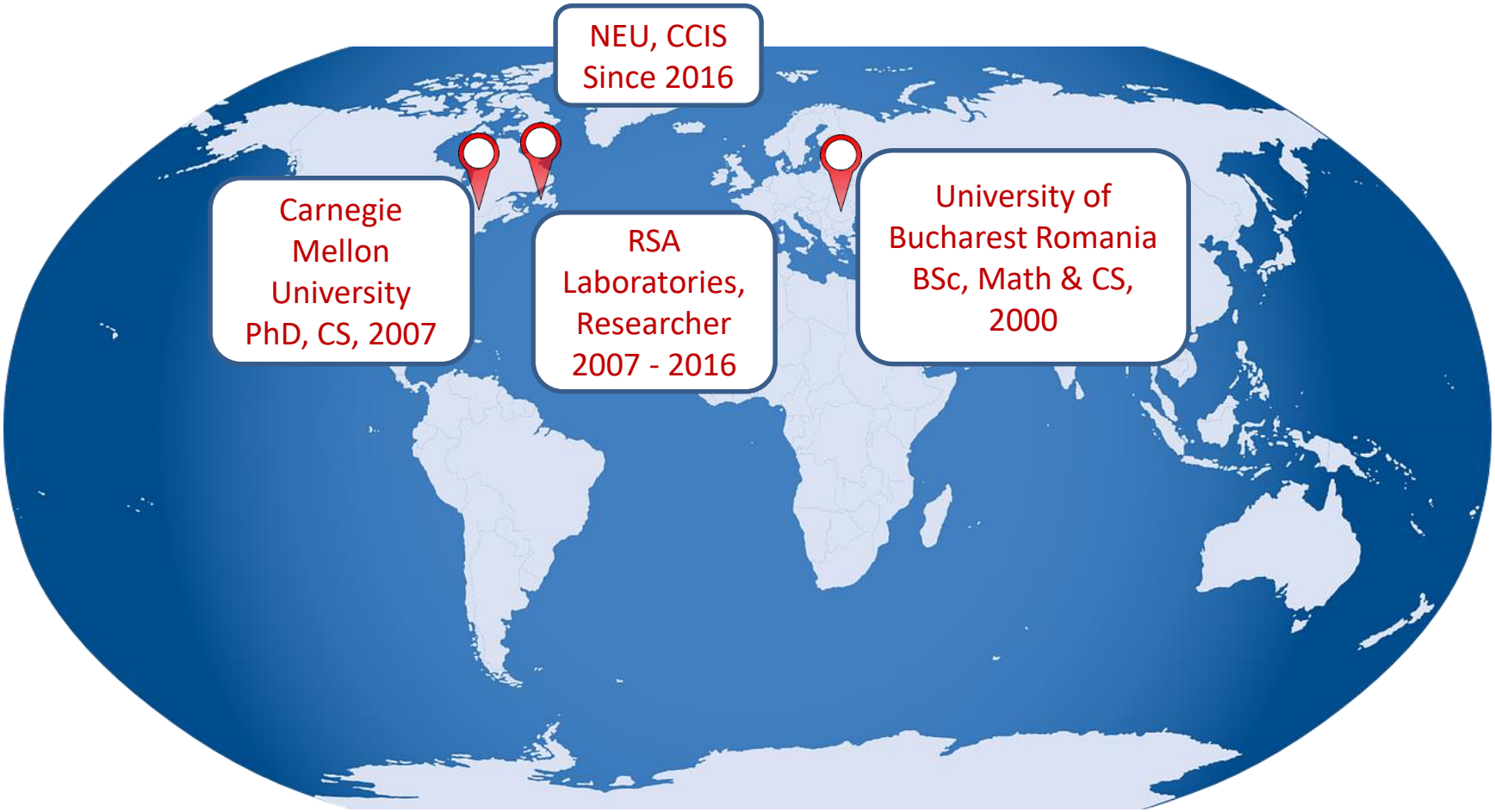
CS 4770: Cryptography

CS 6750: Cryptography and Communication Security

Alina Oprea
Associate Professor, CCIS
Northeastern University

January 8 2018

Introductions



Background

- Ph.D. at CMU
  - Research in storage security
- RSA Laboratories
  - Cloud security, applied cryptography, security analytics
  - Worked with Prof. Rivest (“R” in RSA)
  - Practical research in industry lab
- NEU CCIS – joined Fall 2016
  - Machine learning for security
  - Adversarial machine learning
  - Privacy-preserving analytics

Class Introductions

- Graduate class – enrollment 22
- Undergraduate class – enrollment 6

Cryptography

- What is *cryptography*?
  - *The art of writing or solving codes* (Concise Oxford Dictionary)
  - Historically adopted only by military organizations and governments
- Modern cryptography
  - Ubiquitous on the Internet
  - Secure communication and systems across the globe
  - *The study of mathematical techniques for securing digital information, systems, and distributed computations against adversarial attacks*

Historical cryptography

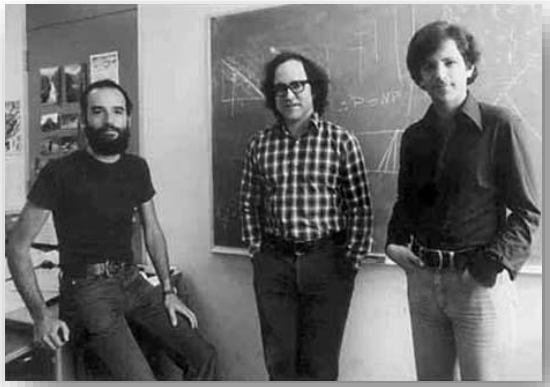
Cryptography ≈ Encryption

Main applications: **military and diplomacy**

ancient times → **world war II**

Modern cryptography

Cryptography based on rigorous science/math



information theory

public-key cryptography

signature schemes

rigorous definitions

multiparty-computations

zero-knowledge

threshold crypto

electronic auctions

electronic voting

crypto currencies

private info retreival

computation in cloud

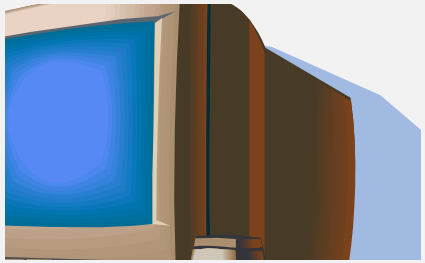
...

post-war **sevenites** **now**

What happened?

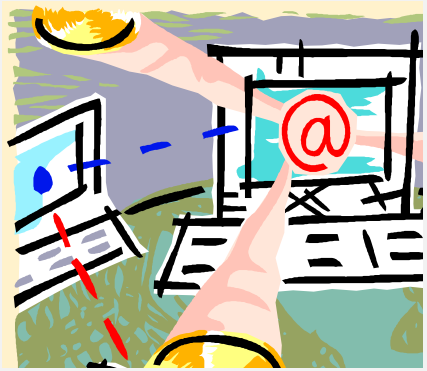
Technology

Afforadable
Increased computational power
Cloud computing



Demand

Companies and individuals start to do business electronically



Theory

Information theory
+
computational complexity

Can reason about security in a **formal way**.

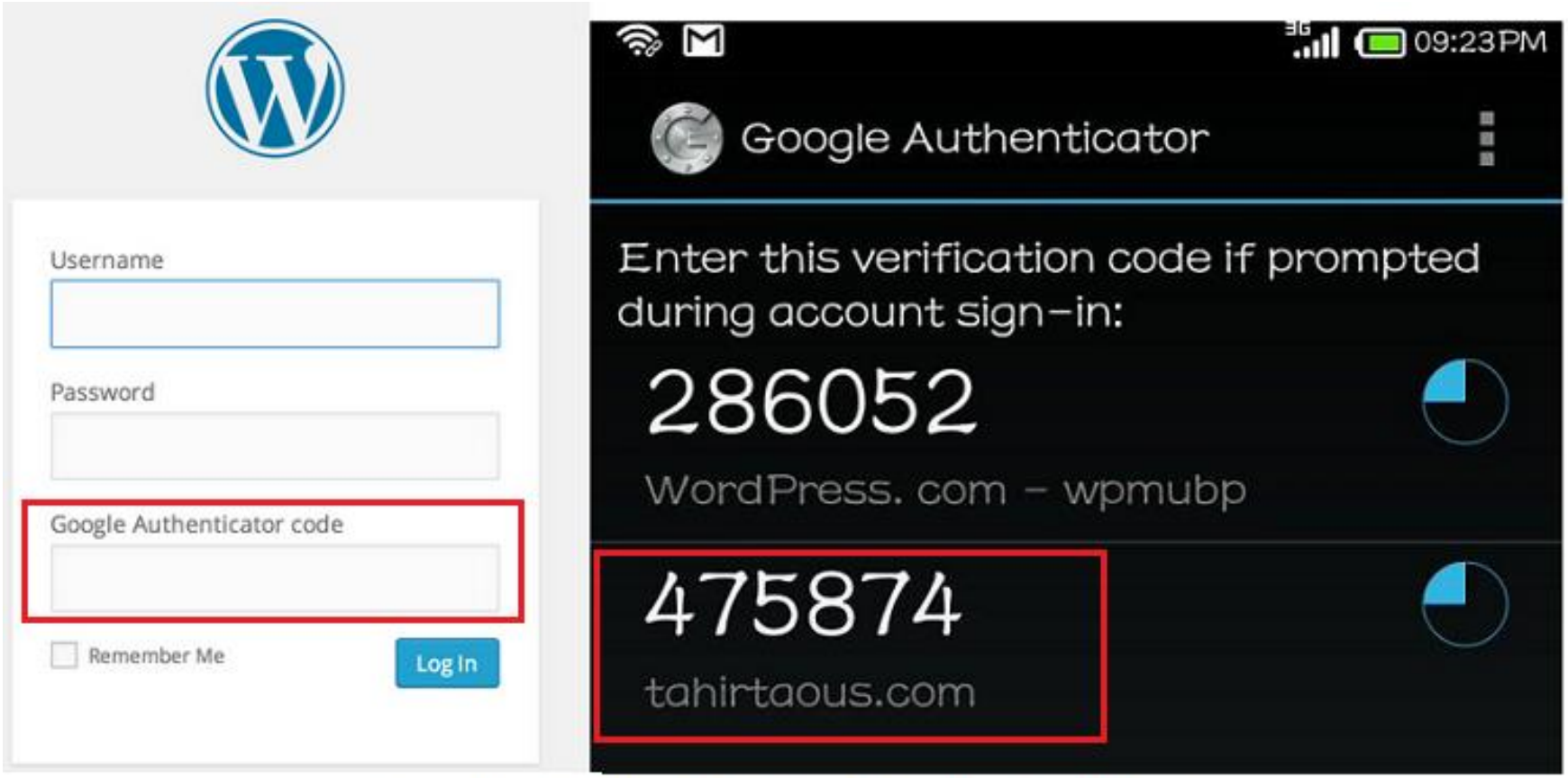
Modern cryptography

- Rigorous definitions of what it means to have secure encryption, signatures, authentication
- Precise assumptions
  - Factoring large numbers is “computationally hard”
  - Vetted over years
- Proofs of security
  - Construction satisfies definition under assumption
- Challenges
  - Create realistic models of adversarial capabilities
  - Find minimal assumptions
- Enables a number of emerging applications

Electronic commerce

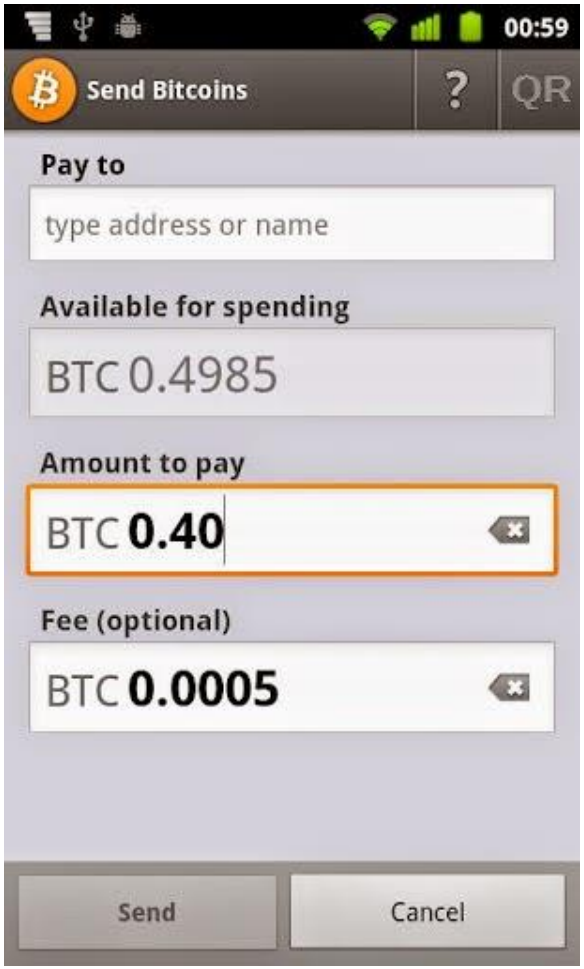
- https invokes the Secure Socket Layer (SSL) communication security protocol to securely transmit your credit card number to the server
- SSL uses cryptography

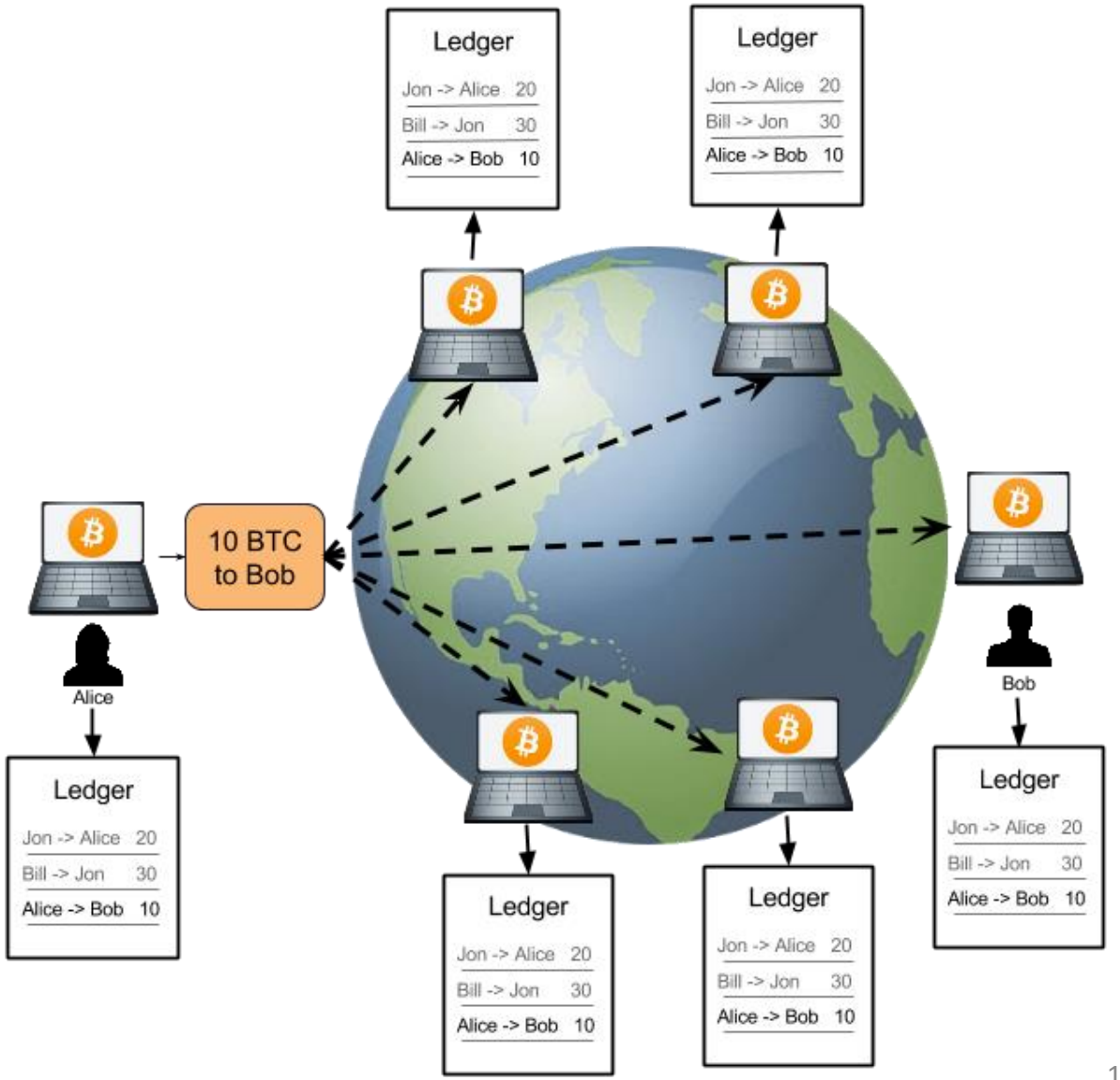
Multi-factor authentication



Google Authenticator WordPress Plugin

Crypto-currencies





Course objectives

- Introduction to basic cryptographic primitives
  - Secret-key cryptography
  - Public-key cryptography
  - Threat models
- Modern cryptographic protocol design
  - Sound, rigorous proofs of security
  - Understand fundamental assumptions
- Applications
  - Secure network communication, secure computation, crypto currencies

http://www.ccs.neu.edu/home/alina/classes/Spring2018

CS 4770, CS 6750: Syllabus

- Symmetric-key primitives
  - Block ciphers, symmetric-key encryption
  - Pseudorandom functions and pseudorandom generators
  - MACs and authenticated encryption
- Hash functions
  - Integrity schemes
- Public-key cryptography
  - Public-key encryption and signatures
  - Key exchange
- Applications
  - Secure network communication, secure computation, crypto currencies

Textbook: Introduction to Modern Cryptography.
J. Katz and Y. Lindell

Policies

- Instructors
  - Alina Oprea
  - TA: Sourabh Marathe
- Schedule
  - Mon, Thu 11:45am – 1:25pm, Robinson 107
  - Office hours:
    - Alina: Thu 4:00 – 6:00 pm (ISEC 625)
    - Sourabh: Tue 2-3pm (TBD)
- Your responsibilities
  - Please be on time and attend classes
  - Participate in interactive discussion
  - Submit assignments/ programming projects on time
- Late days for assignments
  - 5 total late days, after that loose 20% for every late day
  - Assignments are due at 11:59pm on the specified date
- Respect university code of conduct
  - No collaboration on homework / programming projects
  - http://www.northeastern.edu/osccr/academic-integrity-policy/

Grading

- Written problem assignments – 25%
  - 3-4 theoretical problem assignments based on studied material in class
- Programming projects – 20%
  - 3 programming projects
  - Language of your choice (Java, C/C++, Python)
  - In-person grading with instructor/TA
- Exams – 50%
  - Midterm – 25%
  - Final exam – 25%
- Class participation – 5%
  - Participate in class discussion and on Piazza

Outline

- Encryption setting
- Kerckhoff's principle
- Classical (traditional) cryptography
  - Shift cipher
  - Substitution cipher
  - Vigenere
  - Enigma

The cast

The good players

Alice

Bob

Introduced in the original RSA paper

The bad players

Eve

Eavesdropper

Mallory

Malicious

More application-specific malicious players

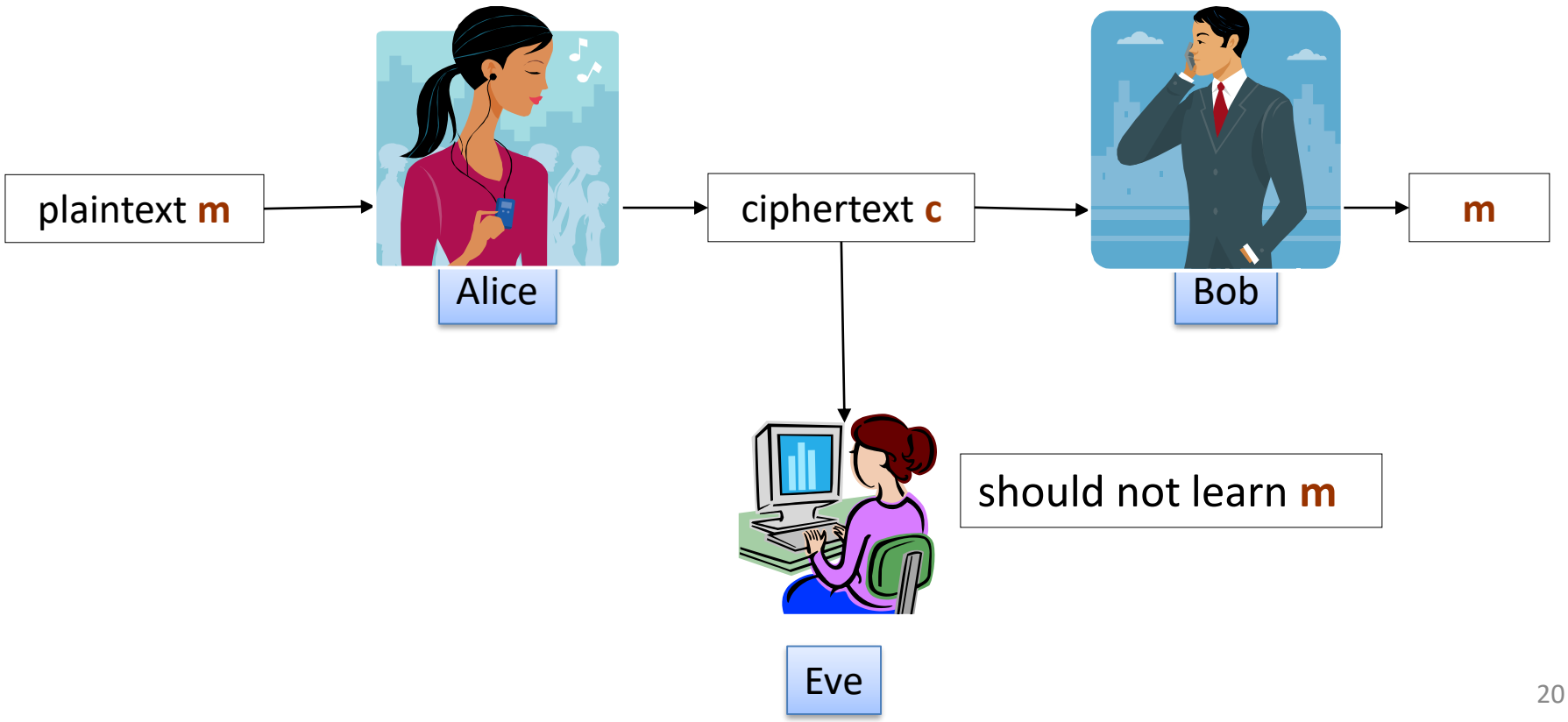
Goals and objectives

- Most basic problem
  - Ensure security of communication between parties over an insecure medium

- Basic security goals
  - *Confidentiality* (secrecy)
    - Only the intended recipient can see the communication
  - *Authenticity* (integrity)
    - Communication is generated by the alleged sender

Encryption Schemes (a very general picture)

Encryption scheme = encryption & decryption procedures



Kerckhoffs' principle

Auguste Kerckhoffs (1883):

The enemy knows the system

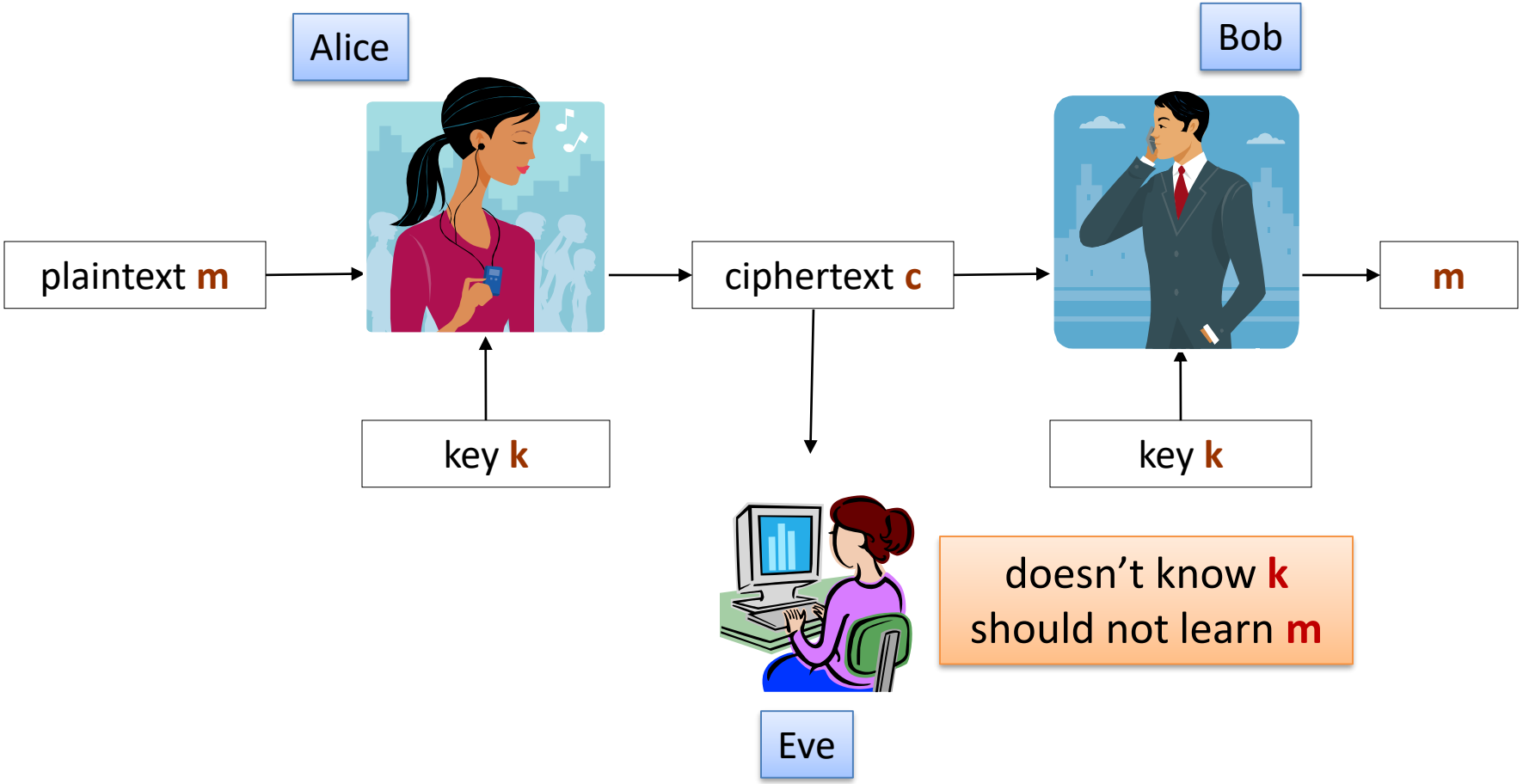
The cipher should remain secure even if **the adversary knows the specification of the cipher.**

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

key **k**

that is **usually chosen uniformly at random**

A more refined picture



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**”.

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

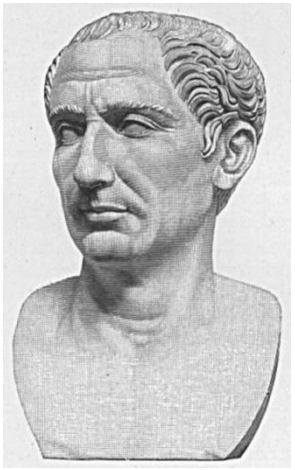
for every **k, m** we should have $\mathbf{Dec_k(Enc_k(m)) = m}$.

Idea 1: Shift cipher

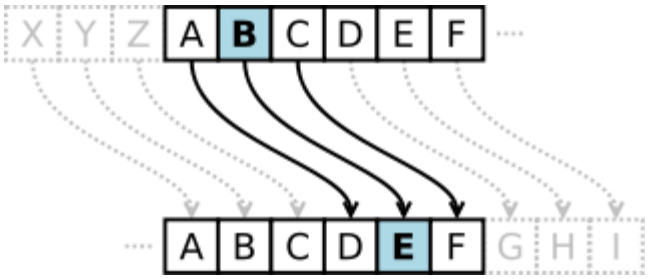
$\mathcal{M}$ = words over alphabet {A,…,Z} ≈ {0,…,25}

$\mathcal{K}$ = {0,…,25}

$Enc_k(m_0,\ldots,m_n) = (m_0+k \bmod 26,\ldots, m_n+k \bmod 26)$



Cesar: **k = 3**



Example 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
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

P = CRYPTOGRAPHYISFUN

K = 11

C = NCJAVZRCLASJTDQFY

C → 2; 2+11 mod 26 = 13 → N

R → 17; 17+11 mod 26 = 2 → C

…

N → 13; 13+11 mod 26 = 24 → Y

How to decrypt?

$Dec_k(c_0,\ldots,c_n) = (c_0\text{-}k \bmod 26,\ldots, c_n\text{-}k \bmod 26)$

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.

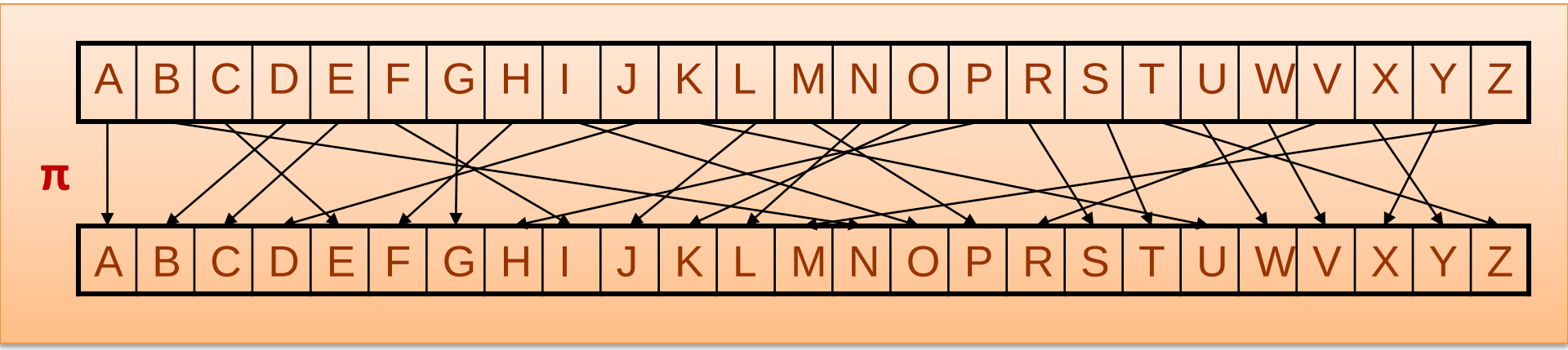
This is called a **brute force attack**.

Moral: the key space needs to be large!

Idea 2: Substitution cipher

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

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



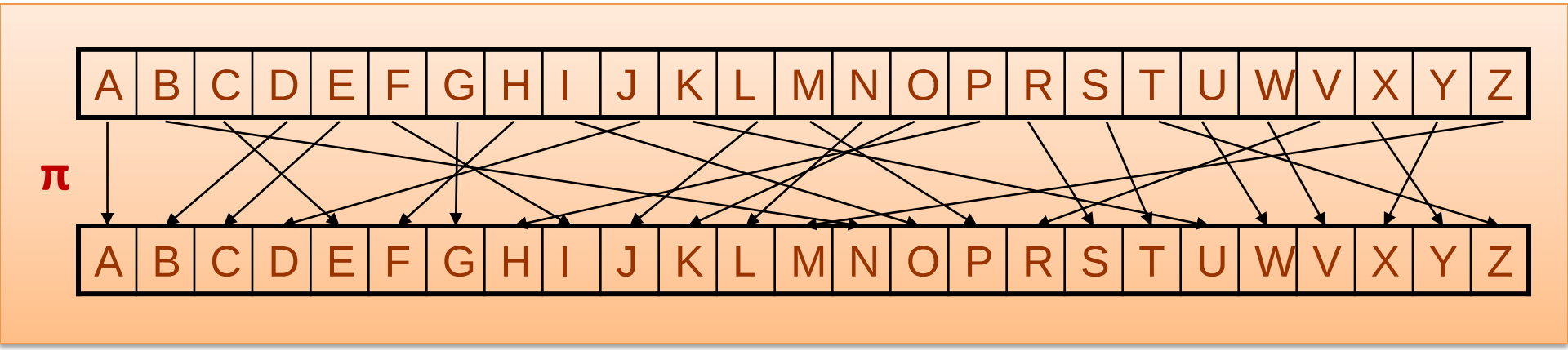
$Enc_\pi(m_0,\ldots,m_n) = (\pi(m_0),\ldots, \pi(m_n))$

$Dec_\pi(c_0,\ldots,c_n) = (\pi^{-1}(c_0),\ldots, \pi^{-1}(c_n))$

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

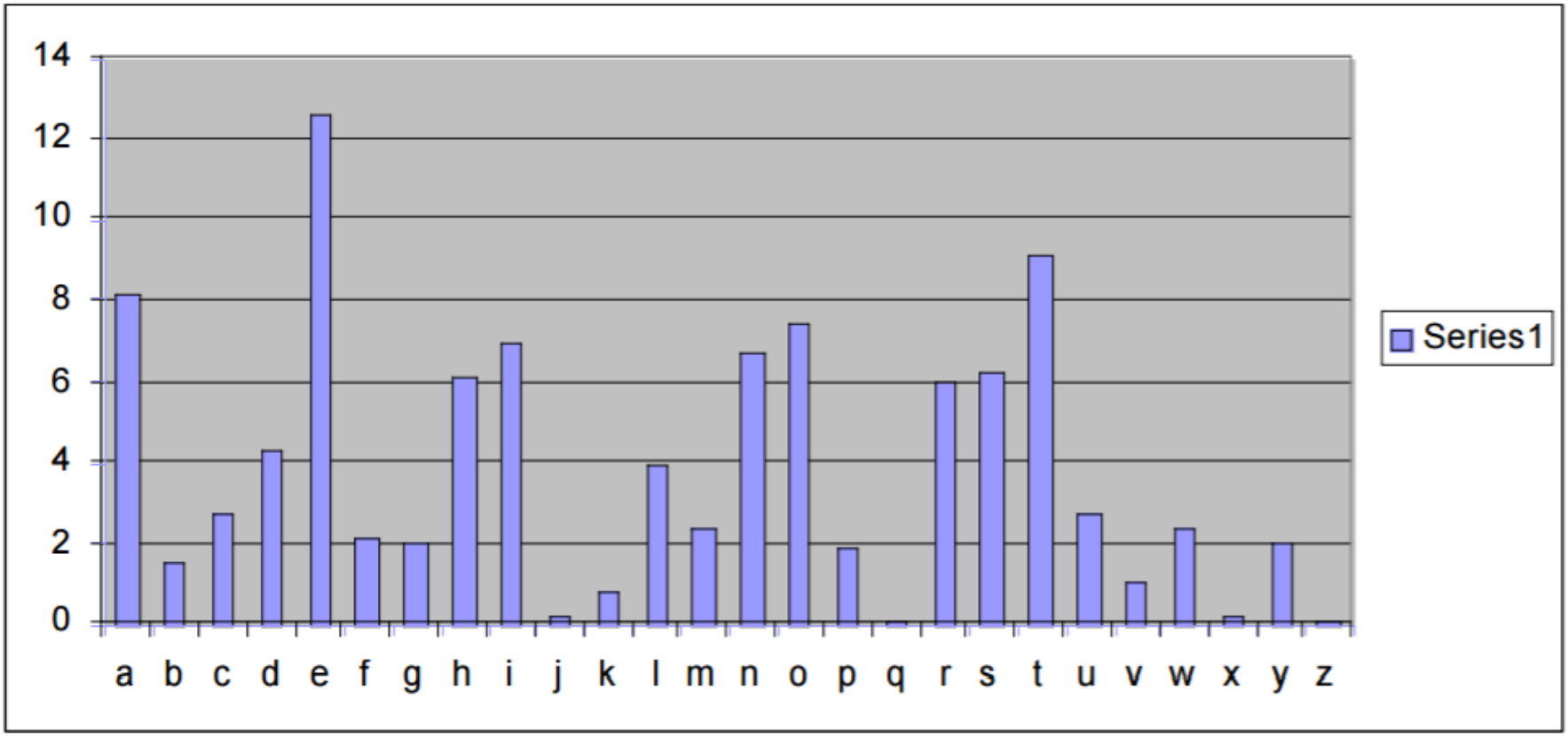
How to break the substitution cipher?

- Exhaustive search is infeasible
  - Key space size is $26! \approx 2^{88}$!
- Dominated the art of secret writing throughout the first millennium A.D.
- Thought to be unbreakable by many back then
- Until …. frequency analysis

History of frequency analysis

- Discovered by the Arabs
  - Earliest known description of frequency analysis is in a book by the ninth-century scientist Al Kindi
- Rediscovered or introduced in the Europe during the Renaissance
- Key insights
  - Each language has distinctive features: frequency of letters and groups of two or more letters
  - Substitution ciphers preserve the language features
  - Substitution ciphers are vulnerable to frequency analysis attacks

Frequency of English letters



Other languages

French

E	16.7%	T	7.3%	C	3.5%	G	1.1%	J	0.3%
S	8.2%	O	5.8%	P	3.0%	Q	1.1%	Y	0.2%
A	8.0%	U	5.5%	M	2.9%	B	0.7%	Z	0.2%
N	7.9%	L	4.9%	V	1.4%	X	0.6%	K	0.1%
I	7.6%	D	3.9%	F	1.2%	H	0.5%	W	0.0%
R	7.4%								

German

E	18.0%	T	5.7%	G	3.2%	F	1.6%	P	0.8%
N	10.6%	D	5.4%	O	2.7%	W	1.5%	J	0.3%
I	8.1%	U	4.6%	C	2.7%	K	1.3%	Y	0.0%
R	7.2%	H	4.1%	M	2.3%	Z	1.1%	X	0.0%
S	6.9%	L	3.3%	B	1.7%	V	0.9%	Q	0.0%
A	6.0%								

Other characteristics of English

- Vowels, which constitute 40% of plaintext, are often separated by consonants
- Letter A is often found in the beginning of a word or second from last
- Letter I is often third from the end of a word
- Letter Q is followed only by U
- Most common tri-gram is “the”
- And more …

Frequency analysis in action

- The number of different ciphertext characters or combinations are counted to determine the frequency of usage
- The ciphertext is examined for patterns, repeated series, and common combinations
- Replace ciphertext characters with possible plaintext equivalents using known language characteristics
- Guesses are made until plaintext “makes sense”
- Drawback: needs longer ciphertext

Moral: large key space is not enough for secure cipher

Example

- http://crypto.interactive-maths.com/frequency-analysis-breaking-the-code.html

Example: The Culper spy ring

- American revolutionary war in 1778
- Gather information on British troop movements
- Developed Culper code book

52	but	271	hesitate	488	produce		Y
53	buy	272	history	489	prison	706	yet
54	bring	273	horrible	490	progress	707	you
55	boat	274	hospital	491	promise	708	your
56	barn	275	hurrican	492	proper	709	yesterday
57	banish	276	hypocrite	493	prosper		Z
58	baker	277	*DOCUMENT DAMAGE*	494	prospect	710	zeal
59	battle	278	*DOCUMENT DAMAGE*	495	punish		PROPER NAMES
60	better	279	*DOCUMENT DAMAGE*	496	pertake	711	Gen Washington

Poly-alphabetic substitution ciphers

- Main weaknesses of monoalphabetic substitution ciphers
  - Each letter in the ciphertext corresponds to only one letter in the plaintext letter
- Idea for a stronger cipher (1460 by Alberti)
  - Use more than one cipher alphabet, and switch between them when encrypting different letters
- Giovani Battista Bellaso published it in 1553
- Developed into a practical cipher by Blaise de Vigenère and published in 1586

Idea 3: Vigenère cipher

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

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

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

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

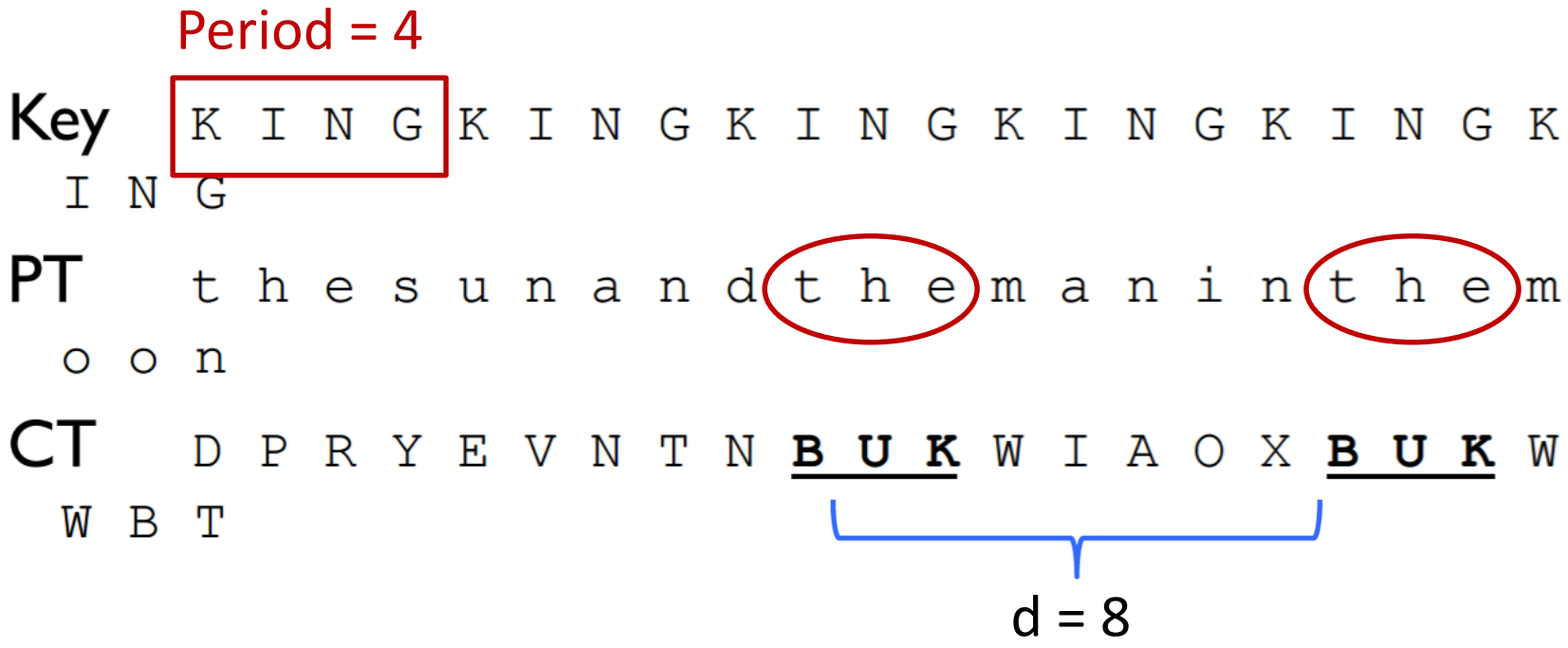
Security of Vigenère cipher

- Vigenère masks the frequency with which a character appears in a language
  - One letter in the ciphertext corresponds to multiple letters in the plaintext
  - Makes the use of frequency analysis more difficult
- Any message encrypted by a Vigenère cipher is a collection of as *many shift ciphers* as there are letters in the key
- How to break the cipher
  - Find the length of the key
  - Divide the message into that many shift ciphers
  - Use frequency analysis to solve the resulting shift ciphers

History of breaking Vigenère

- 1596 - Cipher was published by Vigenère
- 1854 - It is believed that Charles Babbage knew how to break it in 1854, but he did not published the results
- 1863 - Kasisky showed how to break Vigenère
- 1920 - Friedman published ``The index of coincidence and its applications to cryptography''

Method 1: Kasisky test



- Insight
  - The distance between duplicate n-grams in ciphertext is multiple of cipher period (length of secret key)
- Algorithm
  - Search for pairs of identical segments of length at least 3
  - Record distances between the two segments $d_1, d_2, \ldots$
  - Period p divides gcm($d_1, d_2, \ldots$)

Method 2: Index of coincidence

Letter	p_i	Letter	p_i	Letter	p_i	Letter	p_i
A	.082	H	.061	O	.075	V	.010
B	.015	I	.070	P	.019	W	.023
C	.028	J	.002	Q	.001	X	.001
D	.043	K	.008	R	.060	Y	.020
E	.127	L	.040	S	.063	Z	.001
F	.022	M	.024	T	.091		
G	.020	N	.067	U	.028		

$$I_c(x) = \sum_{i=0}^{i=25} p_i^{\,2} = 0.065$$

p_i: frequency of letter i (not considering punctuation)

Finding the key length

- Ciphertext $c_1,\ldots c_n$
- Key $k_1,\ldots k_t$

$$\begin{matrix} k_1 \\ k_2 \\ \ldots \\ k_t \end{matrix} \begin{bmatrix} c_1 & c_{t+1} & \cdots & c_{n-t+1} \\ c_2 & c_{t+2} & \cdots & c_{n-t+2} \\ \cdots & \cdots & \cdots & \cdots \\ c_t & c_{2t} & \cdots & c_n \end{bmatrix} \begin{matrix} y_1 \\ y_2 \\ \ldots \\ y_t \end{matrix}$$

- Each y_i should respect the English language character frequency
  - Assuming sufficiently long text

Guessing the key length

- If t is the key length, then each y_j looks like English text

$$I_c(y_j) = \sum_{i=0}^{25} p_i^2 = 0.065, \forall j \in \{1, \ldots t\}$$

- If t is not the key length, then each y_j looks like random text

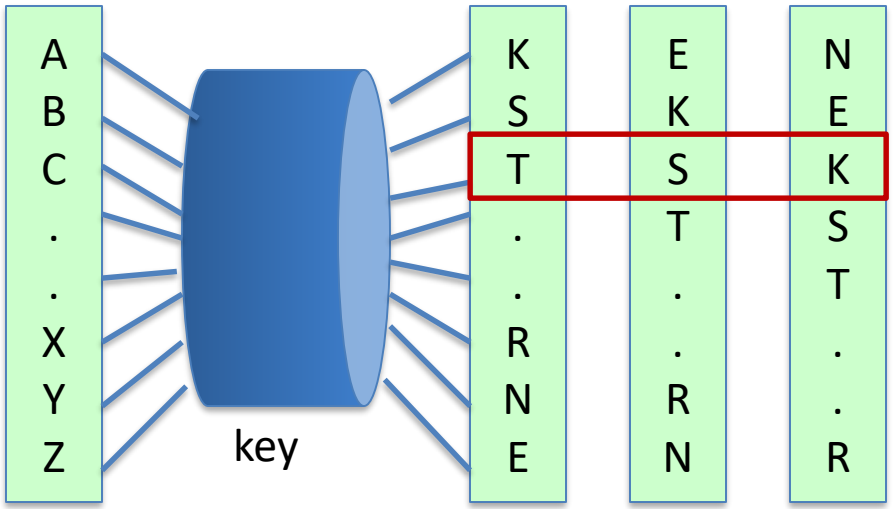
$$I_c(y_j) = \sum_{i=0}^{25} 1/26^2 = 0.038 \; \forall j \in \{1, \ldots t\}$$

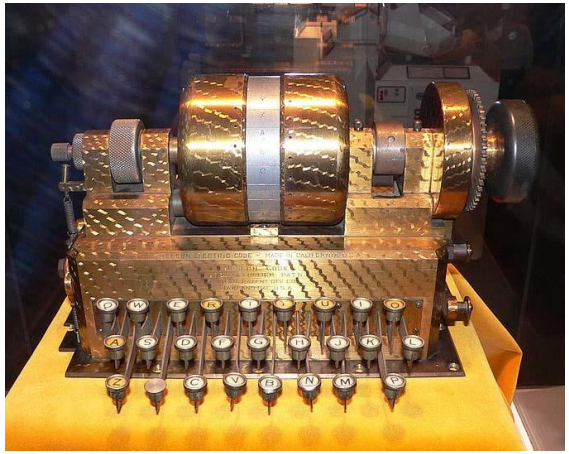
Breaking the Vigenère cipher

- Find the length of the key *t*
  - Method 1: Kasisky test
  - Method 2: Index of coincidence
    - Easier to perform automatically
- Divide the message into *t* shift ciphers
- Use frequency analysis to solve the resulting shift ciphers
  - Need longer message to perform frequency analysis
- What is the right key size?
  - Short key periods provide an advantage in breaking cipher
  - Longer keys harder to store/manage

Rotor Machines (1870-1943)

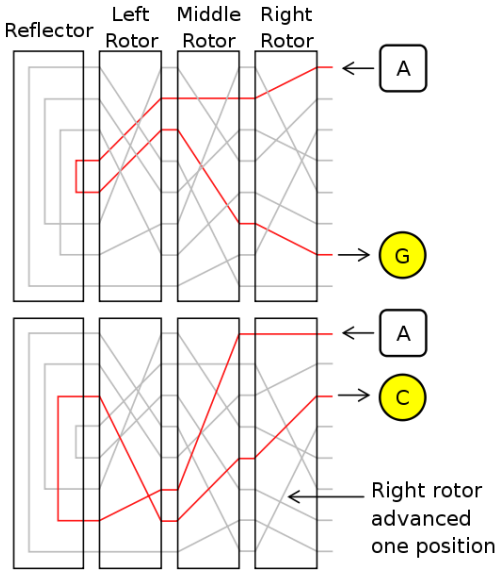
Early example: the Hebern machine (single rotor)





Rotor Machines (cont.)

Most famous: the Enigma (3-5 rotors)



keys = $26^4 = 2^{18}$ (actually 2^{36} due to plugboard – swap pairs of letters)

Key takeaways

- Cryptography has evolved from narrow field to science of securing communication and interaction in digital world
- Historically, cryptography used by military and governments
  - Shift ciphers, substitution ciphers, Vigenere cipher
  - All historical ciphers have been broken
  - Security by obscurity not effective
- Modern cryptography designs rigorous primitives
  - Define realistic security model
  - Find minimal assumptions
  - Modularity in designing higher-level protocols
  - Emerging applications rely heavily on cryptographic primitives

Acknowledgement

Some of the slides and slide contents are taken from
http://www.crypto.edu.pl/Dziembowski/teaching
and fall under the following:

We have also used materials from Prof. Dan Boneh online cryptography course at Stanford University:

http://crypto.stanford.edu/~dabo/courses/OnlineCrypto/