



CY2550: Foundations of Cybersecurity

Section 03

Crypto Module: Attack Models and Classical Cryptography
(Shift, Monoalphabetic substitution, Vigenere, OTP)

Outline

- ▶ Definitions and attacker models
- ▶ Classical cryptography:
 - ▶ Shift cipher
 - ▶ Monoalphabetic substitution cipher
 - ▶ Polyalphabetic substitution: Vigenere
 - ▶ Perfect secrecy: One-time pad (OTP)



Definitions and attacker models

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

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



Cryptographic protocols

- ▶ Protocols that
 - ▶ Enable parties to ... **communicate securely**
 - ▶ Achieve goals to ... **protect message confidentiality and integrity**
 - ▶ In an environment where boundaries and interaction with it are well defined
 - ▶ **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 is the trusted computing base, i.e. what is trusted
 - ▶ What are the capabilities of the adversaries? **Threat model**

Crypto's famous cast

► <http://cryptocouple.com/>

The good players

Alice



Bob



The bad players

Eve

Eavesdropper

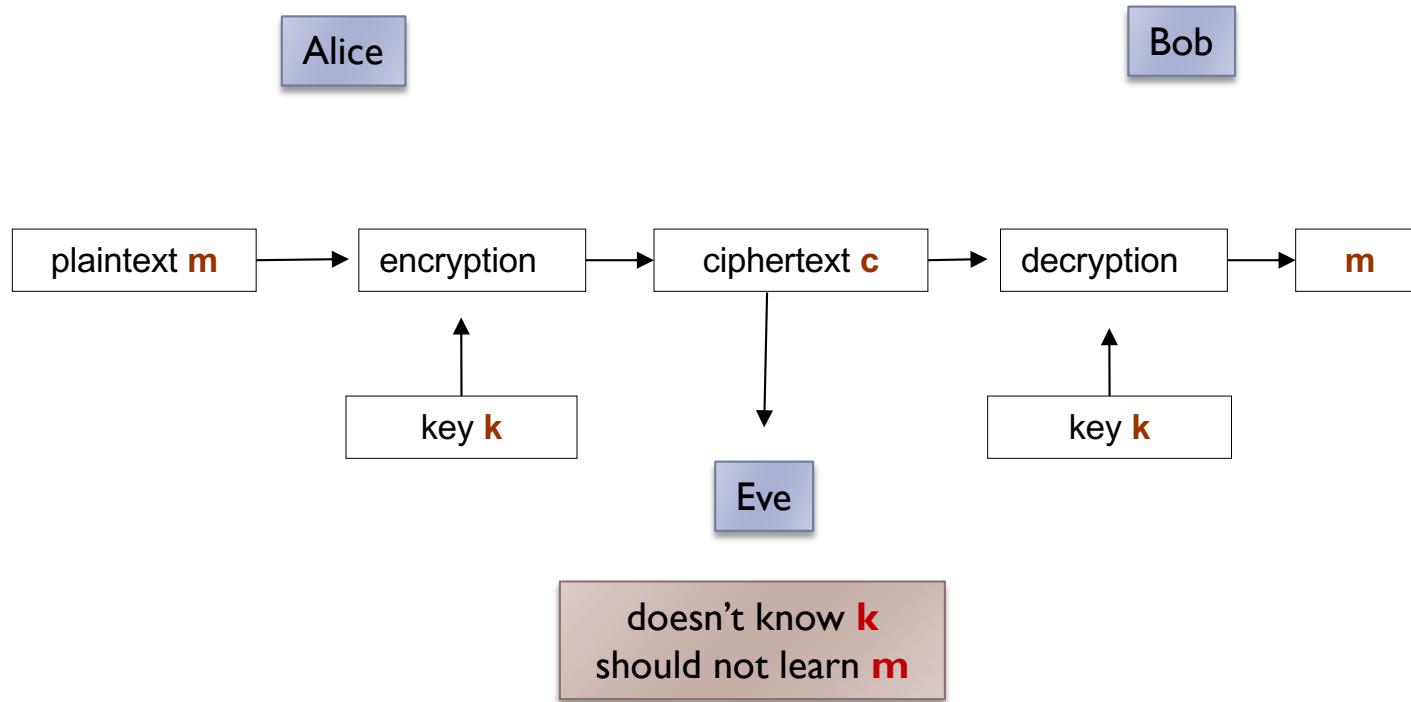


Mallory
Malicious



1978

Encryption terminology



Encryption scheme (cipher, cryptosystem) = encryption & decryption

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



Dutch linguist and cryptographer
1835 – 1903

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

- I. It is unrealistic to assume that the design details remain secret. Too many people need to know
 - I. One of the Enigma designs was sold, other machines were captured
2. Software/hardware can be **reverse-engineered!**
3. Pairwise-shared keys are easier to **protect, generate** and **replace**.
4. 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 is referred to as
``**security by obscurity**''.

Attacker threat model

I. Knowledge about the cipher (cryptosystem)

- ▶ **Kerchhoff's Principle**

- ▶ A cryptosystem should be secure even if everything about the system, except the key, is public knowledge
- ▶ Attacker is assumed to have full knowledge of the chosen cryptographic algorithm; **No security through obscurity**

2. Interaction with messages and the protocol

- ▶ **Passive:** only observes and attempts to decrypt messages
 - ▶ Only threatens confidentiality
- ▶ **Active:** observes, modifies, injects, or deletes messages
 - ▶ Threatens confidentiality, integrity, and authenticity

Attacker threat model (2)

3. Interaction with the encryption 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. Resources available (storage and/or **computation**)

- ▶ Unlimited resources
- ▶ Finite resources – Computational security
 - ▶ to calculate, typically polynomial running time
 - ▶ to store things



Classical cryptography

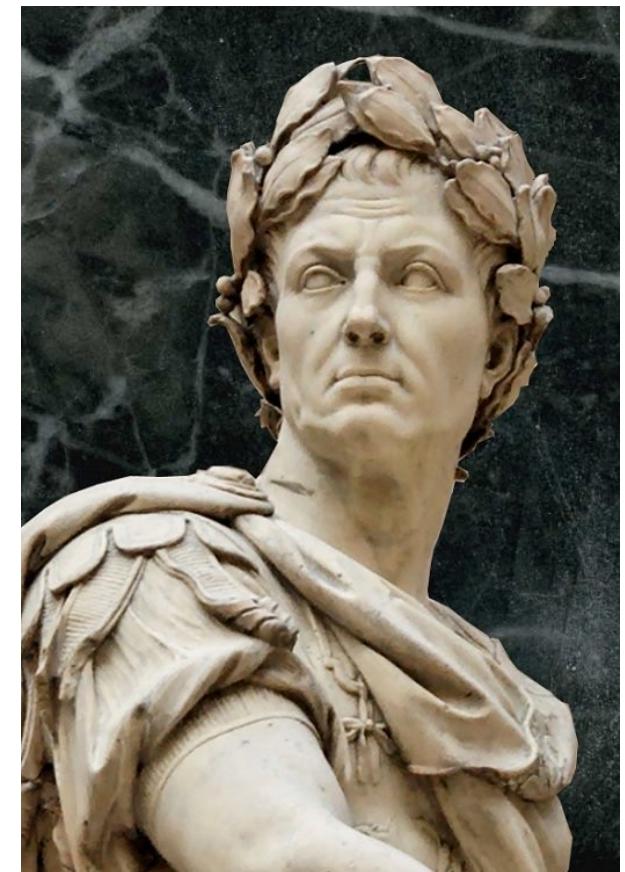
Caesar Shift Cipher

- ▶ Symmetric substitution cipher
 - ▶ Key is a number k ; (*for Caesar shift $k=3$*)
 - ▶ 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 EUWXV EULQJ D NQLIH WR WKH SDUWB



A mathematical view

\mathcal{K} – key space

\mathcal{N} - natural numbers

\mathcal{M} – plaintext space

\mathcal{C} - ciphertext space

An **encryption scheme** is a pair $(\mathbf{Gen}, \mathbf{Enc}, \mathbf{Dec})$, where

- **Gen** : $\mathbf{N} \rightarrow \mathbf{K}$ is a **key generation** algorithm,
- **Enc** : $\mathbf{K} \times \mathbf{M} \rightarrow \mathbf{C}$ is an **encryption** algorithm,
- **Dec** : $\mathbf{K} \times \mathbf{C} \rightarrow \mathbf{M}$ is an **decryption** algorithm.

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

Correctness

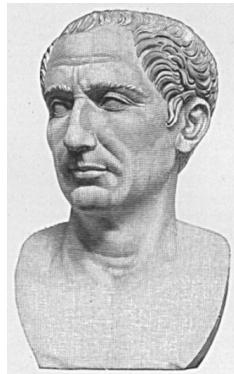
for every k, m we should have $\mathbf{Dec}_k(\mathbf{Enc}_k(m)) = m$.

Shift cipher: Mathematical View

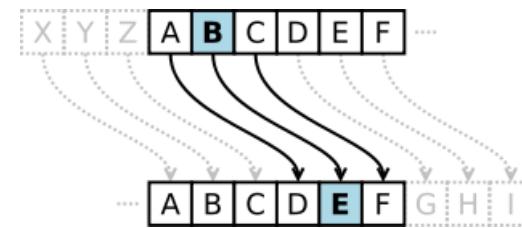
\mathcal{M} = words over alphabet $\{A, \dots, Z\} \approx \{0, \dots, 25\}$

$\mathcal{K} = \{0, \dots, 25\}$

$$\text{Enc}_k(m_1, \dots, m_n) = (m_1 + k \bmod 26, \dots, m_n + k \bmod 26)$$



Cesar: $k = 3$



$$\text{Dec}_k(c_1, \dots, c_n) = (c_1 - k \bmod 26, \dots, c_n - k \bmod 26)$$

Security of the shift cipher

How to break the shift cipher?

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

What is the decryption of FWPG?

Let **c** be a ciphertext.

For every $k \in \{0, \dots, 25\}$ check if $\text{Dec}_k(c)$ “makes sense”.

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

Thus $\text{Dec}_k(c)$ is the message.

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

How many keys are possible in the alphabet above?

Moral: the key space needs to be large!

Monoalphabetic substitution cipher

- ▶ Replace each letter X with $\pi(X)$ 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
 $\pi =$ 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

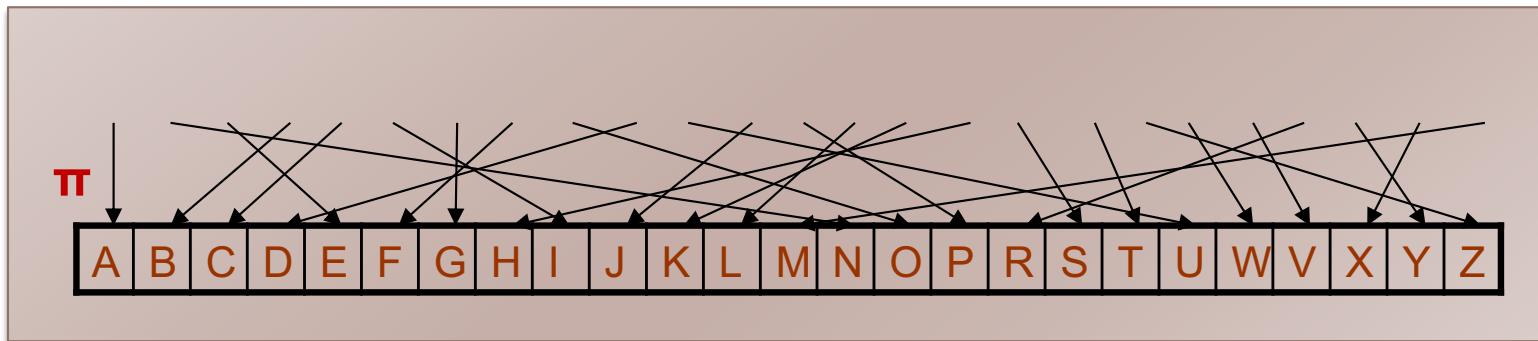
HELLO WORLD
↓
 π
YZXXT PTMXO

Substitution cipher

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

\mathcal{M} = words over alphabet {A,...,Z} $\approx \{0,...,25\}$

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



$$\text{Enc}_\pi(m_1, \dots, m_n) = (\pi(m_1), \dots, \pi(m_n))$$

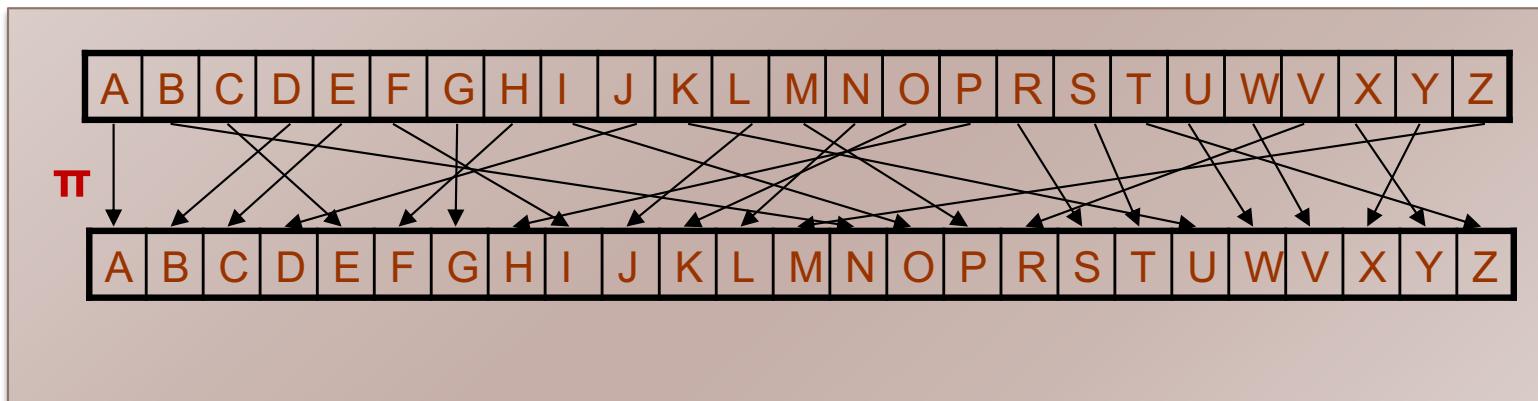
$$\text{Dec}_\pi(c_1, \dots, c_n) = (\pi^{-1}(c_1), \dots, \pi^{-1}(c_n))$$

Example substitution cipher

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

\mathcal{M} = words over alphabet {A,...,Z} $\approx \{0,...,25\}$

\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 ($26! = 4 \times 10^{26}$ possible keys)
 - ▶ Broken by frequency analysis
 - ▶ Al-Kindi 800 AD

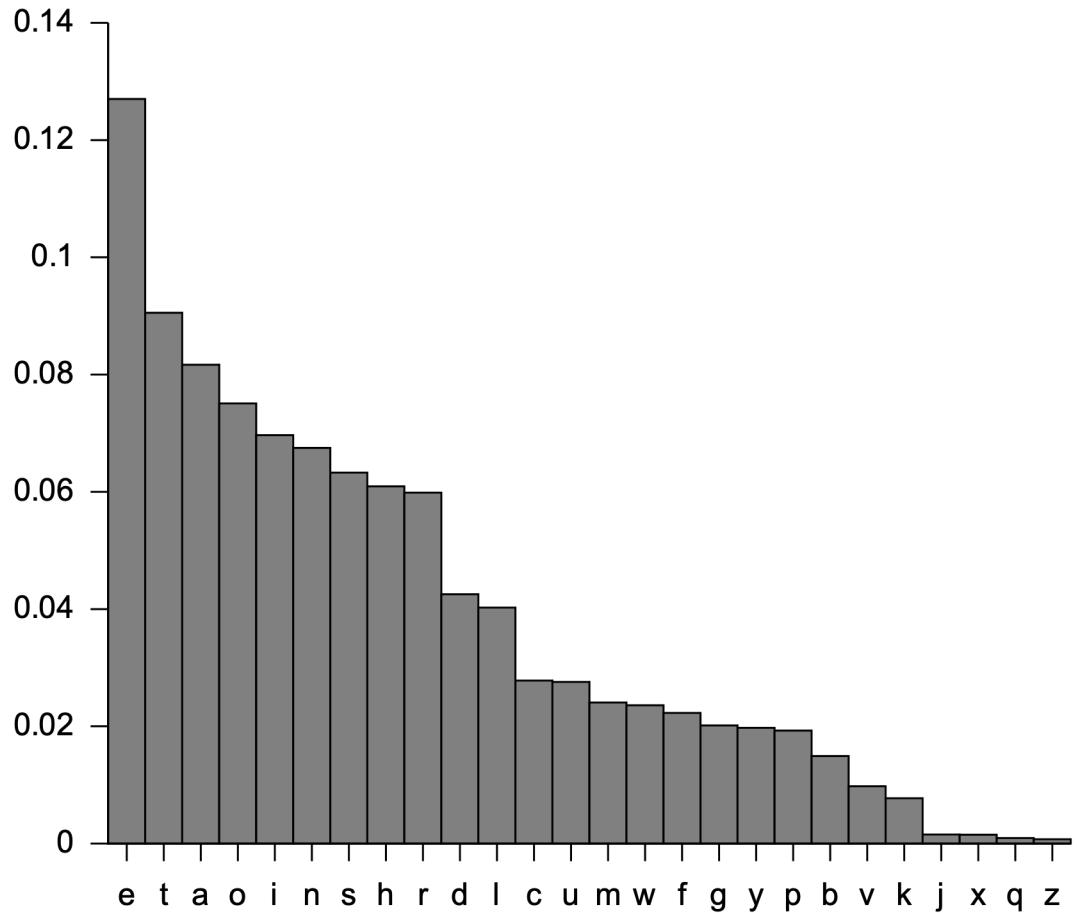
مداده - ولله در العالی رحیمه اندیشم میر محمد ولی الله

الله اول الا سر الراجم رحمة الله وحده
رسالة الرسالة يعود الى اخر الدرر استفتح العرش بالراجم
فهي رسالة ربانية لغرض حل مأزق زمانه وكانت تصرح بـ احلام الراجم
الكل العجماء وللصلوة والتعوذ من الشيطان: فلما رأى الله الرسالة سار على الناس العظيم
عندها لفظ اسلامي: اسلامي واحمد لله رب العالمين انت تحيي الناس العجم وسرد الفضل العجم
الاسلامي واسمع عن دار الاسلام وبعدها: فلما رأى الله العجم ارتضى له ابا الحسن العجم

Frequency analysis

- ▶ Human languages have patterns
 - ▶ Frequency of letter usage
 - ▶ Frequency of n -letter combinations (bigrams, trigrams)
- ▶ These patterns survive substitution
- ▶ To decipher, map the letters from the ciphertext based on their frequency

Letter frequency for English



Example of frequency analysis

- ▶ Language with 3 letters: a b c, with frequency of 0.2, 0.3, 0.5
- ▶ Assume key is permutation (a b c) -> (c b a)
- ▶ Now consider that you are given an encrypted text E
- ▶ Compute f the frequency of the letters in E, how many times a letter appeared/total characters in E text
- ▶ We will probably get f_a close to 0.5, f_b close to 0.3 and f_c close to 0.2; *This will indicate that the permutation left b unchanged and switched a with c, so key is (c,b,a)*
- ▶ Notes:
 - ▶ More encrypted text is better, less will make it hard to guess
 - ▶ In real languages the difference between the frequency of letters are not so large so it will take many iterations and guessing

Cryptanalysis of monoalphabetic substitution: lessons

- ▶ Having more encrypted text helps with frequency analysis, so limit data encrypted with the same key
- ▶ Use large blocks of data: 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 $K = (k_1, k_2, \dots, k_m)$,
 - ▶ 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 $\{A, \dots, Z\} \approx \{0, \dots, 25\}$

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

$$\begin{aligned}\text{Enc}_k(m_1, \dots, m_n) = & (m_1 + k_1, \dots, m_t + k_t, \\ & m_{t+1} + k_1, \dots, m_{2t} + k_t, \\ & \dots \\ &) \bmod 26\end{aligned}$$

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: CRYPTOGRAPHY

Key: LUCKLUCKLUCK

Ciphertext: NLAZEIIIBLJJI

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 **Kasiski test** or **index of coincidence**
 2. Divide the ciphertext into x shift cipher encryptions
 3. Use frequency analysis on each shift cipher



Kasisky Test

Plaintext	T H E S U N A N D T H E M A N I N T H E M O O N
Key	K I N G K I N G K I N G K I N G K I N G K I N G
Ciphertext	D P R Y E V N T N B U K W I A O X B U K W W B T

Distance = 8

- ▶ 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 Kasiski 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 would not be breakable

One Time Pad (1920s)

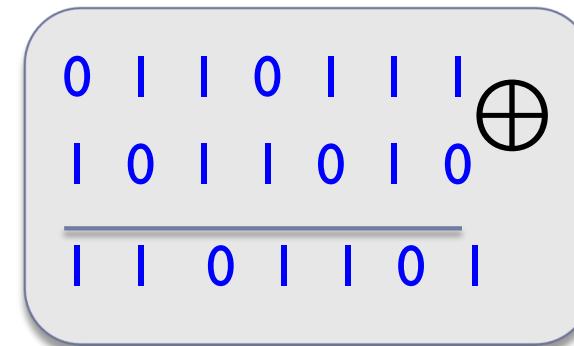
- ▶ Fixes the vulnerability of the Vigenère cipher by using very long keys
- ▶ Key is a random string that is at least as long as the plaintext
- ▶ Similar encryption as with Vigenère (different shift per letter)



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



One-time pad

ℓ – a parameter
 $\mathcal{K} = \mathcal{M} = \{0,1\}^\ell$

component-wise **xor**

Vernam's cipher:

$$\begin{aligned}\text{Enc}_k(m) &= k \oplus m \\ \text{Dec}_k(c) &= k \oplus c\end{aligned}$$



Gilbert Vernam
(1890 – 1960)

Correctness:

$$\text{Dec}_k(\text{Enc}_k(m)) = k \oplus (k \oplus m) = m$$

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

consider the following experiment

(\mathbf{m} – a message)

1. the key \mathbf{K} is chosen uniformly at random
2. $\mathbf{C} := \mathbf{Enc}_{\mathbf{K}}(\mathbf{m})$ is given to the adversary

how to
define
security

?

Idea 1

(m – a message)

1. the key K is chosen uniformly at random
2. $C := \text{Enc}_K(m)$ 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”:

$$\text{Enc}_K(m) = m$$

satisfies this definition!



Idea 2

(m – a message)

1. the key K is chosen uniformly at random
2. $C := \text{Enc}_K(m)$ 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)

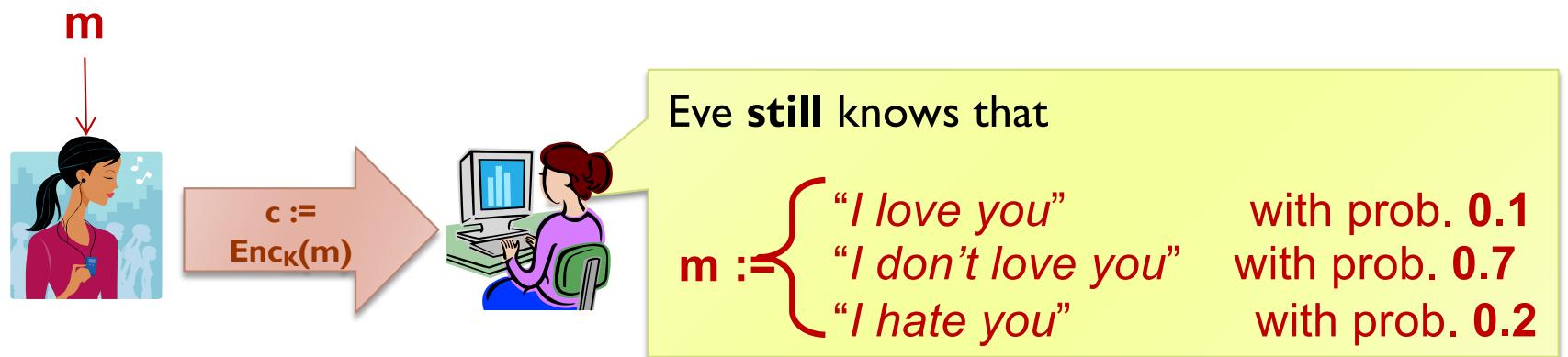
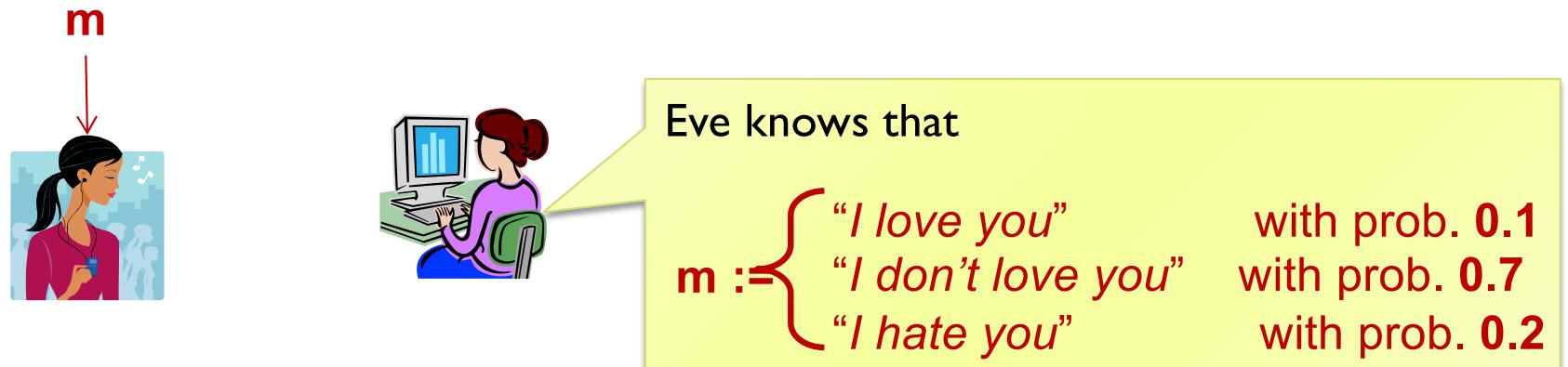
1. the key K is chosen uniformly at random
2. $C := \text{Enc}_K(m)$ is given to the adversary

An idea

“The adversary should not learn any information about
 m .”

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

Example



“The adversary should not learn any information about m .”

An encryption scheme is **perfectly secret** if
for every distribution of M
and every $m \in \mathcal{M}$ and $c \in C$
 $\Pr[M = m] = \Pr[M = m | C = c]$

Ciphertext-only attack (passive)
Unlimited computational power

In English

- ▶ The adversary believes the probability that the plaintext is m is $\Pr(M=m)$ **before seeing the ciphertext**
 - ▶ Maybe they are very sure, or maybe they have no idea
- ▶ The adversary believes the probability that the plaintext is m is $\Pr(M=m \mid C=c)$ **after seeing that the ciphertext is c**
- ▶ $\Pr(M=m \mid C=c) = P(M= m)$ means that after knowing that the ciphertext is c , the adversary's belief does not change
 - ▶ Intuitively, the adversary learned **nothing** from the ciphertext

NOTE: there are no computational assumptions about the attacker, this is why this is also called unconditional security or perfect security

Put Another Way

- ▶ Imagine you have a ciphertext c where the length $|c| = 1000$
- ▶ I can give you a key k_i with $|k_i| = 1000$ such that:
 - ▶ The decrypted message m_i is the first 1000 characters of Hamlet
- ▶ Or, I can give you a key k_j with $|k_j| = 1000$ such that:
 - ▶ The decrypted message m_j is the first 1000 characters of the US Constitution
- ▶ If an algorithm offers perfect secrecy then:
 - ▶ For a given ciphertext of length n
 - ▶ All possible corresponding plaintexts of length n are possible decryptions

Is Shift Cipher Perfectly Secure?

An encryption scheme is **perfectly secret** if

for every distribution of \mathbf{M}

and every $\mathbf{m} \in \mathcal{M}$ and $\mathbf{c} \in \mathcal{C}$

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

- Perfectly secure for 1 letter message:
 - $\Pr[\mathbf{M} = \mathbf{m}] = 1/26$
 - $\Pr[\mathbf{M} = \mathbf{M} \mid \mathbf{C} = \mathbf{c}] = \Pr[\mathbf{K} = \mathbf{c} - \mathbf{m} \bmod 26] = 1/26$

- ▶ Counterexample (2-letter message):
 - ▶ $\mathbf{M}_1 = \text{AB}; \mathbf{M}_2 = \text{AZ}; \mathbf{c} = \text{BC}$
 - ▶ $\Pr[\mathbf{M} = \mathbf{M}_1 \mid \mathbf{C} = \mathbf{c}] = \Pr[\mathbf{k} = 1] = 1/26$
 - ▶ $\Pr[\mathbf{M} = \mathbf{M}_2 \mid \mathbf{C} = \mathbf{c}] = 0$

Cryptanalysis of OTP

- ▶ Intuitively, the key is random, so ciphertext is also random (because of properties of XOR)
- ▶ **OTP achieves Perfect Secrecy**
 - ▶ Shannon or Information Theoretic Security
 - ▶ Basic idea: ciphertext reveals no “information” about plaintext
- ▶ **Caveats**
 - ▶ If the length of the OTP key is less than the length of the message...
 - ▶ It's not a OTP anymore, not perfectly secret!
 - ▶ If you reuse the OTP key...
 - ▶ It's not a OTP anymore, not perfectly secret!
- ▶ **Major issue with OTP in practice?**
 - ▶ How to securely distribute the key books to both parties

Why the one-time pad is not practical?

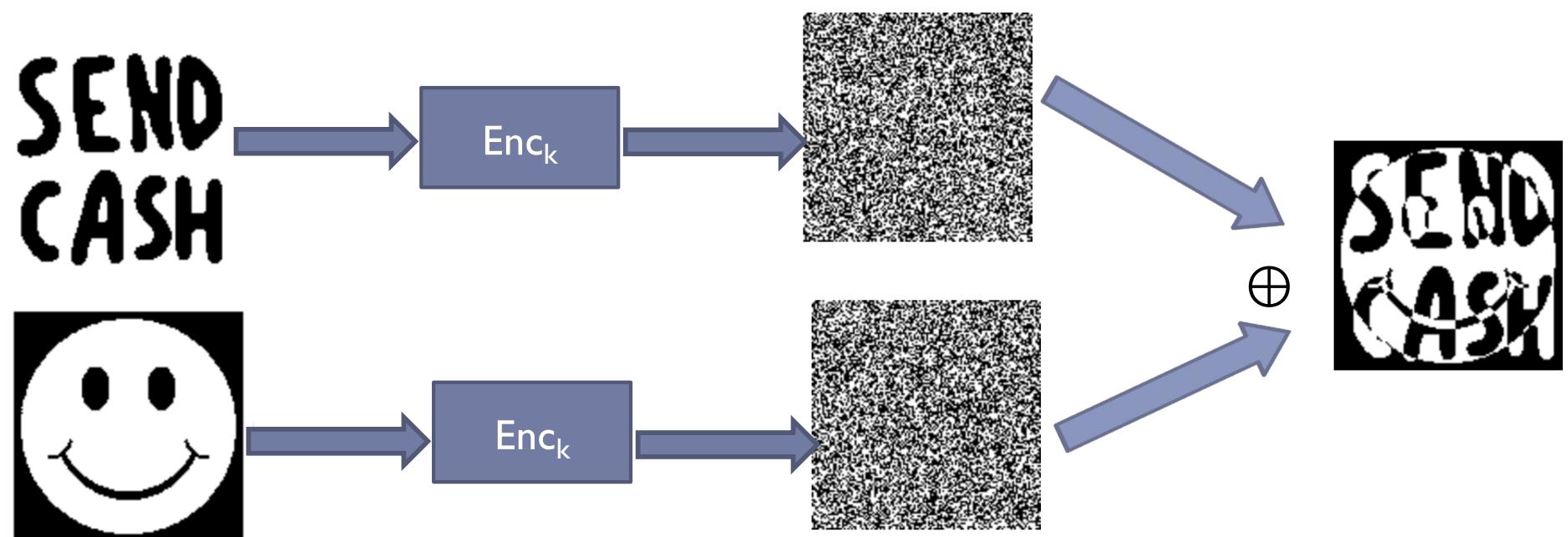
1. **The key is as long as the message.**
2. **The key cannot be reused.**
3. **Alice and Bob must share a new key every time they communicate**

All three are necessary for perfect secrecy!

This is because:

$$\begin{aligned}\mathbf{Enc}_k(m_1) \text{ xor } \mathbf{Enc}_k(m_2) &= (k \text{ xor } m_1) \text{ xor } (k \text{ xor } m_2) \\ &= m_1 \text{ xor } m_2\end{aligned}$$

Example: key reuse



Venona project (1946 – 1980)



American **National Security Agency** decrypted **Soviet** messages that were transmitted in the 1940s.

That was possible because the Soviets reused the keys in the one-time pad scheme.

Ethel and Julius Rosenberg

Key takeaways

- ▶ **Historical methods for encryption are not secure**
 - ▶ Shift cipher, mono-alphabetic substitution cipher, Vigenere
 - ▶ Attacks: Brute force (small key space), frequency analysis
- ▶ **Defining security for encryption is difficult**
 - ▶ Perfect secrecy is one of the first rigorous notion of security
- ▶ **One-time pad is perfectly secure for an eavesdropper that has unbounded computation power**
 - ▶ But many practical drawbacks
 - ▶ Still has been used in critical military applications
- ▶ **Modern cryptography relies on computational assumptions to become practical**
 - ▶ E.g., it is computationally hard to factor large numbers; adversary has limited computational resources