Lecture 5: Classical Encryption Part 1

COSC362 Data and Network Security

Book 1: Chapter 3 – Book 2: Chapter 2

Spring Semester, 2021

Motivation

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

Some books:

- "The Codebreakers" by D. Kahn about history of cryptography
- "The Code Book" by S. Singh about classical and modern cryptography

Outline

Introduction
- Basic Definitions
- Cryptanalysis
- Statistics of Natural Language

Transposition Ciphers

Simple Substitution Ciphers
- Caesar Cipher
- Random Simple Substitution Cipher

Outline

Introduction
- Basic Definitions
- Cryptanalysis
- Statistics of Natural Language

Transposition Ciphers

Simple Substitution Ciphers
- Caesar Cipher
- Random Simple Substitution Cipher

Terminology

The science of cryptology has two facets:

- *Cryptography:* the study of designing cryptosystems
- *Cryptanalysis:* the study of breaking cryptosystems

These facets are generally studied together.
Another facet (not covered in this course) could be:

- *Steganography:* the study of concealing information

Confidentiality and Authentication

- Cryptography is the science of *secret writing:*
  - Transformations of data depending on a secret called the *key*.
- Cryptography used to provide *confidentiality* and *authentication* (or *integrity*):
  - Confidentiality: a key is needed to *read* the message.
  - Authentication: a key is needed to *write* the message.

Cryptosystems

A cryptosystem consist of:

- a set of *plaintexts* (holding the original message)
- a set of *ciphertexts* (holding the encrypted message)
- a set of *keys*
- a function, called *encryption* or *encipherment*, which transforms the plaintext into a ciphertext
- an inverse function, called *decryption* or *decipherment*, which transforms the ciphertext back into the plaintext

The ciphertext is sometimes called *cryptogram*.

Symmetric and Asymmetric Cryptography

- Symmetric key cipher (secret key cipher):
  - Encryption and decryption keys are known ONLY to the sender and receiver.
  - Secure channel for transmission of the keys.
- Asymmetric key cipher (public key cipher):
  - Each participant has a public key AND a private key.
  - Possibly working for both encryption of messages and creation of digital signatures.

Notation for Symmetric Encryption Algorithms

- Encryption function E
- Decryption function D
- Message or plaintext M
- Cryptogram or ciphertext C
- Shared secret key K

Encryption is denoted as $C = E(M, K)$
Decryption is denoted as $M = D(C, K)$

Methods of Cryptanalysis

An adversary has access to many methods to break a cryptosystem, such as:

- What are the resources available to the adversary? *Examples:* computational capability, inputs/outputs of the system.
- What is the adversary aiming to achieve? *Examples:* retrieving the whole secret key, distinguishing two messages (e.g. YES and NO). Why is this important?

Exhaustive Key Search

- Basic method: exhaustive key search (or brute force attack) where the adversary tries ALL possible keys.
- NO ONE can prevent such attack:
  - All cryptosystems must have ENOUGH keys to make exhaustive search too difficult computationally.
- The adversary may find the key without trying exhaustive search!
- The adversary may break the cryptosystem without finding the key!

Prevention of exhaustive key search is a *minimum* standard.

Attack Classification

- *Ciphertext Only Attack:* the attacker has access to ONLY intercepted ciphertexts.
- *Known Plaintext Attack:* the attacker knows a small amount of plaintexts and their corresponding ciphertexts.
- *Chosen Plaintext Attack:* the attacker can obtain the ciphertext from some plaintext that it has selected (the attacker has an “inside encryptor” available).
- *Chosen Ciphertext Attack:* the attacker can obtain the plaintext from some ciphertext that it has selected (the attacker has an “inside decryptor” available).

Which Attacks Should Be Prevented?

- A cryptosystem is seen as *highly insecure* if it can be practically attacked using only intercepted ciphertexts.
- A cryptosystem should be secure against chosen plaintext and chosen ciphertext attacks (modern standard).
- History shows that chosen ciphertext attacks are practical to set up for an attacker.

Kerckhoffs' Principle

Kerckhoffs' Principle: The attacker has complete knowledge of the cipher (i.e. the decryption key is the only item UNKNOWN to the attacker):

- History has shown that it is a reasonable assumption. Can we think of any examples?
- Using a secret, non-standard algorithm can cause severe problems:
  - This would be an example of *security through obscurity*.

Alphabets

- Historical ciphers: defining the alphabet for the plaintext and ciphertext (usually the same).
- Roman alphabet: $A, B, C, \cdots, Z$.
  Sometimes are included: space, upper and lower case, punctuation.
- Sometimes, alphabet is mapped to numbers:
  $A = 0, B = 1, C = 2, \cdots, Z = 25$. And space is 26.
- Real-world attackers would need to work out the alphabet.

Statistical Attacks

- Statistical attacks depend on using the redundancy of the alphabet. Can you read this? TDY S VRY CLD
- Information from distribution of single letters, digrams (double letters) and trigrams (triple letters) helps in the attack.
- Exact statistics of a language vary according to what sample is taken.

Sample Statistics for English

- The following statistics give a typical distribution of *English text* (calculated on a text passage of 143000 characters).
- Simplifying the statistics: the text is restricted to a plaintext alphabet of 27 characters:
  - ABCDEFGHIJKLMNOPQRSTUVWXYZ∇ with ∇ being space.
- Proportions shown are relative:
  - *Example:* ∇ accounts for 14.6% of all characters while E∇ accounts for 2.3% of all digrams.

Single Character Percentage Frequencies

∇ 14.6	A 7.0	H 2.6	V 1.3	Z 0.1
E 10.1	R 5.2	M 2.5	B 1.3	J 0.1
N 7.8	S 5.1	P 2.5	Y 0.8	Q 0.1
T 7.5	L 3.7	U 2.4	W 0.6	
I 7.1	C 3.5	G 1.7	K 0.2	
O 7.0	D 3.5	F 1.6	X 0.1	

Most Common Digram Percentage Frequencies

E∇ 2.3	D∇ 1.7	ES 1.3	RE 1.1
∇A 2.1	TI 1.7	AT 1.3	IO 1.1
ON 1.9	AN 1.6	ND 1.3	∇I 1.1
IN 1.9	EN 1.6	N∇ 1.3	ME 1.0
∇T 1.8	TH 1.6	AL 1.2	ER 0.9
S∇ 1.7	NT 1.4	HE 1.2	∇O 0.9

Outline

Introduction
- Basic Definitions
- Cryptanalysis
- Statistics of Natural Language

Transposition Ciphers

Simple Substitution Ciphers
- Caesar Cipher
- Random Simple Substitution Cipher

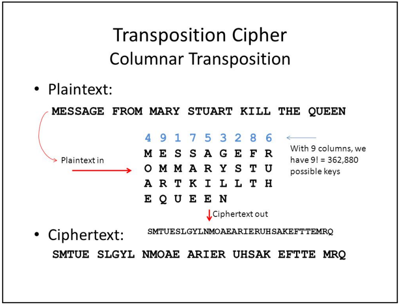
Basic Cipher Operations

Historical ciphers combine two basic operations:

Transposition: characters in the plaintext are mixed up with each other (permuted).

Substitution: each character (resp. set of characters) is replaced by a different character (resp. set of characters).

Transposition Ciphers



- Permuting characters in a fixed period d and permutation f.
- Plaintext seen as a matrix of rows of length d.
- Permuting rows/columns and outputting in row/column order.
- Here, considering permutation of rows and outputting in column order.

Simple Transposition Cipher

- Key is (d, f)
- Each block of d characters is re-ordered using permutation f
- $d!$ permutations of length d:
  - $d! = d \times (d-1) \times (d-2) \times \cdots \times 2 \times 1$
- *Example:* $d = 10$ gives $3,628,800$ possible keys

Cryptanalysis of a Transposition Cipher

- Frequency distribution of ciphertext characters $=$ Frequency distribution of plaintext characters:
  - Helping to identify a transposition cipher.
- If *d* is small then transposition ciphers solved by hand using *anagramming*:
  - Restoring disarranged characters to their original positioning.
- Guessing value of *d* and writing the ciphertext in columns s.t. there are *d* rows.
- Knowledge of plaintext language digrams and trigrams to optimise trials.
- Automating this process.

Outline

Introduction
- Basic Definitions
- Cryptanalysis
- Statistics of Natural Language

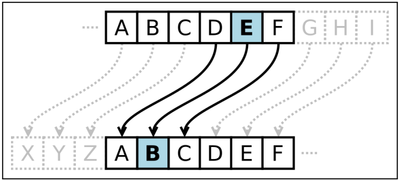
Transposition Ciphers

Simple Substitution Ciphers
- Caesar Cipher
- Random Simple Substitution Cipher

Simple Substitution Ciphers

- Each character in the plaintext alphabet replaced by a character in the ciphertext alphabet following a substitution table.
- Also called *monoalphabetic* substitution ciphers.
- Transposition ciphers permutes PLAINTEXT characters while substitution ciphers permute ALPHABET characters.
- Some special cases:
  - Caesar cipher
  - Random simple substitution cipher

Caesar Cipher



- Moving the ith letter of an alphabet to $(i+j)$th letter, s.t. the key is j.

- One can write the substitution table or simply:
  - Encryption: $C_i = (M_i + j) \mod n$
  - Decryption: $M_i = (C_i - j) \mod n$
  - either $n = 26$ or $n = 27$ (size of alphabet)
- *Example:* $j = 1$ then CIPHER $\rightarrow$ DJQIFS

Cryptanalysis of the Caesar Cipher

- Finding where one of the most frequent characters is shifted to.
- *Example:* given the ciphertext PACGHJUHHCRICGRFWRUCRICPHGLFLQH
  - Counting the characters
  - Most common characters: H and C (frequency is 5 each)
- Let ∇ be in the alphabet ($n = 27$), finding where it is mapped to?
  - Trial 1: $\nabla \to$ H, i.e. $j = 8$, thus HTV∇BM$\nabla\nabla$VJA$\cdots$, so incorrect
  - Trial 2: $\nabla \to$ C, i.e. $j = 3$, thus MY∇DEGREE∇OF∇DOCTOR∇OF∇MEDICINE

Random Simple Substitution Cipher

- Assigning a random character of the alphabet to another character of the alphabet.
- Encryption and decryption defined by substitution table that randomly permutes the alphabet.
- If the alphabet has 26 characters, then 26! keys. Why?
  - Greater than 10^{26}
  - Too many keys to search even with modern computers
- Caesar cipher is a special case.

Example

Message: THE∇EVENING∇AND∇THE∇MORNING

Substitution table (key):

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

Message substitution (encryption):

T → O	N → B	D → I
H → A	I → Z	∇ → E
E → N	N → B	T → O
∇ → E	G → W	H → A
E → N	∇ → E	E → N
V → D	A → S	∇ → E
E → N	N → B	. . .

Cryptogram: OANENDNBZBWESBIEOANEKURBZBW

Cryptanalysis of a Random Substitution

- Using frequency analysis on alphabet characters
- Deciphering the following ciphertext:

FJLTXCFWKOVLHKJVKCBCOTEEVLPKCKJVJSTWTJYVKJVOJSTSBPLVITWCWPVDBIT

WICKTKQLVPHYTPRBJSTQLVYTKKJSCJETSCGTUHKJPTKYLFRTPETXCBTKJFXCJTJ

STGCZHTVOCGVZJCXTJTLCJCSHWPLTPOLCWYKFOJSTCQQCLCJHKVQTLCJTKEFJSV

HJCQQLTYFCRZTETCLJSTCXVLJFATXTWJKSVHZPRTYCZYHZCJTPCJCGTLBZVEOFI

HLTKCBQTLYTWJESFYSFKZCLITFWYVWJFWHVHKVQTLCJFVWFJEVHZPQLVPHYTXVL

TJSCWYHRFYXTJTLKVOICKCBTCLKCBCZFJJZTZTKKJSCWVWTYTWJFXTQTLYHRFYX

TJTLJSTYCHKJFYKVPCFKYVWKJCWJZBLTYHQTLCJTPCWPFKWTGTLPTKJLVBTPJST

KVZTQLVPHYJJSCJPFKCQQTCLKFKJSTPFKJFZZTPECJTLWVEVWTYHRFYXTJTLVOE

CJTLQLVPHYTKXVLTJSCWYHRFYXTJTLKVOICKJSTTDQTWKTFWECJTLJSTWPVTKWV

JCXVHWJJVCYTWJFXTQTLYHRFYXTJTLJSTILTCJOCYJVLVOJSTTDQTWKTLTKFPTK

FWJSTTZTYJLFYTWTLIBJSTYVKJVOKHLGTFZZCWYTEFZZRTXFWFXHXCWPJSTITWT

LCZTDQTWKTKCPZFRFJHX · · ·

Frequency Analysis of Ciphertext

No.	Character	%	Frequency
1	T	15.4	110
2	J	10.2	73
3	C	8.3	59
4	K	6.7	48
5	L	6.7	48
6	V	6.3	45

- E and T are the most frequent characters in English:
  - E $\rightarrow$ T and T $\rightarrow$ J in the substitution table
- Looking for English words such as THE and other common trigrams.

Using Cryptool

- Solving random substitution by hand is tedious and requires many trials and errors.
- Using Cryptool (freeware package):
  - Some utilities to help us, such as frequency counts.
  - Cryptool has a ciphertext-only tool to solve simple random substitution (needs a bit of help to get the full answer).

Substitution Table (Key)

Using Cryptool, the substitution table (key) is:

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

The plaintext starts with: ITREMAINSFORUSTOSAY$\cdots$