Introduction

CS 5158/6058 Data Security and Privacy

Spring 2018

Instructor: Boyang Wang

About Me

- University of Cincinnati, since Fall 2017
- Ph.D. ECE, The University of Arizona, August 2017
- Ph.D. Crypto, Xidian University, June 2014

- Tucson AZ
- Logan UT
- Pittsburgh PA
- Toronto, Canada
- Xi’an, China

About Me

- After *Ice and Fire*, and I survived!!!

About This Course

- CS 5158/6058 Data Security and Privacy
- Time: TuTh 12:30pm - 1:50pm
- Location: Baldwin 645

- Instructor: Boyang Wang
- Email: boyang.wang@uc.edu
- Office: ERC 532
- Office Hours: Tu 2:30pm - 4:30pm (or by appointment)

Textbooks

- Textbooks
  - *Introduction to Modern Cryptography (By Drs. J. Katz and Y. Lindell, 2nd edition, recommended)*
  - *The Joy of Cryptography (By Dr. M. Rosulek, free & available online)*

- Prerequisites (by topics)
  - Probability
  - Programming (C/C++, Python or Java)

Topics

- Fundamental Crypto Techniques (7 weeks)
  - E.g. encryption, signatures, hash functions
  - *Covered by textbooks & slides*

- Advanced Topics in Data S&P (7 weeks)
  - E.g., differential privacy, crypto currency, searchable encryption
  - *Covered by slides & additional references*

Assignments & Exams

- No midterm or final exams

- 3 individual programming assignments (30%)
- 5 homeworks (30%)

- 1 group programming assignment (20%)
- 1 final group project (20%)
  - A presentation & a final paper
  - Presentations will be held in Week 13 & 14.

Encryption

(hiding information)

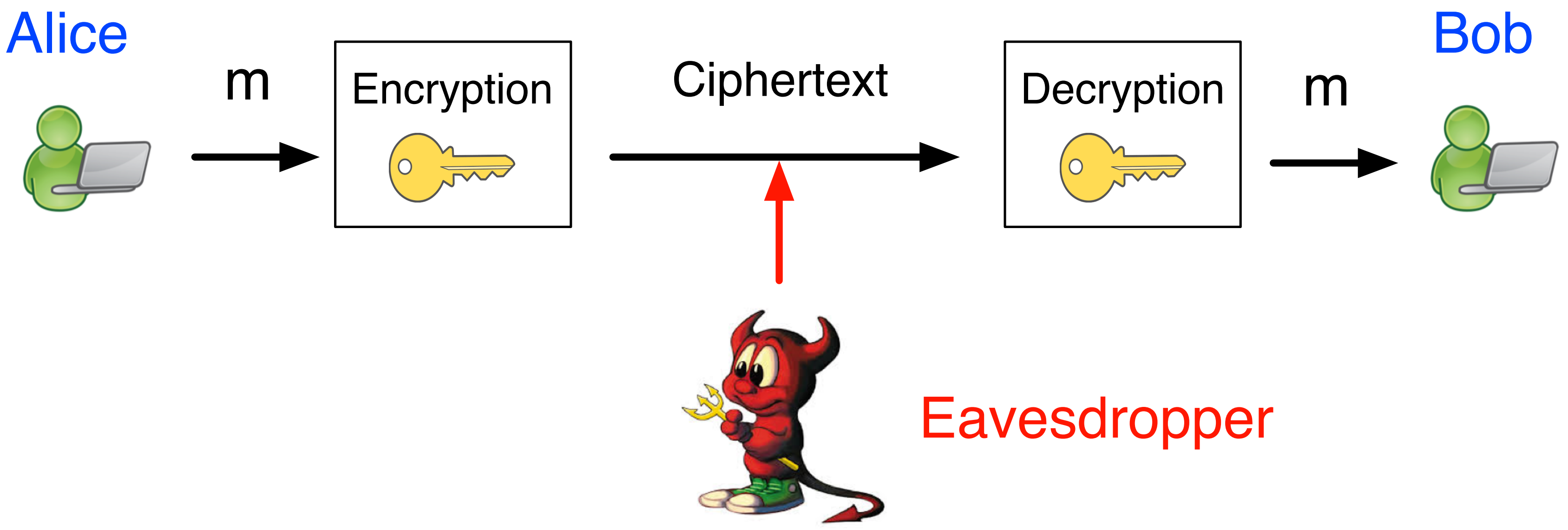
Why Do We Need Encryption?

- Channels are public.
  - Classroom (broadcast)
  - Social networks (Facebook, Twitter), Emails
  - Other examples?

- Communications are private.
  - I won $1,000,000 in Vegas!
  - Social security numbers
  - Your final grade in this course

Encryption Model

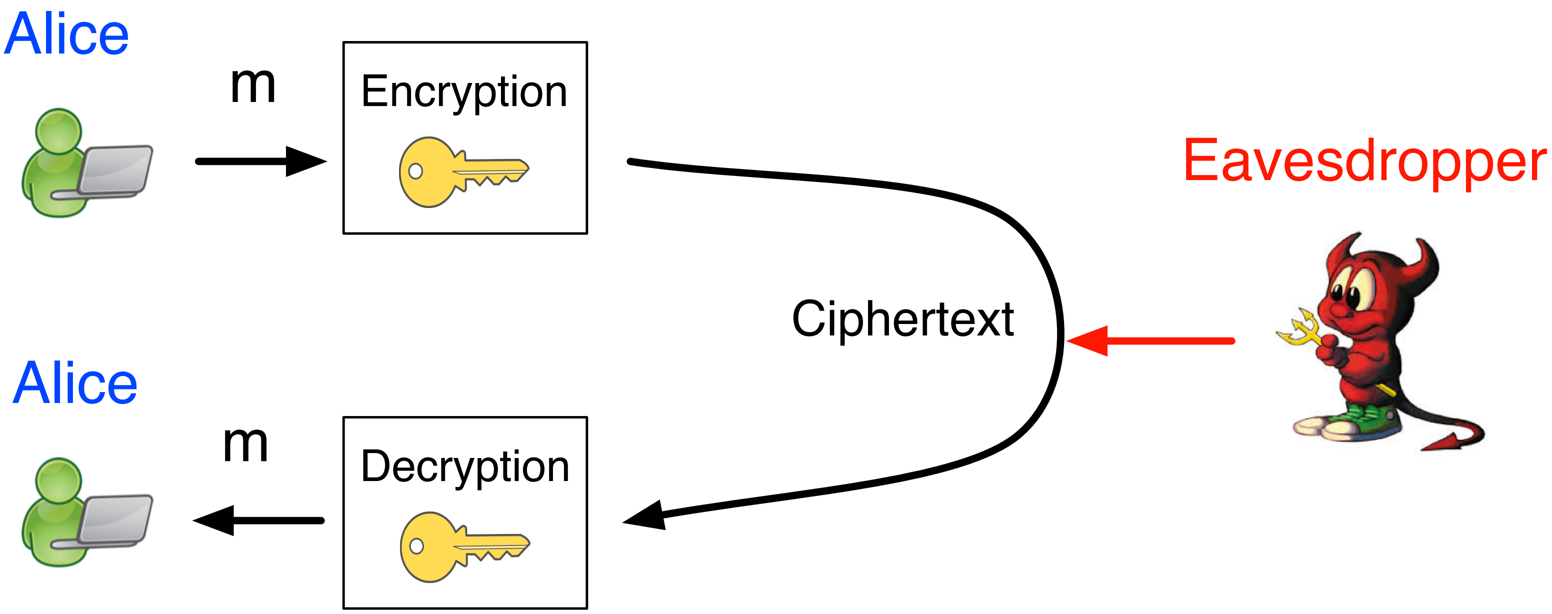
- Alice, Bob, Eavesdropper



- Plaintext **m**, ciphertext **c**, key **k**
- Alice and Bob share key k in advance.

Encryption Model

- Alice, Alice, Eavesdropper



- Alice keeps key k private.

Algorithms

- KeyGen: a probabilistic algorithm that outputs a key k
- Enc: takes a key k and a plaintext (message) m as input, and outputs a ciphertext c
- Dec: takes a key k and a ciphertext c as input, and outputs a plaintext m

Write as $\mathsf{Enc}_k(m)$, $\mathsf{Dec}_k(c)$

Correctness

- Symmetric-Key Encryption
    - Enc and Dec use a same key

- Correctness

For every key $k \in \mathcal{K}$ output by KeyGen and every message $m \in \mathcal{M}$, it holds that

$$\mathsf{Dec}_k(\mathsf{Enc}_k(m)) = m$$

Kerckhoffs's Principle

- Auguste Kerckhoffs (Dutch, 19th century)

> ***The cipher method must not be required to be secret, and it must be able to fall into the hands of the enemy without inconvenience.***

- Security rely solely on secrecy of the key
  - Enc algorithms can be public
  - Change keys is easier than changing algos
  - Increase key length is easier

Historical Ciphers

(The ones that are not secure)

Shift Cipher

- Plaintext in lower case, ciphertext in UPPER CASE
- Each char (in `a`:0~`z`:25) shifts to right by a key

```
abcdefghijklmnopqrstuvwxyz
EFGHIJKLMNOPQRSTUVWXYZABCD
```

- Enc: `a` + 4 = `E` (or 0 + 4 = 4)
- Dec: `E` - 4 = `a` (or 4 - 4 = 0)
- Key: 4
- **`data`** —> **`HEXE`**

Shift Cipher

- Is correct, but is it secure? No!
- Key space is small.
  - Brute-force attack (try every possible key)
  - Only has 26 keys, try each one and check the results of which key "make sense"

  - Example: `HEXE`
    - shift left by 1: `gdwd`; shift left by 2: `fcvc`; shift left by 3: `ebub`; shift left by 4: `data`

Shift Cipher

- Practice: given ciphertext **EKPEKPPCVK**, recover the key and plaintext using brute-force attacks.
  - Shift left by 1: **djodjoobuj**
  - Shift left by 2: **cincinnati**

	abcdefghijklmnopqrstuvwxyz
left by 1:	BCDEFGHIJKLMNOPQRSTUVWXYZA
left by 2:	CDEFGHIJKLMNOPQRSTUVWXYZAB

- We need a large key space!
  - Necessary but not sufficient

Substitution Cipher

- A char in plaintext maps to a char in ciphertext.
- One-to-one mapping (bijection)

```
abcdefghijklmnopqrstuvwxyz
EXAUNDKBMVORQCSFHYGWZLJITP
```

- Enc: `b` <—> `X`
- Dec: `X` <—> `b`
- Key: a permutation
- `data` —> `UEWE`

Substitution Cipher

```
abcdefghijklmnopqrstuvwxyz
EXAUNDKBMVORQCSFHYGWZLJITP
```

- Practice 1: What is the ciphertext of “`drmarccahay`”
- Answer: `UYQEYAAEBET`

- Practice 2: What is the size of key space?
- Answer: 26! (approximately 2^{88}, brute force is hard)

Substitution Cipher

- Is correct, has large key space, is it secure? No!

- What can we learn from Practice 1?

```
drmarccahay
UYQEYAAEBET
```

- Leak Frequency!
    - Enc is deterministic (`a` always outputs `E`)

Frequency Leakage

- Frequency distribution of English chars is known.

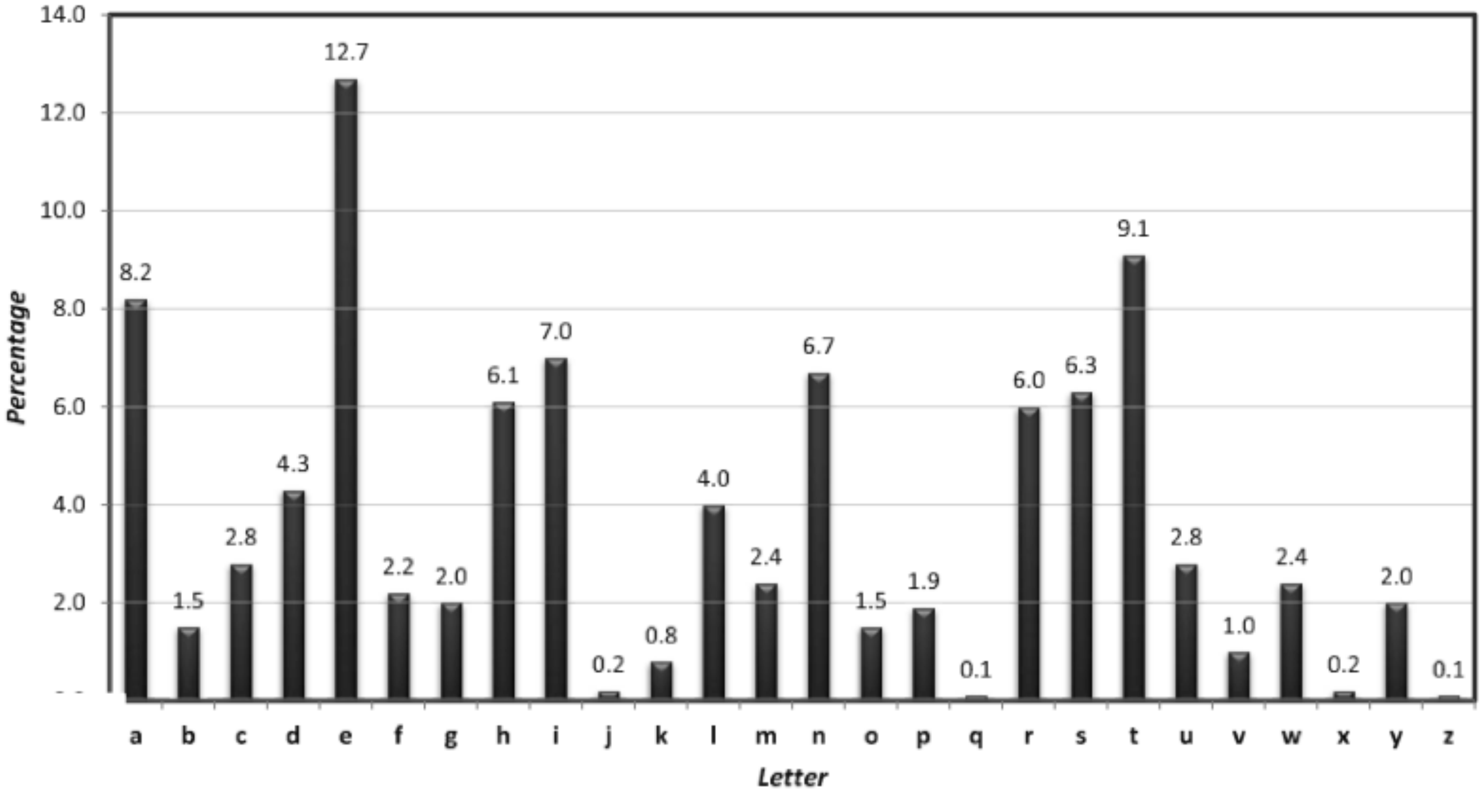


FIGURE 1.3: Average letter frequencies for English-language text.

This figure is from K&L textbook

Frequency Leakage

- Assume ciphertext is very long

- Count the frequency of each char in ciphertext
  - If count(**`N`**) is the greatest (i.e. around 13% of the length of ciphertext), than **`N`** <—> **`e`**
  - Some guess may need more tries

- We need to hide frequency!

Vigenere Cipher

- Preserve frequency, e.g. `r` could map to `F` or `X`
- Several independent instances of Shift Cipher
- Key is a string, e.g. `gouc`

`plaintext:`	`drmarccahy`
`key:`	`goucgoucgo`
`ciphertext:`	`JFGCXQWCNM`

  - Enc: 3 (`d`) + 6 (`g`) = 9 (`J`)
  - Dec: 9 (`J`) - 6 (`g`) = 3 (`d`)
  - 1st, 5th, 9th chars are encrypted by “g”(“6”)

Vigenere Cipher

- <u>Practice:</u>

a	b	c	d	e	f	g	h	i	j	k	l	m
0	1	2	3	4	5	6	7	8	9	10	11	12
n	o	p	q	r	s	t	u	v	w	x	y	z
13	14	15	16	17	18	19	20	21	22	23	24	25

```
plaintext:  security

key:        goucgouc

ciphertext: ????????
```

Vigenere Cipher

- Practice:

a	b	c	d	e	f	g	h	i	j	k	l	m
0	1	2	3	4	5	6	7	8	9	10	11	12
n	o	p	q	r	s	t	u	v	w	x	y	z
13	14	15	16	17	18	19	20	21	22	23	24	25

security	18	4	2	20	17	8	19	24
goucgouc	6	14	20	2	6	14	20	2
????????	24	18	22	22	23	22	13	0

- Answer: Y S W W X W N A

Vigenere Cipher

- Key space **IKI=26**^{t}, t is the length of a key string
    - t independent instances of Shift Ciphers
    - If t = 20, |K| is approximately 2^{94}

- Is correct, key space could be large and frequency is preserved, is it secure? No!

Historical Cipher

- Practice: if the message space has 3000 different characters:
  - What is the size of key space for Shift?
  - What is the size of key space for Substitution?
  - If key length is 4, what is the size of key space for Vigenere?

- Shift: 3000; Sustitution: 3000!; Vigenere: 3000^4

Additional Reading

Chapter 1, *Introduction to Modern Cryptography, Drs. J. Katz and Y. Lindell, 2nd edition*