CY 2550 Foundations of Cybersecurity

Cryptography, Part 1

January 16, 2020

Alina Oprea

Associate Professor, Khoury College

Northeastern University

Logistics

- Project 0 is due tomorrow
- Project 1 will have a written homework component
- No quiz from History of Security or Linux Basics
- Holiday on Monday, January 20
- TA office hours cancelled on Monday, January 20
- TA office hours: every day, schedule is posted on Piazza
- Alina’s office hours
  - Thursday, 2-4pm

The Science of Secrets

- **Cryptography**: the study of mathematical techniques to providing aspects of information security services
  - Creating secrets
- **Cryptanalysis:** the study of mathematical techniques for attempting to defeat information security services
  - Breaking secrets
- **Cryptology:** the study of cryptography and cryptanalysis

Cryptographic Protocols

- Protocols that
  - Enable parties to … communicate securely
  - Achieve goals to … protect message confidentiality and integrity
  - Overcome adversaries
- Need to understand
  - Who are the parties and the context in which they act?
  - What are the security goals of the protocols?
  - What are the capabilities of the adversaries? Threat model

The cast

The good players

Alice

Bob

Introduced in the original RSA paper

The bad players

Eve

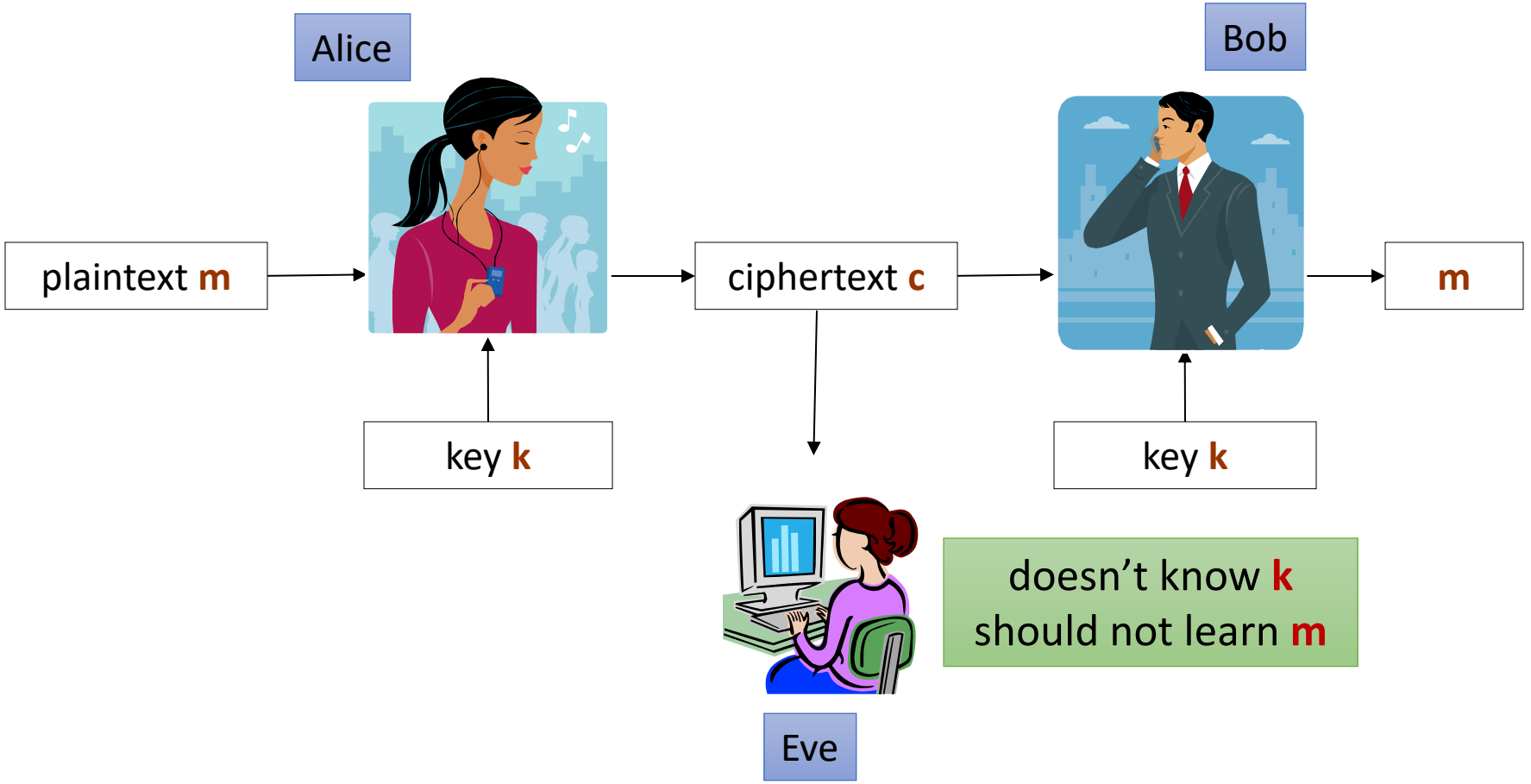
Eavesdropper

Mallory

Malicious

Malicious players

Encryption Terminology



Encryption scheme = encryption & decryption procedures

Goals and objectives

- Objective
  - Ensure security of communication between parties over an insecure medium

- Security goals
  - *Confidentiality* (secrecy)
    - Only the intended recipient can see the communication
  - *Authenticity*
    - Communication is generated by the alleged sender
  - *Integrity* – no unauthorized modifications to messages
  - *Non-repudiation* – no disclaiming of authorship

Kerckhoffs' principle

Auguste Kerckhoffs (1883):

The enemy knows the system

The cipher should remain secure even if **the adversary knows the specification of the cipher.**

The only thing that is **secret** is a

key **k**

that is **usually chosen uniformly at random**

Kerckhoff‘s principle: motivation

1. It is unrealistic to assume that the design details remain secret. Too many people need to know. Software/hardware can be **reverse-engineered!**
2. Pairwise-shared keys are easier to **protect**, **generate** and **replace**.
3. The design details can be discussed and **analyzed in public**.
   - Public competition for selection of block cipher (AES) and hash functions (SHA3)

Not respecting this principle
=
``**security by obscurity**”.

9

Attacker Threat Model

1. Interaction with messages and the protocol
   - Passive: only observes and attempts to decrypt messages
     - Only threatens confidentiality
   - Active: observes, modifies, or deletes messages
     - Threatens confidentiality, integrity, and authenticity
2. Full knowledge of the chosen cryptographic algorithm
   - Kerchhoff's Principle
     - A cryptosystem should be secure even if everything about the system, except the key, is public knowledge
   - **No security through obscurity**

Attacker Threat Model

3. Interaction with cipher algorithm
   - Ciphertext-only attack: attacker only sees encrypted messages
   - Chosen-plaintext attack (CPA)
     - Attacker may choose a number of messages and obtain the ciphertexts for them
   - Chosen-ciphertext attack (CCA)
     - Attacker may choose a number of ciphertexts and obtain the plaintexts
   - Both CPA and CCA attacks may be adaptive
     - Choices may change based on results of previous requests
4. Computationally bounded
   - Finite resources to calculate and store things
   - Polynomial running time

Classical Cryptography

Approaches to Secure Communication

Steganography

- “covered writing”
- hides the existence of a message
- depends on secrecy of method

Cryptography

- “hidden writing”
- hide the meaning of a message
- depends on secrecy of a short key, not method

A mathematical view

$\mathcal{K}$ – **key** space

$\mathcal{M}$ – **plaintext** space

$\mathcal{N}$ - **natural numbers**

$\mathcal{C}$ - **ciphertext** space

An **encryption scheme** is a pair **(Gen,Enc,Dec)**, where

- **Gen : N → K** is a **key generation** algorithm,
- **Enc : K × M → C** is an **encryption** algorithm,
- **Dec : K × C → M** is an **decryption** algorithm.

We write $\mathbf{Enc_k(m)}$ and $\mathbf{Dec_k(c)}$ instead of **Enc(k,m)** and **Dec(k,c)**.

Correctness

for every **k, m** we should have $\mathbf{Dec_k(Enc_k(m)) = m}$.

Caesar Shift

- Simple symmetric substitution cipher
  - Key is a number *k*
  - To encrypt, “shift” each letter by *k* positions
  - To decrypt, “shift” each letter back by *k* positions

HEY BRUTUS BRING A KNIFE TO THE PARTY

K = 3

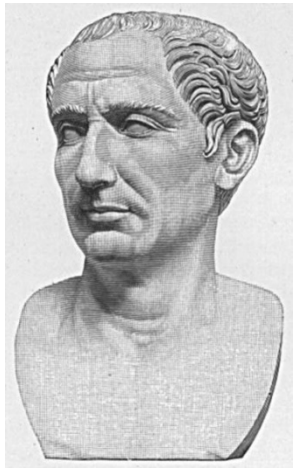
KHB EUXWXV EULQJ D NQLIH WR WKH SDUWB

Shift cipher: Mathematical View

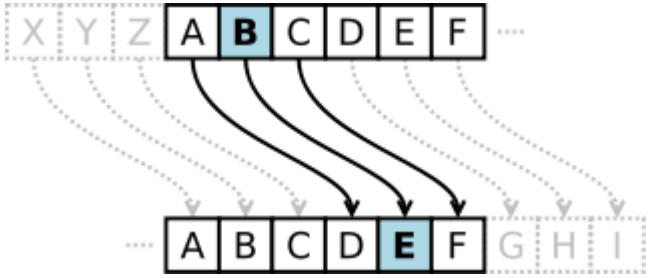
$\mathcal{M}$ = words over alphabet {A,…,Z} ≈ **{0,…,25}**

$\mathcal{K}$ = **{0,…,25}**

$Enc_k(m_1,\ldots,m_n) = (m_1+k \bmod 26,\ldots, m_n+k \bmod 26)$



Cesar: **k = 3**



$Dec_k(c_1,\ldots,c_n) = (c_1-k \bmod 26,\ldots, c_n-k \bmod 26)$

Security of the shift cipher

How to break the shift cipher?

Check all possible keys!

Let **c** be a ciphertext.

For every **$k \in \{0,\ldots,25\}$** check if **$Dec_k(c)$** "makes sense".

Most probably only one such **k** exists.

Thus **$Dec_k(c)$** is the message.

This is called a **brute force attack**.

Moral: the key space needs to be large!

Monoalphabetic Substitution Cipher

- Replace each letter *X* with π(X) where π is a permutation
- In this cipher, the key is the permutation π
  - Key space is all possible permutations

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

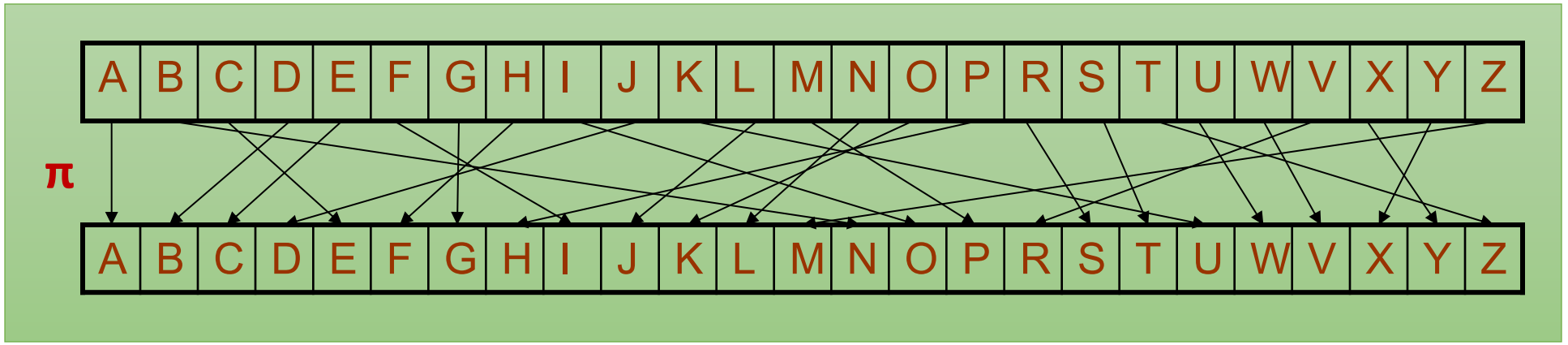
HELLO WORLD

YZXXT PTMXO

Substitution cipher

$\mathcal{M}$ = words over alphabet {A,…,Z} ≈ {0,…,25}

$\mathcal{K}$ = a set of permutations of {0,…,25}



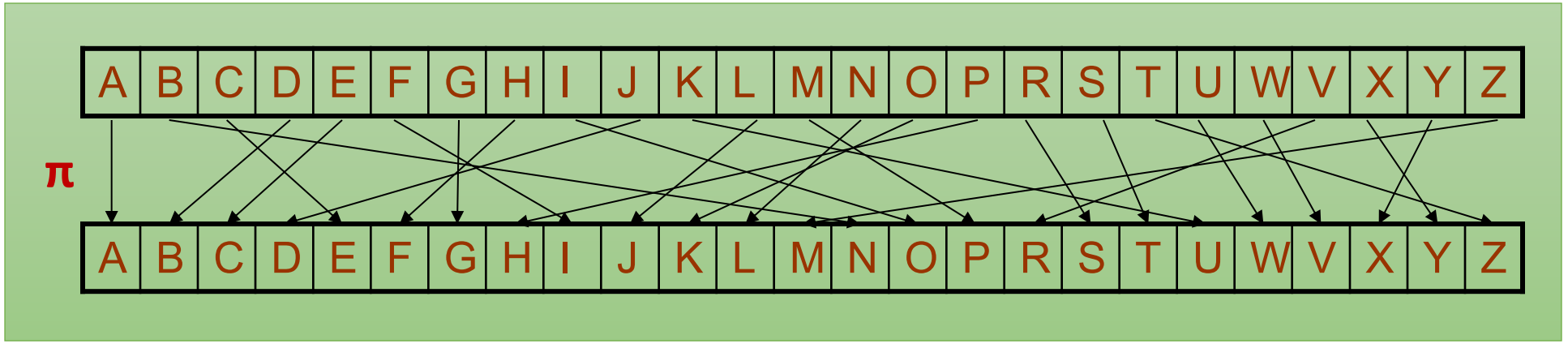
$Enc_\pi(m_1,\ldots,m_n) = (\pi(m_1),\ldots, \pi(m_n))$

$Dec_\pi(c_1,\ldots,c_n) = (\pi^{-1}(c_1),\ldots, \pi^{-1}(c_n))$

Example substitution cipher

$\mathcal{M}$ = words over alphabet {A,…,Z} ≈ {0,…,25}

$\mathcal{K}$ = a set of permutations of {0,…,25}



P = CRYPTOGRAPHY

C = ESXHZKGSAHFX

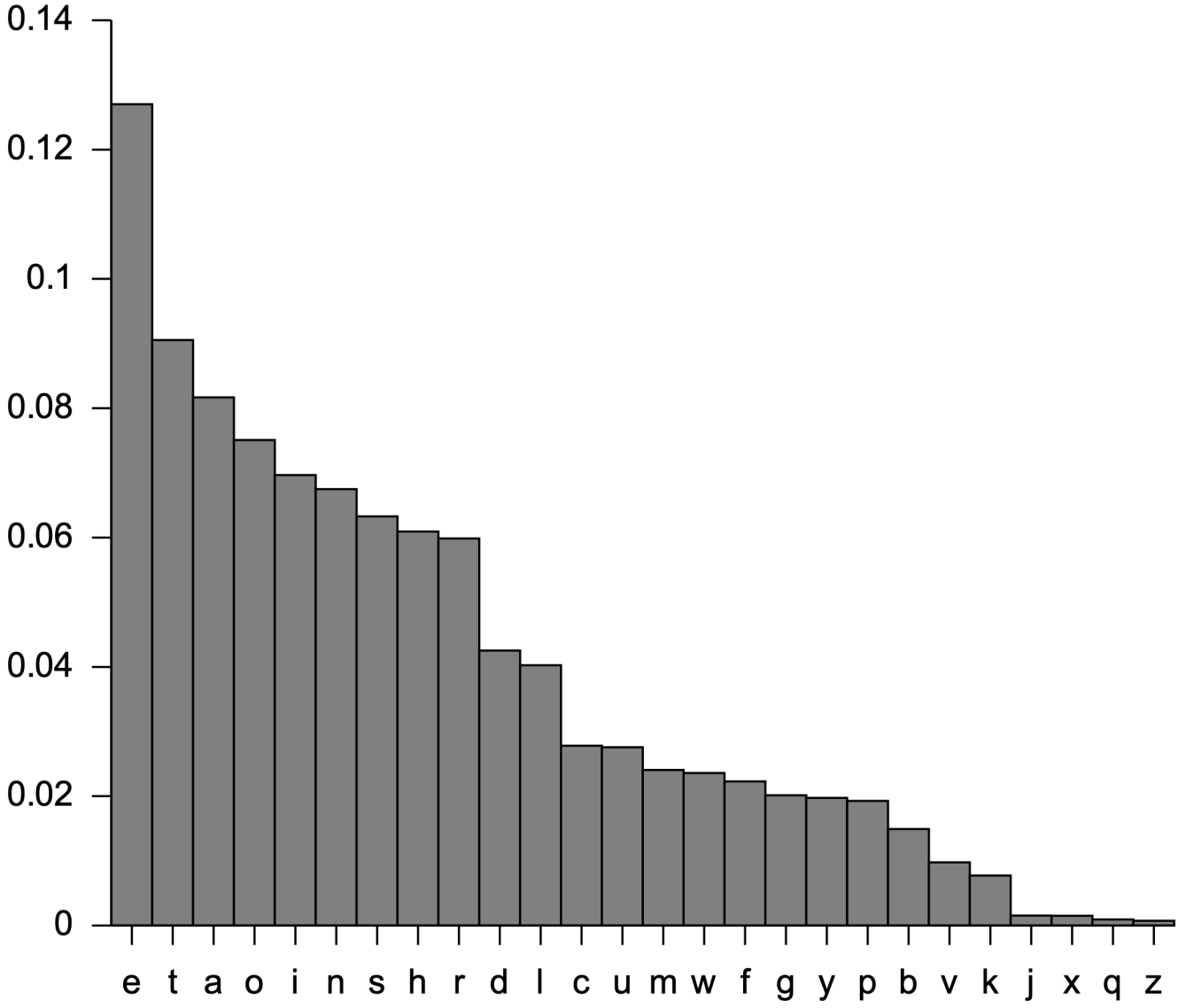
Cryptanalysis of Monoalphabetic Substitution

- Dominates cryptography through the first millennium
- Exhaustive search is infeasible (26! = $4*10^{26}$ possible keys)

- Frequency analysis
  - Remember Al-Kindi from 800 AD

Frequency Analysis

- Human languages have patterns
  - Frequency of letter usage
  - Frequency of *n*-letter combinations (bigrams, trigrams)
- These patterns survive substitution



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

Cryptanalysis of Monoalphabetic Substitution

- Dominates cryptography through the first millennium
- Exhaustive search is infeasible (26! = $4*10^{26}$ possible keys)

- Frequency analysis
  - Al-Kindi from 800 AD

- Lessons?
  - Use large blocks: instead of replacing ~5 bits at a time, replace 64 or 128 bits
    - Leads to block ciphers like DES and AES
  - Use different substitutions to prevent frequency analysis
    - Leads to polyalphabetic substitution ciphers and stream ciphers

Vigenère Cipher (1596)

- Main weakness of monoalphabetic substitution ciphers:
  - Each letter in the ciphertext corresponds to only one letter in the plaintext
- Polyalphabetic substitution cipher
  - Given a key $K = (k_1, k_2, \ldots, k_m)$
  - Shift each letter p in the plaintext by k_i, where i is modulo m
- Somewhat resistant to frequency analysis

Vigenère cipher

$\mathcal{M}$ = words over alphabet {A,…,Z} ≈ {0,…,25}

$\mathcal{K}$ = a set of characters $\{k_1,\ldots k_t\}$

$$Enc_k(m_1,\ldots,m_n) = (m_1+k_1,\ldots, m_t+k_t\,,\; m_{t+1}+k_1,\ldots m_{2t}+k_t,\; \ldots.)\bmod 26$$

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

Example:

Plaintext: C R Y P T O G R A P H Y

Key: L U C K L U C K L U C K

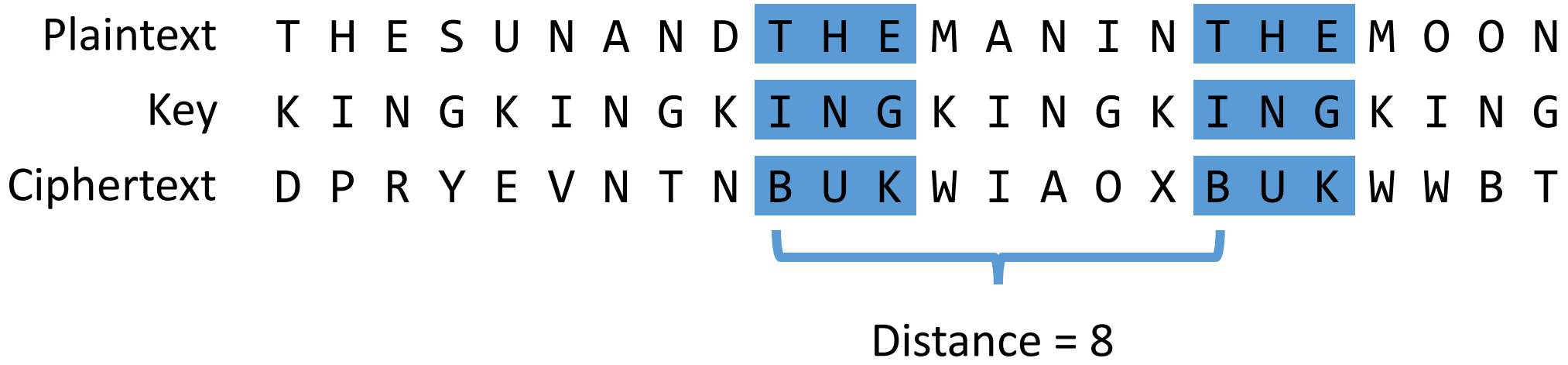
Ciphertext: N L A Z E I I B L J J I

Cryptanalysis of Vigenère Cipher

- Essentially a collection of shift ciphers
  - One letter in ciphertext corresponds to multiple letters in plaintext
  - Can adapt frequency analysis
  - Any message encrypted by a Vigenère cipher is a collection of as *many shift ciphers* as there are letters in the key

- Cracking Vigenère (1854 or 1863)
  1. Guess the key length *x* using **Kasisky test** or **index of coincidence**
  2. Divide the ciphertext into *x* shift cipher encryptions
  3. Use frequency analysis on each shift cipher

Kasisky Test



- Repeating patterns (of length >2) in ciphertext are a tell
  - Likely due to repeated plaintext encrypted under repeated key characters
  - The distance is likely to be a multiple of the key length

Cryptanalysis of Vigenère Cipher

- Cracking Vigenère (1854 or 1863)
  1. Guess the key length *x* using Kasisky test of index of coincidence
  2. Divide the ciphertext into *x* shift cipher encryptions
  3. Use frequency analysis on each shift cipher

- Lessons?
  - As key length increases, letter frequency becomes more random
  - If key never repeated, Vigenère wouldn't be breakable

One Time Pad (1920s)

- Fix the vulnerability of the Vigenère cipher by using very long keys
- Key is a random string that is at least as long as the plaintext
- Similar encryption as with Vigenère (different shift per letter)

Boolean operations: XOR

XOR of two strings in $\{0,1\}^n$ is their bit-wise addition mod 2

X	Y	X $\oplus$ Y
0	0	0
0	1	1
1	0	1
1	1	0

$$\begin{array}{ccccccc} 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ \hline 1 & 1 & 0 & 1 & 1 & 0 & 1 \end{array} \quad \oplus$$

One-time pad

ℓ – a parameter

$\mathcal{K} = \mathcal{M} = \{0,1\}^{\ell}$

component-wise **xor**

Vernam's cipher:

$$\mathbf{Enc_k(m) = k \oplus m}$$
$$\mathbf{Dec_k(c) = k \oplus c}$$

Gilbert Vernam (1890 –1960)

Correctness:

$$\mathbf{Dec_k(Enc_k(m)) = k \oplus (k \oplus m)}$$
$$\mathbf{= m}$$

Defining “security of an encryption scheme” is not trivial.

consider the following experiment

(**m** – a message)

1. the key **K** is chosen uniformly at random
2. **$C := Enc_K(m)$** is given to the adversary

how to define security

?

Idea 1

(**m** – a message)

1. the key **K** is chosen uniformly at random
2. **C := $Enc_K(m)$** is given to the adversary

An idea

"The adversary should not be able to learn **K**."

A problem

the encryption scheme that "doesn't encrypt":

$Enc_K(m) = m$

satisfies this definition!

Idea 2

(**m** – a message)

1. the key **K** is chosen uniformly at random
2. **$C := Enc_K(m)$** is given to the adversary

An idea

"The adversary should not be able to learn **m**."

A problem

What if the adversary can compute, e.g., the first half of **m**?

Idea 3

(**m** – a message)

1. the key **K** is chosen uniformly at randomly
2. **C := $Enc_K(m)$** is given to the adversary

An idea

"The adversary should not learn any information about **m**."

Sounds great! But what does it actually mean?
How to formalize it?

Example

Eve knows that

$$\mathbf{m} := \begin{cases} \text{“I love you”} & \text{with prob. } \mathbf{0.1} \\ \text{“I don’t love you”} & \text{with prob. } \mathbf{0.7} \\ \text{“I hate you”} & \text{with prob. } \mathbf{0.2} \end{cases}$$

$\mathbf{c := Enc_K(m)}$

Eve **still** knows that

$$\mathbf{m} := \begin{cases} \text{“I love you”} & \text{with prob. } \mathbf{0.1} \\ \text{“I don’t love you”} & \text{with prob. } \mathbf{0.7} \\ \text{“I hate you”} & \text{with prob. } \mathbf{0.2} \end{cases}$$

"The adversary should not learn any information about **m**."

An encryption scheme is **perfectly secret** if

for every distribution of **M**

and every **m** ϵ $\mathcal{M}$ and **c** ϵ C

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

Ciphertext-only attack
(passive)