Countdown

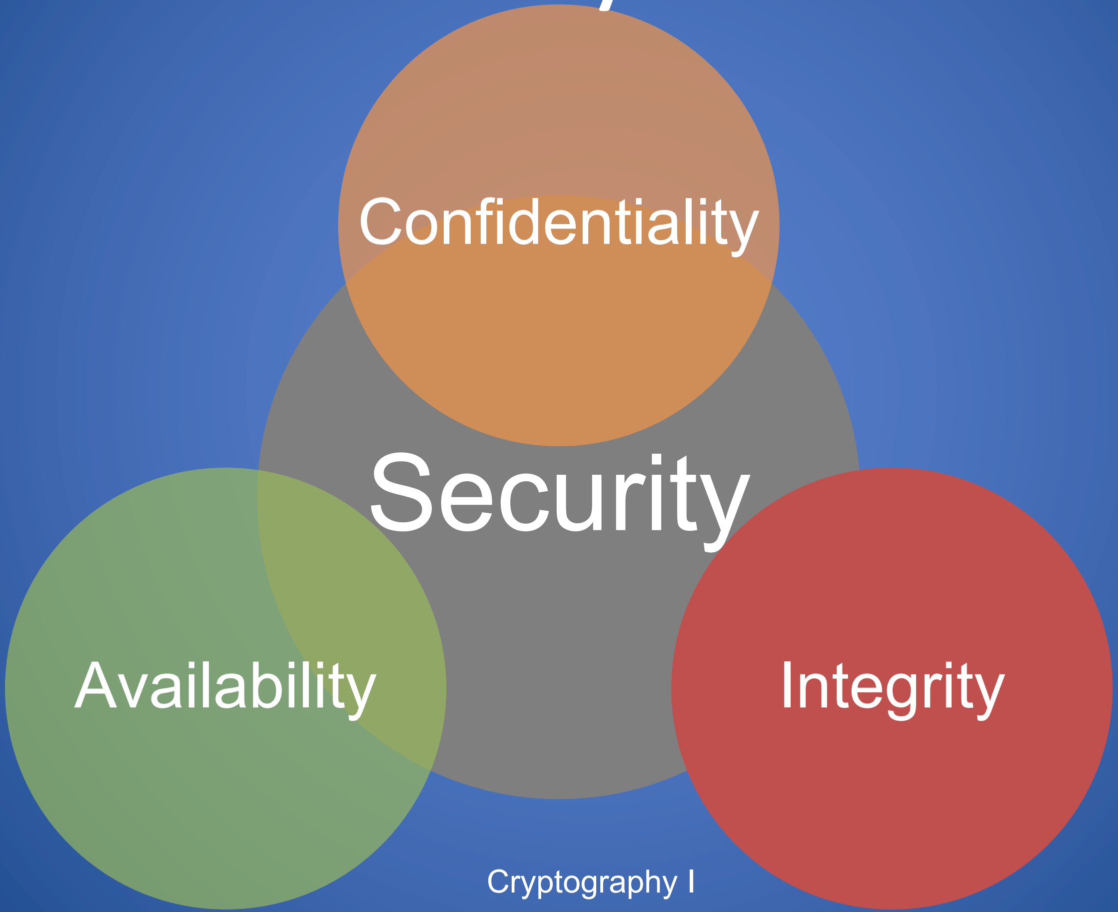
5 4 3 2 1 .5

Class is starting now!

Cryptography I

CS 166: Introduction to Computer Systems Security

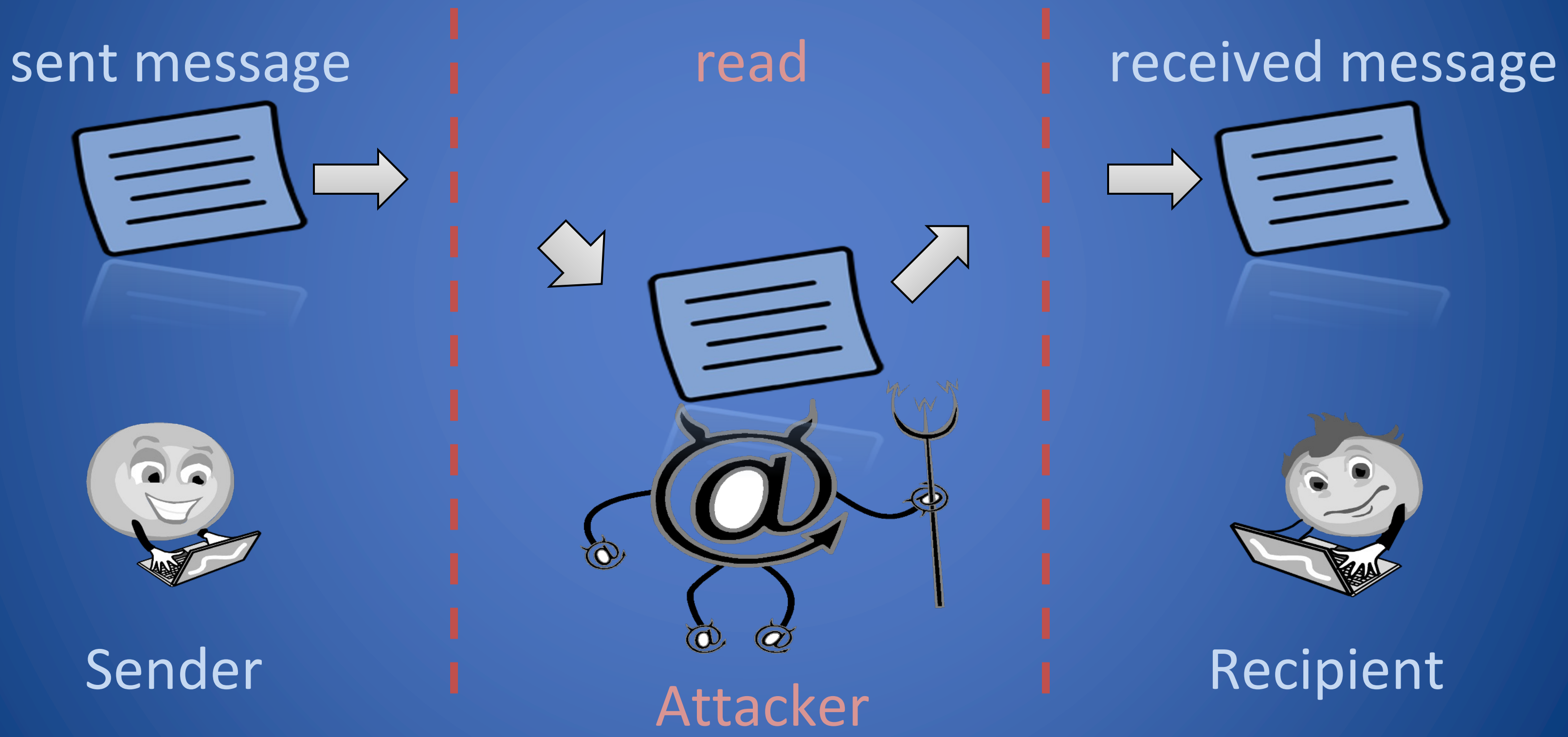
Security Goals



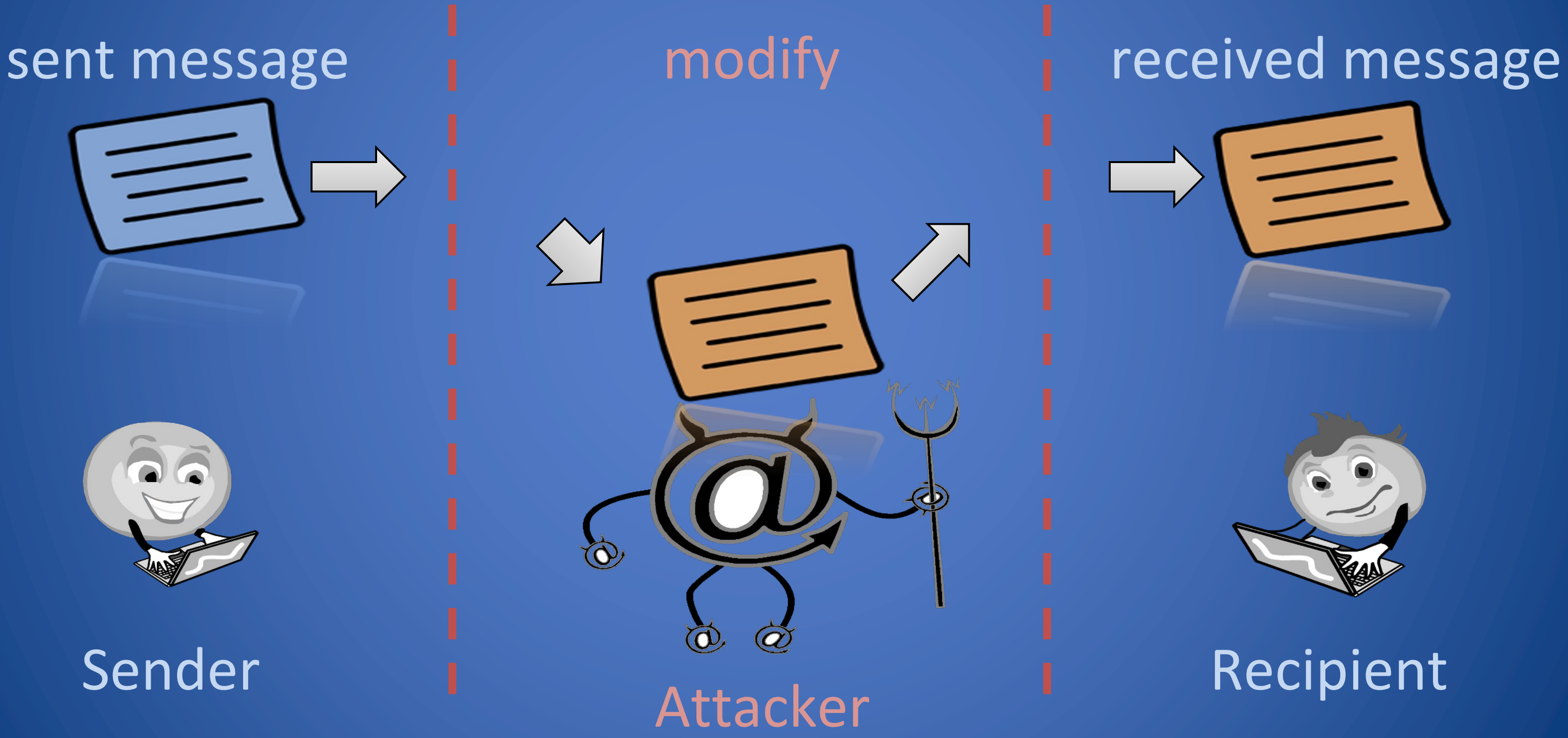
Attacks on Communication

Standard Communication

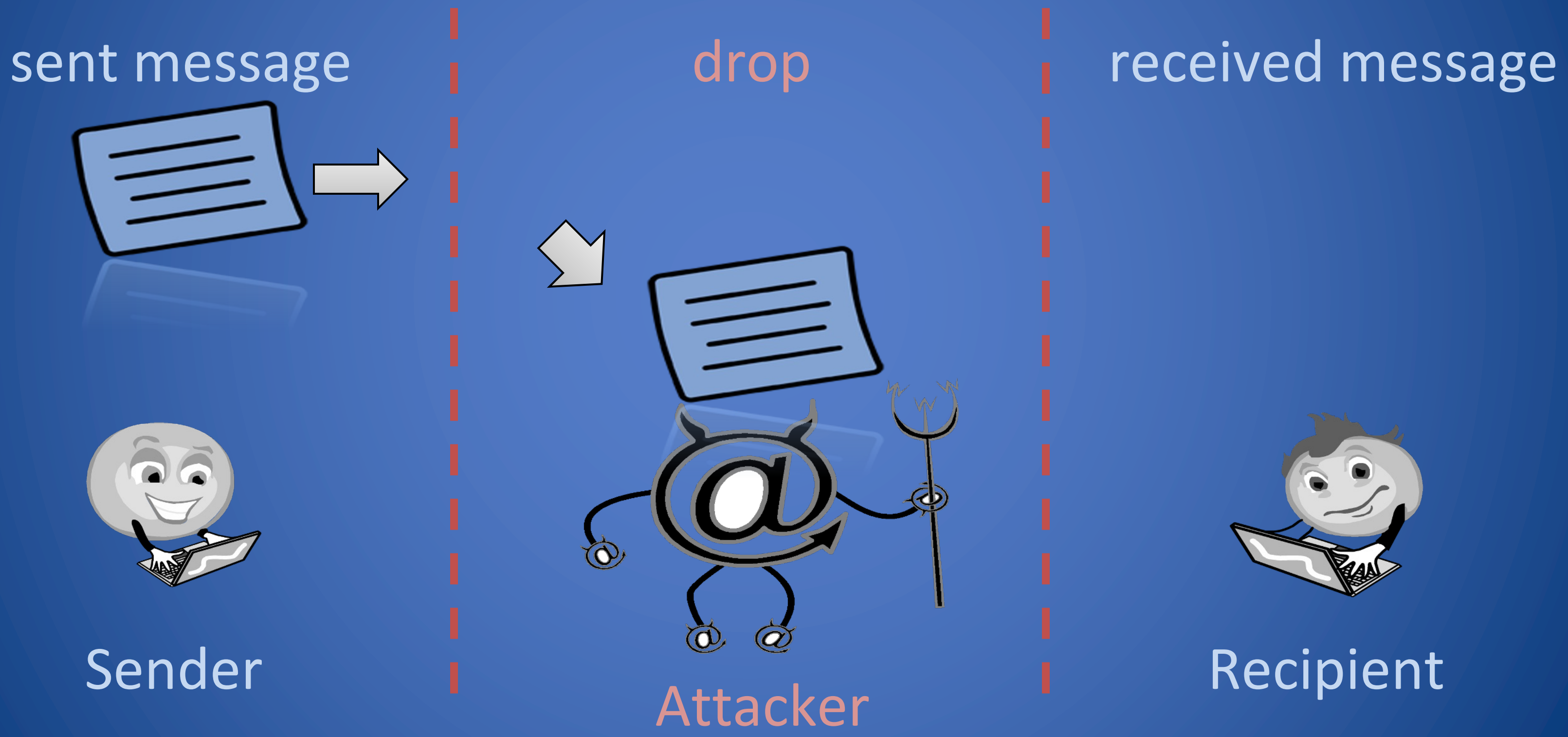
Eavesdropping



Tampering

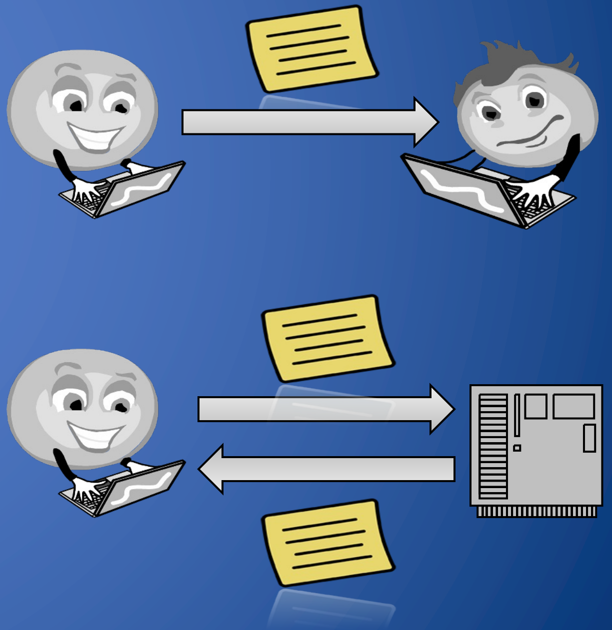

sent message
modify
received message
Sender
Attacker
Recipient

Blocking


sent message
drop
received message
Sender
Attacker
Recipient

Cryptography

- Cryptography provides methods for assuring the confidentiality and integrity of data that is
  - transmitted over communication channels (e.g., web pages and email messages)
  - stored on devices (e.g., files on a laptop or data center)

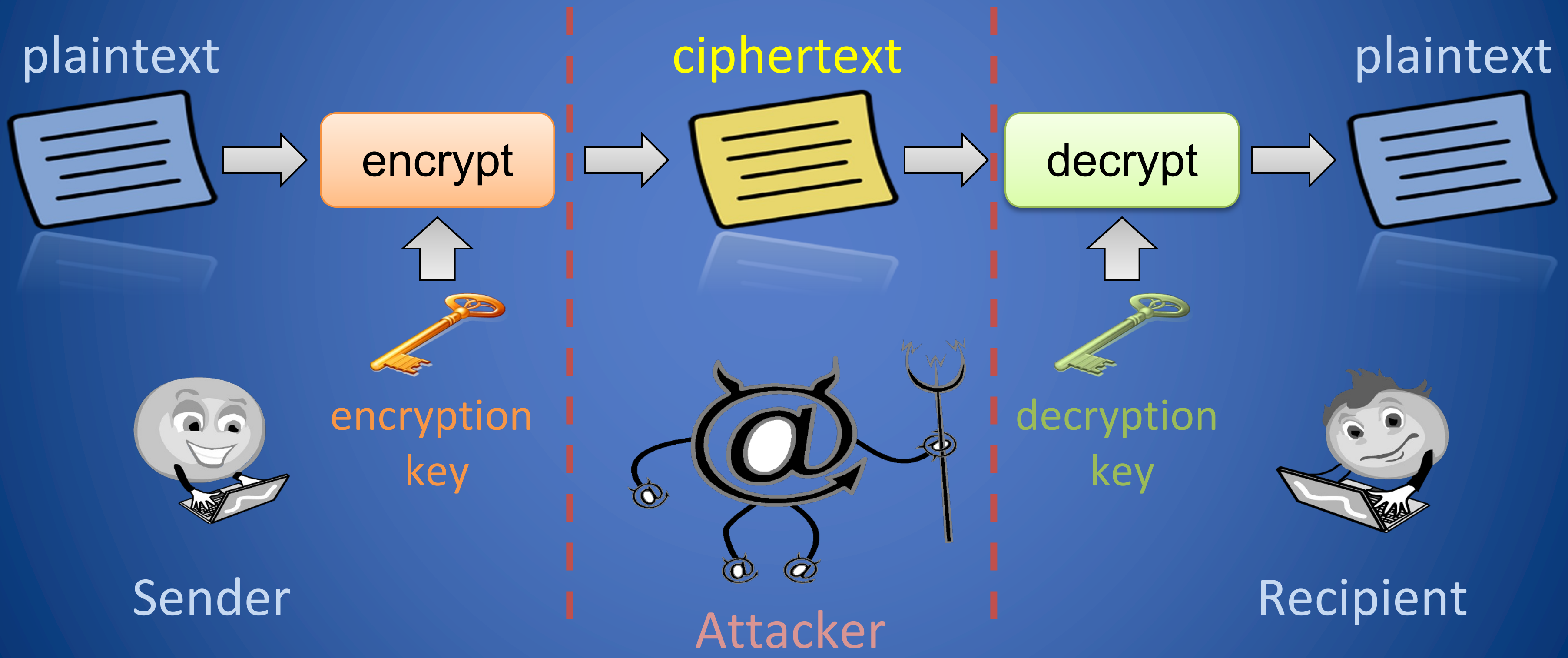


Open Design Principle

- Publicly available system architecture and algorithms
- Security relies solely on keeping keys secret
- Formulated by Auguste Kerckhoffs in 1883
- Opposite of "security by obscurity"
- Claude Shannon in 1949 said *"the enemy knows the system"*:
  - *"one ought to design systems under the assumption that the enemy will immediately gain full familiarity with them"*

Image source: https://en.wikipedia.org/wiki/Auguste_Kerckhoffs#/media/File:Auguste_Kerckhoffs.jpg

Encrypted Communication

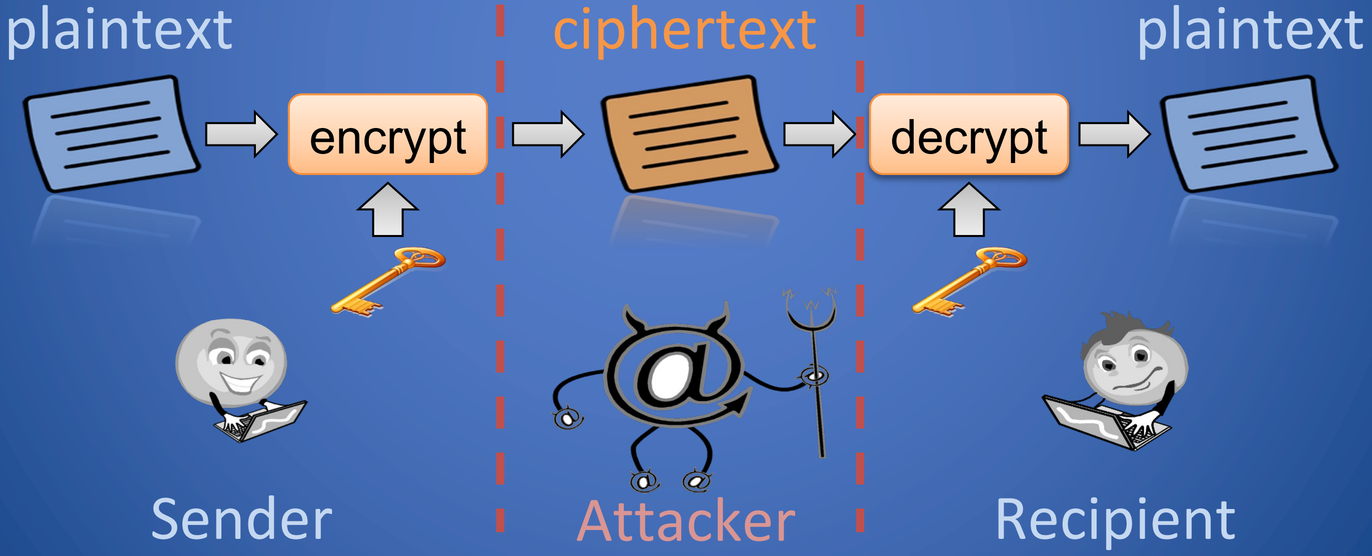


Encryption

- Encryption allows to secure communication
  - Originally focused on confidentiality alone
- The encryption algorithm combines the plaintext with the encryption key to produce the ciphertext 
  - The ciphertext is transmitted instead of the plaintext
- The decryption algorithm combines the ciphertext with the decryption key to return the plaintext 
  - Only the intended recipient should have the secret key
- Encryption and decryption should be computationally infeasible without the corresponding keys

Symmetric Encryption

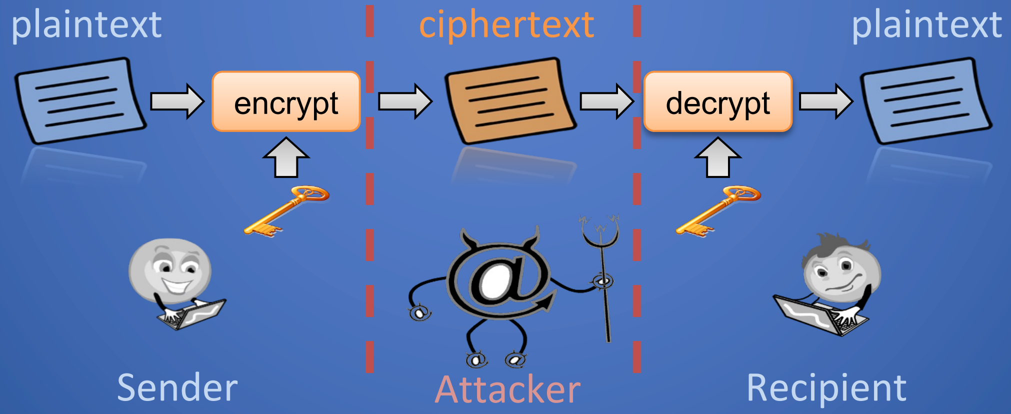
- Same key used for encryption and decryption
- Encryption and decryption algorithms are one the reverse of the other



Symmetric Encryption

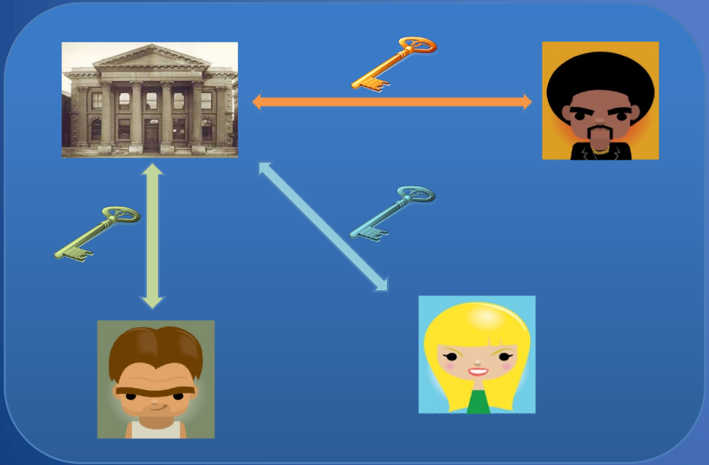
Advantage:
- Conceptual simplicity

Disadvantage:
- Secure channel to set up key

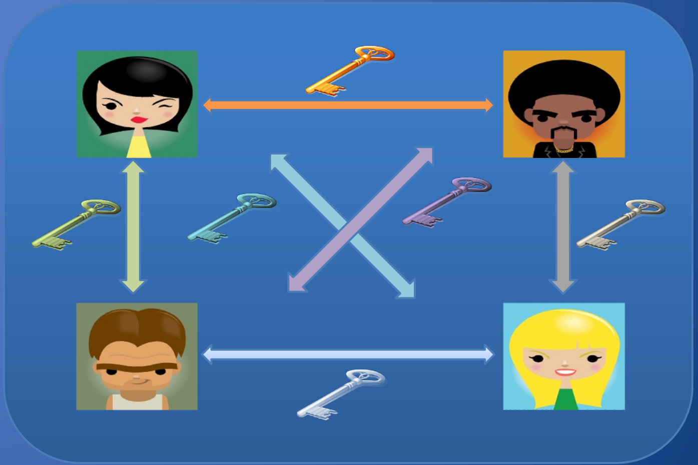


Symmetric Key Distribution

- A distinct keys needs to be set up for each pair of communicating users



- Quadratic number of keys for pairwise communication



Classic Symmetric Encryption

Julius Caesar's Cipher

- Encryption
  - replace A with D
  - replace B with E
  - replace C with F
  - …
  - replace X with A
  - replace Y with B
  - replace Z with C
- Encryption key
  - Forward alphabet shift: +3
- Decryption key
  - Reverse alphabet shift: –3

AVE → DZH

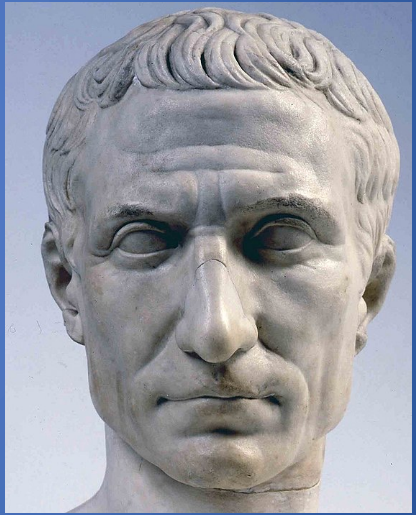


Image source: https://en.wikipedia.org/wiki/Julius_Caesar#/media/File:Gaius_Iulius_Caesar_(Vatican_Museum).jpg

Alphabet Shift Cipher

- Generalization of Caesar's cipher
- Replace each character c of the plaintext with the character k positions after c in the alphabet
- Key for encryption and decryption: number k
- Insecure encryption method
- Can be easily cracked by trying all possible values of k between 1 and the size of the alphabet

Substitution Cipher

- Arbitrary permutation of the characters
  - A → K
  - B → T
  - C → G
  - …

CAB → GKT

- Key: permutation of the alphabet characters (e.g., KTG …)
- Number of possible keys for a 26-character alphabet ≈ 4×10^{26}
- Unfeasible to try all possible keys but …
- Can be cracked by frequency analysis
  - most frequent letters in English: e, t, o, a, n, i, ...
  - most frequent digrams: th, in, er, re, an, ...
  - most frequent trigrams: the, ing, and, ion, ...
- Attack first described in a 9th century book by al-Kindi

Frequency Analysis

PCQ VMJYPD LBYK LYSO KBXBJXWXV BXV ZCJPO EYPD KBXBJYUXJ LBJOO KCPK. CP **LBO** LBCMKXPV XPV IYJKL PYDBL, QBOP KBO BXV OPVOV **LBO** LXRO CI SX'XJMI, KBO JCKO XPV EYKKOV **LBO** DJCMPV ZOICJO BYS, KXUYPD: "DJOXL EYPD, ICJ X LBCMKXPV XPV CPO PYDBLK Y BXNO ZOOP JOACMPLYPD LC UCM LBO IXZROK CI FXKL XDOK XPV **LBO** RODOPVK CI XPAYOPL EYPDK. SXU Y SXEO KC ZCRV XK LC AJXNO X IXNCMJ CI UCMJ SXGOKLU?"

OFYRCDMO, LXROK IJCS **LBO** LBCMKXPV XPV CPO PYDBLK

Example from

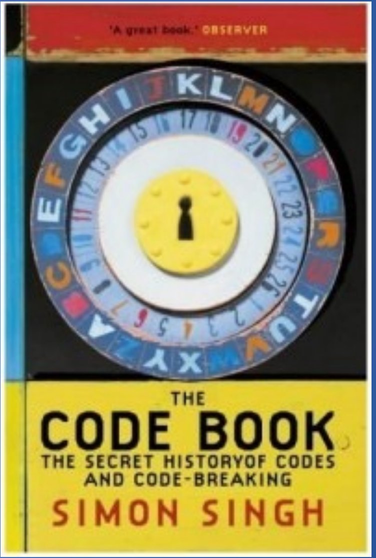
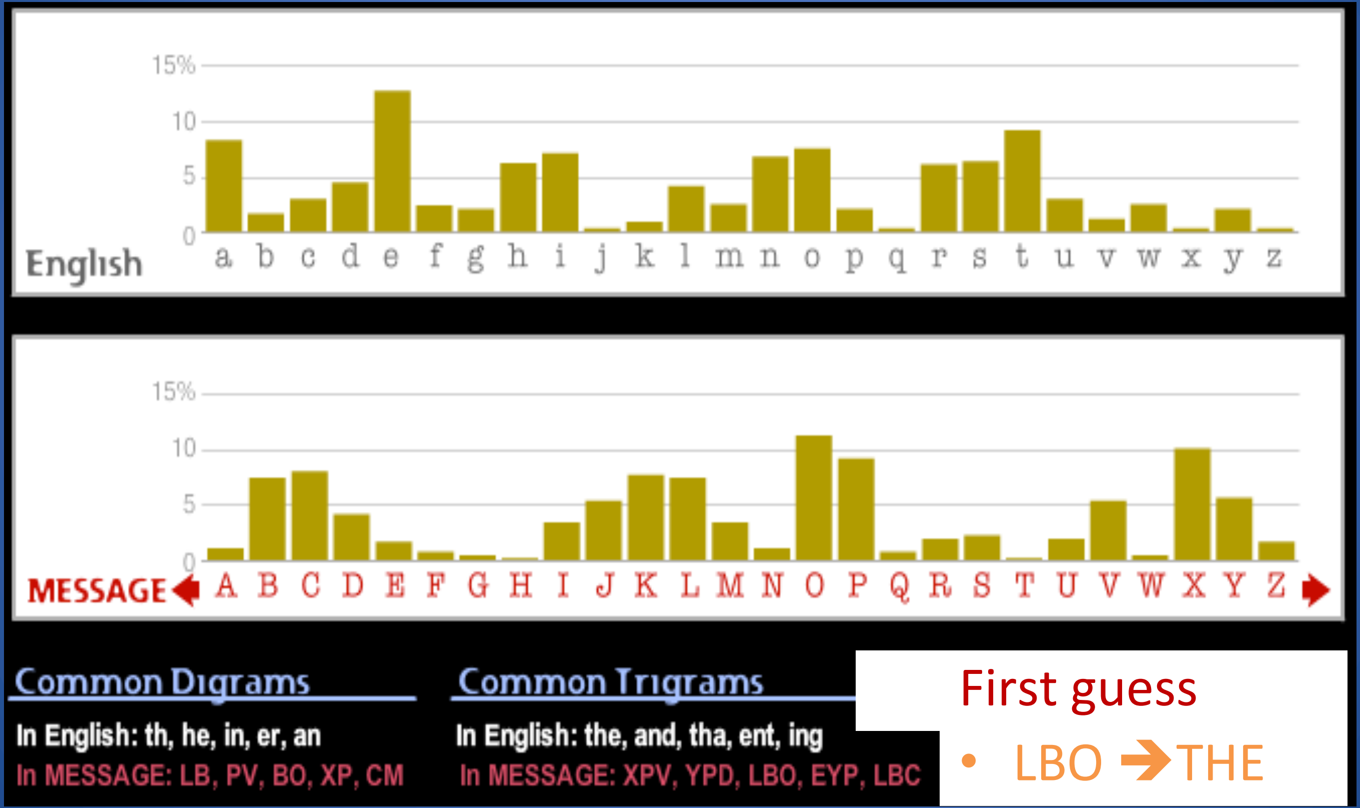


Image source: https://simonsingh.net

Letter Frequencies Graph



First guess

- LBO → THE

Frequency Analysis (cont.)

PCQ VMJYPD **THYK TYSE** KHXHJXWXV **HXV** ZCJPE EYPD KHXHJYUXJ **THJEE** KCPK. CP **THE THCMKXPV** XPV IYJKT PYDHT, QHEP KHE **HXV** EPVEV **THE TXRE** CI SX'XJMI, **KHE** JCKE XPV EYKKEV **THE** DJCMPV ZEICJE **HYS**, KXUYPD: "DJEXT EYPD, ICJ X **THCMKXPV** XPV CPE PYDHTK Y **HXNE** ZEEP JEACMPTYPD **TC** UCM **THE** IXZREK CI FXKT XDEK XPV **THE** REDEPVK CI XPAYEPT EYPDK. SXU Y SXEE KC ZCRV XK **TC** AJXNE X IXNCMJ CI UCMJ SXGEKTU?"

EFYRCDME, **TXREK** IJCS **THE THCMKXPV** XPV CPE PYDHTK

L → T

B → H

O → E

More guesses

J → R

K → S

X → A

Frequency Analysis (cont.)

PCQ VMRYPD THYS TYSE SHAHRAWAV HAV ZCRPE EYPD SHAHRYUAR THREE SCPS. CP THE THCMSAPV APV IYRST PYDHT, QHEP SHE HAV EPVEV THE TARE CI SA'ARMI, SHE RCSE APV EYSSEV THE DRCMPV ZEICRE HYS, SAUYPD: “DREAT EYPD, ICR A THCMSAPV APV CPE PYDHTS Y HANE ZEEP REACMPTYPD TC UCM THE IAZRES CI FAST ADES APV THE REDEPVS CI APAYEPT EYPDS. SAU Y SAEE SC ZCRV AS TC ARANE A IANCMR CI UCMR SAGESTU?”

EFYRCDME, TARES IRCS THE THCMSAPV APV CPE PYDHTS

L → T

B → H

O → E

J → R

K → S

X → A

Decryption

PCQ VMJYPD LBYK LYSO KBXBJXWXV BXV ZCJPO EYPD KBXBJYUXJ LBJOO KCPK. CP LBO LBCMKXPV XPV IYJKL PYDBL, QBOP KBO BXV OPVOV LBO LXRO CI SX'XJMI, KBO JCKO XPV EYKKOV LBO DJCMPV ZOICJO BYS, KXUYPD: “DJOXL EYPD, ICJ X LBCMKXPV XPV CPO PYDBLK Y BXNO ZOOP JOACMPLYPD LC UCM LBO IXZROK CI FXKL XDOK XPV LBO RODOPVK CI XPAYOPL EYPDK. SXU Y SXEO KC ZCRV XK LC AJXNO X IXNCMJ CI UCMJ SXGOKLU?”

OFYRCDMO, LXROK IJCS LBO LBCMKXPV XPV CPO PYDBLK

→

Now during this time Shahrazad had borne king Shahriyar three sons. On the thousand and first night, when she had ended the tale of Ma'aruf, she rose and kissed the ground before him, saying: “great king, for a thousand and one nights I have been recounting to you the fables of past ages and the legends of ancient kings. May I make so bold as to crave a favour of your majesty?”

Epilogue, Tales from the Thousand and One Nights

Clicker Question (TopHat: 821033)

Clicker Question

- Bob is experimenting with different symmetric encryption schemes to securely communicate with Alice
- To test his knowledge, he decides to encrypt the plaintext “HELLO WORLD” using an alphabet shift cipher, where k = 4
- Which of the following ciphertexts is correct?

a. KHOOR ZRUOG

b. MHPOS ARVPH

c. LIQQR WRVOH

d. LIPPS ASVPH

Clicker Question

Answer: D

	H	E	L	L	O	W	O	R	L	D
+4	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
	L	I	P	P	S	A	S	V	P	H

One-Time Pad

Bitwise XOR

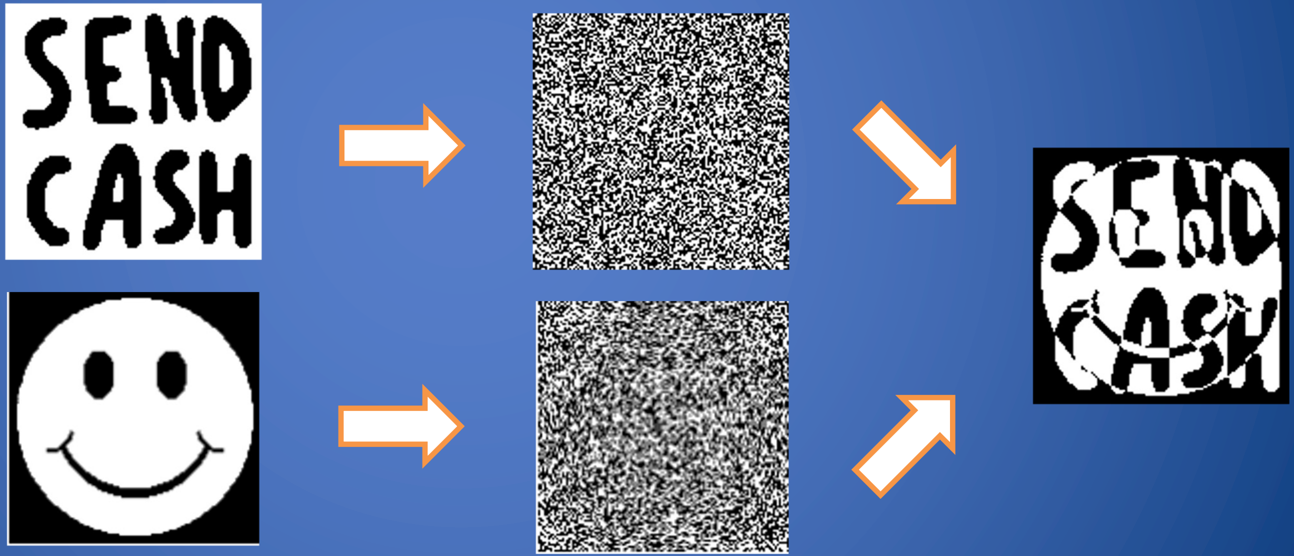
X	Y	$X \oplus Y$
0	0	0
0	1	1
1	0	1
1	1	0

One-Time Pad

- Key
  - Sequence of random bits
  - Same length as plaintext
- Encryption
  - C = K ⊕ P
  - Example
    - P = 01101001
    - K = 10110010
    - C = 11011011
- Decryption
  - P = K ⊕ C
- Advantages
  - Each bit of the ciphertext is random
  - **Fully secure** if key used only once (i.e. Beale's treasure)
- Disadvantages
  - Key as large as plaintext
    - Difficult to generate and share
  - Key cannot be reused

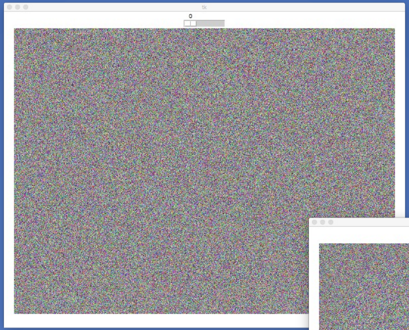
Demo: Pitfalls with One-Time Pads

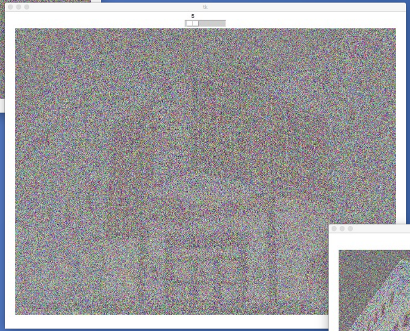
Key Reuse

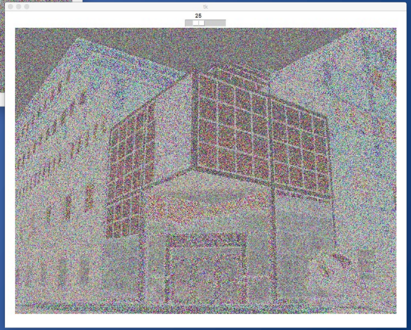


Source: Cryptosmith and David Lowry-Duda, Cryptography Stack Exchange

Imperfect Randomness







Source: Justin Bisignano and Joshua Liebow-Feeser

Modern Symmetric Encryption

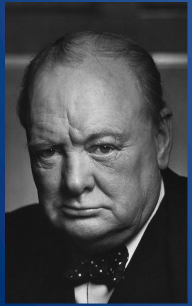
Symmetric Encryption at War

Vigenere Cipher
(American Civil War)

Navajo Code
(WW II US vs Japan)

Enigma machine[3]
(WW II Nazi vs. Allies)

Alan Turing[4] decrypted under the project 'Ultra'



"It was thanks to Ultra that we won the war."

Winston Churchill[5] to King George VI

1: https://en.wikipedia.org/wiki/Vigen%C3%A8re_cipher#/media/File:Confederate_cipher_disk.png
2: https://www.wikitree.com/blog/wp-content/uploads/2019/08/24418587.jpeg.jpg
3: https://en.wikipedia.org/wiki/Enigma_machine#/media/File:Enigma_(crittografia)_-_Museo_scienza_e_tecnologia_Milano.jpg
4: https://www.npg.org.uk/collections/search/use-this-image/?mkey=mw165875
5: https://en.wikipedia.org/wiki/Winston_Churchill#/media/File:Sir_Winston_Churchill_-_19086236948.jpg

Modern Symmetric Encryption

Data Encryption Standard (DES)

- Developed by IBM in collaboration with the NSA
- Became US government standard in 1977
- 56-bit keys
- Exhaustive search attack feasible since late 90s

Advanced Encryption Standard (AES)

- Selected as US government standard in 2001 through open competition
- 128-, 192-, or 256-bit keys
- Exhaustive search attack not currently possible

Image source: https://www.nsa.gov/resources/everyone/digital-media-center/image-galleries/places/

Break!!!!!

60 60 60 60 60

Class is starting now!

Cryptographic Hash Functions

Hash Functions

- A hash function transforms
  - an input message or file of arbitrary length
  - into a fixed-length output value (e.g., 256 bits) called hash value
- A collision occurs when two distinct messages have the same hash value
  - Inevitable because there are more inputs than outputs
  - If two hashes are different, the inputs are different
  - The converse is not true

Cryptographic Hash Functions

- Short output
  - The hash value has small fixed length (e.g., 256 or 512 bits)
- One-way
  - It is hard to find a message with a given hash value
- Collision resistance
  - Given a message, it is hard to find a different message with the same hash value

Cryptographic Hash Functions

- Cryptographic hash function
  - Hash function with special properties
  - Not all properties always required
  - Public function, no secrets
- Only feasible attack to break a property is brute-force search
  - Length of hash value should be at least 256 bits (32 bytes)
- One-way
  - Given a hash value x, it is hard to find a plaintext P such that h(P) = x
- Weak collision resistance
  - Given a plaintext P, it is hard to find a plaintext Q such that h(Q) = h(P)
- Strong collision resistance
  - It is hard to find a pair of plaintexts P and Q such that h(Q) = h(P)
  - Birthday Paradox

Hashing People to Birthdays

- Define the birthday hash function as the mapping of a person to the month and date of birth (e.g., August 15)
  - 366 possible hash values
- Birthday paradox…
  - Suppose there are N students in a classroom
  - To be sure that at least two students have the same birthday N must be at least 367
  - How many people have the same birthday in this classroom? (DEMO)

Demo: Birthday survey

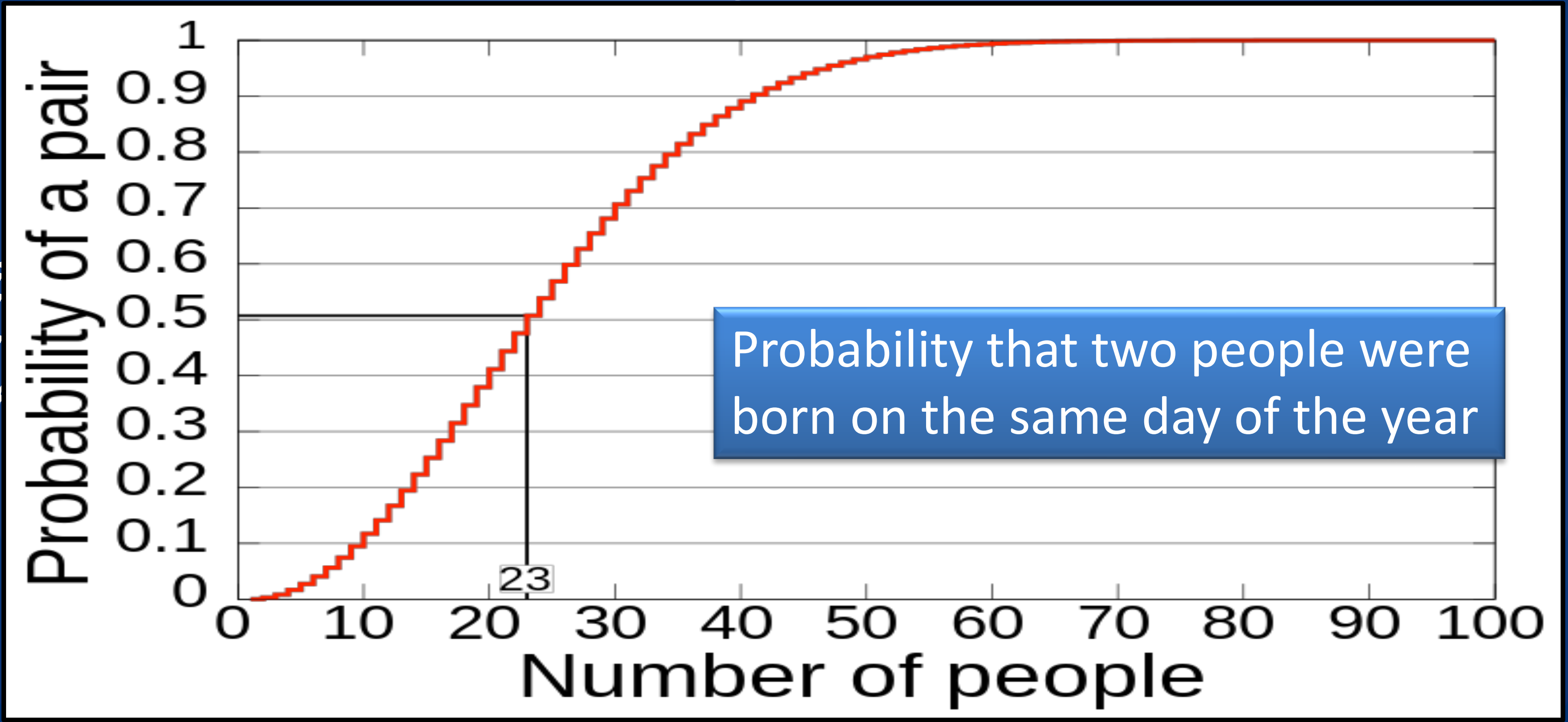
- Enter your birthday here:

https://forms.gle/rDwsU1wjdncDA1856

(Link also on lectures page of website)

- If you don't want to enter your birthday, just pick any random date

Birthday Paradox



Hash Functions vs. Hash table

- Hash function
  - –Function h mapping plaintext P to fixed-length value x = h(P), called hash value or digest of P
  - –Should take time proportional to length of plaintext
- Collision
  - –Pair of plaintexts P and Q that map to the same hash value, h(P) = h(Q)
  - –Collisions are unavoidable

- Hash table
  - –Widely used data structure
  - –Stores items into locations associated with hash values
  - –Chaining or open addressing deal with collisions
  - –Constant expected search time if hash function spreads items uniformly

Applications

- File integrity
  - Alice stores her files on a cloud server managed by Bob
  - She later retrieves the files
  - How can she make sure the files were not corrupted?
  - She wants something more efficient than keeping a copy of all her files
- Solution
  - Alice computes and keeps a crypto hash for each file
  - Security ensured by weak collision resistance
  - Efficient scheme since Alice stores short hashes (e.g., 32 bytes) instead of files

Applications

- Password authentication
  - How to authenticate users without storing passwords?
  - We want to avoid server breach to leak passwords
  - We want to defend against password-guessing attacks
- Solution
  - Store crypto hash of password but not the password
  - One-way makes it difficult to recover password from hash
  - Weak collision resistance makes it hard to guess other password with same hash

Practice

- Practical hash functions
  - –Functions widely believed to perform in practice like a cryptographic hash function
  - –No mathematical proof that they satisfy the three properties
  - –No significant attacks
  - –Standardized by NIST

- MD5 (128 bits)
  - –Developed by Ron Rivest (1991)
  - –Considered insecure, do NOT use
- SHA-1 or RIPEMD160 (160 bits)
  - –SHA-1 NIST 1995
  - –RIPEMD Developed by Hans Dobbertin, Antoon Bosselaers and Bart Preneel (1996)
- SHA-2: different lengths (224, 256, 384, 512 bits)
  - –Developed by the NSA (2002)
- SHA-3: Keccack (different number of bits)
  - –Developed by Guido Bertoni, Joan Daemen, Michaël Peeters, and Gilles Van Assche (2011)
  - –Won SHA-3 Competition **(11/2/2007 – 10/2/2012)**
  - – Not widely used

Let’s try together

- Practicing with different hashes
  - www.tools4noobs.com/online_tools/hash/
- Two different images same hash:
  - https://www.hacksandsecurity.org/posts/two-images-have-same-md5-hash-md5-collision-example
  - https://github.com/sunjw/fhash

Clicker Question (Th:821033)

Bob.com authenticates users by storing a cryptographic hash of each user's password in a server-side database. Which property of hash functions is most important when protecting against an attacker who has direct access to the password database?

A. One-way
B. Weak collision resistance
C. Strong collision resistance
D. All of the above

Clicker Question - Answer

Bob.com authenticates users by storing a cryptographic hash of each user's password in a server-side database. Which property of hash functions is most important when protecting against an attacker who has direct access to the password database?

A. **One-way**
B. **Weak collision resistance**
C. Strong collision resistance
D. All of the above

What We Have Learned

- Security goals and attacks on communication
- Frequency analysis defeats classic encryption
- One-time pads and the importance of randomness
- Use AES (not DES) for symmetric encryption
- Cryptographic hash function
  - –Building block for security protocols
- Entropy
  - –Formal measure of uncertainty in the outcome of a process