CS396: Security, Privacy & Society

Fall 2022

Lecture 6: Cryptographic System I

Instructor: Abrar Alrumayh

September 16, 2022

STEVENS
INSTITUTE OF TECHNOLOGY
1870

Outline

- Classical ciphers and how to break them
- What does it mean for a cipher to be secure?
  - Perfect secrecy

Classical ciphers

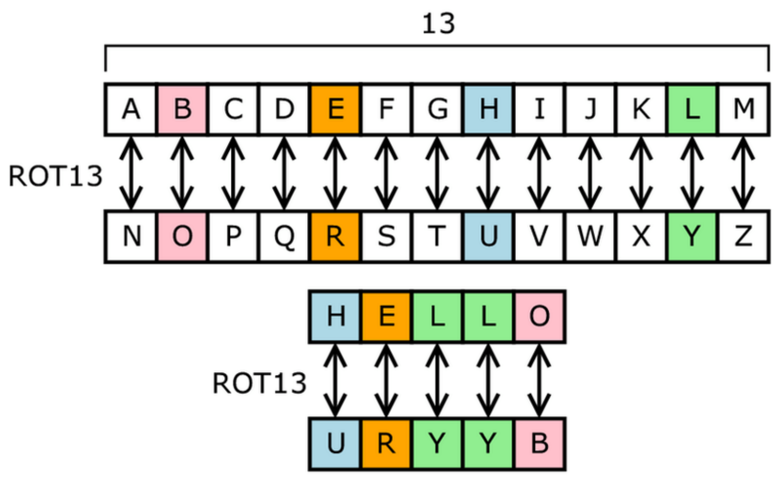
Classical ciphers

- developed prior to the invention of the computer.
- used throughout all of history, up until the early days of World War II.
- From very simple to very complex.
- Ex:
  - Substitution Cipher
  - Caesar Cipher
  - Shift Cipher

Substitution ciphers

Large class of ciphers

- each letter is uniquely (no repeating) replaced by another
- Choose a **random permutation** of English alphabets...
- there are 26! possible substitution ciphers
  - e.g., one popular substitution “cipher” for some Internet posts is ROT13
- historically
  - all classical ciphers are of this type



Example

- Encipher “DROP AT LOCATION”, K = 13

General structure of substitution ciphers

Based on letter substitution

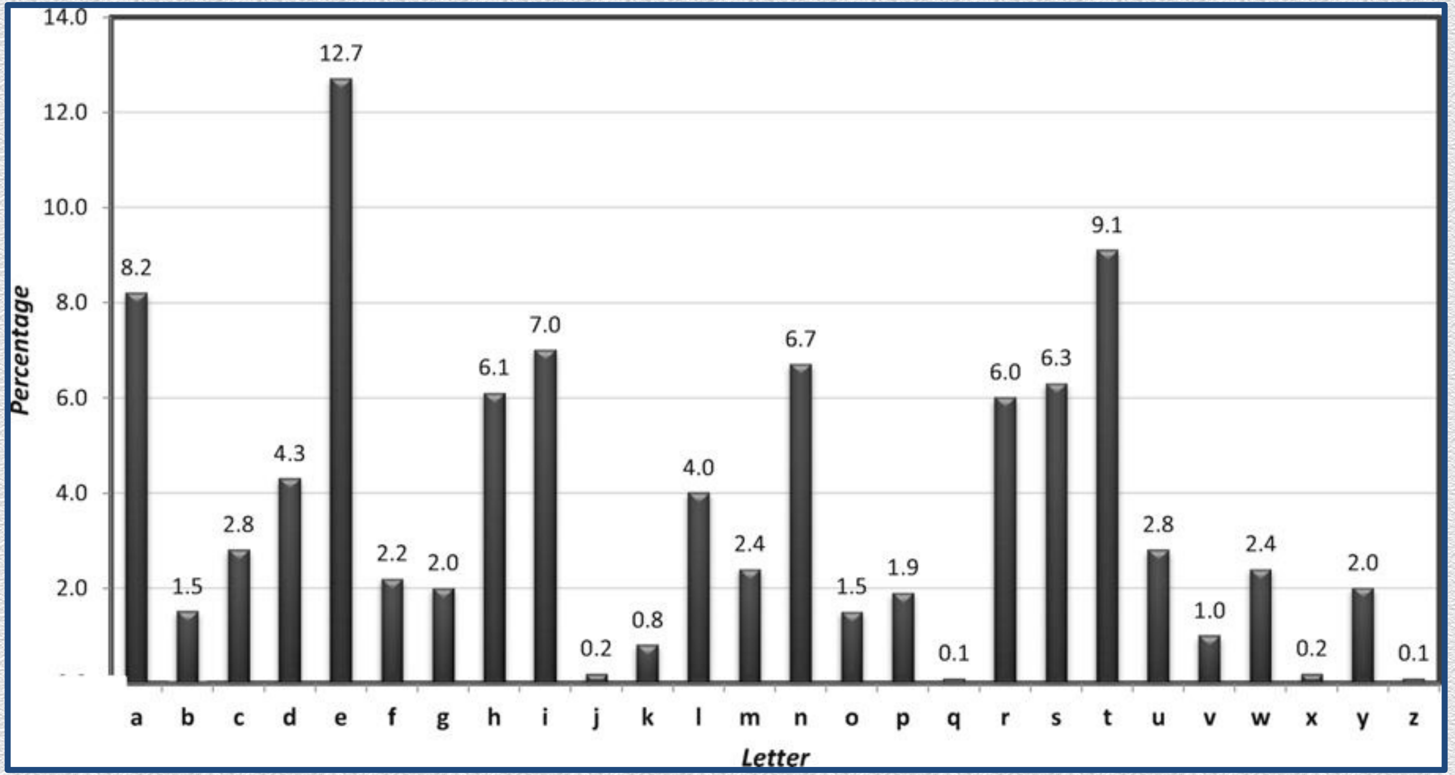
- message space $\mathcal{M}$ is "valid words" from a given alphabet
  - e.g., English text without spaces, punctuation or numerals
  - characters can be represented as numbers in [0:25]
- encryption
  - mapping each plaintext character into another character
  - character mapping is typically defined as a "shift" of a plaintext character by a number of positions in a canonical ordering of the characters in the alphabet
  - character shifting occurs with "wrap-around" (using mod 26 addition)
- decryption
  - undo character shifting with "wrap-around" (using mod 26 subtraction)

Limitations of substitution ciphers

Generally, susceptible to frequency (and other statistical) analysis

- letters in a natural language, like English, are not uniformly distributed
- cryptographic attacks against substitution ciphers are possible
  - e.g., by exploiting knowledge of letter frequencies, including pairs and triples

Letter frequency in (sufficiently large) English text



Frequency analysis

- Breaking substitution cipher (ciphertext only attack):
  - Collect a long ciphertext – frequency patterns will not change.
  - Compute frequencies of various letters
  - Reconstruct the key: most frequent letter represents “E”, second most is “T”, etc. Use bigrams, trigrams, etc.

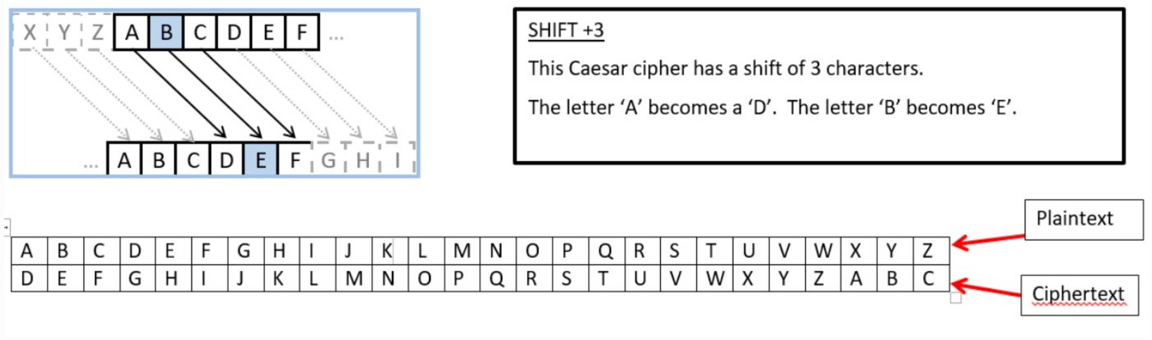
```
wkh sdvvzrug lv vhyhq grqw whoo dqbrqh
```

```
h = 5
v = 4
q = 3
r = 3
g = 3
d = 2
b = 1
k = 1
l = 1
s = 1
y = 1
```

Classical ciphers – examples

<u>Caesar's cipher</u>

- simple substitution cipher
- shift each character in the message by 3 positions
- cryptanalysis
  - **no secret key is used** – based on “security by obscurity”
  - thus the code is trivially insecure once knows Enc (or Dec)



Source: https://upload.wikimedia.org/wikipedia/commons/f/fa/Ascii-proper-color.svg

Classical ciphers – examples (II)

Shift cipher

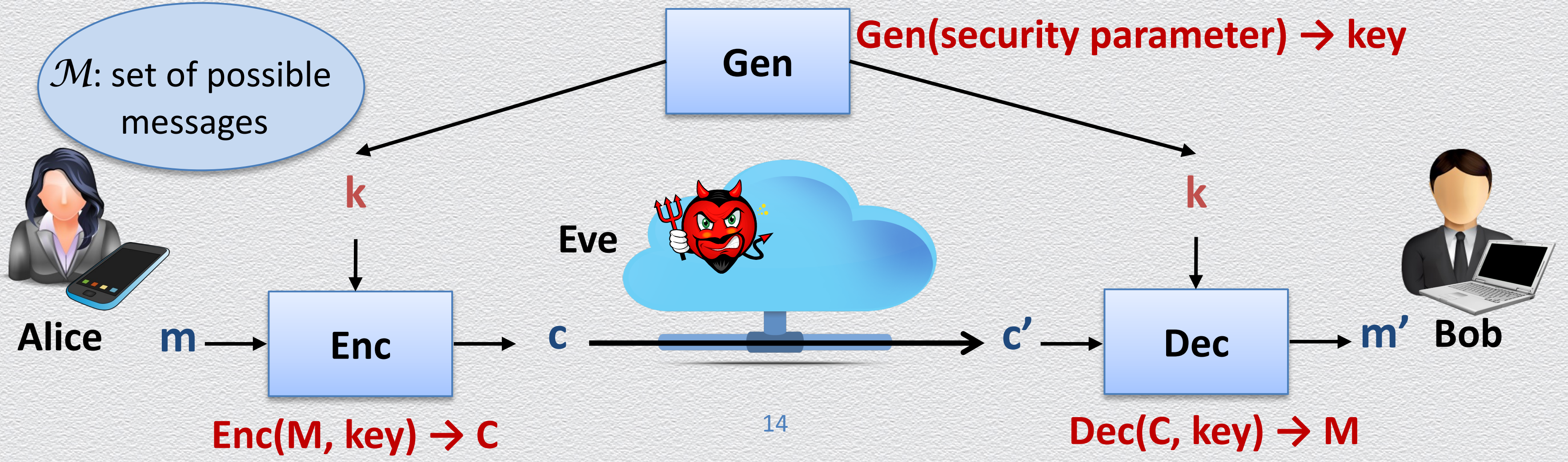
- **keyed extension** of Caesar's cipher
- randomly set key k in [0:25]
  - shift each character in the message by k positions
- cryptanalysis
  - **brute-force attacks** are effective given that
    - **key space is small** (26 possibilities or, actually, 25 as 0 should be avoided)
    - message space M is **restricted to "valid words"**
      - e.g., corresponding to valid English text

Perfect secrecy

Security tool: Symmetric-key encryption scheme

Encryption scheme consists of:

- a **message space $\mathcal{M}$**; and
- a triplet of algorithms **(Gen, Enc, Dec)**
  - Gen: A method for generating random keys k
  - Enc: Encryption algorithm
  - Dec: Decryption algorithm



Perfect correctness

For any **k** $\in \mathcal{K}$,

m $\in \mathcal{M}$,

and any ciphertext **c** output of $\mathrm{Enc}_k(m)$,

it holds that:

$$\Pr[\ \mathrm{Dec}_k(c) = m\ ] = 1$$

Towards defining perfect security

- defining security for an encryption scheme is not trivial
  - e.g., what we mean by << Eve "cannot learn" m (from c) >> ?
- our setting so far is a random experiment
  - a message m is chosen according to $\mathcal{D}_{\mathcal{M}}$
  - a key k is chosen according to $\mathcal{D}_{\mathcal{K}}$
  - $\mathrm{Enc}_k(m) \rightarrow c$ is given to the adversary

how to define security?

First attempt: Protect the key k!

- Security means that

  the adversary should **not** be able to **compute the key k**

- Intuition
  - it'd better be the case that the key is protected!... ⇨ necessary condition
- Problem
  - this definition fails to exclude clearly insecure schemes ⇨ but not sufficient condition!
  - Example from Caesar Cipher:
    - ATTACK = BUUBDL and DEFEND = EFGFOE, k =1
    - Broken by checking patterns! don't need the key!

Second attempt: hide the message!

- Security means that

  the adversary should **not** be able to **compute the message m**

- Intuition
  - it'd better be the case that the message m is not learned...
- Problem
  - this definition fails to exclude clearly undesirable schemes
  - what if the ciphertext reveals the frequency of the alphabets in the plaintext?
  - e.g., those that protect m partially, i.e., they reveal the least significant bit of m

Third attempt: hide everything about the message

- Security means that

  the adversary should **not** be able to **learn any information about m**

- Intuition
  - it seems close to what we should aim for perfect secrecy…
- Problem
  - this definition ignores the adversary's prior knowledge on $\mathcal{M}$
  - Something about the message may already be known
  - e.g., distribution $\mathcal{D}_{\mathcal{M}}$ may be known or estimated
    - m is a valid text message, or one of "`attack`", "`no attack`" is to be sent

Fourth attempt: hide everything that is not already known!

- Security means that

  the adversary should **not** be able to **learn any <u>additional</u> information on m**

- We **cannot** hide what may be **a priori known** about the message.
- Adversary should not learn any **NEW** information about the message after seeing the ciphertext
- How can we formalize this?

Fourth attempt: hide everything that is not already known!

- How can we formalize this?

Alice m

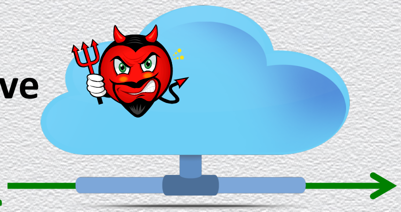
$\mathbf{Enc_k(m) \rightarrow c}$

Eve

$$m = \begin{cases} \texttt{attack} & \text{w/ prob. } 0.8 \\ \texttt{no attack} & \text{w/ prob. } 0.2 \end{cases}$$

Eve's view remains the same!

Eve



c

$$m = \begin{cases} \texttt{attack} & \text{w/ prob. } 0.8 \\ \texttt{no attack} & \text{w/ prob. } 0.2 \end{cases}$$

Fourth attempt: hide everything that is not already known!

- Messages come from some distribution
  - D be a random variable for sampling the messages from the message space M.
- Distribution D is known to the adversary (a **priori** information)
- The ciphertext $c = Enc_k(m)$, depends on:
  - m chosen according to D
  - k is chosen randomly
  - Enc may also use some randomness
  - These induce a distribution C over the ciphertexts c.
- The adversary only observes c

Two equivalent views of perfect secrecy

a posteriori = a priori $\sim$ **C is independent of M**

For every $\mathcal{D}_{\mathcal{M}}$, m ∈ $\mathcal{M}$ and c ∈ $\mathcal{C}$, for which Pr [C = c] > 0, it holds that

$$\mathbf{Pr[\ M = m\ |\ C = c\ ] = Pr[\ M = m\ ]}$$

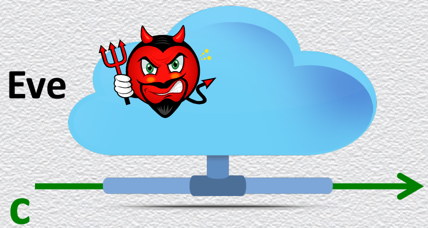
For every m, m′ ∈ $\mathcal{M}$ and c ∈ $\mathcal{C}$, it holds that

$$\mathbf{Pr[\ Enc_K(m) = c\ ] = Pr[\ Enc_K(m') = c\ ]}$$

random experiment
$\mathcal{D}_{\mathcal{M}} \rightarrow$ **m = M**
$\mathcal{D}_{\mathcal{K}} \rightarrow$ k = K
$\mathrm{Enc_k(m)} \rightarrow$ c = C

m = $\begin{cases} \texttt{attack} & \text{w/ prob. 0.8} \\ \texttt{no attack} & \text{w/ prob. 0.2} \end{cases}$

Eve's view remains the same!



m = $\begin{cases} \texttt{attack} & \text{w/ prob. 0.8} \\ \texttt{no attack} & \text{w/ prob. 0.2} \end{cases}$

Perfect secrecy (or information-theoretic security)

Definition 1

A symmetric-key encryption scheme (Gen, Enc, Dec) with message space $\mathcal{M}$,
is **perfectly secret** if for every $\mathcal{D}_{\mathcal{M}}$, every message m ∈ $\mathcal{M}$ and every ciphertext c ∈ $\mathcal{C}$
for which Pr [C = c] > 0, it holds that

$$\Pr[\, M = m \mid C = c \,] = \Pr[\, M = m \,]$$

- intuitively
  - the *a posteriori* probability that any given message m was actually sent is the **same** as the *a priori* probability that m would have been sent
  - observing the ciphertext reveals **nothing (new)** about the underlying plaintext

Alternative view of perfect secrecy

Definition 2

A symmetric-key encryption scheme (Gen, Enc, Dec) with message space $\mathcal{M}$, is **perfectly secret** if for every messages m, m' ∈ $\mathcal{M}$ and every c ∈ $\mathcal{C}$, it holds that

$$\Pr[\,\mathrm{Enc}_K(m) = c\,] = \Pr[\,\mathrm{Enc}_K(m') = c\,]$$

- intuitively
  - the probability distribution $\mathcal{D}_C$ **does not depend** on the plaintext
  - i.e., M and C are **independent** random variables
  - the ciphertext contains "**no information**" about the plaintext
  - "**impossible to distinguish**" an encryption of **m** from an encryption of **m'**