Lecture 6: Classical Encryption Part 2

COSC362 Data and Network Security

Book 1: Chapter 3 – Book 2: Chapter 2

Spring Semester, 2021

Motivation Reminder

Studying historical ciphers in order to:

- Establish basic notation and terminology
- Introduce basic cryptographic operations still used as building blocks for modern cryptographic algorithms
- Explore typical attacks and adversary capabilities that cryptosystems should defend against

Outline

Outline

Defining Polyalphabetic Substitution

- Using multiple mappings from plaintext to ciphertext.
- The effect with multiple alphabets is to smooth frequency distribution:
  - Direct frequency analysis should no longer be effective.
- Typical polyalphabetic ciphers are periodic substitution ciphers based on a period d.
- Given d ciphertext alphabets $C_0, C_1, \cdots, C_{d-1}$, let $f_i : A \rightarrow C_i$ be a mapping from the plaintext alphabet A to the ith ciphertext alphabet C_i for $0 \leq i \leq d-1$.

Encryption Process

A plaintext message

$$M = M_0 \cdots M_{d-1} M_d \cdots M_{2d-1} M_{2d} \cdots$$

is encrypted to

$$E(K, M) = f_0(M_0) \cdots f_{d-1}(M_{d-1}) f_0(M_d) \cdots f_{d-1}(M_{2d-1}) f_0(M_{2d}) \cdots$$

Special case with $d = 1$: the cipher is monoalphabetic (simple substitution cipher)

Random Polyalphabetic Substitution Cipher

- *Key Generation:*
  - Select a block length d
  - Generate d random simple substitution tables
- *Encryption:*
  - Encrypting the ith character by using the substitution table number j such that $i \equiv j \pmod{d}$
- *Decryption:*
  - Using the same substitution table as in encryption in order to reverse the simple substitution

Example

Let $d = 3$, thus there are 3 ciphertext alphabets.

Pltxt char.	ABC	DEF	GHI	JKL	MNO
C_1	UWY	SX∇	TVZ	CEI	AFG
C_2	QLM	PJO	RKN	∇XS	YUW
C_3	MLQ	RNQ	GFA	ZVT	YWU

Pltxt char.	PQR	STU	VWX	YZ∇	
C_1	BDH	KNR	JOP	LMQ	
C_2	ZVT	FGA	HDB	EIC	
C_3	POJ	HDB	IEC	∇XS	

If the plaintext is IT∇IS∇A∇BEAUTIFUL∇DAY then the ciphertext is ZGSZFSUCLXQBNNKRSSSQ∇.

Vigenère Cipher

- Popular form of periodic substitution ciphers based on *shifted* alphabets.
- The key K is a sequence of characters:
  - $K = K_0 K_1 \cdots K_{d-1}$
  - Let M be the plaintext character
  - For $0 \leq i \leq d-1$, K_i gives the amount of shift in the ith alphabet, i.e. $f_i(M) = (M + K_i) \mod n$
  - $n = 27$ when including space in the alphabet
- In the 19th century, it was believed to be unbreakable.

Example

Message M	AT∇T	HE∇T	IME∇
Key K	LOCK	LOCK	LOCK
$E(K,M)$	LGBC	SSBC	T∇GJ

- Numbering the alphabet:
  $\mathrm{A}=0, \mathrm{B}=1, \cdots, \mathrm{Z}=25, \nabla=26$.
- In particular, $\mathrm{L}=11, \mathrm{O}=14, \mathrm{C}=2, \mathrm{K}=10$:
  - the 1st character of each 4-character group is shifted by 11,
  - the 2nd character is shifted by 14,
  - the 3rd character is shifted by 2,
  - the 4th character is shifted by 10.
- Shifting is computed modulo 27 (the alphabet "wraps around").

Cryptanalysis of Vigenère Cipher

- Identify the period length
  Different techniques such as:
  - Kasiski method (illustrated below)
  - Cryptool uses autocorrelation to estimate the period automatically
- Attack separately *d* substitution tables
  Each substitution is just a shift (Caesar cipher):
  - If there is sufficient ciphertext then it is straightforward

Identifying the Period Using Autocorrelation

- Method used to find the period length d of any periodic polyalphabetic cipher.
- Given a ciphertext C, computing the correlation between C and its shift C_i for all plausible values i of the period.
- English is non-random:
  - Better correlation between two texts with the same size shift than between two texts with different size shifts.
- Seeing peaks in the value of C_i when i is a multiple of the period.
- Plotting results on a histogram and then identifying the period.

Example

The first characters of a ciphertext C are:

AUVHSGE**PELPEK**QTEDKSFNYJYATCTCCKFTSUTEFVBVVHPNMFUHBFNPV

YFVRFVUSPEEVHFNAOFLBFYJPFPTMFFMFVHBVHFJAENEGVTIGHPWSFU

HPTTMAAGVESGIHJT**PELPEK**JPTIGMPTNJPGJUAUFOXPBFUIEGTIGFJTEIQ

WFXESYIUJTIGIOVEOVIPPOGCWBKTJPGIKMIQWFXESNOOIHFOIHJTCGIXC

SBNRFCDZFEFRLZKNUGRFUTFFIOJITKNRWISAFPTTIQUHJIUYATUUSTOVP

DFFBZPOOGOGVHFIRJOAOFSUTAOIEGGAUWRFUWIKCIYESGATUODKAUG

DXKTIVHFVWPERJOETYHJEHJJAWGAMTEBFYSGCPTDFFSUKLMVHFPAUW

RFQFUJEDCSFCNEVHFGXBNTFFSUCTJQNPHHJUCMKEOVGBXEJVADJASC

CUGRPHIUUOXPIOFEFFAQCRUHRPOTIGNBVUSGOGVHFKNWGSUKGBVIP

P**WIKCIO**YGTIFPDICDPPHBPDUJESGWBUSPOEUJIOIIOJITOATVESNYHTATR

OGCSJVUBVIPPAOFHJUKFGNJPCJUIWGRFCSPPIOI**WIKCIO**AEGIUCPMGAT

WRFVONGTPUTVFYIKSTASUGMPHWPTKBPDUQFPNLPYTIGQVKCLUUCVL

FOEUJOEUBZYHJEHIGDJUEOVAOILFFTIGMPUTJPEYVRJEACNENASUGRJ

Example Step 1

Identifying the period length *d*:

- Noting that sequences **PELPEK** and **WIKCIO** occur multiple times.
- Positions of some pairs of such strings are separated by 117 and 93 characters.
- Period is almost certain to be 1 or 3:
  - the only common divisors of 117 and 93 are 1 and 3.

This is the Kasiski method. One can also automate the process by plotting the autocorrelation (Cryptool).

Example Step 2

Attacking separately 3 different alphabets:

- Finding the shift for each alphabet as in Caesar cipher.
- Looking for character with the largest frequency, assuming this is shifted from E.
- Turning out that:
  - the first has key A (shift of 0)
  - the second has key B (shift of 1)
  - the third has key C (shift of 2)

The plaintext starts with:
ATTHREEOCLOCKPRECISELYIWASATBAKERSTREET· · ·

Other Ciphers Designed for Use by Hand

- *Beaufort cipher:* similar to Vigenère cipher but using the substitution $f_i(M) = (K_i - M) \mod n$.
- *Autokey cipher:* starting off as Vigenère cipher but using the plaintext to define subsequent alphabets once the alphabets defined by the key have been used.
  - this cipher is NOT periodic
- *Running key cipher:* using a (practically) infinite set of alphabets from a shared key.
  - in practice, the shared key is an extract from a book called *book cipher*

Rotor Machines

- Early 20th century
  Electromechanical machines developed for encryption using *rotors* as moving alphabets.
- World War II
  The famous *Enigma* machine used by the Germans:
  - each character is encrypted using a different alphabet
  - the period is of about 17,000 so in practice it would never repeat the same message
  - nice simulation in Cryptool

Outline

Hill Cipher

- The American mathematician Lester S. Hill published his cipher in 1929.
- *Polygram cipher* (also polygraphic cipher):
  - simple substitution cipher on an extended alphabet consisting of multiple characters
  - *Example:* digram substitution in which the alphabet consists of all pairs of characters
- Major weakness: its linearity, hence known plaintext attacks are easy.

Definition

Performing a linear transformation on d plaintext characters to get d ciphertext characters:

- Encryption involves multiplying a $d \times d$ matrix K by the block of plaintext M.
- Decryption involves multiplying the matrix K^{-1} by the block of ciphertext C.

Encryption: $C = KM$
Decryption: $M = K^{-1}C$

Encryption Example

- Let $d = 2$ so encryption takes digrams as input and outputs blocks.
- Each plaintext pair is written as a column vector, letters are encoded as numbers.
- Suppose the 1st pair for encryption is EG:
  - E $= 4$ and G $= 6$ in our encoding
  - represented as $\binom{4}{6}$
- If insufficient letters to fill a block then padding:
  - it can be done with uncommon letter such as Z
- Computations take place modulo 27.

Encryption and Decryption

$d = 2$, $K = \begin{pmatrix} 4 & 6 \\ 1 & 7 \end{pmatrix}$, $K^{-1} = \begin{pmatrix} 4 & 12 \\ 11 & 10 \end{pmatrix}$

One can check that $KK^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Plaintext: $M = (\text{EG}) = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$

Encryption: $C = KM = \begin{pmatrix} 4 & 6 \\ 1 & 7 \end{pmatrix} \begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 25 \\ 19 \end{pmatrix} = (\text{ZT})$

Decryption: $M = K^{-1}C = \begin{pmatrix} 4 & 12 \\ 11 & 10 \end{pmatrix} \begin{pmatrix} 25 \\ 19 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} = (\text{EG})$

Cryptanalysis of Hill Cipher

- Known plaintext attacks possible given d plaintext-ciphertext matching blocks.
- Given blocks (column vectors) M_i and C_i for $0 \leq i \leq d-1$:
  - $C = [C_0 C_1 \cdots C_{d-1}]$
  - $M = [M_0 M_1 \cdots M_{d-1}]$
  - Solving $C = KM$ for K
  - $M = K^{-1} C$

Cryptanalysis Example

- Let $d = 2$ be known
- Ciphertext is PE∇TBEDLSTE∇HNFQTBRLHIDB
- Known plaintext is the 2 first blocks FR and OM (the first word is FROM)

Example Step 1

Encoding the plaintext and ciphertext:

$M_0 = (\mathrm{FR}) = \begin{pmatrix} 5 \\ 17 \end{pmatrix}$, $M_1 = (\mathrm{OM}) = \begin{pmatrix} 14 \\ 12 \end{pmatrix}$

$C_0 = (\mathrm{PE}) = \begin{pmatrix} 15 \\ 4 \end{pmatrix}$, $C_1 = (\nabla\mathrm{T}) = \begin{pmatrix} 26 \\ 19 \end{pmatrix}$

Therefore $M = [M_0 M_1] = \begin{pmatrix} 5 & 14 \\ 17 & 12 \end{pmatrix}$ and $C = [C_0 C_1] = \begin{pmatrix} 15 & 26 \\ 4 & 19 \end{pmatrix}$

Example Step 2

Recovering encryption matrix K:

$C = KM$ with $K = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

Hence $\begin{pmatrix} 15 & 26 \\ 4 & 19 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 5 & 14 \\ 17 & 12 \end{pmatrix}$

And so $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 15 & 26 \\ 4 & 19 \end{pmatrix} \begin{pmatrix} 5 & 14 \\ 17 & 12 \end{pmatrix}^{-1} = \begin{pmatrix} 13 & 5 \\ 2 & 6 \end{pmatrix}$

Example Step 3

Computing K^{-1} and decrypting the ciphertext:

$K = \begin{pmatrix} 13 & 5 \\ 2 & 6 \end{pmatrix}$ and thus $K^{-1} = \begin{pmatrix} 12 & 17 \\ 23 & 26 \end{pmatrix}$

$M = K^{-1}C$ with $C = \begin{pmatrix} \mathrm{B} & \mathrm{D} \\ \mathrm{E} & \mathrm{L} \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 4 & 11 \end{pmatrix}$ (3rd and 4th blocks)

$M = \begin{pmatrix} 12 & 17 \\ 23 & 26 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 4 & 11 \end{pmatrix} = \begin{pmatrix} 26 & 7 \\ 19 & 4 \end{pmatrix} = \begin{pmatrix} \nabla & \mathrm{H} \\ \mathrm{T} & \mathrm{E} \end{pmatrix}$

The plaintext is FROM∇THE∇REMINISCENCES∇O

Comments on Cryptanalysis of Hill Cipher

- In known plaintext attacks, equations may not be fully determined:
  - Step 2 will fail since matrix not invertible
  - Further plaintext/ciphertext characters can be examined
- Ciphertext only attacks follow known plaintext attacks with extra task of finding probable blocks of matching plaintext-ciphertext:
  - *Example:* when $d = 2$, frequency distribution of non-overlapping pairs of ciphertext characters can be compared with distribution of pairs of plaintext characters
- Automated cryptanalysis in Cryptool, assuming encoded alphabet with $\mathrm{A} = 1, \mathrm{B} = 2, \cdots, \mathrm{Z} = 26, \nabla = 27$.

Do we have some time left?

Yes, then let's start Lecture 7!