Historical Cryptography

CS 458: Information Security

Kevin Jin

Reading

- Chapter 2 and Chapter 20
- Applied Cryptography, Bruce Schneier (optional)
- *Handbook of Applied Cryptography* (optional) *http://www.cacr.math.uwaterloo.ca/hac/*

Overview

- Classical Cryptography
  - Substitution Ciphers
    - Cæsar cipher
    - Vigènere cipher
    - One Time Pad
    - Book cipher
  - Transposition Ciphers

Cryptosystem components

- Plaintext (p) – original message
- Ciphertext (c) – encrypted message
- Key (k) – private information
- Encryption algorithm – $c = E(p,k)$
- Decryption algorithm – $p = D(c,k)$

Attacks

- Opponent whose goal is to break cryptosystem is the *adversary*
  - Standard cryptographic practice: Assume adversary knows algorithm used, but not the key
- Three types of attacks:
  - **ciphertext only**: adversary has only ciphertext; goal is to find plaintext, possibly key
  - **known plaintext:** adversary has ciphertext, corresponding plaintext; goal is to find key
  - **chosen plaintext:** adversary may supply plaintexts and obtain corresponding ciphertext; goal is to find key

Basis for Attacks

- Mathematical attacks
  - Based on analysis of underlying mathematics
- Statistical attacks
  - Natural language contains particular distribution of letters, pairs of letters (digrams), triplets of letters (trigrams), *etc.*
    - Called models of the language
    - In English, ‘e’ appears the most frequently (65 times more frequently than the least frequent ‘z’ and ‘q’)
  - Encryption may not fully destroy the distribution, so observe the ciphertext for related properties

Classical Cryptography

- Sender and receiver share **common** key
  - Keys may be the same, or trivial to derive from one another
  - Sometimes called *symmetric cryptography*
- Two basic types
  - Transposition ciphers
  - Substitution ciphers
  - Combinations are called *product ciphers*

Transposition Cipher

- Rearrange letters in plaintext to produce ciphertext
- Example (Rail-Fence Cipher or 2-columnar transposition)
  - Plaintext is `HELLO WORLD`
  - Write the plaintext on alternating "rails"
  - ```
    H E
    L L
    O W
    O R
    L D
    ```
  - Ciphertext is `HLOOL ELWRD`
    ```

Transposition Cipher

- Generalize to n-columnar transpositions
- Example 3-columnar
  - HEL
LOW
ORL
DXX
  - HLODEORXLWLX

How to attack the cipher?

- 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

- Ciphertext: `HLOOLELWRD`
- Frequencies of 2-grams beginning with H
  - HE 0.0305 ← the winner!
  - 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

HLOOL ELWRD          ⇨            OW

                                  OR

                                  LD
```

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

Substitution Ciphers

- Change characters in plaintext to produce ciphertext
- Example (Cæsar cipher)
  - Plaintext is `HELLO WORLD`
  - Change each letter to the third letter following it (X goes to A, Y to B, Z to C)
    - Key is 3, usually written as letter ‘D’
  - Ciphertext is `KHOOR ZRUOG`
  - Mono-alphabetic substitution

`KHOOR ZRUOG`

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

Cæsar cipher

- $M = \{$ sequences of letters $\}$
- $K = \{ i \mid i$ is an integer and $0 \leq i \leq 25 \}$
- $E = \{ E \mid k \in K$ and for all letters m,

  $$E(m, k) = (m + k) \bmod 26 \}$$
- $D = \{ D \mid k \in K$ and for all letters c,

  $$D(c, k) = (26 + c - k) \bmod 26 \}$$

M – plain text; K – key; E – encryption function;
D – decryption function

How to attack the cipher?

- Exhaustive search
  - If the key space is small enough, try all possible keys until you find the right one
  - Caesar cipher has 26 possible keys
- Statistical analysis
  - The right key should let decrypted message match the 1-gram model of English
  - CryptoQuote techniques

Statistical Attack

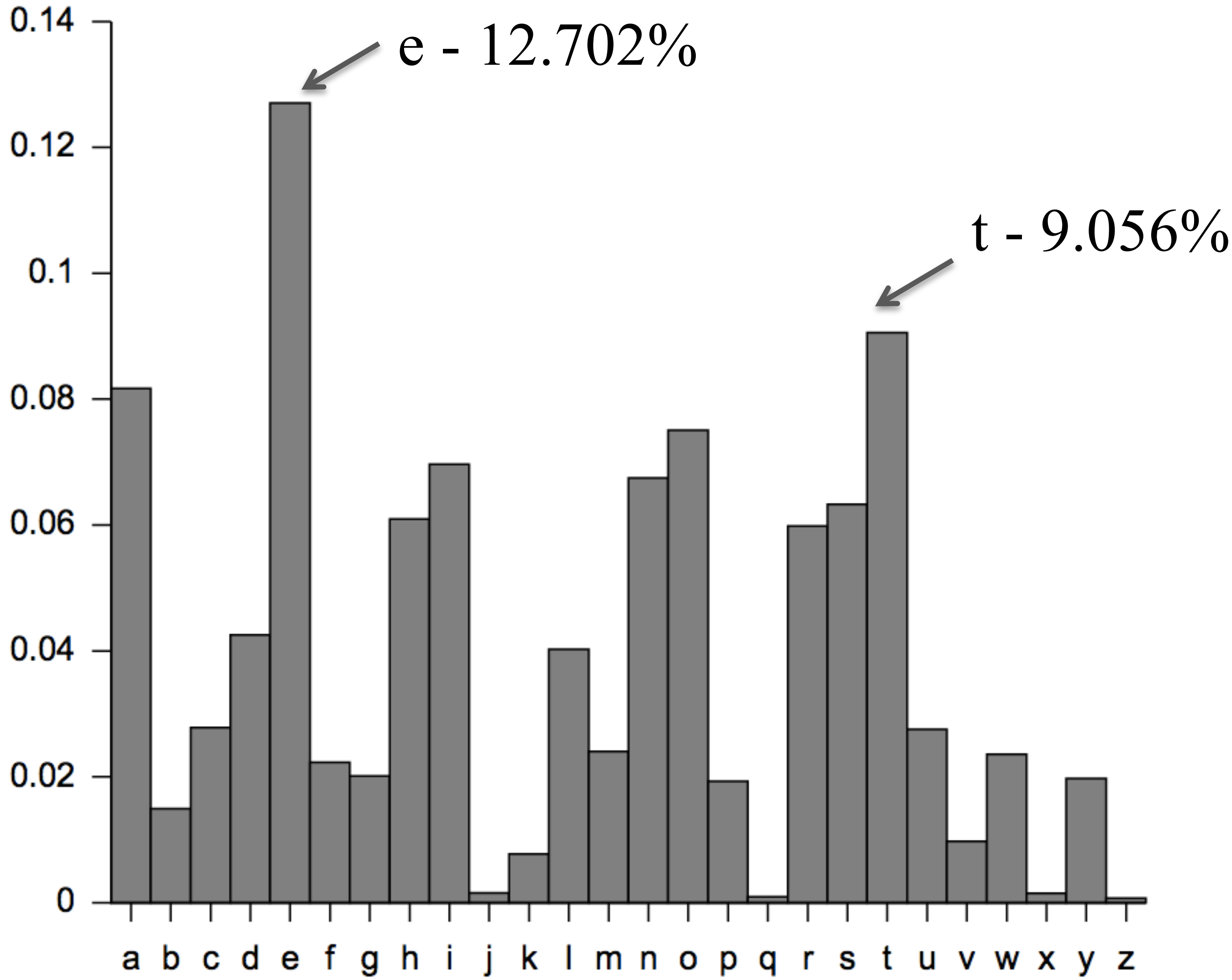
- 1-grams of the ciphertext `KHOOR ZRUOG`

  G 0.1 H 0.1 K 0.1 O 0.3

  R 0.2 U 0.1 Z 0.1

- Apply 1-gram model of English
  - Letter frequencies http://math.ucsd.edu/~crypto/java/EARLYCIPHERS/Vigenere.html

Character Frequencies



Source: http://en.wikipedia.org/wiki/Letter_frequency#Relative_frequencies_of_letters_in_the_English_language

Statistical Attack on Caesar Cipher

- 1-grams of the ciphertext `KHOOR ZRUOG`

  g 0.1 h 0.1 k 0.1 o 0.3

  r 0.2 u 0.1 z 0.1

- 1-grams of English probability *p*:

a	0.080	h	0.060	n	0.070	t	0.090
b	0.015	i	0.065	o	0.080	u	0.030
c	0.030	j	0.005	p	0.020	v	0.010
d	0.040	k	0.005	q	0.002	w	0.015
e	0.130	l	0.035	r	0.065	x	0.005
f	0.020	m	0.030	s	0.060	y	0.020
g	0.015					z	0.002

- Question: how to choose the right key such that the two 1-grams match the best?

Statistical Analysis

- $\varphi(i) = \Sigma_{0 \leq c \leq 25} f(c)\, p(c - i)$
  - assuming key is i
  - $\varphi(i)$ correlation of frequency of letters in ciphertext with corresponding letters in English
  - $f(c)$ frequency of character c in ciphertext
  - $p(x)$ frequency of character x in English

$$\begin{aligned}\varphi(i) = {} & 0.1p(6 - i) + 0.1p(7 - i) + 0.1p(10 - i) \\ & + 0.3p(14 - i) + 0.2p(17 - i) + 0.1p(20 - i) \\ & + 0.1p(25 - i)\end{aligned}$$

Correlation of Frequency

'Match the best' means …

The right key $0 \le i \le 25$ should maximize

$$\varphi(i) = 0.1 \bullet p\,(6 - i) + 0.1 \bullet p\,(7 - i) + 0.1 \bullet p\,(10 - i) + 0.3 \bullet p\,(14 - i) + 0.2 \bullet p\,(17 - i) + 0.1 \bullet p\,(20 - i) + 0.1 \bullet p\,(25 - i)$$

$\varphi(k)$ is maximum iff the two sides match (having similar relative percentage)

frequency of 'o' in ciphertext $\approx$ frequency of decrypted 'o' in english

$$0.3 \bullet \overbrace{p\,(14 - i)}$$

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

Correlation: φ(i) for 0 ≤ i ≤ 25

i	φ(*i*)	*i*	φ(*i*)	*i*	φ(*i*)	*i*	φ(*i*)
0	0.0482	7	0.0442	13	0.0520	19	0.0315
1	0.0364	8	0.0202	**14**	**0.0535**	20	0.0302
2	0.0410	9	0.0267	15	0.0226	21	0.0517
3	**0.0575**	**10**	**0.0635**	16	0.0322	22	0.0380
4	0.0252	11	0.0262	17	0.0392	23	0.0370
5	0.0190	12	0.0325	18	0.0299	24	0.0316
6	**0.0660**					25	0.0430

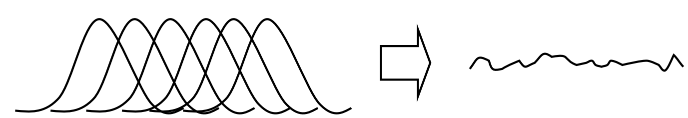
The Results

- Most probable keys, based on φ:
  - $i = 6$, $\varphi(i) = 0.0660$, plaintext EBIIL TLOLA
  - $i = 10$, $\varphi(i) = 0.0635$, plaintext AXEEH PHKEW

  - $i = 3$, $\varphi(i) = 0.0575$, plaintext HELLO WORLD
  - $i = 14$, $\varphi(i) = 0.0535$, plaintext WTAAD LDGAS
- Only English phrase is for $i = 3$
  - That's the key (3 or 'D')
  - Why ranked #3?

What is the problem of Cæsar cipher?

- Key is too short
  - Can be found by exhaustive search
  - Statistical frequencies not concealed well
    - They look too much like regular English letters
- 1-grams are not changed (only shifted)
- So make key longer
  - Use a sequence as key: k_1 k_2 k_3… k_n(key space 26^n)
- Conceal statistical frequencies through diffusion
  - Use k_i to encrypt the i^{th} letter of plaintext
  - Statistical patterns average out



Key the Mapping

- Caesar mapping (shift 3)
  - ABCEDFGHIJKLMNOPQRSTUVWXYZ
  - XYZABCEDFGHIJKLMNOPQRSTUVW
- Key mapping
  - ABCEDFGHIJKLMNOPQRSTUVWXYZ
  - SECURABDFGHIJKLMNOPQTVWXYZ
- Poor mapping at the end
- Still only one mapping of a character across whole message
  - Just a crypto quote

Vigènere Cipher

- Like Cæsar cipher, but use a phrase as key
- Example
  - Message `THE BOY HAS THE BALL`
  - Key `VIG`
  - Encipher using Cæsar cipher for each letter:

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

	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

Relevant Parts of Tableau

plaintext \ key	*G*	*I*	*V*
A	G	I	V
B	H	J	W
E	K	M	Z
H	N	P	C
L	R	T	G
O	U	W	J
S	Y	A	N
T	Z	B	O
Y	E	H	T

- Tableau shown has relevant rows, columns only
- Example
  - key V, letter T: follow V column down to T row (giving "O")
  - key I, letter H: follow I column down to H row (giving "P")

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

Useful Terms

- **period:** length of key
  - In earlier example, period is 3
- **tableau:** table used to encipher and decipher
  - Vigènere cipher has key letters on top, plaintext letters on the left
- **polyalphabetic:** the key has several different letters
  - Cæsar cipher is mono-alphabetic

Attacking the Cipher

- Approach
  - Establish period; call it *n*
  - Break message into *n* parts, each part being enciphered using the same key letter
  - Solve each part
- We will show each step
- Automated in applet
  - http://math.ucsd.edu/~crypto/java/EARLYCIPHERS/Vigenere.html

The Target 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 x 3 = 6

Index of Coincidence

- Index of coincidence (IC)
  - The probability that any two randomly chosen letters from ciphertext are the same
- A measure of variation in frequencies of letters
  - IC of aaaaaaabc (> or <) IC of aabcdefghi ?
- This variation depends on the period of key
  - Longer key tends to average out statistical patterns that exist in English (and thus in plaintext)

Index of Coincidence

- Tabulated for different periods (Known results)

Period	Index of coincidence
1	0.066
2	0.052
3	0.047
4	0.045
5	0.044
10	0.041
Large	0.038

Compute Index of Coincidence

- IC $= \Sigma_{0 \le i \le 25} \dfrac{[F_i\,(F_i - 1)]}{n\,(n-1)}$
  - n is length of ciphertext
  - F_i the number of times character i occurs in ciphertext
  - e.g., letter A appears 3 times in the cipher text of length n: 3/n * 2/(n-1)
- Here, IC = 0.043
  - Indicates a key of slightly more than 5
  - This is a statistical measure, so it can be an error, but it agrees with the previous estimate (which was 6)

The Target 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

Splitting Into Alphabets

alphabet 1: `AIKHOIATTOBGEEERNEOSAI` IC 0.069
alphabet 2: `DUKKEFUAWEMGKWDWSUFWJU` IC 0.078
alphabet 3: `QSTIQBMAMQBWQVLKVTMTMI` IC 0.078
alphabet 4: `YBMZOAFCOOFPHEAXPQEPOX` IC 0.056
alphabet 5: `SOIOOGVICOVCSVASHOGCC` IC 0.124
alphabet 6: `MXBOGKVDIGZINNVVCIJHH` IC 0.043

- 1,2,3,5 indicate period 1
- 4 and 6 don't (well, statistics)
- Step 2 done; now we are dealing with **6 Caesar ciphers**!

```
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
```

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** **I**F**T**AG **PA**U**E**F V**AT**A**S** CI**IT**W

EOC**NO** E**I**OO**L** **B**M**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 **E**EG**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 I into A:

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), meaning fourth alphabet maps O into A:

ALIMS **RICKP** O**CKSL** **A**I**GHS** **AN**O**TO**

MICO**L** **INTO**G **PACET** V**ATIS** **Q**I **ITE**

ECC**NO** **MIC**O**L** **BUTT**V **EGOOD** C**NESI**

VS**SEE** **NS**C**SE** **LDO**A**A** **RECL**S **ANAND**

H**HECL** **E**O**NON** **ES**G**OS** **ELD**C**M** **ARE**C**C**

MICAL

Got It !

- QI means that U maps into I, as Q is always followed by U…So we get the key for the fifth alphabet:

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

A LIMERICK PACKS LAUGHS ANATOMICAL INTO
SPACE THAT IS QUITE ECONOMICAL BUT THE
GOOD ONES IVE SEEN SO SELDOM ARE CLEAN
AND THE CLEAN ONES SO SELDOM ARE COMICAL

One-Time Pad

- A Vigenère cipher with a random key at least as long as the message
  - Provably unbreakable
  - 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
  - Warning: keys *must* be random, or you can attack the cipher by trying to regenerate the key
    - Approximations, such as using pseudorandom number generators to generate keys, are *not* random
  - Remember the key must be transmitted via a secure channel

Book Cipher

- Approximate one-time pad with book text
  - Sender and receiver agree on text to pull key from
  - Bible, Koran, Phone Book
- Problem is that book text is not random
  - Combine English with English
  - Can still perform language based statistical analysis

Key Points

- Two basic types of ciphers
  - Transposition ciphers and substitution ciphers
  - Product ciphers combine them
- Caesar cipher uses one key
- Vigenère cipher uses a sequence of keys
- Cryptanalysis
  - Exhaustive search
  - Statistical analysis