COS 433/Math 473: Cryptography

Mark Zhandry
Princeton University
Fall 2020

What is Cryptography?

What is Cryptography

Concise Oxford English Dictionary: "the art of writing or solving codes"

Merriam-Webster: "the enciphering and deciphering of messages in secret code or cipher"

Wikipedia: "the practice and study of techniques for secure communication in the presence of third parties called adversaries"

None of these capture the true breadth of the field

My Definition

Cryptography is about using secrets to solve interesting tasks

(still doesn't capture everything)

A Long & Rich History

Dates back almost 4000 years

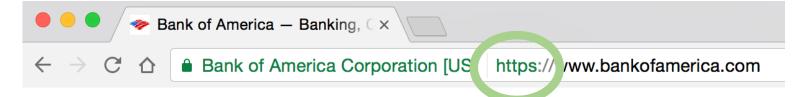
Important historical consequences

- 1587 Babington Plot
- WWI Zimmermann Telegram
- WWII Enigma

Intimately tied to development of modern computer

- First program written for Atlas supercomputer
- First magnetic core memories, high-speed tape drives, all-transistor computers, desktop-sized computers, remote workstations all built based on NSA orders

Cryptography Is Everywhere





Sign in to add another account

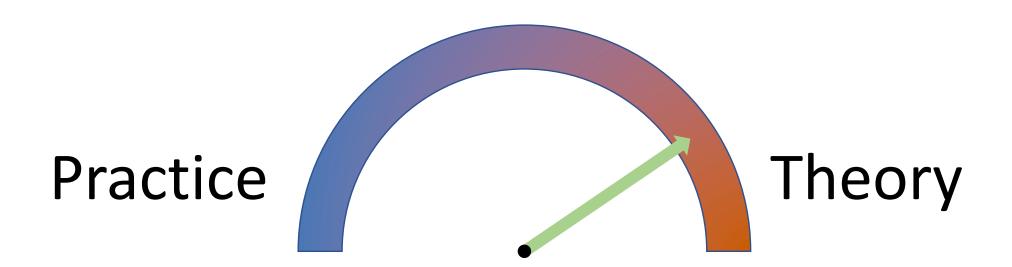








COS 433



Inherent to the study of crypto

- Working knowledge of fundamentals is crucial
- Cannot discern security by experimentation
- Proofs, reductions, probability are necessary

COS 433

What you should expect to learn:

- Foundations and principles of modern cryptography
- Core building blocks
- Applications

Bonus:

- Debunking some Hollywood crypto email me scenes from movies/shows!
- Better understanding of crypto news

COS 433

What you will **not** learn:

- Hacking
- Implementing crypto
- How to design secure systems
- Viruses, worms, buffer overflows, etc

Administrivia

Course Information

Instructor: Mark Zhandry

TAs: Udaya Gai

Anunay Kulshrestha

Lectures: TuTh 11:00-12:20pm, Zoom

Webpage: cs.princeton.edu/~mzhandry/2020-Fall-COS433/

Office Hours: please fill out HW0 poll

Canvas

Main channel of communication

- Course announcements
- Submit assignments

Ed Discussion:

- Discuss homework problems with other students
- Ask content questions to instructors, other students
- Find project teams (≤4 people)

Prerequisites

- Ability to read and write mathematical proofs
- Familiarity with algorithms, analyzing running time, proving correctness, O notation
- Basic probability (random variables, expectation)

Helpful:

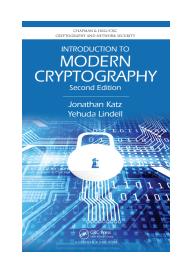
- Familiarity with NP-Completeness, reductions
- Basic number theory (modular arithmetic, etc)

Reading

No required text

But highly recommend:

Introduction to Modern Cryptography by Katz, Lindell



For each lecture, page numbers for 2nd edition will be posted on course website

Grading

40% Homeworks

- ~1 every two weeks (6 total)
- 1 dropped homework
- 2 late days per assignment
- Only typed solutions, submission instructions TBA
- Collaboration encouraged, but write up own solutions

30% Projects

More details soon

30% Take-home Final

Individual

Classroom Policies

Please stop me if you have any questions

(Preferably by "raising hand")

Lectures/slides will be recorded and made available

• I don't take attendance

Feel free to call me "Mark", "Professor", "Hey You", etc, though "Mark" is preferred

Approximate Course Outline

Week 1: Pre-modern crypto (≤ ~1950s)

Weeks 2-6: Foundations of modern cryptography

- Crypto theory
- Symmetric key cryptography

Weeks 7-12: Public key cryptography

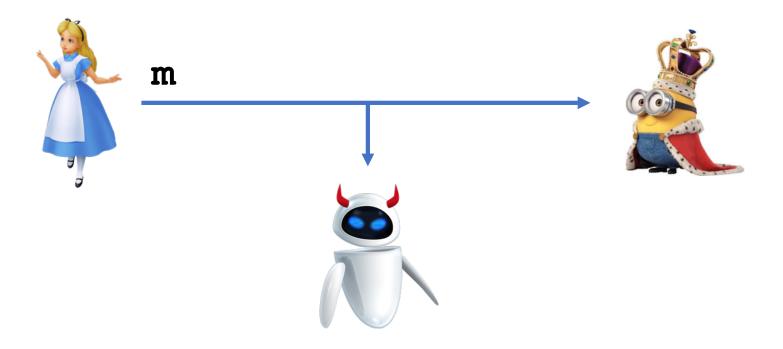
Number theory

Today "Pre-modern" Crypto Part I: Pencil & Paper Ciphers

Pre-modern Cryptography

1900 B.C. – mid 1900's A.D.

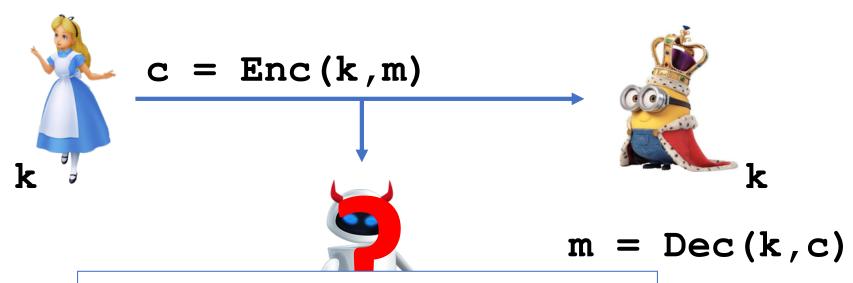
With few exceptions, synonymous with encryption



Pre-modern Cryptography

1900 B.C. – mid 1900's A.D.

With few exceptions, synonymous with encryption



For our discussions, assume **Enc**, **Dec** known, only **k** is secret

Ancient Crypto

1900 BC, Egypt



1500 BC, Mesopotamia



50 B.C. – Caesar Cipher

Used by Julius Caesar

Alphabet shift by 3

```
b c d e f g h i j
                     k | 1
                              op
                         m n
                                        S
                                          t
                                   q
                                     r
                                            u
                                                   X
                                                        Z
       H
         Ι
              K
                L
                  M
                     N
                       0
                          P
                              R
                                 S
                                   T
                                     U
                                          W
                                            X
                            Q
                                        V
```

Example:

plaintext: super secret message

ciphertext: VXSHU VHFUHW PHVVDJH

Caesar not a true cipher: what's the secret key?

Generalization: Shift Ciphers

Shift by fixed, secret increment (k = 0, ..., 25)

Some examples:

- Shift by 1: Augustus Caesar; Jewish mezuzah
- Shift by 3: Caesar Cipher
- Shift by 13: ROT13

Sometimes also called "Caesar ciphers"

Security of Shift Ciphers?

Problem: only 26 possibilities for key

"Brute force" attack:

- Try all 26 possible shifts
- For each shift, see if something sensible comes out

Example Brute Force Attack

Ciphertext: HJETG HTRGTI BTHHPVT

Key	Plaintext
0	HJETG HTRGTI BTHHPVT
1	IKFUH IUSHUJ CUIIQWU
2	JLGVI JVTIVK DVJJRXV
3	KMHWJ KWUJWL EWKKSYW
4	LNIXK LXVKXM FXLLTZX
5	MOJYL MYWLYN GYMMUAY
6	NPKZM NZXMZO HZNNVBZ
7	OQLAN OAYNAP IAOOWCA
8	PRMBO PBZOBQ JBPPXDB
9	QSNCP QCAPCR KCQQYEC
10	RTODO RDBODS LDRRZFD
11	SUPER SECRET MESSAGE
12	TVQFS TFDSFU NFTTBHF

Key	Plaintext
13	UWRGT UGETGV OGUUCIG
14	VXSHU VHFUHW PHVVDJH
15	WYTIV WIGVIX QIWWEKI
16	XZUJW XJHWJY RJXXFLJ
17	YAVKX YKIXKZ SKYYGMK
18	ZBWLY ZLJYLA TLZZHNL
10	ACXMZ AMKZMB UMAAIOM
20	BDYNA BNLANC VNBBJPN
21	CEZOB COMBOD WOCCKQO
22	DFAPC DPNCPE XPDDLRP
23	EGBQD EQODQF YQEEMSQ
24	FHCRE FRPERG ZRFFNTR
25	GIDSF GSQFSH ASGGOUS

Security of Shift Ciphers?

Problem: only 26 possibilities for key

"Brute force" attack:

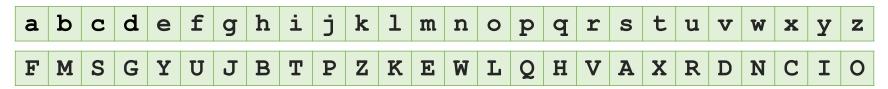
- Try all 26 possible shifts
- For each shift, see if something sensible comes out

To avoid brute force attacks, need large key space

On modern hardware, typically need #(keys) ≥ 2⁸⁰
 (Usually choose at least 2¹²⁸, 2²⁵⁶)

Generalization: Substitution Ciphers

Apply fixed permutation to plaintext letters



Example:

plaintext: super secret message

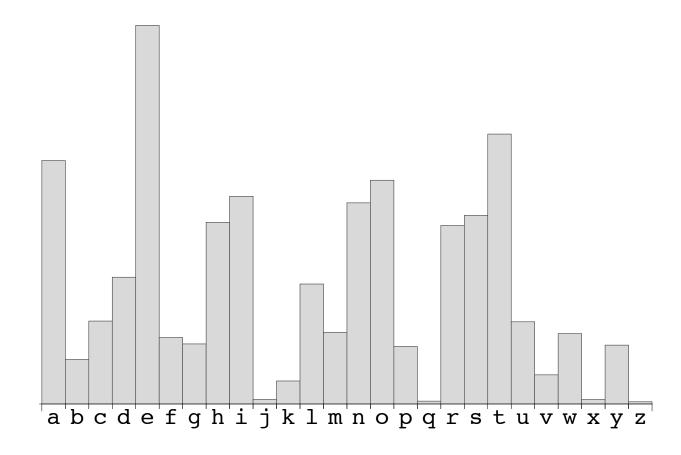
ciphertext: ARQYV AYSVYX EYAAFJY

Number of possible keys?

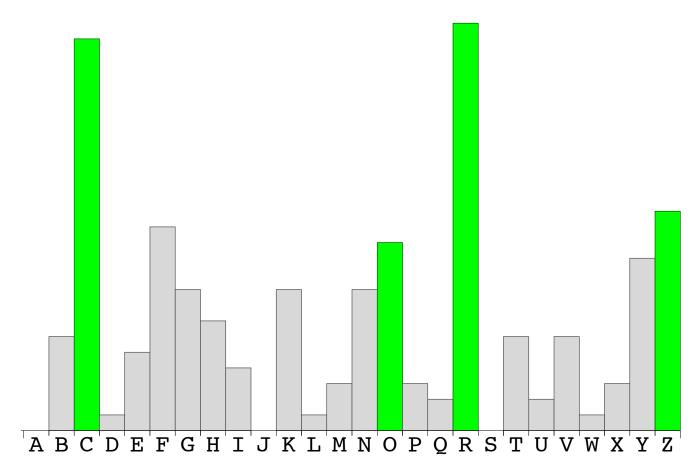
26! $\approx 2^{88}$ \Rightarrow brute force attack expensive

800's A.D. – First Cryptanalysis

Al-Kindi – Frequency Analysis: some characters are more common than others



BOFC HNR Z NHMNCYCHCYOF KYIVRG CO RFKOBR NRFNYCYPR BZCZ, RPRF CVOHXV CVRE ZGR GRNYTYRFC CO Z MGHCR WOGKR ZCCZKU.
YFBRRB, ME KOHFCYFX TRCCRGN ZFB KODIZGYFX CO CEIYKZT CRQC, EOH KZF GRKOPRG CVR ITZYFCRQC ZN LRTT ZN CVR URE



Reasonable conjecture:

 $e \rightarrow R$, $t \rightarrow C$, $a \rightarrow Z$, $o \rightarrow O$

Boft HNe a NHMNtYtHtyof KylveG to efkoBe
NefnytyPe(Bata) ePef tVoHXV tVeE aGe
Genytyeft to a MGHte WoGKe(attaKU)
YFBeeB, ME KoHFtyfX TetteGN afb KoDlaGyfX
to tElykat teQt, EoH Kaf GeKoPeG(tVe)
ITayfteQt an LeTT an tVe UeE

Maybe "data"? Maybe "attack"?

Probably "the"

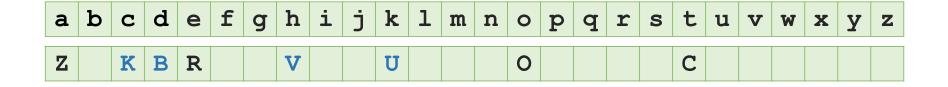
a	b	С	d	е	f	g	h	i	j	k	1	m	n	0	p	q	r	s	t	u	v	W	x	У	Z
Z				R										0					С						

```
doft HNe a NHMNtYtHtyof cylheG to efcode
NefnytyPe data, ePef thoHXh theE aGe
GeNYTYeFt to a MGHte WoGce attack.
Yfdeed, ME coHftyfX TetteGN afd coDIaGYfX
to tElycaT teQt, EoH caf GecoPeG the
ITayfteQt an LeTT an the keE

"encode"?

"as"?

"are"?
```



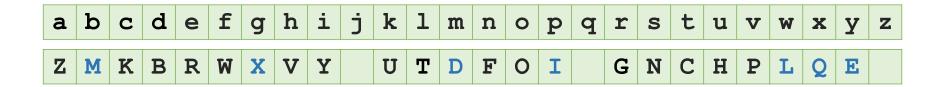
```
"use"?
dont (Hse) a sHMstYtHtYon cYIher to encode
sensYtYPe data, ePen thoHXh theE are
resYTYent to a MrHte Worce attack.
Yndeed ME coHntYnX TetteGs and coDIarYnX
to tEIYcaT teQt / EoH can recoPer the
ITaYnteQt as LeTT as the keE
  "indeed"? "even"?
                             "recover"?
                 "force"?
```

f h i k m n S u r X Z Z R V K B U F 0 N

dont use a suMstitution ciIher to encode sensitive data, even thouXh theE are resiTient to a Mrute force attack. indeed, ME countinX Tetters and coDIarinX to tEIicaT teQt, Eou can recover the ITainteQt as LeTT as the keE

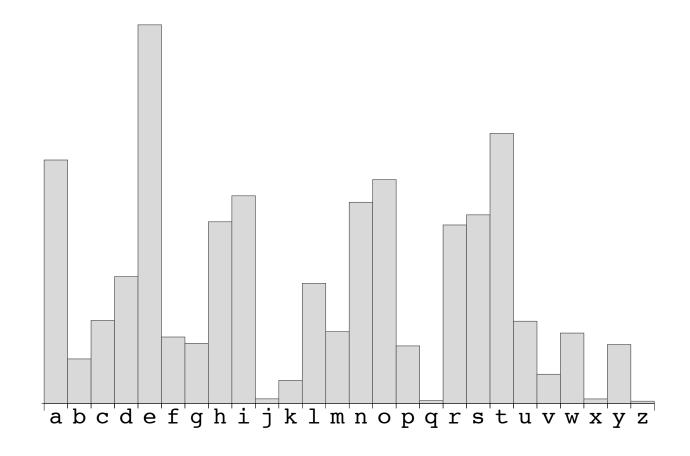
a	b	С	d	е	f	g	h	i	j	k	1	m	n	0	p	q	r	s	t	u	v	W	x	У	Z
Z		K	В	R	W		V	Y		U			F	0			G	N	С	Н	P				

dont use a substitution cipher to encode sensitive data, even though they are resilient to a brute force attack. indeed, by counting letters and comparing to typical text, you can recover the plaintext as well as the key



Problem with Substitution

Differing letter frequencies reveal a lot



Substitution Cipher Variants

Polybius Square

```
1 2 3 4 5
1 a b c d e
2 f g h ij k
3 l m n o p
4 q r s t u
5 v w x y z
```

plaintext: super secret message ciphertext: 4345351542 431513421544 32154343112215

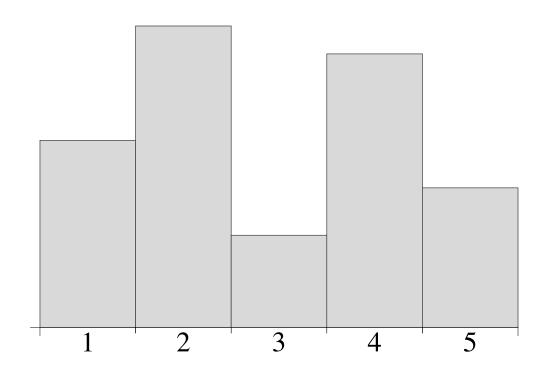
Problem?

Keyed Polybius Square

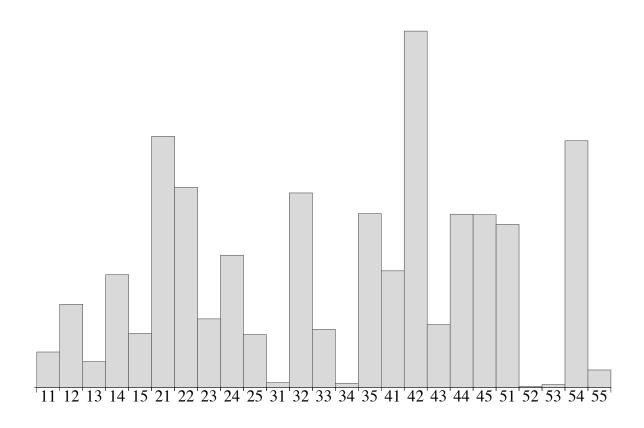
	1	2	3	4	5
1	У	n	r	b	f
2	d	1	W	0	g
3	S	p	a	t	k
4	h	v	ij	x	С
5	q	u	Z	е	m

plaintext: super secret message ciphertext: 3152325413 315445135434 55543131332554

Frequency of Polybius?



Frequency of Polybius?



General Alphabets

Ptxt and ctxt symbols need not be the same

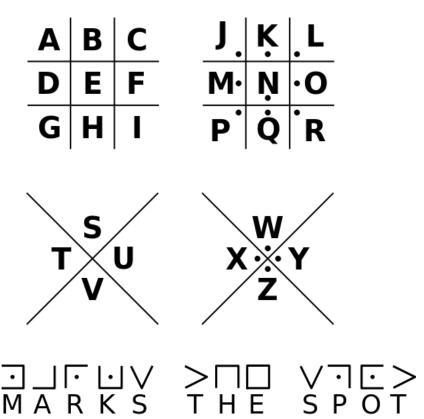
- ctxt symbols can be letters, (tuples of) numbers, etc.
- ptxt symbols can also numbers, bits, bytes

In general, changing ctxt alphabet doesn't affect security of cipher

Keyed Polybius = Un-keyed Polybius + Substitution

Other reasons to change ciphertext alphabet?

Pigpen Cipher



Frequency analysis requires seeing many copies of the same character/group of characters

Idea: encode d=2,3,4, etc characters at a time

- New alphabet size: 26^d
- Symbol frequency decreases:

```
• Most common digram: "th", 3.9% trigram: "the", 3.5%
```

quadrigram: "that", 0.8%

 Require much larger ciphertext to perform frequency analysis

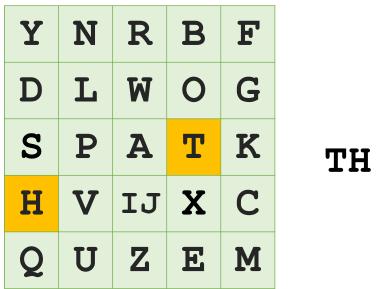
Example: Playfair cipher

- Invented by Sir Charles Wheatstone in 1854
- Used by British until WWII

Y	N	R	В	F
D	L	W	0	G
S	P	A	T	K
Н	V	IJ	X	С
Q	U	Z	E	M

Example: Playfair cipher

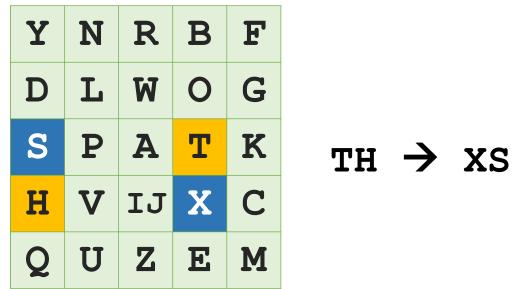
- Invented by Sir Charles Wheatstone in 1854
- Used by British until WWII



To encode, choose opposite corners of rectangle

Example: Playfair cipher

- Invented by Sir Charles Wheatstone in 1854
- Used by British until WWII



- To encode, choose opposite corners of rectangle
- Additional rules for repeats, digrams in same row, etc.

Limitations:

- For small d, frequency analysis still possible given enough ciphertext material
- For large d, need > 26^d bits to write down general substitutions
 - Impractical to use arbitrary permutations for large d
 - Some tricks (like Playfair) possible to reduce key size while minimizing risk of frequency analysis

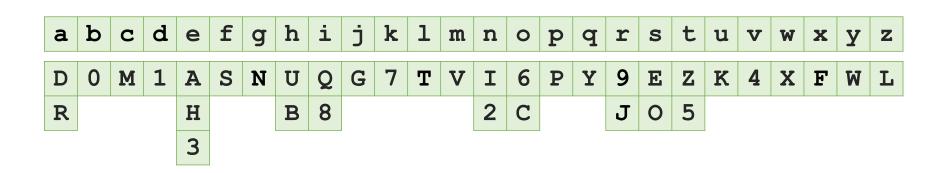
Ciphertexts use a larger alphabet

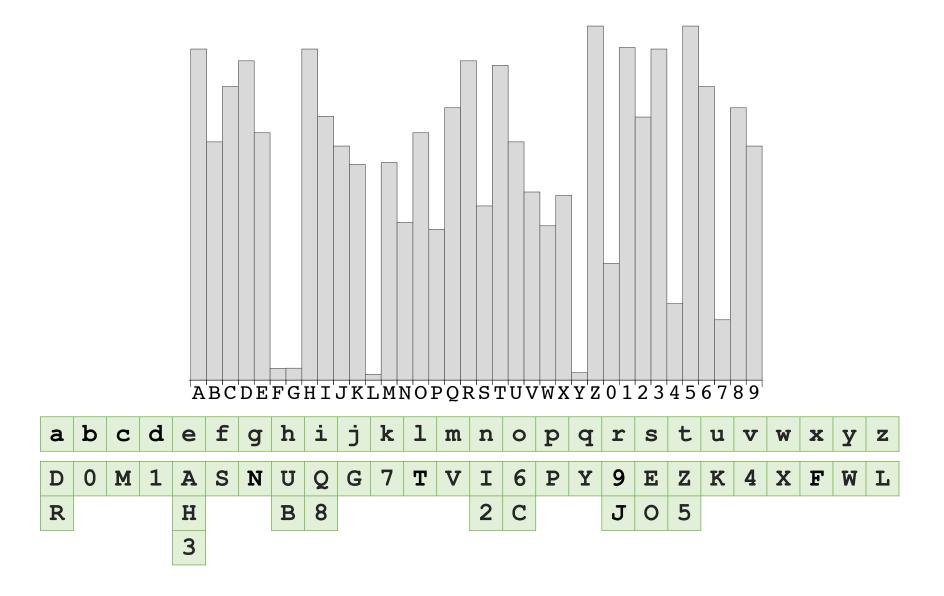
Common letters have multiple encodings

To encrypt, choose encoding at random

plaintext: super secret message

ciphertext: EKPH9 O3MJ3Z VAOEDNH





In principle, by using sufficiently large ciphertext alphabet, character frequencies can be made ≈uniform ⇒ Thwarts vanilla frequency analysis

However, still possible to cryptanalyze

 Frequency analysis on tuples of letters will still be non-uniform

Example: "Grand Chiffre" (Great Cipher)

NO 2	Q.R.S	J 9 26 9 70 6
811 238 219	407 511 355	340 141 205 518 820 448
702 359 338	595 723 527	618 284 436 639 615
genera lua 15	tieu, x 668	06 19 presque 801
gend	limites 708	obei 39 preten dre, tion 30
ges	le Roy de 758	obliger ation . 85 protexte
gla 155	le Zrince, de . 798	
gle 215		observ, er, ation 179 principa Lua . 32
gli 175	le Marquis de . 858	obstacle 179 prisonnier 132 pro
glo, ire		oc, casion 249 prochain 202
gna	le Sieur de	ocup.er 184 profit, er 161
gne 845 435		of 3kg projec 181
gni 485		office, ier, s 439 propos ition . 38
.gno 505	tord 189	offre
gowern, or, ment . 16	luy 848 259	dient Agg prouv 441
gra, a	~11	oir
grand 125		oire
gre 585		oit 6tg puid, sance 5.3
gri		06
gua 695	mo	
gue 285	magasin s	1
querre		
gui. de . S 8 45		
6		ordinaire .s. \$39 quantite 361
Ba 16	mal, ade s.ic.s. bia	ordonnice de la
be 56	mand, et 679	ordro.s. he quates
bo 156	maniere, s 7.19	or of the total and
bu		130 quel le c
baut		out, 160 querhon c. 14
babi, t, le, tane . 486		viere
	The state of the s	ourr 240 ou'll 34
beur, e, s 650		(TP quinze 153
himor	meilleur 8.79	2a 170 quo, n . 140 153

Example: "Grand Chiffre" (Great Cipher)

- Developed in 1600's, used by Louis XIV
- Remained unbroken for 200 years
- Combination of polygraphic and homophonic
- 1890's finally cracked by Étienne Bazeries
 - Guessed that "124-22-125-46-345" stood for "les ennemies"
 - From there, things unraveled

Example: Copiale cipher

oglinamoranvzilgjimis/érzneengyra=rzp+lonkmis/ luak · uraspit-fryanche On flyart pakt puca fujogowbooija zufgirzinlingris lobing i Coather listent semipen luippi , pish finguing rood malliquit tagan jortati egester , zvindergangunistrud xmasgespeggue Gradura nemi nei Suntagiczniej Santjinulophywniowych fetyno zyz x helzin seglafnist mzuvnimzynjaurzhwa avenip · 6+7 8c. crouszilas falsygopoeto Kent YEINDY+per lufs nzato vynuh Vi asktil mijhpucahpengthadicken pyrgrouf dem: sairtefelim lezmuna lakapahaharzelfiniheta Ni = ijabr sellengenerig sentengin ur +prosestign maina by fraginty for umbaghzungzadrerdezajg 000 minfinagrestiasthirzhi= Szut Sajminizgze djy stroce prikejenbelia grogh-faluz quo vrudeahuyrupmulinzny · piùngzió jútchtit kritic ky potraphy z nine » *แบร์ง่ะกฤตันกับผิงผิกแปะกฎการสุดมสารากับกักรอยู่แปะกับ organiorkers mugnintoubdai figefr alu perice az nepona · xinxicapay odenhazismi ojme. องค์อการกลังสุราธารรมหาในอาการกระบางสุราธาราชา 24 rurioy diorza bellafithed ovzha: wrurzy icie simlenti suforthis existinguinamt roothonap que czgwichjernintzy offipozrzbahyzutlaernyjutaniu cydier mulyzycke polisy nezofferendycjąza Inpotore set talgitti=mi tizilerinah lojunoswafung la MILTIMAGENICAMI. Applijy Morat+ 3 Verszásfry indisuzá Alzupokozon Philadeli smanuzich Veringendrung 12 AUVZit+TAFlacifr: IAOSabiz w+IAXZUbritorAuf: ง พระมีผู้เกี่ยง คนในนักมาของ เอกละ เกาะ เลือกเลย เรายนอยกลั Horas Spines fragencla dinotic participation of zlifpriges poedy liablion mich ugarzatienne Vomalio . + Hickori mackes-getoutroj fentenum Oo-walyist

Example: Copiale cipher

- 105-page encrypted book written in 1730's
- Secret society of German ophthalmologists
 - Believed to be Freemasons whose rites had been banned by the pope
- Not broken until 2011 with help of computers

Polyalphabetic Substitution

Use a different substitution for each position

Example: Vigenère cipher

Sequence of shift ciphers defined by keyword

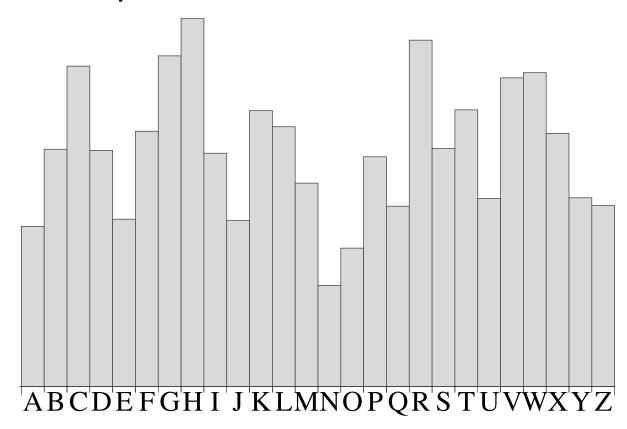
keyword: crypt ocrypt ocrypto

plaintext: super secret message

ciphertext: ULNTK GGTPTM AGJQPZS

Polyalphabetic Substitution

Vanilla frequency analysis gives average of several substitution ciphers



Cryptanalysis of Vigenère

Suppose we know keyword length

- Group letters into n buckets, each bucket encrypted using the same shift
- Perform frequency analysis on each bucket

Suppose we don't know keyword length

- Brute force: try several lengths until we get the right one
- Improvement: superposition

Compare shifts of ciphertext, looking for shifts containing many matches

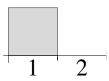
Example: shift by 1

CTYCGS TYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG

CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG

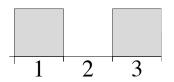
Compare shifts of ciphertext, looking for shifts containing many matches

Example: shift by 2
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG



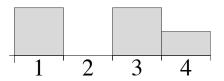
Compare shifts of ciphertext, looking for shifts containing many matches

Example:shift by 3
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG



Compare shifts of ciphertext, looking for shifts containing many matches

Example: shift by 4
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG

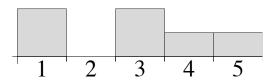


Compare shifts of ciphertext, looking for shifts containing many matches

Example: shift by 5

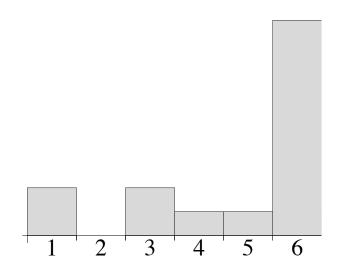
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG

CTY CGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG



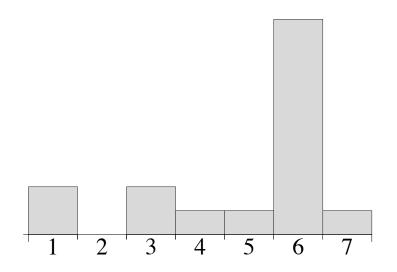
Compare shifts of ciphertext, looking for shifts containing many matches

Example: shift by 6
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGJAPGQCKAPGG
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGJAPGQCEAPGG



Compare shifts of ciphertext, looking for shifts containing many matches

Example: shift by 7
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGQCEAPGG
CTYCGSTTYCVOPRQBTBATYCLOURAPGBGIAPGCCEAPGG



Why does it work?

For shifts that are multiplies of key size:

- Both bottom and top ciphertexts encrypted with same key
- ++(ctxt matches) = ++(ptxt matches)≈ |ptxt| * col. prob. for English≈ |ptxt| * 0.065

Why does it work?

For shifts that are NOT multiplies of key size:

- Both bottom and top ciphertexts encrypted with "independent" shifts
- Probability of a match at any position is 1/26 ≈ 0.038
- #(ctxt matches) ≈ |ptxt| * 0.038

The One-Time Pad

Vigenère on steroids

- Every character gets independent substitution
- Only use key to encrypt one message,
 key length ≥ message length

keyword: agule melpqw gnspemr

plaintext: super secret message

ciphertext: SAIPV EINGUP SRKHESR

No substitution used more than once, so frequency analysis is impossible

The One-Time Pad

1882: described by Frank Miller for the telegraph

- Words and phrases first converted to 5-digit numbers using a codebook
- Key = sequence of "shift-numbers" to be added to resulting digits

1919: Patent for Vernam cipher

- Map characters to 5-bit strings using Baudot code
- Bitwise XOR with key = random bit string

Limitations of One-time Pad

Need extremely large random keys and secure way to transmit them!

5-UCO British OTP system (WWII)

 Key tape for single unit cost £5,000 a year (~\$300k in 2020 dollars)

German GEE (WWII)

Key's not truly random, cryptanalyzed by US Army

Russian diplomatic OTP (WWII, Cold War)

 Tapes occasionally re-used, successful cryptanalysis by US and UK intelligence

Cryptanalysis of OTP

Try to encrypt two messages, security will fail

Enc(k,
$$m_0$$
) - Enc(k, m_1)
= (k + m_0) - (k + m_1)
= m_0 - m_1

Enough redundancy in English text to usually recover messages from difference

Transposition Ciphers

Transposition Ciphers

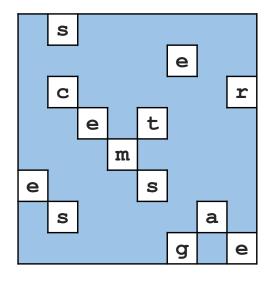
Shuffle plaintext characters

Greek Scytal (600's B.C.)



https://commons.wikimedia.org/wiki/File:Skytale.png

Grille (1500's A.D.)



a	Ø	h	0	e	v	q	k
g	i	р	U	Ø	Ø	£	j
е	С	n	i	d	Z	W	r
g	i	ø	b	t	e	b	0
k	С	d	m	i	Z	d	р
е	b	i	d	S	h	ø	r
n	s	d	u	r	е	a	v
h	k	Ø	g	u	g	a	e

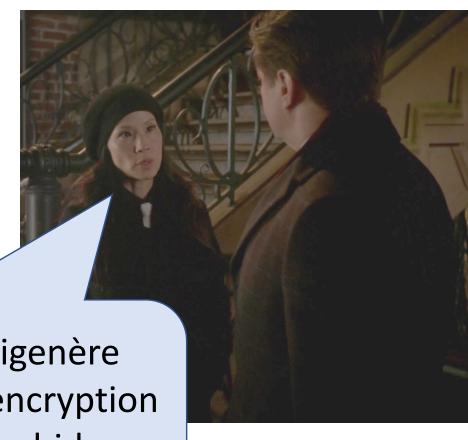
Aside: Steganography

Hiding the fact that a message is even being sent

Many examples

- Invisible ink
- Microdots
- Blinking Morse-code
- Images in low-order color bits
- Delays in network packets
- Differing typefaces

Holiwudd Criptoe!



Do you know what a Vigenère cipher is? It's a form of encryption that allows a person to hide messages inside regular texts.

Column Transposition

key: crypto

ptxt: supersecremessage

Encryption:

С	r	У	p	t	0	Sort by first row	C	0	p	r	t	У
S	u	р	Ф	r	Ŋ		S	S	Ф	u	r	р
е	С	r	е	t	m		е	m	е	С	t	r
е	S	S	а	g	Φ		е	е	а	S	g	S

ctxt: SEESMEEEAUCSRTGPRS (read off columns)

Cryptanalysis:

- Guess key length, reconstruct table
- Look for anagrams in the rows

Double Column Transposition

key: graphy

ctxt0: SEESMEEEAUCSRTGPRS

Encryption:

g	r	a	p	h	У	Sort by first row	a	g	h	p	r	У
S	е	Ф	S	m	M		Ф	S	m	S	Ф	е
е	е	а	u	С	S		а	е	С	u	е	S
r	t	g	р	r	S		g	r	r	р	t	S

ctxt: **EAGSERMCRSUPEETESS**

Example: Germany, WWI

 French were able to decrypt after seeing several messages of the same length

Bifid Cipher

Polybius square + Transposition + Inverse Polybius

	1	2	3	4	5
1	У	n	r	b	f
2	d	1	W	0	g
3	s	p	a	t	k
4	h	v	ij	x	С
5	q	u	Z	е	m

plaintext: super secret message

Polybius: 35351 354153 5533325

12243 145344 5411354

Transpose: 353513541535533325122431453445411354

Inv.Polybius:kkrefkzagnosctchre

Bifid Cipher

Polybius square + Transposition + Inverse Polybius

Invented in 1901 by Felix Delastelle

Each ctxt character depends on two ptxt characters

Still possible to break using frequency analysis

Next Time

"Pre-modern" Crypto Part II: Enter technology

Reminders

By Thursday September 3rd:

HW0: Fill out OH poll

Homework 1, Project 1 to be released soon

Start looking for project teams (≤4)

Send me Hollywood Crypto examples!