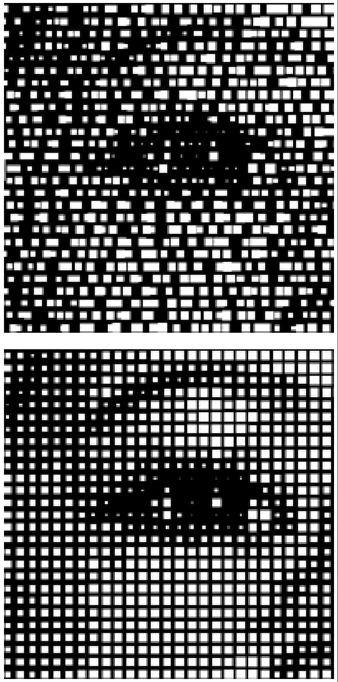
MODERN STEGANOGRAPHY

- Hiding one message within another (“container”)
- Most containers are rich media
  - Images, audio, video are very redundant, can be tweaked without affecting human eye/ear
  - US argued that Bin Laden implanted instructions within taped interviews
- Copyright notices embedded in digital art
  - Prove ownership
  - Serial number embedded to prevent replication
  - Seek infringements on the web using spiders
- Digital cameras EXIF tags
  - Not secretive, but hidden from the eye
  - Embed info such as camera type, date, shutter speed, focal length,..
- Similarly, possible to embed messages in invisible parts of html pages

HIDING A MESSAGE IN AN IMAGE

- Example: use 1-2 Least Significant Bits (LSB) in each pixel
  - human eye wont notice the difference
  - message can be compressed to reduce number of bits needed
  - only half the bits are likely to change on average
  - prefer "containers" with a lot of variation

- Check out Steganos (www.steganos.com), Digimarc (www.digimarc.com)



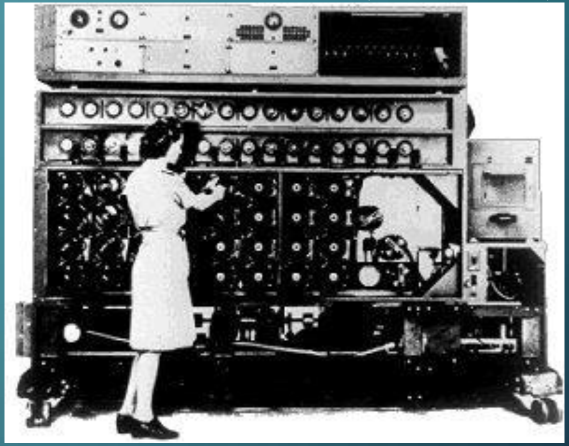
The top image shows an enlarged area of a printed picture with an image encoded in it. The bottom image shows the same printed picture without the concealed image.

Source: Ben Gurion University

EXAMPLE (STEGANOS)

Original Picture

Embedded Picture (bombe)



With embedded picture

JPG version

STEGANALYSIS

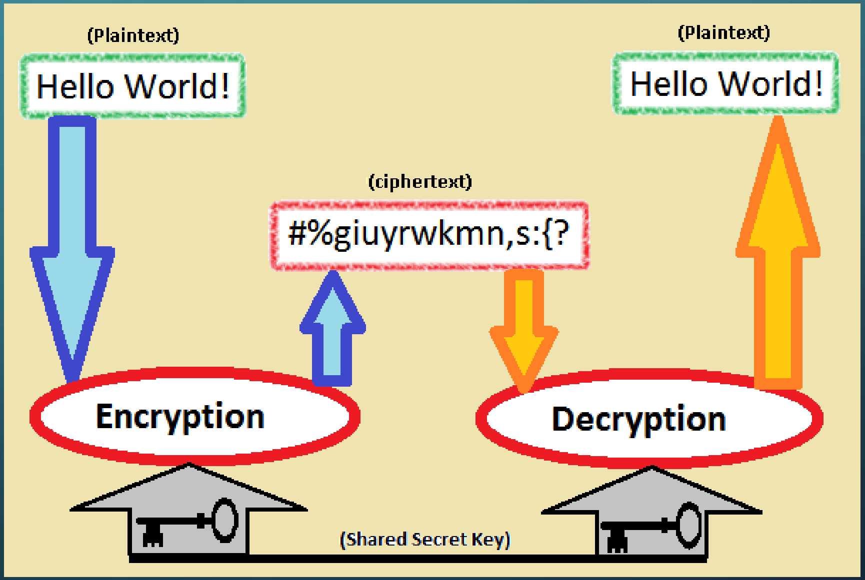
- Detection: is there a hidden message?
  - Develop signatures for known steganographic tools, e.g. in LSB method, expect local homogeneity
  - When content is encrypted, the message should have a high entropy ("white noise")
  - Promising results: high detection rates
- Decoding: recover hidden message
  - No significant work in this area !
- Prevention: destroy or remove a hidden message
  - Most steganographies not robust to image alterations
  - Short messages (e.g. copyright) can be encoded redundantly and survive an alternation

STEGANOGRAPHY (SUMMARY)

- Steganography is arguably weaker than cryptography because the information is revealed once the message is intercepted
- On the other hand, an encrypted message that is not hidden may attract attention, and in some cases may itself incriminate the messenger
- In any event, steganography can be used in conjunction with cryptography

CRYPTOGRAPHY

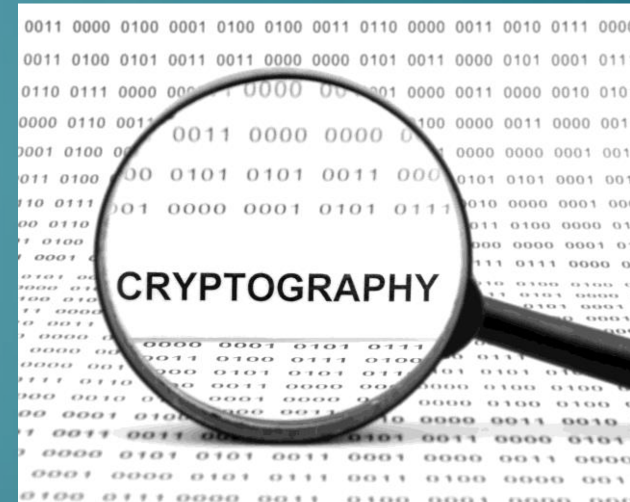
- Cryptography (or cryptology; from Greek κρυπτός, "hidden, secret"; and γράφειν, graphein, "writing", or -λογία, -logia, "study", respectively) is the practice and study of techniques for secure communication in the presence of third parties (called adversaries).
- Modern cryptography intersects the disciplines of mathematics, computer science, and electrical engineering. Applications of cryptography include ATM cards, computer passwords, and electronic commerce.



http://en.wikipedia.org/wiki/Cryptography

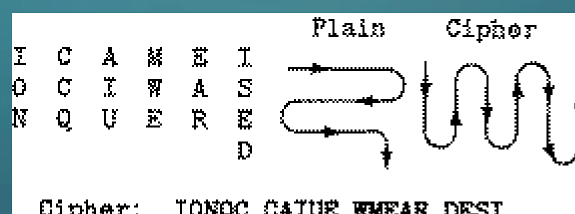
BASIC TERMINOLOGY

- **Plaintext** - the original message
- **Ciphertext** - the coded message
- **Cipher** - algorithm for transforming plaintext to ciphertext
- **Key** - info used in cipher known only to sender/receiver
- **Encipher (encrypt)** - converting plaintext to ciphertext
- **Decipher (decrypt)** - recovering ciphertext from plaintext
- **Cryptography** - study of encryption principles/methods
- **Cryptanalysis (codebreaking)** - the study of principles/ methods of deciphering ciphertext *without* knowing key
- **Cryptology** - the field of both cryptography and cryptanalysis

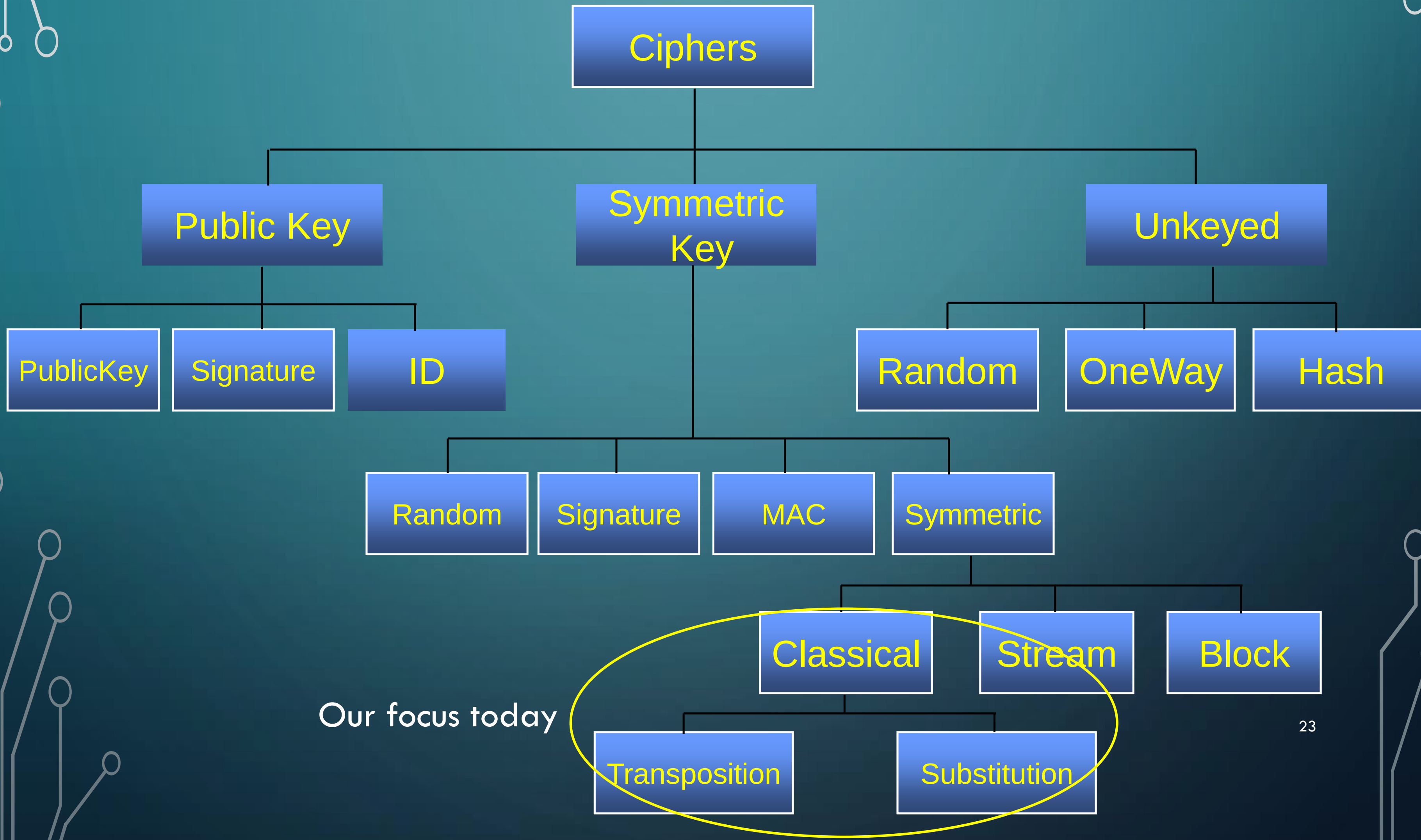


HISTORY OF CRYPTOGRAPHY

I CAME I SAW I CONQUERED

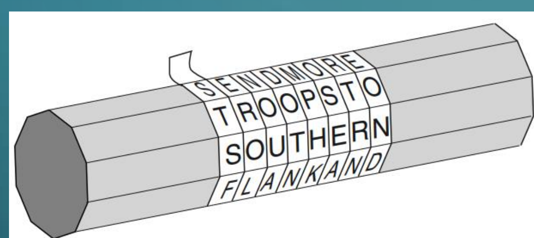


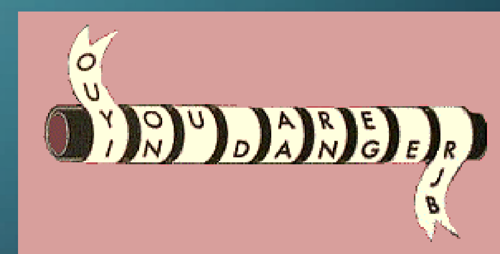
CIPHER CLASSIFICATION



CRYPTOGRAPHY

- *transposition* of letters
  - Scytale is first cryptographic device (Lysander of Sparta 500 BC)
    - Message written on a leather strip (the inside of a servant's belt), which is then unwound to scramble the message.
    - Unreadable without proper diameter of wooden rod
    - The message warned Lysander that Persia was about to go to war against him. He immediately set sail and defeated the Persians.





- *substitution*
  - Hebrew ATBASH (אתבש) https://en.wikipedia.org/wiki/Atbash
  - Kama-Sutra suggests that women learn to encrypt their love messages by substituting pre-paired letters (4^{th} Century AD)
  - Cipher – replace letters
  - Code – replace words

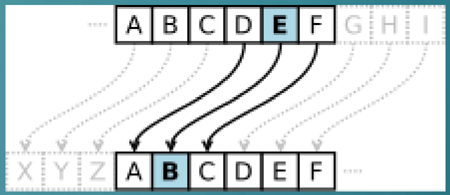
POLYBIUS SQUARE

- The Greeks also invented a code which changed letters into numbers. A is written as 11, B is 12, and so on. So WAR would read 52 11 42. A form of this code was still being used two thousand years later during the First World War. Modern variation:

	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

https://en.wikipedia.org/wiki/Bifid_cipher

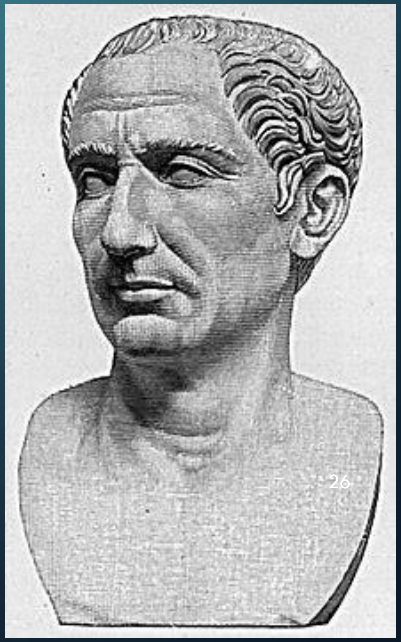
CAESAR CIPHER



- Simplest and earliest known use of a substitution cipher
- Used by Julius Caesar
- Involves replacing each letter of the alphabet with the letter standing three places further down the alphabet
- Alphabet is wrapped around so that the letter following Z is A
- Plain: ABCDEFGHIJKLMNOPQRSTUVWXYZ
- Cipher: XYZABCDEFGHIJKLMNOPQRSTUVW

plain: meet me after the toga party

cipher: PHHW PH DIWHU WKH WRJD SDUWB



CAESAR CIPHER ALGORITHM

- Mathematically give each letter a number

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
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

- Algorithm can be expressed as:

$$c = E(3, p) = (p + 3) \bmod (26)$$

- A shift may be of any amount, so that the general Caesar algorithm is:

$$C = E(k, p) = (p + k) \bmod 26$$

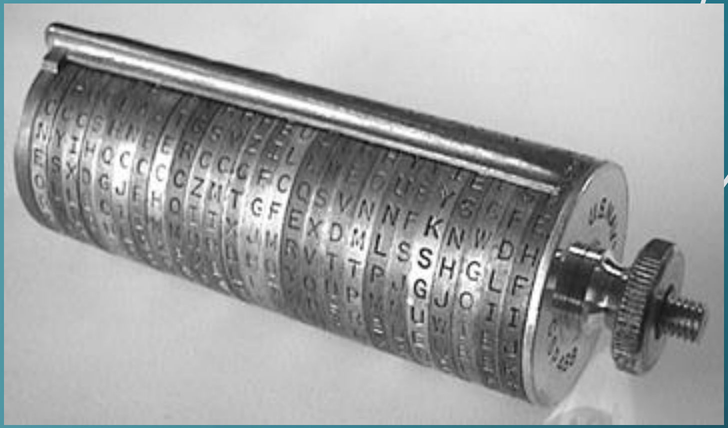
- Where k takes on a value in the range 1 to 25; the decryption algorithm is simply:

$$p = D(k, C) = (C - k) \bmod 26$$

Brute-Force Cryptanalysis of Caesar Cipher

KEY	PHHW PH DIWHU WKH WRJD SDUWB
1	oggv og chvgt vjg vqic rctva
2	nffu nf bgufs uif uphb qbsuz
3	meet me after the toga party
4	ldds ld zesdq sgd snfz ozqsx
5	kccr kc ydrcp rfc rmey nyprw
6	jbbq jb xcqbo qeb qldx mxoqv
7	iaap ia wbpan pda pkcw lwnpu
8	hzzo hz vaozm ocz ojbv kvmot
9	gyyn gy uznyl nby niau julns
10	fxxm fx tymxk max mhzt itkmr
11	ewwl ew sxlwj lzw lgys hsjlq
12	dvvk dv rwkvi kyv kfxr grikp
13	cuuj cu qvjuh jxu jewq fqhjo
14	btti bt puitg iwt idvp epgin
15	assh as othsf hvs hcuo dofhm
16	zrrg zr nsgre gur gbtn cnegl
17	yqqf yq mrfqd ftq fasm bmdfk
18	xppe xp lqepc esp ezrl alcej
19	wood wo kpdob dro dyqk zkbdi
20	vnnc vn jocna cqn cxpj yjach
21	ummb um inbmz bpm bwoi xizbg
22	tlla tl hmaly aol avnh whyaf
23	skkz sk glzkx znk zumg vgxze
24	rjjy rj fkyjw ymj ytlf ufwyd
25	qiix qi ejxiv xli xske tevxc

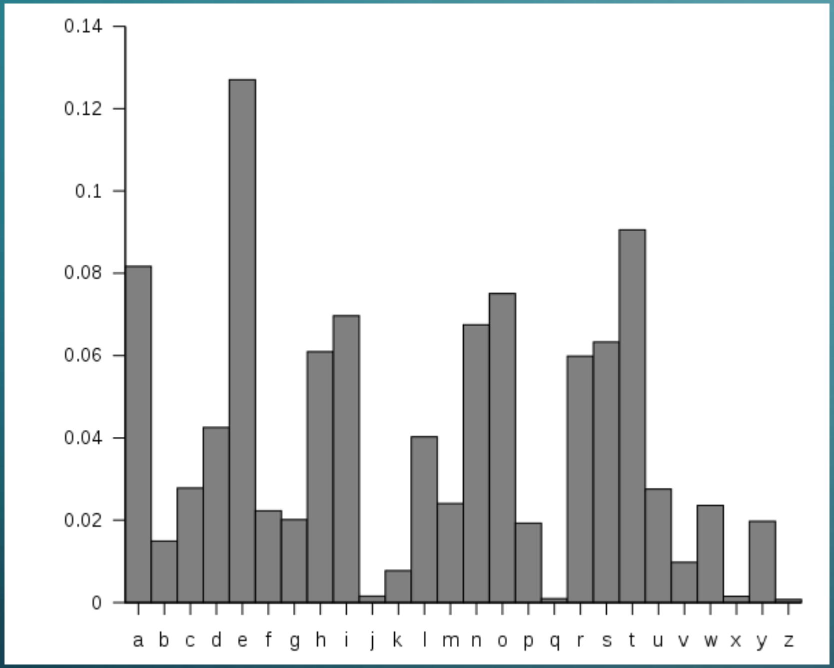
MONOALPHABETIC CIPHERS

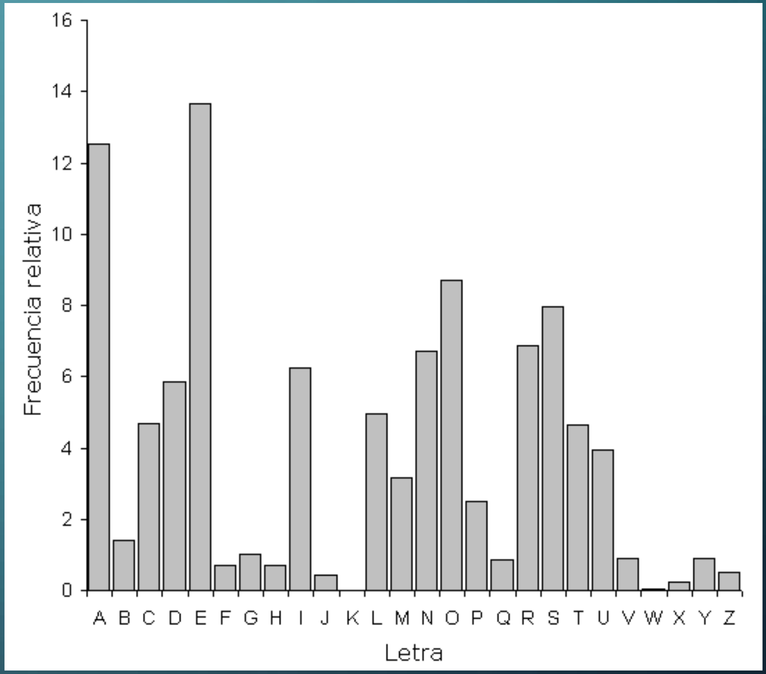


- Jefferson wheel implementation
  https://en.wikipedia.org/wiki/Jefferson_disk
  - Set the message across the wheels
  - Select another line (in random) as cipher
- Substitution based on key phrase
  - Substitution key consists of phrase's letters (uniquely) followed by rest of the alphabet in order
    - Phrase: THIS IS ALICE AND BOB'S KEY
    - Key: THISALCENDBOKY-FGJMPQRUVWXZ
  - 26!=403291461126605635584000000 (4.03291×10^{26}) monoalphabetic substitution ciphers

BREAKING MONOALPHABETIC CIPHERS

- Relative Frequency of Letters in English Text (or native language)
- https://en.wikipedia.org/wiki/Frequency_analysis





- The Arabs broke monoalphabetic substitution using frequency analysis
- Thus, letters ciphering e, t, and a are easily discovered
- Subsequently can look for the rest of the letters and letter pairs

BREAKING MONOALPHABETIC CIPHERS

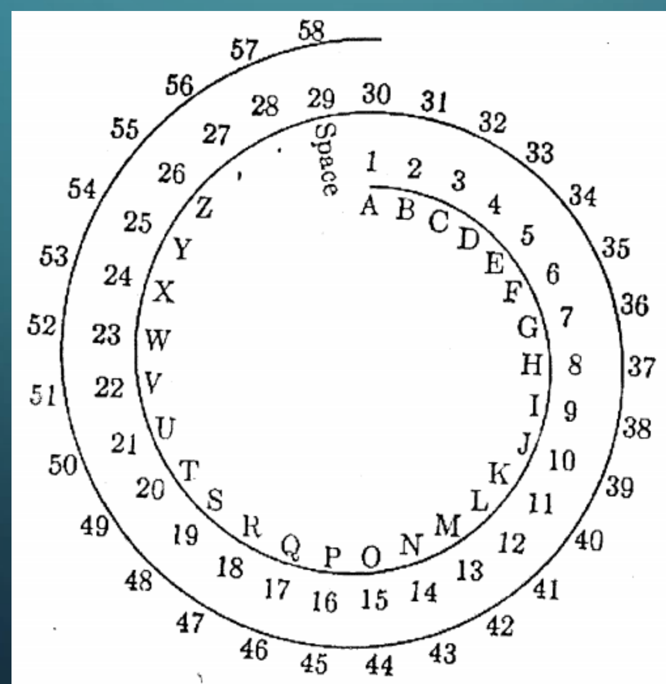
- According to the unicity distance of English, 27.6 letters of ciphertext are required to crack a mixed alphabet simple substitution.
- In practice, typically about 50 letters are needed, although some messages can be broken with fewer if unusual patterns are found.
- https://en.wikipedia.org/wiki/Unicity_distance

CÆSAR'S PROBLEM

- Key is too short
  - Can be found by exhaustive search
  - Statistical frequencies not concealed well
    - They look too much like regular English letters
- So make it longer
  - Multiple letters in key
  - Idea is to smooth the statistical frequencies to make cryptanalysis harder
  - Use alternate symbols for alphabet like https://en.wikipedia.org/wiki/Pigpen_cipher

HOMOPHONIC SUBSTITUTION

- Homophonic substitution cipher can be used to foil frequency analysis
- In these ciphers, plaintext letters map to more than one ciphertext symbol.



HOMOPHONIC SUBSTITUTION

- Usually, the highest-frequency plaintext symbols are given more equivalents than lower frequency letters. In this way, the frequency distribution is flattened, making analysis more difficult.
  - Keyed 2-digit substitution

	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
T	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	00	01	02	03	04	05
H	43	44	45	46	47	48	49	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
E	71	72	73	74	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70
K	90	91	92	93	94	95	96	97	98	99	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89

  - Reverse frequency

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
06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	00	01	02	03	04	68	05
43	44	45	46	47	48	49	25	26			29	30	31	32	33		35	36	37	38		40		87	
71		73	74	50			53	54			57		59	60			63	64	65	66					
90			93	94			97	98			76		78	79			82	83	84						
72				51			56	58					61	34			39	86	42						
91				95			81	77					80	62			67	88	70						
92				52										85					89						
75				96															41						
				27															69						
				55																					
				99																					
				28																					

POLYALPHABETIC CIPHERS

- Polyalphabetic substitution cipher
  - Improves on the simple monoalphabetic technique by using different monoalphabetic substitutions as one proceeds through the plaintext message

All these techniques have the following features in common:

- A set of related monoalphabetic substitution rules is used
- A key determines which particular rule is chosen for a given transformation

VIGENERE POLYALPHABETIC CIPHER

- Vigenere's polyalphabetic cipher (19th century) generalizes Caesar's shift cipher
  - Use keyword to select encrypting rows

♦ The Vigenere cipher is not amenable to simple frequency analysis

♦ Actually invented earlier (Giovan Battista Bellaso [16th century](#))

♦ Called "le chiffre indéchiffrable" (The Unbreakable Cipher)

Vigenere Tableau

	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	S
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

VIGENÈRE CIPHER

- Best known and one of the simplest polyalphabetic substitution ciphers
- In this scheme the set of related monoalphabetic substitution rules consists of the 26 Caesar ciphers with shifts of 0 through 25
- Each cipher is denoted by a key letter which is the ciphertext letter that substitutes for the plaintext letter a
- Used as the standard field system by the British Army in World War I and the U.S. Army and other Allied forces during World War II

EXAMPLE OF VIGENÈRE CIPHER

- To encrypt a message, a key is needed that is as long as the message
- Usually, the key is a repeating keyword
- For example, if the keyword is *deceptive*, the message “we are discovered save yourself” is encrypted as:

key: deceptivedeceptivedeceptive

plaintext: wearediscoveredsaveyourself

ciphertext: ZICVTWQNGRZGVTWAVZHCQYGLMGJ

(STRONGER) VIGENÈRE *AUTOKEY* SYSTEM

- A keyword is concatenated with the plaintext itself to provide a running key
- Example:

key: deceptivewearediscoveredsav

plaintext: wearediscoveredsaveyourself

ciphertext: ZICVTWQNGKZEIIGASXSTSLVVWLA

- Since the key is as long as the message, the “Friedman and Kasiski” tests no longer work, as the key is not repeated.
- Even this scheme is vulnerable to cryptanalysis
  - Because the key and the plaintext share the same frequency distribution of letters, a statistical technique can be applied

PLAYFAIR CIPHER

- Best-known multiple-letter encryption cipher
- Treats digrams in the plaintext as single units and translates these units into ciphertext digrams
- Based on the use of a 5 x 5 matrix of letters constructed using a keyword
- Invented by British scientist Sir Charles Wheatstone in 1854
- Used as the standard field system by the British Army in World War I and the U.S. Army and other Allied forces during World War II

PLAYFAIR KEY MATRIX

- Fill in letters of keyword (minus duplicates) from left to right and from top to bottom, then fill in the remainder of the matrix with the remaining letters in alphabetic order
- Using the keyword MONARCHY:

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

PLAYFAIR CIPHER

Plaintext is encrypted two letters at a time, according to the following rules:

1. Repeating plaintext letters that are in the same pair are separated with a filler letter, such as x, so that balloon would be treated as ba lx lo on.

2. Two plaintext letters that fall in the same row of the matrix are each replaced by the letter to the right, with the first element of the row circularly following the last. For example, ar is encrypted as RM.

3. Two plaintext letters that fall in the same column are each replaced by the letter beneath, with the top element of the column circularly following the last. For example, mu is encrypted as CM.

4. Otherwise, each plaintext letter in a pair is replaced by the letter that lies in its own row and the column occupied by the other plaintext letter. Thus, hs becomes BP and ea becomes IM (or JM, as the encipherer wishes).

Assume one wants to encrypt the digram OR. There are five general cases:

1)

*	*	*	*	*
*	O	Y	R	Z
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*

Hence, OR → YZ

2)

*	*	O	*	*
*	*	B	*	*
*	*	*	*	*
*	*	R	*	*
*	*	Y	*	*

Hence, OR → BY

3)

Z	*	*	O	*
*	*	*	*	*
*	*	*	*	*
R	*	*	X	*
*	*	*	*	*

Hence, OR → ZX

4)

*	*	*	*	*
*	*	*	*	*
*	O	R	C	*
*	*	*	*	*
*	*	*	*	*

Hence, OR → RC

5)

*	*	*	*	*
*	*	R	*	*
*	*	O	*	*
*	*	I	*	*
*	*	*	*	*

Hence, OR → IO

https://en.wikipedia.org/wiki/Playfair_cipher

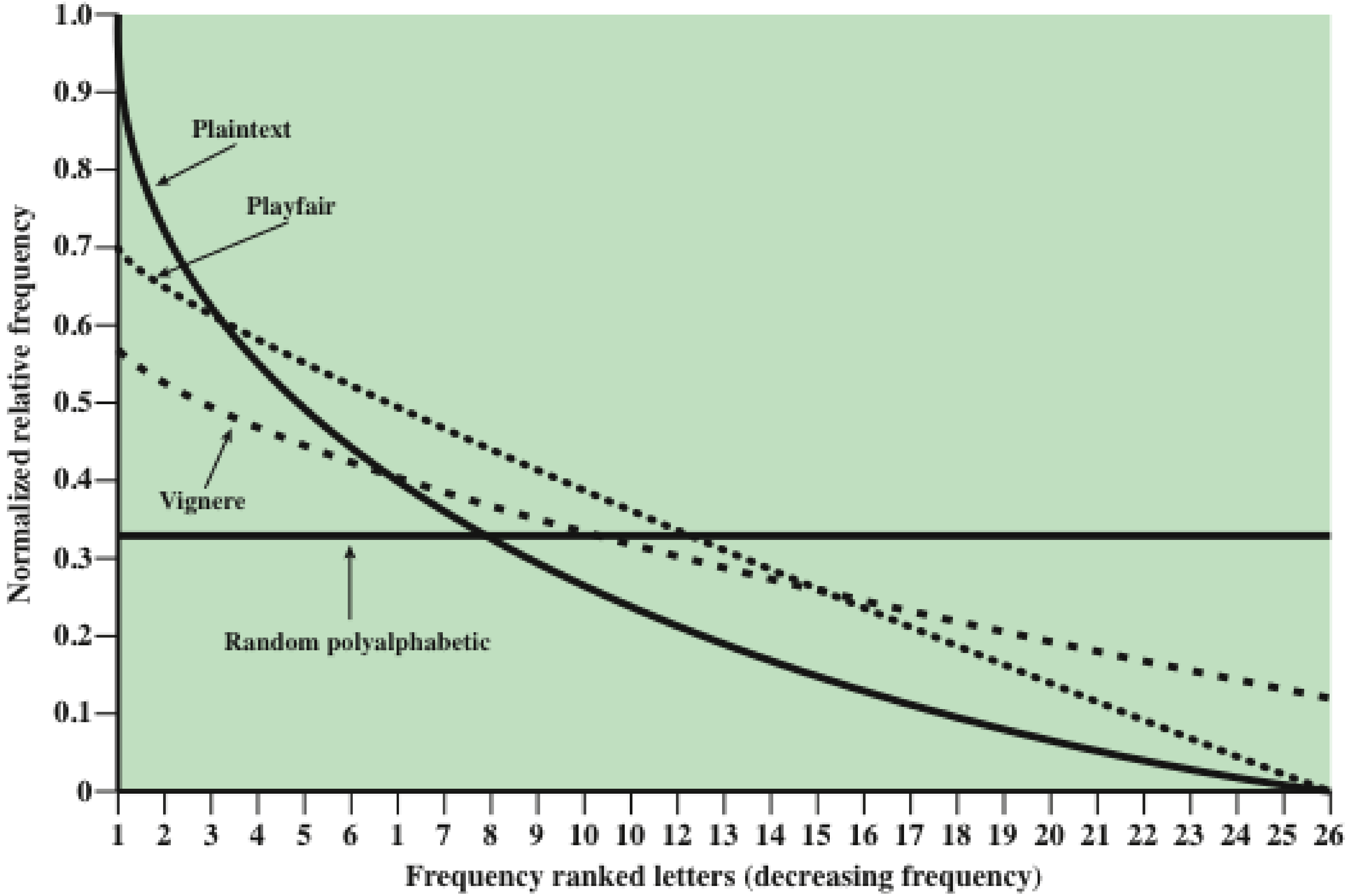


Figure 3.6 Relative Frequency of Occurrence of Letters

TRANSPOSITION CIPHER

- Rearrange letters in plaintext to produce ciphertext
- Example (Rail-Fence Cipher)
  - Plaintext is `HELLO WORLD`
  - Rearrange as

```
HLOOL
ELWRD
```

  - Ciphertext is `HLOOL ELWRD`

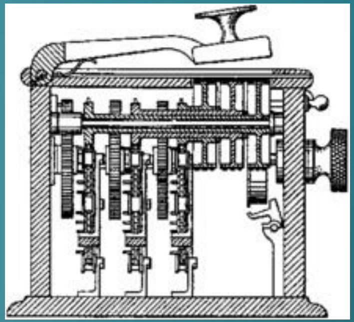
TRANSPOSITION CIPHERS

- Railfence with large cycles: TRHCEEIETGSSMAIAEASS

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 (by key): IETGIAESHCEESSMATRSS
- Columnar
  - IEEIRSHSMESCSTATGSEA

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



HILL CIPHER

- Developed by the mathematician Lester Hill in 1929
- Strength is that it completely hides single-letter frequencies
  - The use of a larger matrix hides more frequency information
  - A 3 x 3 Hill cipher hides not only single-letter but also two-letter frequency information
- Strong against a ciphertext-only attack but easily broken with a known plaintext attack

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

HISTORICAL CODING

- Louis XIV's Great Cipher (Rossignols) used one symbol (3-digit number) per syllable (held 200 years)
- https://en.wikipedia.org/wiki/Great_Cipher
- Mary Queen of Scots used a combination of cipher and coded words https://en.wikipedia.org/wiki/Babington_Plot
  - Referred to as a nomenclature because many codes were for names
- e.g,

assassinate	= D	general	= Σ	immediately	= 08
blackmail	= P	king	= Ω	today	= 73
capture	= J	minister	= ψ	tonight	= 28
protect	= Z	prince	= θ	tomorrow	= 43

Plain message = assassinate the king tonight
Encoded message = D–Ω–28

- US Army used Navajo language as code in WWII https://en.wikipedia.org/wiki/Code_talker

ATTACKING THE CIPHER, PRINCIPLES IN DECRYPTION

- Anagramming
  - If 1-gram frequencies match English frequencies, but other *n*-gram frequencies do not, probably transposition
  - Rearrange letters to form *n*-grams with highest frequencies

EXAMPLE, DECODING RAIL-FENCE

- Ciphertext: `HLOOLELWRD`
- Frequencies of 2-grams beginning with H
  - HE 0.0305
  - HO 0.0043
  - HL, HW, HR, HD < 0.0010
- Frequencies of 2-grams ending in H
  - WH 0.0026
  - EH, LH, OH, RH, DH ≤ 0.0002
- Implies E follows H

EXAMPLE

- Arrange so the H and E are adjacent

```
HE
LL
OW
OR
LD
```

- Read off across, then down, to get original plaintext

BREAKING VIGENERE CIPHER

- Babbage broke Vigenere's Cipher (1854, Crimean war)
  - Stage 1: Discover key length
    - Look for repeated sequences, and measure their distance
    - The key length is a factor of these distances
  - Stage 2: Identify the key itself
    - Compare distributions for each of the key letters with the standard distribution, to identify the shift
- Babbage could not publish his work
  - Similar techniques developed independently by Kasiski (a Prussian officer); Kerckhoff (French cryptographer)

- Check out an applet that breaks Vigenere: http://math.ucsd.edu/~crypto/java/EARLYCIPHERS/Vigenere.html

DECRYPTING THE VIGENERE CIPHER

- We want to break this cipher:

```
ADQYS MIUSB OXKKT MIBHK IZOOO
EQOOG IFBAG KAUMF VVTAA CIDTW
MOCIO EQOOG BMBFV ZGGWP CIEKQ
HSNEW VECNE DLAAV RWKXS VNSVP
HCEUT QOIOF MEGJS WTPCH AJMOC
HIUIX
```

ESTABLISH PERIOD

- Kaskski: *repetitions in the ciphertext occur when characters of the key appear over the same characters in the plaintext*
- Example:

```
key    VIGVIGVIGVIGVIGV
plain  THEBOYHASTHEBALL
cipher OPKWWECIYOPKWIRG
```

Note the key and plaintext line up over the repetitions (underlined). As distance between repetitions is 9, the period is a factor of 9 (that is, 1, 3, or 9)

REPETITIONS IN EXAMPLE

Letters	*Start*	*End*	*Distance*	*Factors*
MI	5	15	10	2, 5
OO	22	27	5	5
OEQOOG	24	54	30	2, 3, 5
FV	39	63	24	2, 2, 2, 3
AA	43	87	44	2, 2, 11
MOC	50	122	72	2, 2, 2, 3, 3
QO	56	105	49	7, 7
PC	69	117	48	2, 2, 2, 2, 3
NE	77	83	6	2, 3
SV	94	97	3	3
CH	118	124	6	2, 3

ESTIMATE OF PERIOD

- OEQOOG is probably not a coincidence
  - It's too long for that
  - Period may be 1, 2, 3, 5, 6, 10, 15, or 30
- Most others (7/10) have 2 in their factors
- Almost as many (6/10) have 3 in their factors
- Begin with period of $2 \times 3 = 6$

CHECK ON PERIOD

- Index of coincidence is probability that two randomly chosen letters from ciphertext will be the same
- Tabulated for different periods:

1	0.066	3	0.047	5	0.044
2	0.052	4	0.045	10	0.041
Large	0.038				

COMPUTE IC

- $IC = [n\,(n-1)]^{-1} \sum_{0 \le i \le 25} [F_i\,(F_i - 1)]$
  - where n is length of ciphertext and F_i the number of times character i occurs in ciphertext
- Here, IC = 0.043
  - Indicates a key of slightly more than *5*
  - A statistical measure, so it can be in error, but it agrees with the previous estimate (which was 6)

SPLITTING INTO ALPHABETS

alphabet 1: `AIKHOIATTOBGEEERNEOSAI`
alphabet 2: `DUKKEFUAWEMGKWDWSUFWJU`
alphabet 3: `QSTIQBMAMQBWQVLKVTMTMI`
alphabet 4: `YBMZOAFCOOFPHEAXPQEPOX`
alphabet 5: `SOIOOGVICOVCSVASHOGCC`
alphabet 6: `MXBOGKVDIGZINNVVCIJHH`

- ICs (#1, 0.069; #2, 0.078; #3, 0.078; #4, 0.056; #5, 0.124; #6, 0.043) indicate all alphabets have period 1, except #4 and #6; assume statistics off

FREQUENCY EXAMINATION

```
  ABCDEFGHIJKLMNOPQRSTUVWXYZ
1 31004011301001300112000000
2 10022210013010000010404000
3 12000000201140004013021000
4 21102201000010431000000211
5 10500021200000500030020000
6 01110022311012100000030101
```

Letter frequencies are (H high, M medium, L low):

```
  HMMMHMMHHMMMMHHMLHHHMLLLLL
```

BEGIN DECRYPTION

- First matches characteristics of unshifted alphabet
- Third matches if I shifted to A
- Sixth matches if V shifted to A
- Substitute into ciphertext (bold are substitutions)

AD**I**YS **RI**U**K**B O**CK**K**L** MI**GH**K **A**ZO**TO** E**I**OO**L**
IF**T**AG **PA**U**E**F V**AT**A**S** CI**IT**W **E**OC**NO** E**I**OO**L**
BM**T**FV **EG**G**O**P C**NE**K**I** HS**SE**W **N**EC**SE**
D**D**AA**A** **R**W**C**XS **AN**S**N**P H**HE**U**L** QO**NO**F
EEG**OS** W**L**PC**M** **A**J**E**OC **MI**U**A**X

LOOK FOR CLUES

- **A**J**E** in last line suggests "are", meaning second alphabet maps A into S:

  ALIYS **RICK**B O**CKSL** MI**GHS** **A**ZO**TO**

  MIOO**L** **INT**AG **PACE**F V**ATIS** CI**ITE**

  EOC**NO** **MI**OO**L** **BUT**FV **EGOO**P C**NESI**

  HS**SEE** **N**EC**SE** **LD**AA**A** **REC**XS **ANAN**P

  H**HECL** QO**NON** **E**EG**OS** **EL**PC**M** **ARE**OC

  MICAX

NEXT ALPHABET

- **MICA**X in last line suggests "mical" (a common ending for an adjective, like comical), meaning fourth alphabet maps O into A:

```
ALIMS RICKP OCKSL AIGHS ANOTO MICOL
INTOG PACET VATIS QIITE ECCNO MICOL
BUTTV EGOOD CNESI VSSEE NSCSE LDOAA
RECLS ANAND HHECL EONON ESGOS ELDCM
ARECC MICAL
```

GOT IT!

- QI means that U maps into I, as Q is always followed by U:

ALIME RICKP ACKSL AUGHS ANATO
MICAL INTOS PACET HATIS QUITE
ECONO MICAL BUTTH EGOOD ONESI
VESEE NSOSE LDOMA RECLE ANAND
THECL EANON ESSOS ELDOM ARECO
MICAL

THE REAL MESSAGE IS

- A LIMERICK PACKS LAUGHS ANATOMICAL INTO SPACE THAT IS QUITE ECONOMICAL BUT THE GOOD ONES I'VE SEEN SO SELDOM ARE CLEAN, AND THE CLEAN ONES SO SELDOM ARE COMICAL

- Feinberg, Leonard. The Secret of Humor. Rodopi, 1978. ISBN 9789062033706. p102

UNBREAKABLE ENCRYPTION

- One time pads
  - Sender and receiver use a pre-arranged *random* stream of letters
  - Encryption=addition modulo 26
    - XOR when binary
  - Every letter in the key used only once

M	E	S	S	A	G	E
T	H	I	S	K	E	Y
F	L	A	K	K	K	C

- One time pads provide for the only perfectly secure encryption algorithms
  - Why? Look at ciphertext DXQR. Equally likely to correspond to plaintext DOIT (key AJIY) and to plaintext DONT (key AJDY) and any other 4 letters
  - All the rest are only computationally secure
  - Used by Soviet spies, and also for US-USSR hotline.
  - https://en.wikipedia.org/wiki/One-time_pad

EXAMPLE OF ONE-TIME PAD

ciphertext:	ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS
key:	*pxlmvmsydoftyrvzwc tnlebnecvgdupahfzzlmnyih*
plaintext:	mr mustard with the candlestick in the hall

ciphertext:	ANKYODKYUREPFJBYOJDSPLREYIUNOFDOIUERFPLUYTS
key:	*mfugpmiydgaxgoufhklllmhsqdqogtewbqfgyovuhwt*
plaintext:	miss scarlet with the knife in the library

ONE-TIME PAD FUNDAMENTAL DIFFICULTIES

1. There is the practical problem of making large quantities of random keys. Any heavily used system might require millions of random characters on a regular basis. Supplying truly random characters in this volume is a significant task.
2. Even more daunting is the problem of key distribution and protection. For every message to be sent, a key of equal length is needed by both sender and receiver. Thus, a mammoth key distribution problem exists.

- Relies on randomness of key. Otherwise you can attack the cipher by trying to regenerate the key
- Approximations, such as using pseudorandom number generators to generate keys, are not random
- Limited utility: is useful primarily for low-bandwidth channels

SUMMARY OF HISTORICAL CRYPTO

- Encryption Algorithms and Keys
  - Substitution : letters (bits), words
  - Transposition
- Decryption Algorithms
  - Reversed process
  - Require knowledge of the algorithm and the key
- Cryptanalysis
  - Identify algorithm
  - Obtain *as many* plaintext-ciphertext pairs
  - Use systematicity (patterns)
  - Use hints (cribs)

MODERN CRYPTOGRAPHY

RISE OF THE MACHINES

Main source: Network Security Essentials / Stallings

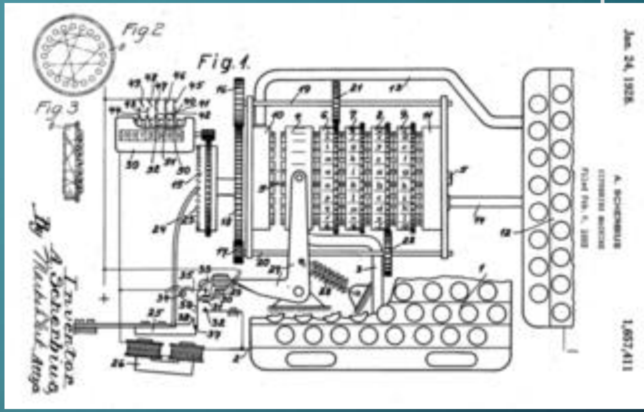
KERCKHOFFS PRINCIPLES: SYSTEM + KEYS

1. The system must be substantially, if not mathematically, undecipherable;
2. The system must not require secrecy and can be stolen by the enemy without causing trouble;
3. It must be easy to communicate and remember the keys without requiring written notes, it must also be easy to change or modify the keys with different participants;
4. The system ought to be compatible with telegraph communication;
5. The system must be portable, and its use must not require more than one person;
6. Finally, regarding the circumstances in which such system is applied, it must be easy to use and must neither require stress of mind nor the knowledge of a long series of rules.

August Kerckhoffs, Journal of Military Science, 1883

THE GERMAN ENIGMA

- Invented as a commercial machine (Scherbius at end of WWI), and failed
  - Electrical typewriter-like encryption machine
  - Each keystroke lights a letter



- Performing substitutions
  - Letter-pairs are switched
  - Pulse goes through scramblers
  - Hits reflector and goes back

- Original Enigma (M3) based on commercial version
  - Reconfigurable 6 swapped letter-pairs
  - 3 rotating scramblers (26^3 orientations)
  - scramblers can be configured in 6 (3!) ways
  - Later, up to 5 scramblers to choose from

- Theoretical key space = a total of 10^{17} combinations
  https://en.wikipedia.org/wiki/Enigma_machine

THE GERMAN ENIGMA

- Before war broke out in 1939 the German army, navy and air force were told to encode their messages using the enigma machine
- The Germans believed that no one could crack the Enigma code. But the Allies knew that if they could, they would be able to find out their enemy's military secrets.
- At least once a day the Germans changed the order of the rotors, their starting positions and the plugboard connections. To decipher a message sent using Enigma, you had to work out exactly how all of these had been set

- Try out an enigma emulator
- http://users.telenet.be/d.rijmenants/en/enigmasim.htm

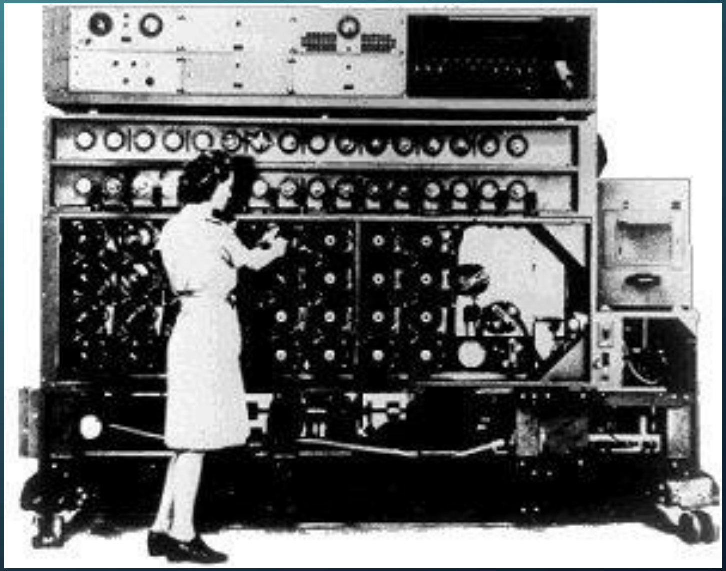
POLES CRACK THE ENIGMA

- Polish obtained an Enigma from a German spy (1933)
  - Hans-Thilo Schmidt sold to French intelligence
- Obtained information on its usage
  - daily code book indicated rotors and orientation
  - a different *orientation* key for each message

- Rejewski focused on the repetitions
  - Message key encrypted twice in the message header
  - Formalized relationships between 1^{st}-4^{th}, 2^{nd}-5^{th}, and 3^{rd}-6^{th} letters
    - ABCDEFGHIJKLMNOPQRSTUVWXYZ
    - FQHPLWOGBMVRXUYCZITNJEASDK
  - Built chains
    - (AFW), (BQZKVELRI), (CHGOYDP), (JMXSTNU)
  - Chains depend only on scrambler orientation, not on pairs swaps
    - Thus need to consider only $3! \times 26^3 = 105456$ configurations
  - Built a catalog of characteristic chains for all configurations

POLES CRACK THE ENIGMA

- Rejewski's algorithm to discover the day key
  - First, use catalog to identify the scrambler setting and orientation
  - Then, run the ciphertext through an Enigma and look at the text to identify swapped letter pairs
    https://en.wikipedia.org/wiki/Cryptanalysis_of_the_Enigma
- Bombe machines were constructed to mechanize the search
- A one-ton machine that essentially stringed together 6 Enigmas (one for each ordering of the scramblers)



BRITISH CRACK IMPROVED ENIGMA

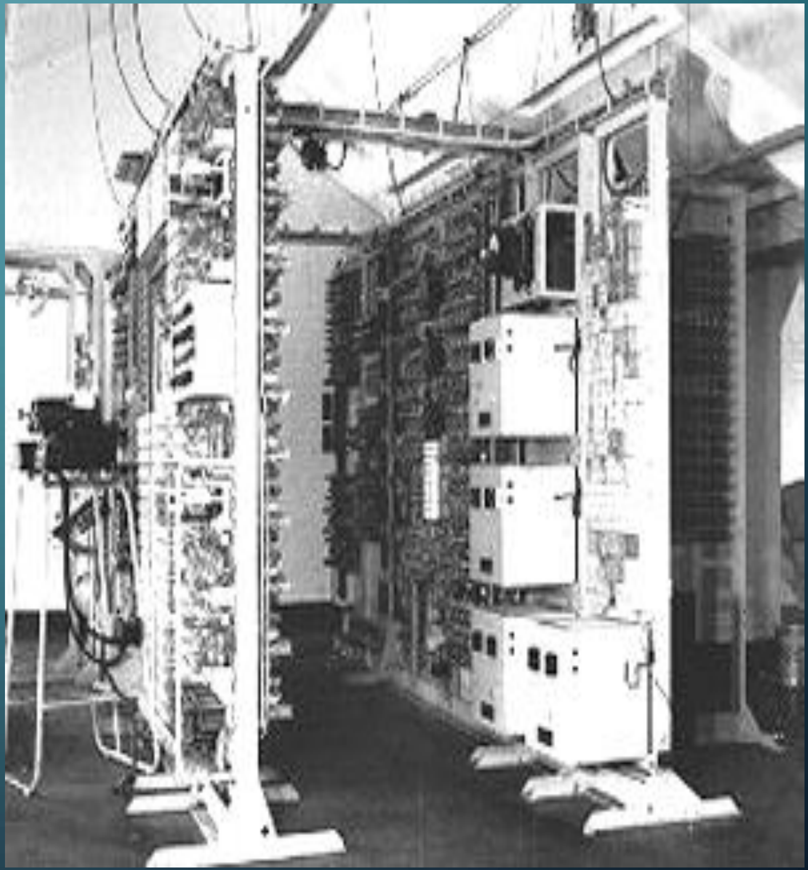
- In 1939, Germans increased Enigma security
  - Navy admiral added 2 extra scramblers to choose from – 10x arrangements (5 choose 3, times the 3! orderings)
  - Hitler used a more complex version – Lorenz Cipher
  - increased to 10 letter pair swaps
- British (Bletchley Park) continued where the Polish left
  - Recruited best Mathematicians (Turing) and large staff (7000)
  - Did not make much progress until received Bombes from Polish
- Used human weaknesses. Provided hints and *cribs*
  - Trivial message keys (key sequences, names initials)
  - Artificial restrictions on scramblers selection/orientation
  - Standard messages (weather) sent with 4^{th} scrambler neutralized
  - Some German codebooks were captured

BRITISH CRACK IMPROVED ENIGMA

- Turing built swap-independent chains (a la Rejewski)
  - First British Bombe (*Victory*) delivered in 1940
  - Search still required significant human help
- In 1942, Germans add 4th active scrambler (M4)
  - Bletchley Park could not decipher M4's messages for 10 months
- Could only break it when info was captured in u-boats
  - Captured machines, rotors, weather manuals, providing cribs
- Later in the war, US Navy also constructed even faster and more sophisticated bombes
  - Japanese used PURPLE, a machine modeled after Enigma
  - Pearl Harbor Attack was broken hours before the attack
- The British ULTRA – broken German, Italian and Japanese communications were crucial to winning the war

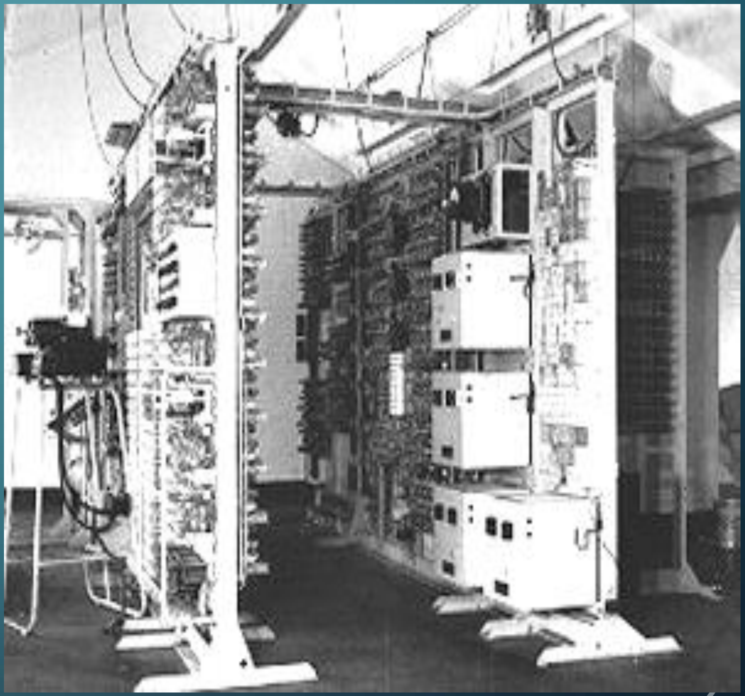
COMPUTER AND CODE BREAKING

- Colossus was built for the code-breakers at Bletchley Park by post office engineers in 1943.
- One of the earliest digital computers.



Computer and Code Breaking

- The computer was as big as a room 5 metres long, 3 metres deep and 2.5 metres high - and was made mainly from parts used for post office telephone and telegraph systems.



SIGABA

- It was suited for fixed station secure communications, and used by U.S. for high-level communications, was the only machine system used by any participant to remain completely unbroken by an enemy during World War II.

B-21 MACHINE BY BORIS HAGELIN

- Patterned on the Enigma and produced for the Swedish General staff, Boris Hagelin of Sweden developed the B-21 machine in 1925. It also had the capability to be connected to an electric typewriter.

BC-38 BY CRYPTO AG ZUG

- Boris Hagelin of Sweden developed a long line of cipher systems, beginning with the B-21, B-211, C-35, C-36, C-38 (which later became America's M-209).

BID 590 (NOREEN)

- The BID 590 was a British built crypto machine and was used by Canada's foreign service communicators at various diplomatic missions to communicate with various government departments.

H-4605 (CRYPTO AG)

- The Crypto AG H4605 was designed as an off-line, keyboard operated cipher machine with twin printing (of cipher and plain text) system with automatic 5-letter grouping. It's a solid piece of equipment, almost 'battleship grade'.

JAPANESE "ENIGMA" ROTOR CIPHER MACHINE

Produced by Germans for Japanese, code name “green” by Americans

Japanese Purple machine

- Electromechanical stepping switch machine modeled after enigma https://en.wikipedia.org/wiki/Type_B_Cipher_Machine
- Used telephone stepping switches instead of rotots
- Purple was broken with the help of MAGIC
- Pear Harbor attack preparations encoded in Purple, decoded hours before attack

KY-28 (NESTOR)

- The KY-28 was an analog, voice encryption device based on transistor circuitry and was the shipboard/airborne member of the NESTOR family of equipment.

RACAL-MILGO 64-1027C DATACRYPTOR

- The Racal-Milgo 64-1027C Datacryptor was used to send and receive secure data via computer. This is the commercial version of the KG-84, and has ability to be loaded via the KYK-13 Fill device.

THE “CLOCK CRYPTOGRAPH”

- It is basically a nicely implemented Wheatstone cipher disk. It was in active use in the Danish armed forces from 1934 (or a little earlier) until around 1948.

COMPUTER AND CODE BREAKING

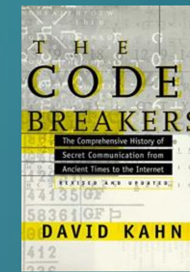
- This Cray XMP was donated to the museum by Cray Research, Inc. It denotes the newest era of partnership between NSA and the American computer industry in the employment of computers for cryptologic processes.

NEVER-ENDING CIPHERS

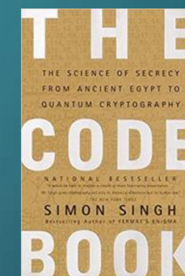
- Many more historical machines than time permits, but visit http://users.telenet.be/d.rijmenants/en/timeline.htm
- And many other classical ciphers, continue your studies at: http://practicalcryptography.com/ciphers/classical-era/
- http://williamstallings.com/Extras/Security-Notes/lectures/classical.html
- For a good series of video, look at the Khan academy course:
- https://www.khanacademy.org/computing/computer-science/cryptography
- Ready to jump to the deep end of the pool, take the free "mooc" course:
- https://www.coursera.org/learn/crypto

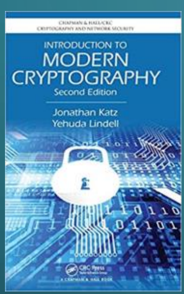
FURTHER READING

The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet / Kahn

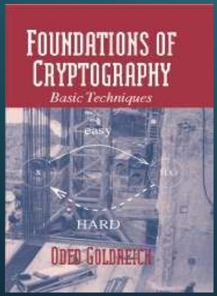


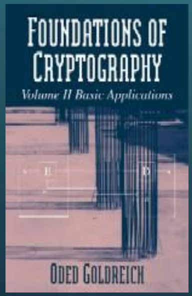
The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography/Singh





Introduction to Modern Cryptography / Katz & Lindell





Foundations of Cryptography / Goldreich.
Graduate-level text.