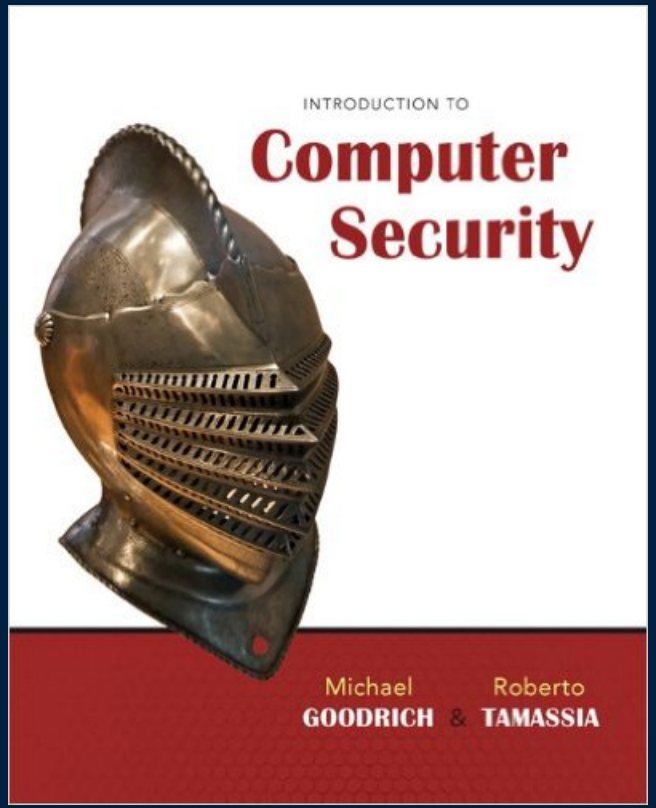
Privacy, Security and Professional Ethics

UBCO Master of Data Science – DATA 553

Introduction to Security



Acknowledgements:

- Goodrich, Michael T. and Tamassia, Roberto.
  - *Introduction to Computer Security*, Addison-Wesley, 2017.
  - *Computer Security: Art and Science book*, 2011.

Some slides are from DSCI 541 (MDS Vancouver)

Some slides are from DATA 553 (MDS Okanagan)

Definitions

A privacy incident occurs when personal information is accessed, collected, used, disclosed or disposed of without proper authorization

An information security incident is the suspected or actual unauthorized access, use, disclosure , modification or destruction of electronic information or interference with information technology

Facts

95% of breached records came from 3 industries in 2016

- Government, Retail, Technology

In 2019, business sector was most touched, followed by healthcare sector

On average there is 1 cyberattack every 39 seconds

Globally 1 ransomware attack occurs every 14 seconds

- The first reported death due to a ransomware attack took place at a hospital in Germany

2020 - 43% of cyberattacks target small businesses

- 64% of companies experienced web-based attacks
- 62% phishing and social engineering attacks
- 59% malicious code and botnets
- 51% denial of service attacks (DoS)

Facts

One of the first 2020 data leak involved 250 million records

Discovery time for 60% of data breaches is months or longer

- It takes an average of 206 days to detect a breach, and an average of 72 days to contain once detected.

Lost business due to a 1 million record data breach costs on average $1.42 million

Enterprise ransomware attacks are on the rise

Phishing is the number one type of threat action involved in data breaches. (Verizon's 2020 Data Breach Investigation Report)

Facts

Gartner report 2022:

- 88% of companies now consider cybersecurity a business risk
- Cyberattacks caused by exposure to third parties increasing

Verizon 2021

- 28 percent of data breaches affected small business victims
- In over 60% of cases, stolen credentials were used

Resources

Explore privacymatters.ubc.ca

While targeted to UBC employees, it contains a large amount of resources on how to protect your data, including for a poster information for students after we all moved to online teaching.

Major Inter-related Security Goals: C.I.A

Allows the *right* people to:

- Read data
- Modify data
- Access data or systems in a timely manner

C.I.A

Confidentiality is the notion of preventing unauthorized disclosure of information, including anything about the data which might allow an intruder to infer information.

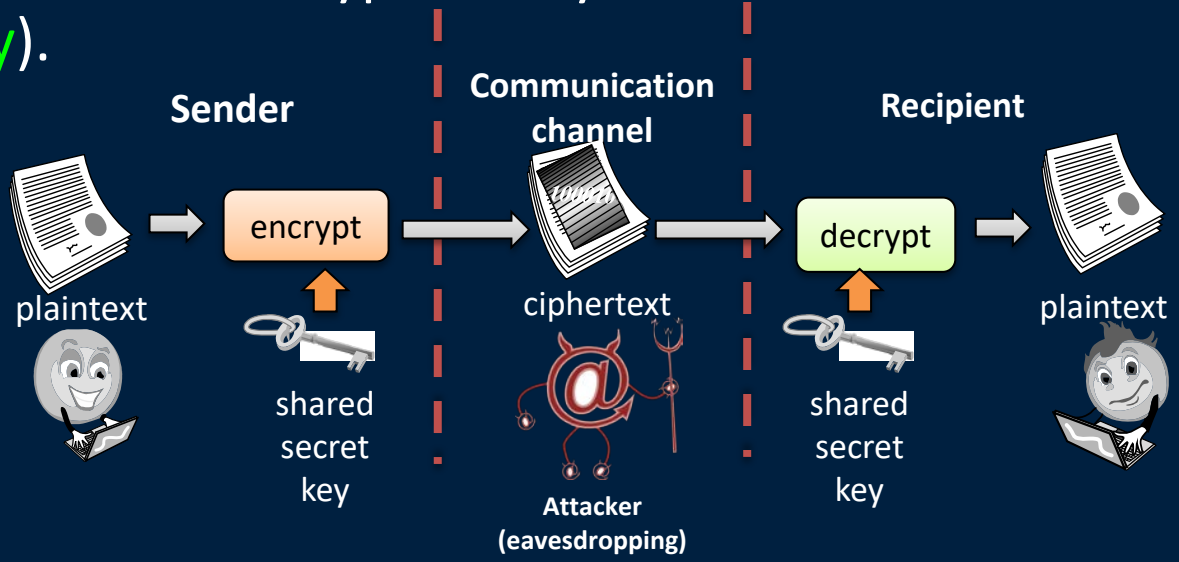
- Metadata is related to content.
  - Metadata is “data about data”. It describes the data, and in particular, it provides attributes about the data. (More about this later in this unit.)

Integrity means that data can only be changed by authorized people, and in authorized ways. Internal consistency checks should be used to help ensure data quality.

Availability is the ability to provide timely access to data and systems, for authorized users.

Tools for confidentiality

Encryption: the transformation of information using a secret, called an encryption key, so that the transformed information can only be read using another secret, called the decryption key (which may, in some cases, be the same as the encryption key – that would be called symmetric cryptography).



Tools for confidentiality

Access Control: Rules and policies that limit access to confidential information to those people and/or systems with a "need-to-know"

- The need-to-know status may be determined by an identity, such as a person's name, userid, or a computer's serial number; or by a role that a person has, such as being a manager or a computer security specialist.

- Identity is important in both computing and society because it's a way to give you access to resources and services.
  - e.g., banking, government services, health care, building access, air travel
  - Outside of the online world, it's also used to run an orderly society (e.g., driver's license, insurance, ownership of items, gym access, liquor store, academia).
    - e.g., Your UBC card or a fob on your keychain can be linked to UBC buildings and doors.

Tools for confidentiality

Authentication: After providing a userid, this process determines if someone really is who they say they are. This determination can be done in a number of different ways, but it is usually based on one or more of:

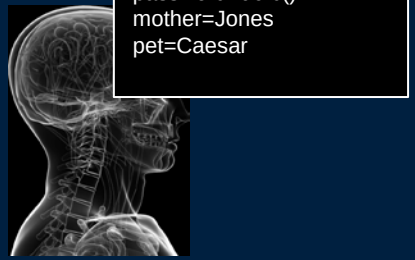
- something the person has,
- something the person knows,
- something the person "is".

Two-Factor Authentication (2FA):

The person needs two of:

- something he/she knows (e.g., password), a physical token (e.g., possession of a bank card), a challenge/ response calculator, etc.

Something you are



Something you know

Something you have

Tools for confidentiality

Authorization: the process of determining whether or not a person (or system) is allowed access to certain resources, including access to certain applications and certain functions, based on an access control policy

Physical Security: the establishment of physical barriers to limit access to protected resources

- Such barriers include locks on cabinets and doors, the placement of computers in windowless rooms, the use of sound-dampening materials, and even the construction of buildings or rooms with walls incorporating copper meshes (called Faraday cages) so that electromagnetic signals cannot enter or exit the enclosure.
  - Note: This may be overkill for many systems.

Integrity

Integrity: the property that information has not been altered in an unauthorized way

Some Integrity Tools:

- Audit Logs: regular, periodic, or random reviews
- Backups: the periodic archival of data; a snapshot of a point-in-time
  - VERY important! Database systems and file systems take backups frequently.
  - You should create backups of your own files, photos, etc., too.
  - Keep some backups offsite, too. Why?
  - Existing, but reasonably current, backups are critical when trying to recover from a ransomware attack.

Integrity

Some Integrity Tools (cont.):

- Checksums: using a function to maps the contents of a file to a numeric value. A checksum function depends on the entire contents of a file and is designed in a way that even a small change to the input file (such as flipping a single bit) is highly likely to result in a different output value.
  - e.g., One-way hash functions from your data structures course
  - e.g., Barcode scanners in a library, or at a grocery-store checkout counter
- Data Correcting Codes (aka Error Correcting Codes): methods for storing data in such a way that small changes can be easily detected and automatically corrected.
  - e.g., RAID systems (file management, database systems)—not part of this course

Availability

Availability: the property that information is accessible and modifiable in a timely fashion by those authorized to do so

Tools:

- Physical protection of the infrastructure is needed to keep information available even in the event of physical challenges.
- Redundancy: computers and storage devices that serve as fallbacks in the case of failures
- Regular backups can help to bring a crashed system back up ASAP, via appropriate recovery of data.
  - Database systems are known for their high availability and will automatically perform crash recovery upon restart to confirm that the last shutdown was orderly. If there were any in-progress transactions at the time of the crash, then those transactions will be backed out (so, it's as if they never entered the system).

Confidentiality and Metadata

A photo that you take with your cell phone or digital camera has metadata associated with it.

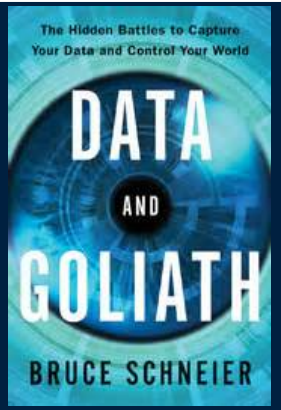
- When you post the photo on the Internet (or e-mail it), some metadata goes with it. This often includes: picture size in MB, date, time, camera settings, camera serial number, location of the picture if you have GPS, etc.

E-mail

- Every e-mail message on the Internet has routing information; but, some of it may be spoofed (i.e., deliberately changed/faked by a hacker).
- Metadata includes information such as: the "from" address, the "to" address; the date, time, and routing (location) information for the path that the e-mail took during transmission; the Message ID (it is unique across the Internet); the message size; and other header information.

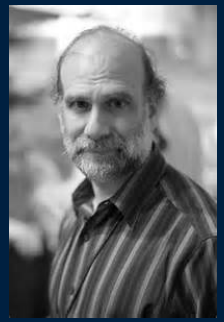
Confidentiality and Metadata

"As former NSA general counsel Stewart Baker said, 'Metadata absolutely tells you everything about somebody's life. If you have enough metadata you don't really need content. In 2014, former NSA and CIA director Michael Hayden remarked, 'We kill people based on metadata.'" [Schneier, 2015, *Data and Goliath*, p. 23]



Source: schneier.com

Food for thought, regarding a person's connections or associations: "If the local police department required us to notify it whenever we made a new friend, the nation would rebel. Yet we notify Facebook." [p. 47]



Source: www.amazon.com

Confidentiality and Privacy

Cell Phones

- Often send GPS information, thus recording your location.
- When you press a button on your phone, that executes a sequence of instructions (code) to perform a function.
  - For example, some program code (possibly rogue software) can turn on the phone's camera or microphone.
- Hackers might do this behind the scenes (without you pressing a button).

The same applies to your laptop or home computer.

Obviously, this is a threat to privacy.

Confidentiality and Privacy

E-mail

- Suppose a huge Internet-related organization, government, hacker, etc., says it will make all e-mails from the past k years public?
  - What would be your reaction?
- Be careful about what you write in e-mail or post on Web sites, including social media like Facebook and thousands of other such sites.
- Even if your e-mail provider, and you, are careful with keeping your sent/received e-mails safe (and even if you delete some), remember: the person to whom you sent that e-mail has a copy and might not be as concerned with security.

Your Web browser remembers which websites you've visited (history).

Confidentiality and Privacy

Social media influences public opinion.

Lots of personal information gets shared.

Abuses? Plenty.

- Read some of Bruce Schneier's work. He's written numerous books and has an excellent monthly blog called "Crypto-Gram". Informative. Free. Easy to read.
  - https://www.schneier.com/crypto-gram.html

Computer Security:

The bad guys only need to find one loophole. The good guys need to close all potential loopholes.

Anonymity

Pseudonyms: fictional identities that can fill in for real identities in communications and transactions, but are otherwise known only to a trusted entity

- Commonly used with undergrad students at UBC (e.g., in *Piazza*—a discussion board for students): Some of our course tools have data that is retained by non-Canadian servers; hence, the caution.
  - The instructor usually needs to know who's posting; other students don't.
- *TurnItIn*—students can register with a pseudonym, but the pseudonym must be registered with the instructor

Proxies: trusted agents that are willing to engage in actions on behalf of an individual in a way that cannot be traced back to that person

- e.g., Tor, DuckDuck Go; private mode in browser clients

https://www.torproject.org/

Tor means "The Onion Router". It's an anonymizing tool used by journalists, human rights activists, hackers, law enforcement, etc.

- Tor bounces encrypted traffic among a number of voluntarily-provided servers around the world. They conceal the originating IP address, making it almost impossible to track.

Deep Web and Dark Web

The Deep Web refers to all of the Web pages that search engines cannot find/index—and this may account for 90-99% of the data out there. It includes databases, webmail pages, registration-required forums, etc.

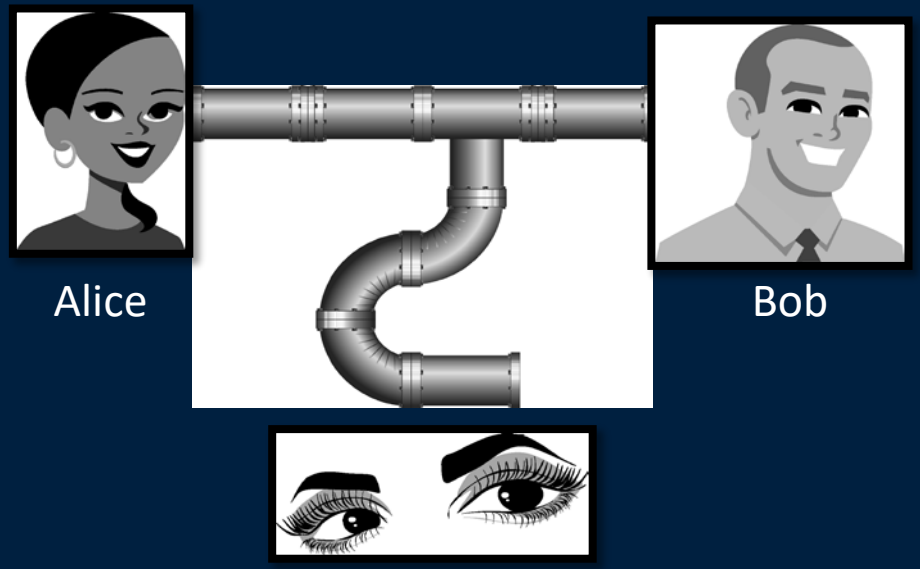
- e.g., Finding the prices of flights, hotels, etc. requires entering user data of some sort, and not just clicking on hyperlinks. Often, the user data (e.g., source, destination, price range, type of hotel) gets typed into a search box and is looked-up in a backend database system which returns the results.
- Client ↔ Application Web Server ↔ Database System

The Dark Web is a small part of the Deep Web and it usually deals with illegal things.

- Specialized software (e.g., Tor) is required to access these pages.
- An example is Silk Road—an online black market used for selling drugs, pornography, fake passports and other government documents, hackers' services, hit men, etc., and you could only get there by using Tor. Payment was often by Bitcoin.
  - It was shut down by the FBI in October 2013.

Threats and attacks

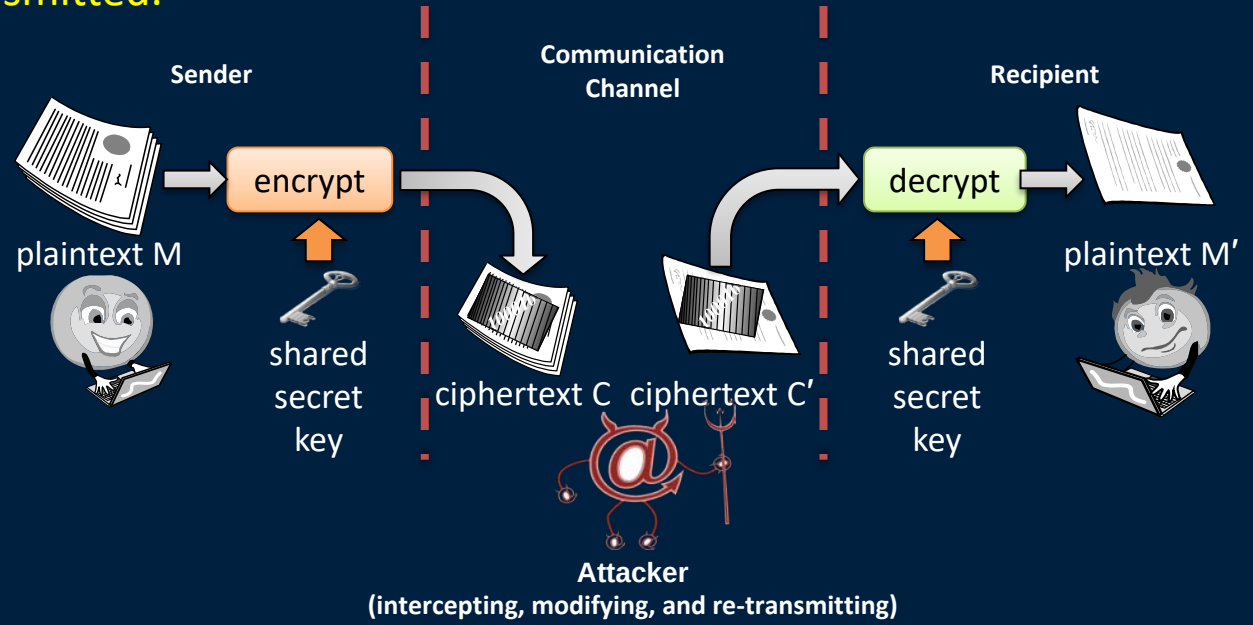
Eavesdropping: the interception of information intended for someone else during its transmission over a communication channel. (If the intent is to modify information, we would call this a man-in-the-middle attack.)



Threats and Attacks

Alteration: unauthorized modification of information

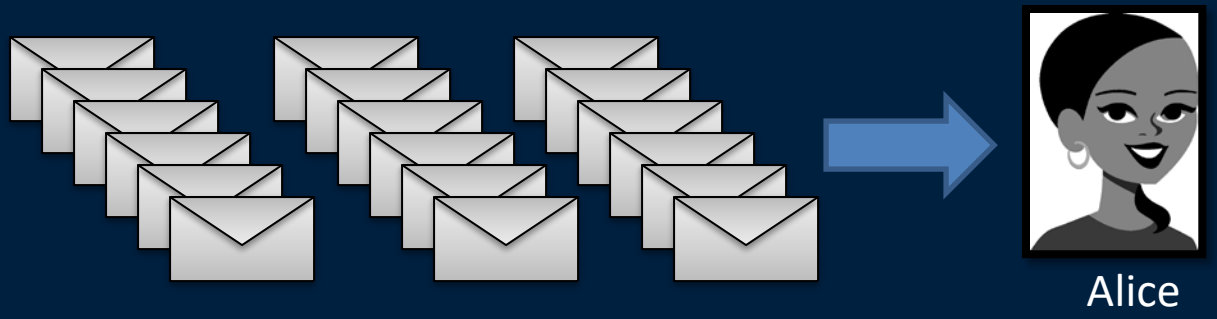
- **Example:** a man-in-the-middle attack where a network stream is intercepted, modified, and retransmitted.



Threats and attacks

Denial-of-service (DoS): the interruption or degradation of a data service or information access

- Example 1: E-mail spam, to the degree that it is meant to simply fill up a mail queue and slow down an e-mail server
- Example 2: Repeatedly sending http requests to a Web site, so that the server is overwhelmed, and maybe crashes.



Threats and attacks

Masquerading: the fabrication of information that is purported to be from someone who is not actually the author.

- Example: Spoofing (e.g., IP address, sender address) in an e-mail

"From: Alice"
(but it is actually from Eve)

Threats and attacks

Repudiation: the denial of a commitment or receipt of data

- This involves an attempt to back out of a real contract or a protocol, with the sender claiming he/she never sent (or received) it, or that the message was changed along the way.
- To prevent this, the different parties must provide receipts acknowledging that the data (contract) has been received, and that it's the sender's real message.

Data Provenance and Audit

Data Provenance: tracking the lineage of a data source and the parameters, assumptions, etc., of a workflow. For example, we may need to track the source of input files, data streams, software, scripts, etc.

- We can capture the source URL, contact person, lab name, timestamp, version of software, database schemas, documentation, parameters, assumptions, and anything to help others to consistently replicate results (reliability) or to convince them of data and processing validity (the right inputs and process).
  - e.g., X-rays (validation of digital images: subject, date, lab, technician, authorizing doctor, radiologist's comments)
  - e.g., files and parameters used to create a simulation (e.g., modeling of salmon migration in the Fraser River: location, weather, time of day, water temperature, water velocity, water turbidity)

Audit Records: extremely important in many domains

- Related to provenance

Principle of Least Privilege

This principle states that the default configuration of a system should have a conservative protection scheme.

- Operating systems and applications often have default options that favor usability over security.

The principle of least privilege:

- Each program and user of a computer system should operate with the bare minimum privileges necessary to function properly.
- If this principle is enforced, abuse of privileges is restricted, and the damage caused by the compromise of a particular application or user account is minimized.
- The military concept of 'need-to-know' information is an example of this principle.

Open Design Principle

According to this principle, the security architecture and design of a system should be made publicly available.

- Security should rely only on keeping cryptographic keys secret, and not keeping the algorithm secret.

Open design allows for a system to be scrutinized by multiple parties, which leads to the early discovery and correction of security vulnerabilities caused by design errors.

The open design principle is the opposite of the approach known as security by obscurity, which tries to achieve security by keeping cryptographic algorithms secret and which historically has been used without success by several organizations.

Access control concerns

Users and Groups

Authentication

Passwords

File Protection

Access Control Lists (ACLs)

- Which users can read/write which files?
- Are my files really safe from others?
- What does it mean to have “root” authority?
- What do we really want to control?

Access Control Matrix

This is a table that defines permissions.

Each row of the table is associated with a subject

- a user, group, or system that can perform actions.

Each column of the table is associated with an object

- a file, directory, document, device, resource, or any other entity for which we want to define access rights.

Each cell of the table is then filled with the access rights for the associated combination of subject and object.

- Access rights can include one or more actions such as reading, writing, copying, executing, deleting, and annotating.
- An empty cell means that no access rights are granted.

Access Control Matrix

objects

subjects

	/etc/passwd	/usr/bin/	/u/roberto/	/admin/
root	read, write	read, write, exec	read, write, exec	read, write, exec
mike	read	read, exec		
roberto	read	read, exec	read, write, exec	
backup	read	read, exec	read, exec	read, exec
...	...	...	...	...

Access right

Capabilities Lists

This is a subject-centered approach to access control. For each subject *s*, there is a list of objects for which *s* has rights, together with the specific rights for each such object.

- Disadvantage: If you want to see who has rights to a particular object, you would have to search through all the lists.
- A capabilities list is essentially a subset of the rows of an access control matrix.

A variant of this is the "Access Control List" which lists, for each object, all of the subjects that have access to that object.

- This is equivalent to a subset of the access control matrix's columns.

Role Based Access Control

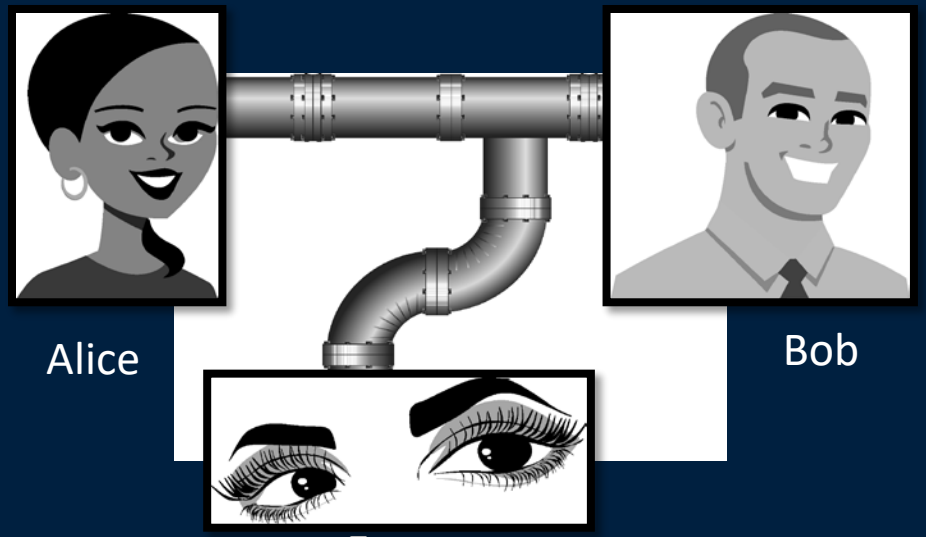
Define **roles** and then specify access control rights for these roles, rather than for the subjects directly.

Faculty member

Department Head

Cryptography Concepts

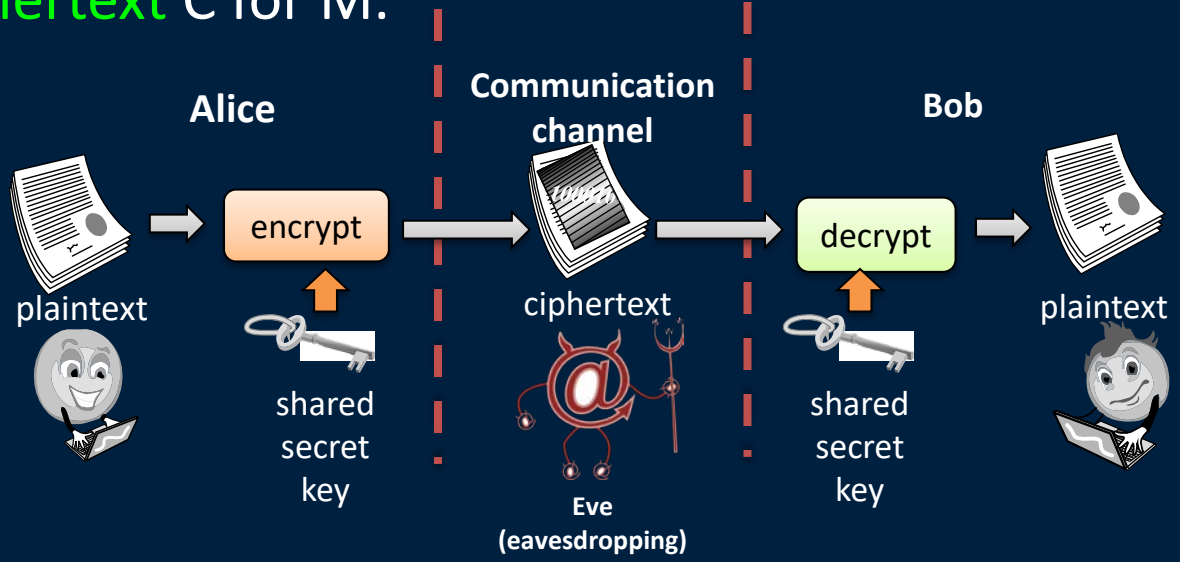
Encryption: a means to allow two parties, traditionally called Alice and Bob, to establish confidential communication over an insecure channel that is subject to eavesdropping. e.g., home <-> work



Encryption and Decryption

The message M is called the plaintext.

Alice will convert plaintext M to an encrypted form using an encryption algorithm E (e.g., Advanced Encryption Standard (AES)) that outputs a ciphertext C for M.



Encryption and Decryption

As equations:

$$C = E(M)$$

$$M = D(C)$$

The encryption and decryption algorithms are chosen so that it is infeasible (i.e., almost impossible) for someone other than Alice and Bob to determine the plaintext M from the ciphertext C. Thus, ciphertext C can safely be transmitted over an insecure channel that can be eavesdropped by an adversary.

Caesar Cipher

M = {sequence of letters m}

K = { i | i is an integer and $0 \le i \le 25$}

E = { E_k | $k \in K$ and for all letters m, $E_k(m) = (m + k) \bmod 26$ }

D = { D_k | $k \in K$ and for all letters c, $D_k(m) = (26 + c - k) \bmod 26$}

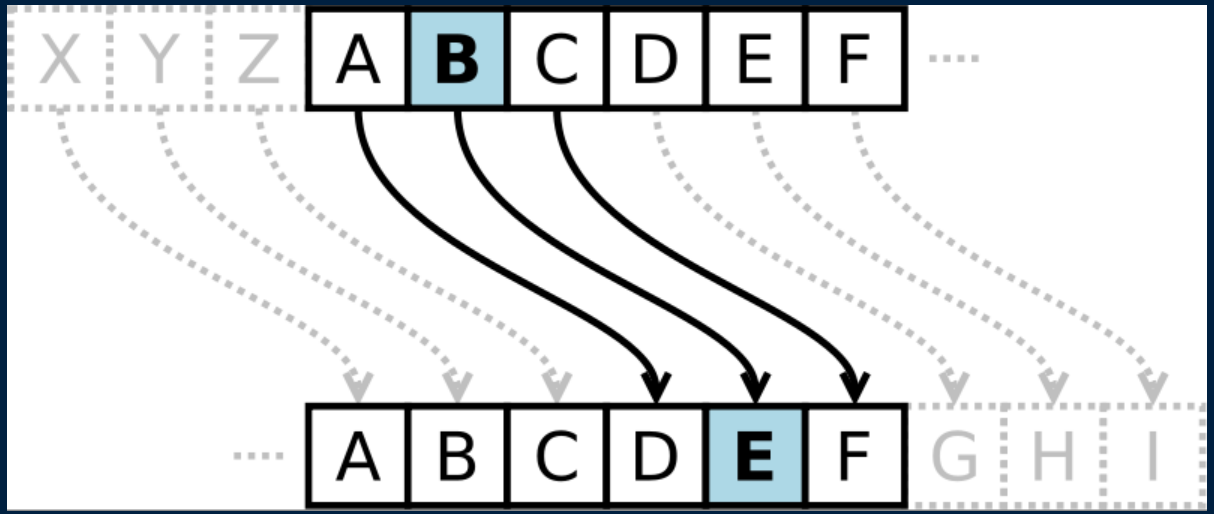
If k = 3, $E_3(m) = (m + 3) \bmod 26$ and $D_k(c) = (26 + c - 3) \bmod 26$

If M = B $\quad E_3(B) = (1 + 3) \bmod 26 = 4 \Rightarrow E$

$D_3(E) = (26 + 4 - 3) \bmod 26 = 27 \bmod 6 = 1 \Rightarrow B$

Caesar Cipher

This substitution cipher is very simple, and very easy to break: Replace each letter with the one “three over” (for example) in the alphabet.



Example[1]

Cyphertext = KHOOR ZRUOG

Computation of the frequency of each letter in cyphertext:

- G -> 0.1
- H -> 0.1
- K -> 0.1
- O -> 0.3
- R -> 0.2
- U -> 0.1
- Z -> 0.1

1-gram model for English language

a	0.080	h	0.060	n	0.070	t	0.090
b	0.015	i	0.065	o	0.080	u	0.030
c	0.030	j	0.005	p	0.020	v	0.010
d	0.040	k	0.005	q	0.002	w	0.015
e	0.130	l	0.035	r	0.065	x	0.005
f	0.020	m	0.030	s	0.060	y	0.020
g	0.060					z	0.002

[1] Example from Matt Bishop - http://nob.cs.ucdavis.edu/book/book-intro/slides/08.pdf

Example[1]

$f(c)$ frequency of character c in ciphertext

$\phi(i) = \sum_{0 \le i \le 25} f(c)p(c-i)$ correlation of frequency in cyphertext

- G -> 0.1
- H -> 0.1
- K -> 0.1
- O -> 0.3
- R -> 0.2
- U -> 0.1
- Z -> 0.1

So, in the example,

$$\varphi(i) = 0.1p(6-i) + 0.1p(7-i) + (0.1p(10-i) + 0.3p(14-i) + 0.2p(17-i) + 0.1p(20-i) + 0.1p(25.i)$$

Where $p(x)$ is the frequency of the character x in English

[1] Example from Matt Bishop - http://nob.cs.ucdavis.edu/book/book-intro/slides/08.pdf

Example[1]

$\phi(i)$ for the cyphertext

i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$	i	$\varphi(i)$
0	0.0482	7	0.0442	13	0.0520	19	0.0315
1	0.0364	8	0.0202	14	0.0535	20	0.0302
2	0.0410	9	0.0267	15	0.0226	21	0.0517
3	0.0575	10	0.0635	16	0.0322	22	0.0380
4	0.0252	11	0.0262	17	0.0392	23	0.0370
5	0.0190	12	0.0325	18	0.0299	24	0.0316
6	0.0660					25	0.0430

[1] Example from Matt Bishop - http://nob.cs.ucdavis.edu/book/book-intro/slides/08.pdf

Example[1]

Most probable keys based on φ

- $i = 6$ $\varphi(i) = 0.0660$ decrypted text: EBIIL TLOLA
- $i = 10$ $\varphi(i) = 00635$ decrypted text: AXEEH PHKEW
- $i = 3$ $\varphi(i) = 0.0575$ decrypted text: HELLO WORLD
- $i = 14$ $\varphi(i) = 0.0535$ decrypted text: WTAAD LDGAS

The key is 3.

[1] Example from Matt Bishop - http://nob.cs.ucdavis.edu/book/book-intro/slides/08.pdf

Caesar's Problem[1]

Key is too short

- Can be found by exhaustive search
- Statistical frequencies not concealed well
  - They look too much like regular English letters

So make it longer

- Multiple letters in key
- Idea is to smooth the statistical frequencies to make cryptanalysis harder

[1] Example from Matt Bishop - http://nob.cs.ucdavis.edu/book/book-intro/slides/08.pdf

Vigenère Cipher[1]

Same as cipher but uses a phrase

Example:

- Message : THE BOY HAS THE BALL
- Key: VIG
- Encipher using Caesar cipher for each letter:

Key	VIGVIGVIGVIGVIGV
Plain	THEBOYHASTHEBALL
Cipher	OPKWWECIYOPKWIRG

[1] Example from Matt Bishop - http://nob.cs.ucdavis.edu/book/book-intro/slides/08.pdf

One-Time Pad

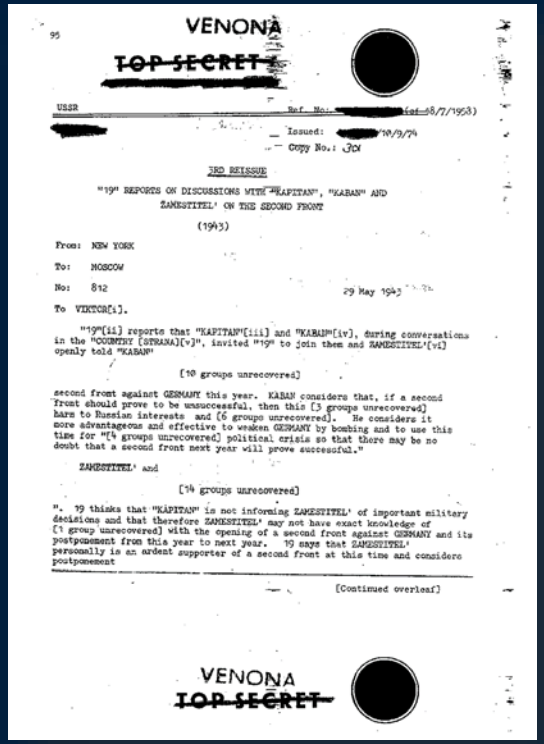
There is one type of substitution cipher that is unbreakable, and it uses a one-time pad. It was invented in 1917.

- We use a block of shift keys, (k1, k2, . . . , kn), to encrypt a plaintext, M, of length n, with each shift key being chosen uniformly at random.

In spite of their perfect security, one-time pads have some weaknesses:

- The key has to be as long as the plaintext.
- How do you store and transmit it?
- Keys can never be reused.

Repeated use of one-time pads allowed the U.S. to break some of the communications of Soviet spies during the Cold War.



VENONA

~~TOP SECRET~~ [redacted]

USSR Ref. No.: [redacted] (of 8/7/1958)

[redacted] Issued: [redacted] 30/9/74

Copy No.: 370

3RD RELEASE

"19" REPORTS ON DISCUSSIONS WITH "KAPITAN", "KABAN" AND ZAMESTITEL' ON THE SECOND FRONT

(1943)

From: NEW YORK

To: MOSCOW

No: 812

29 May 1943

To VIKTOR[i].

"19"[ii] reports that "KAPITAN"[iii] and "KABAN"[iv], during conversations in the "COUNTRY [STRANA][v]", invited "19" to join them and ZAMESTITEL'[vi] openly told "KABAN"

[10 groups unrecovered]

second front against GERMANY this year. KABAN considers that, if a second front should prove to be unsuccessful, then this [3 groups unrecovered] harm to Russian interests and [6 groups unrecovered]. He considers it more advantageous and effective to weaken GERMANY by bombing and to use this time for "[4 groups unrecovered] political crisis so that there may be no doubt that a second front next year will prove successful."

ZAMESTITEL' and

[14 groups unrecovered]

". 19 thinks that "KAPITAN" is not informing ZAMESTITEL' of important military decisions and that therefore ZAMESTITEL' may not have exact knowledge of [1 group unrecovered] with the opening of a second front against GERMANY and its postponement from this year to next year. 19 says that ZAMESTITEL' personally is an ardent supporter of a second front at this time and considers postponement

[Continued overleaf]

VENONA

~~TOP SECRET~~ [redacted]

Symmetric Cryptosystems

When Alice and Bob share a secret key, which is used for both encryption and decryption.

=> symmetric cryptography

Block Ciphers

In a block cipher:

- Plain text and ciphertext have fixed block length b (e.g. 128 bits)
- Plain text of length n is partitioned into a sequence of m blocks
  - $P[0], \ldots, P[m-1]$ where $n \leq bm < n + b$

Each message is divided into a sequence into a sequence of blocks and encrypted or decrypted in terms of its blocks

Block cipher

Block ciphers require that the length of the message be a multiple of the block size

Padding of the last block must be unambiguous

A common padding method is PKCS5

- Sequence of identical bytes, each indicating the length (in bytes) of the padding
- Example:
  - b = 128 (16 bytes)
  - Plain text: “Roberto” (7 bytes)
  - Padded Plain text: “Roberto999999999” (16 bytes) – 9 is the number, not the character.

Block cipher in Practice

Data Encryption Standard (DES)

- Developed by IBM and adopted by NIST in 1977
- 64-bit blocks and 56-bit keys
- Small key space makes exhaustive search (brute force) attacks feasible since late 90s

Triple DES (3DES)

- Nested application of DES with 3 different keys KA, KB, KC
- Effective key length is 168 bits, making exhaustive search attacks unfeasible
- $C = (E_{KC}(D_{KB}(E_{KA}(M); M=D_{KA}(E_{KB}(D_{KC}(C))$
- Equivalent to DES when KA = KB = KC (backward compatible)

Advanced Encryption Standard (AES)

- Selected by NIST in 2001 through open international competition and public discussion
- 128-bit blocks, several possible key lengths: 128, 192 and 256 bits (AES-128, AES-192, AES-256)
- Exhaustive search attack not currently possible
- AES-256 is the symmetric encryption algorithm of choice

AES – Advanced Encryption Standard

AES proceeds in a fixed number of rounds

Each round performs an invertible transformation on a 128-bit array called state.

During a round, the algorithm uses the previous state and performs substitution, permutations and a final XOR operation using a round key derived from the encryption key to generate the next state.

The initial state X_0 is the XOR of the plain text P with the key K

- X_0 = P XOR K

The ciphertext is the output of the final round.

AES

https://youtu.be/O4xNJsjtN6E

Block Cypher Modes

A cypher mode describes how a block cipher encrypts and decrypts a sequence of message blocks

- Electronic Code Book (ECB)
  - Plain text block encrypted and decrypted as separate units, then recombined
  - Very simple, can be parallelized, can tolerate the loss or damage of a block
  - Weakness: patterns in plan text are repeated in cyphertext, so not suitable for documents and images.

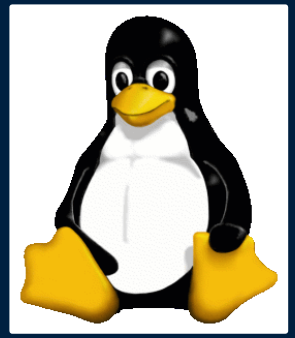
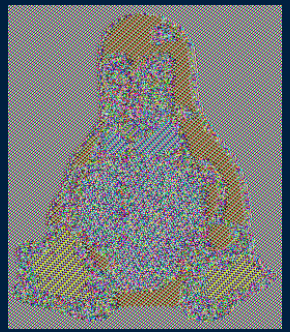


Image by Larry Ewing, lewing@isc.tamu.edu

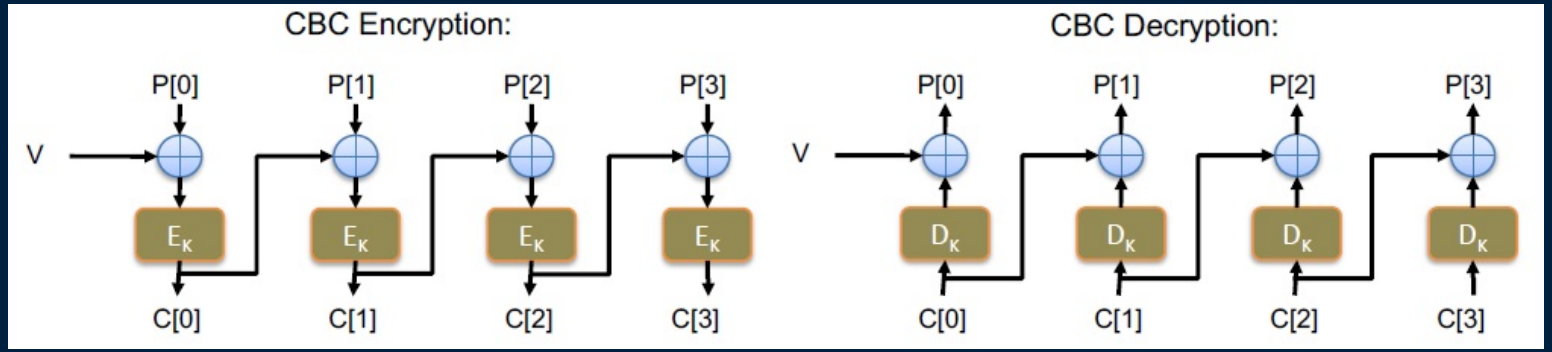


Encrypted using ECB mode
https://github.com/robertdavidgraham/ecb-penguin

Block Cypher Modes

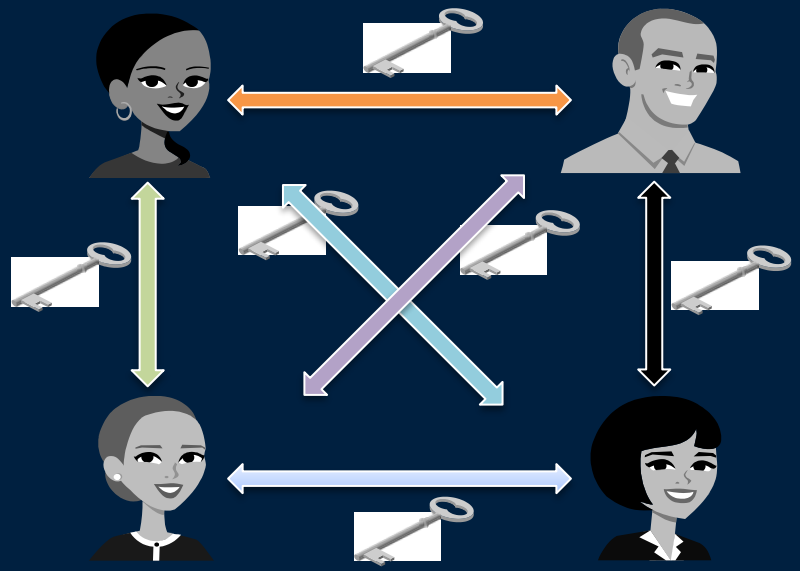
Cypher Block Chaining (CBC) Mode

- Previous ciphertext block is combined with current plaintext mode block
  - $C[i] = E_k(C[i-1]) \oplus P[i]$
- Initialization vector V is a random block separately transmitted encrypted
- Decryption is done in the same manner
  - $P[i] = C[i-1] \oplus D_k(C[i])$



Symmetric Key Distribution

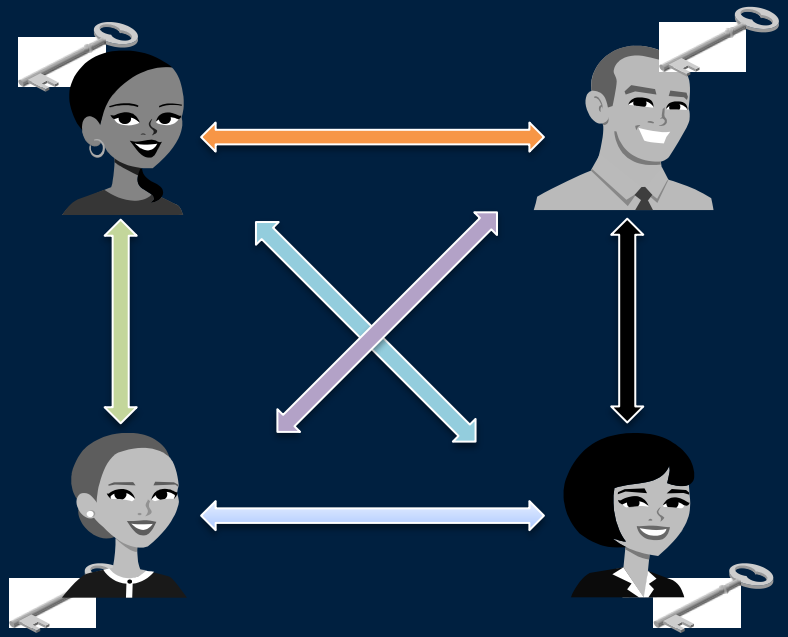
Requires each pair of communicating parties to have a different secret key.



$n(n-1)/2$ keys to be distributed

Public Key Cryptography

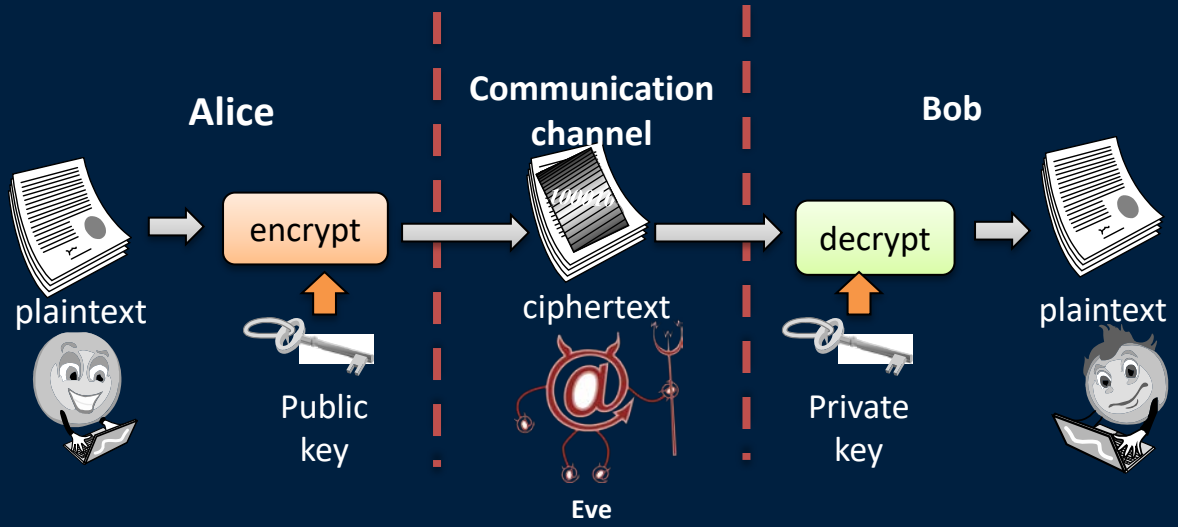
Common public key, one associated private key per party.



n keys to be distributed

Public Key Cryptography

Asymmetric cryptography: Separate keys are used for encryption and decryption



Public cryptography

Interest:

- Confidentiality
  - Encrypt using public key - decrypt using private key
- Integrity / authentication
  - Encrypt using private key - decrypt using public key

Must:

1. be computationally fast
2. be infeasible to derive the private key from the public key

Diffie-Hellman key exchange protocol

Diffie-Helmann key exchange

Multiplicative group of integer modulo p

- p is a prime number
- g is a primitive root modulo p

Alice and Bob publicly agree to use $p = 23$ and $g = 5$.

Alice and Bob use the function $Y = g^x \, mod \, p$ using a number of their choosing x.

- Alice chooses a secret integer $x = 4 \Rightarrow Y = 5^4 \, mod \, 23 = 4$ => sent to Bob $Y = 4$
- Bob chooses a secret integer $x = 3 \Rightarrow Y = 5^3 \, mod \, 23 = 10$ => sent to Alice $Y = 10$

Secret key is compute using $s = Y^x \, mod \, p$

- Alice: $s = 10^4 \, mod \, 23 = 18$ Bob: $s = 4^3 \, mod \, 23 = 18$

Alice and Bob now share a common secret key s!

Diffie-Hellmann key exchange

The protocol is considered secure if p and g are large.

RSA is the algorithm of choice for public/private key encryption generation.

However, the protocol continue to be used to safely exchange keys over the Internet

RSA – River, Shamir, and Adelman

Public key cryptosystem – was publicly described in 1977

Relies on the practical difficulty of factoring the product of two large prime numbers.

Uses exponential functions

Relatively slow algorithm, so can be combined with symmetric key cryptography for large confidential messages.

RSA

Based on

- 2 prime numbers p and q that are privately chosen
- Totient function $\varphi(n)$ or $\lambda(n)$ where $n = pq$
  - can be deleted once the private key has been generated

Public key

- Modulus n, $n = pq$
- Exponent e, chosen so that e and $\varphi(n)$or $\lambda(n)$ are coprime

Private key

- Modulus n
- Exponent d, computed using $d.e \equiv 1$ (mod $\varphi(n)$ or $\lambda(n)$)

Encryption $c \equiv me(mod\ n)$

Decryption $m \equiv c^d\ (mod\ n)$

Example

Euler's totient function $\varphi(n) = (p-1)(q-1)$

- Number of coprime numbers

Let's assume that Alice chooses $p = 7$ and $q = 11 \Rightarrow n = 77$

- $\varphi(77) = 6 \times 10 = 60$

Alice must now choose e to be a coprime of $\varphi(77)$

- $e = 17 \qquad \Rightarrow 17.d \ mod\ (60) \equiv 1 \Rightarrow d = 53$

So Alice sends publicly $n = 77$ and $e = 17$ but keeps d private

Example

Let's assume that Bob want to send a private message to Alice. For example, what to use for shared symmetric encryption.

Bob wants to send AES 128 followed by the shared encryption key.

AES 128 will be transformed as 00 04 18 01 02 08

- $00^{17} \bmod 77 = 00$
- $04^{17} \bmod 77 = 16$
- $18^{17} \bmod 77 = 72$
- $01^{17} \bmod 77 = 01$
- $02^{17} \bmod 77 = 18$
- $08^{17} \bmod 77 = 57$

Bob sends 00 16 72 01 18 57

Example

Alice receives 00 16 72 01 18 57

To decrypt she used her private key $d = 53$

- $00^{53} \bmod 77 = 00$
- $16^{53} \bmod 77 = 04$
- $72^{53} \bmod 77 = 18$
- $01^{53} \bmod 77 = 01$
- $18^{53} \bmod 77 = 02$
- $57^{53} \bmod 77 = 08$

So Alice translate the message to AES 128 and expect the key to follow.

The message is confidential because only Alice can decrypt. She is the only one to have the private key.

Example

Alice wants to send an acknowledgment to Bob

- ACKOK => 00 02 10 14 10 encrypted using her private key
  - $00^{53} \bmod 77 = 00$
  - $02^{53} \bmod 77 = 74$
  - $10^{53} \bmod 77 = 54$
  - $14^{53} \bmod 77 = 01$
  - $10^{53} \bmod 77 = 54$

$\Rightarrow$ Alice sends 00 74 54 01 54

Bob receives 00 74 54 01 54 and decrypt using Alice public key

- $00^{17} \bmod 77 = 00$
- $74^{17} \bmod 77 = 02$
- $54^{17} \bmod 77 = 10$
- $01^{17} \bmod 77 = 01$
- $54^{17} \bmod 77 = 10$

$\Rightarrow$ Bob decrypt 00 02 10 01 10

$\Rightarrow$ Translated as ACKOK

Alice is the only one that can have sent the message (authentication)

And the message has not be modified (integrity)

Cryptographic Hash Functions

Cryptographic hash (digest) is a ‘signature’ for a text or a data file.

It is used to create a checksum on a message M such that it is:

- One-way: easy to compute Y=H(M), but hard to find M given only Y.
- Collision resistant: hard to find two messages M and N such that H(M)=H(N)

Popular Example: Secure Hash Algorithm or SHA-256

- Results in hash values (aka hash messages or digests) that are 256 bits long (= 32 bytes → very strong)
- SHA-512: 512 bits (64 bytes)
- Other cryptographic hash algorithms are MD5 (16 bytes), MD6 (up to 64 bytes), RIPEMD (up to 40 bytes), etc.

Cryptographic Hash Functions

Provides: (a) authentication of the source, and (b) integrity (i.e., that the message hasn't been altered)

- Authentication is often more important than secrecy, in business.
- Used to protect file systems (e.g., used to verify the integrity of files)
- Helps provide integrity of purchased/downloaded software
- Used by the IPSec protocol to ensure that IP packets haven't been altered along the way
- Used in banking protocols
- Used for challenge handshake authentication
  - Send hash of password instead of clear text for validation by a server
- Used to compute Digital Signatures
  - First produce a hash value (digest) from a document, then encrypt the hash value (not the document) using public-key cryptography (RSA). Send in clear both the document and the encrypted hash.

Example

Suppose Alice wants to send Bob a price list for a bunch of products, but she doesn't care if the rest of the world sees it because the price are not confidential. Bob is going to make important decisions based on these prices, and he wants to make sure the price list was really generated by Alice, and that it hasn't changed while in transit.

To do so, Alice:

- Generate a hash value H for the price list using SHA-256
- Alice use Bob's public key to encrypt H, yielding H'. (this is the signature part)
- Alice sends H' to Bob and the price list in plain text
- Bob uses its private key to decrypt H' (to yield H)
- Bob confirms the hash value H by running the same hash algorithm on the price list.

What is wrong?

Example

Suppose Alice wants to send Bob a price list for a bunch of products, but she doesn't care if the rest of the world sees it because the price are not confidential. Bob is going to make important decisions based on these prices, and he wants to make sure the price list was really generated by Alice, and that it hasn't changed while in transit.

To do so, Alice:

- Generate a hash value H for the price list using SHA-256
- Alice use Bob's public key to encrypt H, yielding H'. (this is the signature part)
- Alice sends H' to Bob and the price list in plain text
- Bob uses its private key to decrypt H' (to yield H)
- Bob confirms the hash value H by running the same hash algorithm on the price list.

What is wrong? Should use her private key, Bob uses her public key

Digital certificates

Digital certificates are used to authenticate the sender and her public key

- Provides non-repudiation
- Alice public key really belongs to Alice and not some imposters pretending to be Alice
  - Must be vouched by a trusted higher authority
  - Alice public key is found in a public-key database and contains information about her

Digital certificate is a passport

- Provides identifying information, is tamper-proof and is issued by a trusted authority
- The trusted authority is called a certificate authority (CA)
- Browsers only trust certificates from CAs on their list of trusted CAs

Certificate Authorities

Certificates expire or get revoked

The CA manages a certificate revocation list (CRL)

- Similar to checking for expired or stolen credit cards

Symantec accounts for [from Wikipedia]

- about 1/3 of the digital certificates on the Web
- 44% of the one million busiest website

Verisign used to be the leader but it sold its certification business to Symantec

There are only a small number of CAs and their public keys are well-known

Passwords

Exercise: Suppose passwords must consist of 6 alphanumeric characters (taken from the set of lower-case letters + upper-case letters + digits 0-9).

1. How many unique password combinations are there?

2. How many unique password combinations are there if we require at least 1 digit in the password?

3. Which of (1) or (2) is “more secure”?

Password validity

https://password.kaspersky.com/

Allows you to test your password – and check how long it would take to break it with an average home computer.

- 6 characters => 3 hours
- 5 characters + 1 digit => 3 hours
- 9 alphabetic characters => 4 months
- 9 alphabetic characters (2upper + 7lower + 2 digits) => 4 months

Passwords length

26 UPPER/lower case characters = 52 choices, plus 10 more choices for a digit, plus 32 choices for a special character

=> a pool of 52 + 10 + 32 = 94 choices for each usable character

5 characters: 94^5 =	7,339,040,224 possibilities
6 characters: 94^6 =	689,869,781,056
7 characters: 94^7 =	64,847,759,419,264
8 characters: 94^8 =	6,095,689,385,410,816
9 characters: 94^9 =	572,994,802,228,616,704

Odd characters make password safer

Longer passwords are better

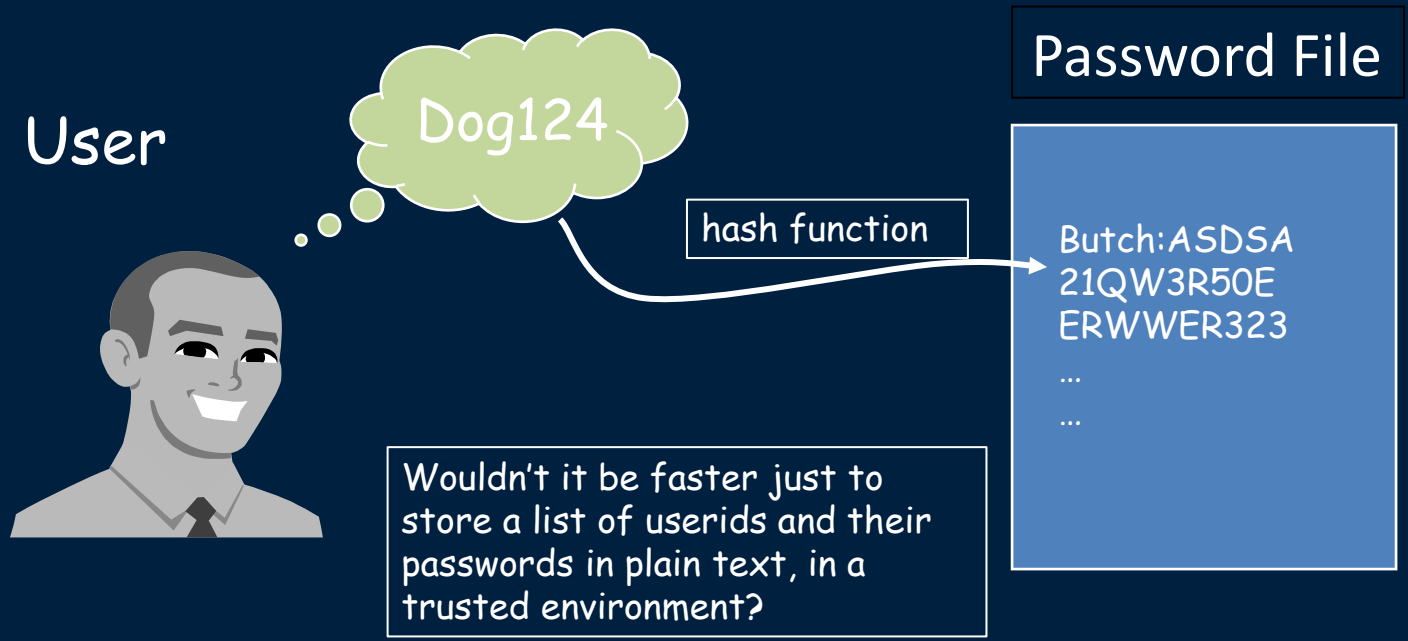
Passwords

A strong password includes characters from at least three of the following groups:

Group	Example
Lowercase letters	a, b, c, ...
Uppercase letters	A, B, C, ...
Numerals	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
Non-alphanumeric (symbols)	( ) ` ~ ! @ # $ % ^ & * - + = | \ { } [ ] : ; " ' < > , . ? /
Unicode characters	€, Γ, ƒ, and λ

Tip: Use pass phrases, e.g., "I re@lly want 2 have 1M$!"

How a password is stored




THE UNIVERSITY OF BRITISH COLUMBIA