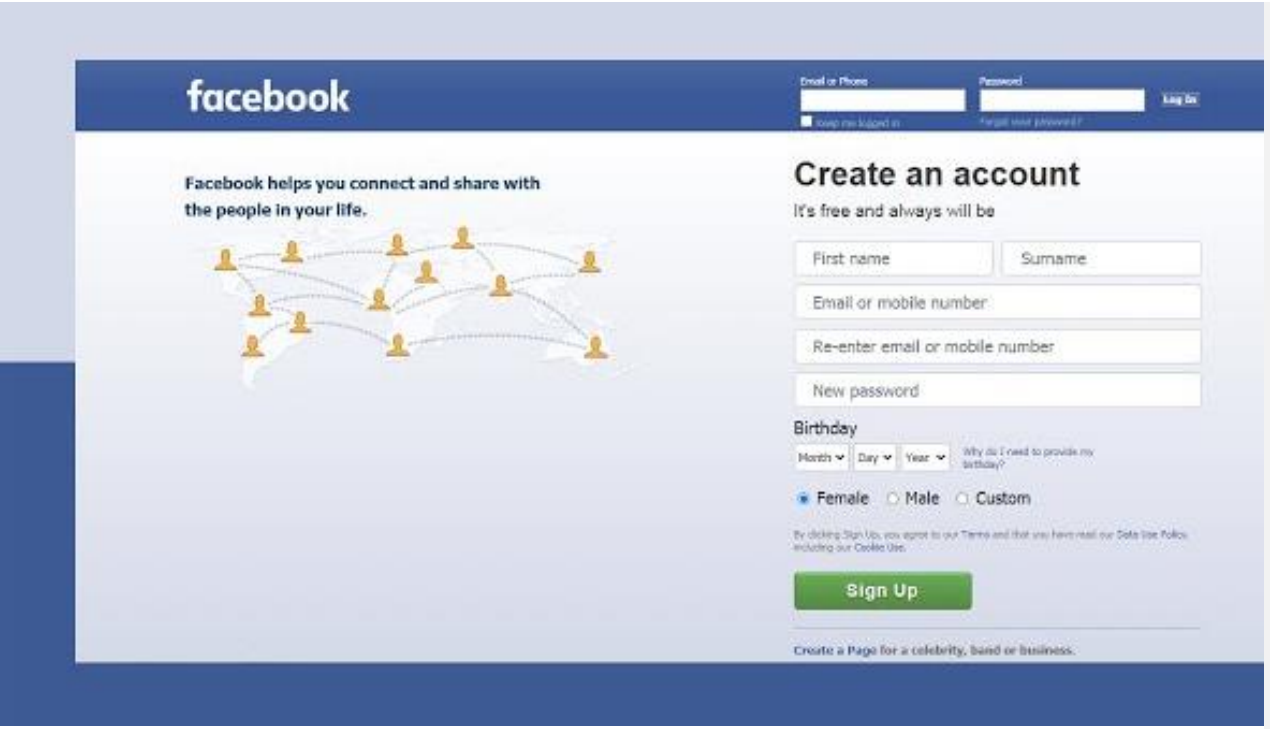
Lecture #3

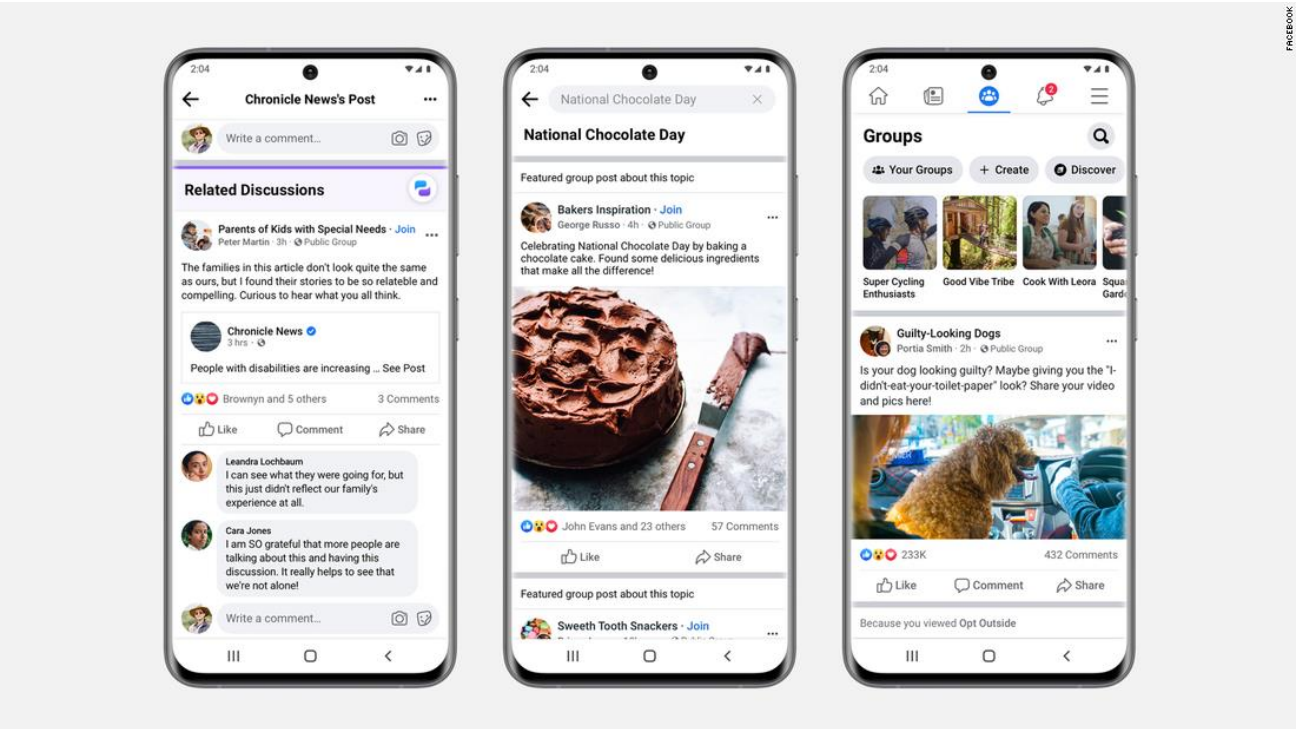
Data Privacy Preservation

Dr. Charnon Pattiyanon
Assistant Director of IT, Instructor
Department of Artificial Intelligence and Computer Engineering
CMKL University

Pervasiveness of Data Collection of Cyber Systems

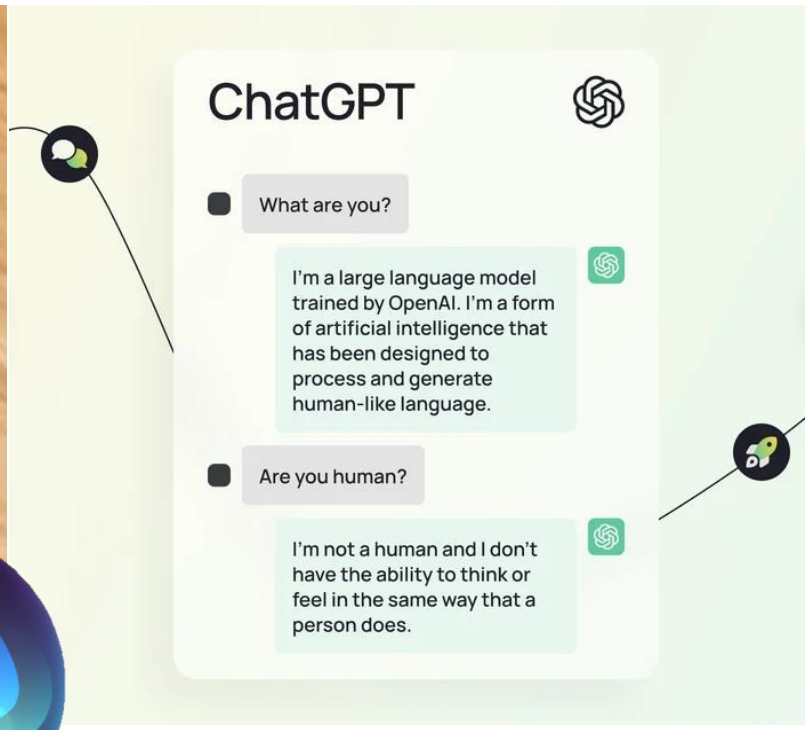
- Do you know how companies, software, hardware, or network collect your personal information?





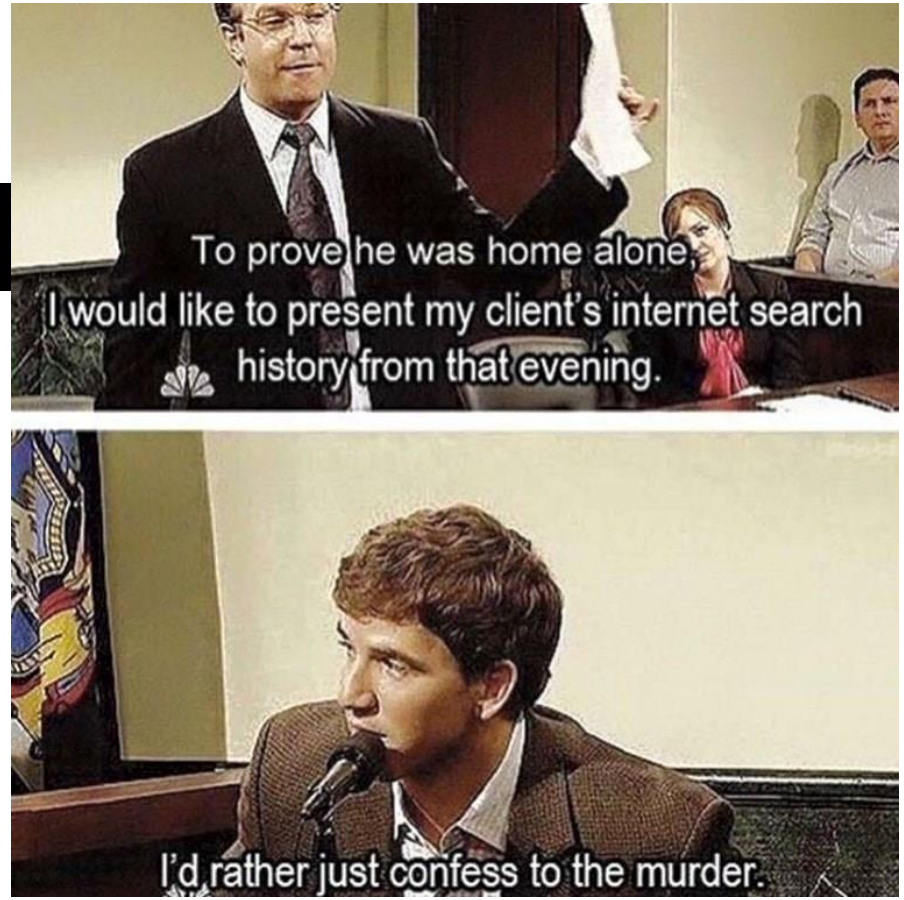
Pervasiveness of Data Collection of Cyber Systems

- Do you know how companies, software, hardware, or network collect your personal information?



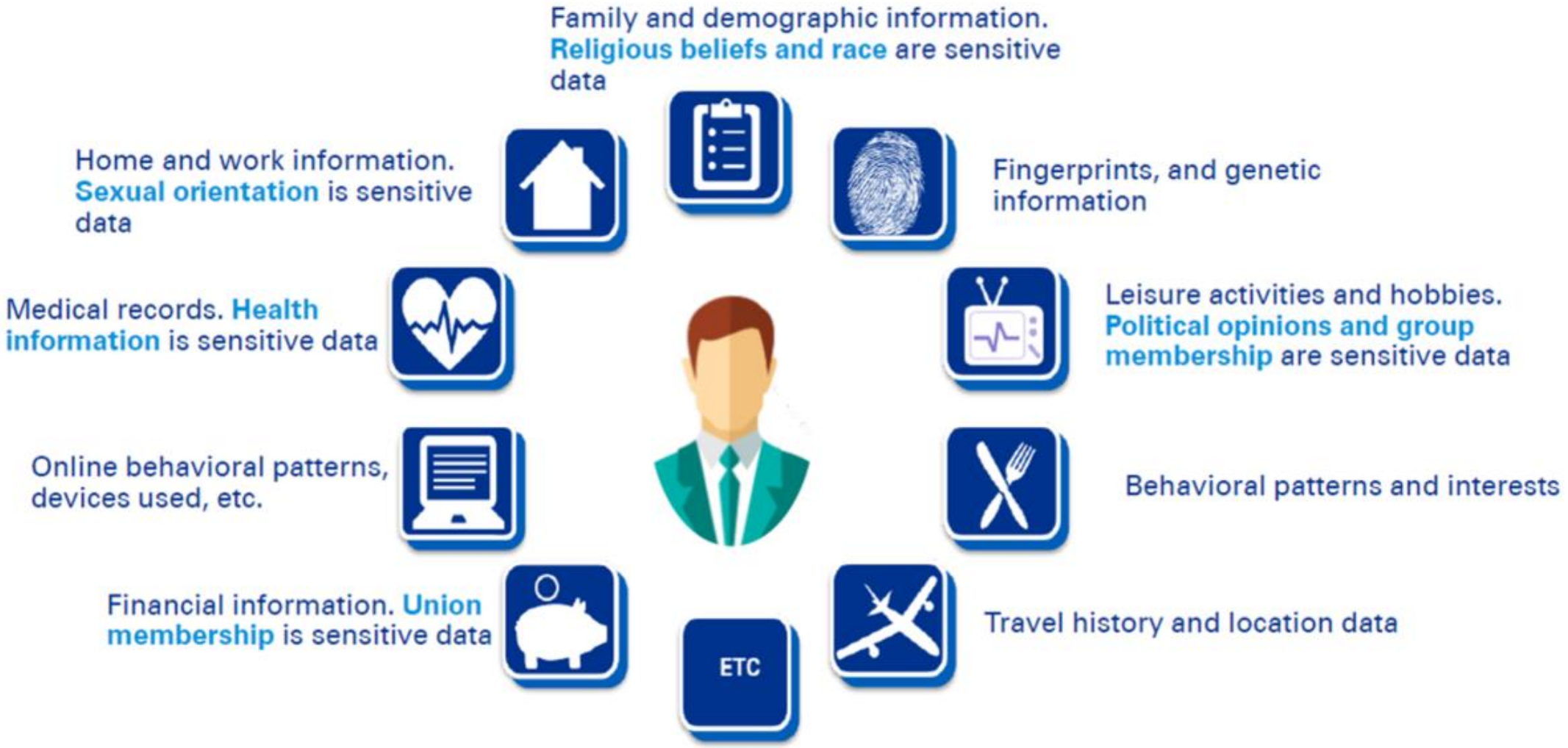
Pervasiveness of Data Collection of Cyber Systems

- Do you know how companies, software, hardware, or network collect your personal information?



CMKL Cybersecurity Competency Registration Form

Once Again, What is Personal Information?



Introducing Data Privacy Preservation

- **Data privacy preservation** is a discipline focused on defining who has access to personal or sensitive data.

Policies to define data protection **strategies** and **processes**

- Does the organization put forward to protect personal data?
- Does the management level personnel recognize the important of personal data protection?
- Does the organization have clear processes with data privacy protection awareness included?

Users have **full control** over their data

- Do users know what personal data they provided to the organization?
- Do users know about the purpose of using or manipulating their data?
- Can users reject or refuse the use of personal data when they are no longer have business with the organization?

Tools & Techniques to preserve data privacy must be in place

- Do IT or cyber systems in the organization employ data protection tools or techniques properly?
- With the tools/techniques employed, does the organization still get insight from the data?

Introducing Data Privacy Preservation

- **Data privacy preservation** is a discipline focused on defining who has access to personal or sensitive data.

Policies to define data protection **strategies** and **processes**

- Does the organization put forward to protect personal data?
- Does the management level personnel recognize the important of personal data protection?
- Does the organization have clear processes with data privacy protection awareness included?

Users have **full control** over their data

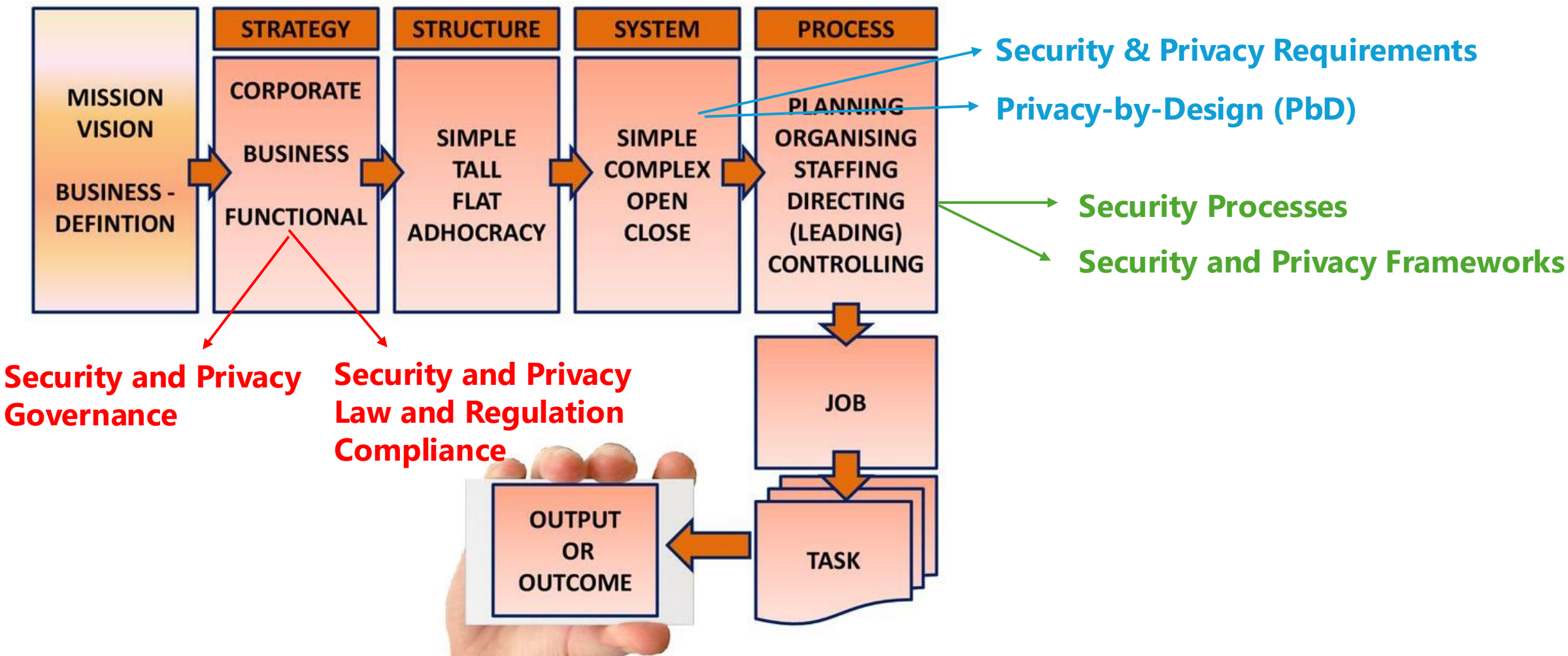
- Do users know what personal data they provided to the organization?
- Do users know about the purpose of using or manipulating their data?
- Can users reject or refuse the use of personal data when they are no longer have business with the organization?

Tools & Techniques to preserve data privacy must be in place

- Do IT or cyber systems in the organization employ data protection tools or techniques properly?
- With the tools/techniques employed, does the organization still get insight from the data?

How Companies Commit To Protect Data Privacy

- Security and privacy should be kept throughout the organizational processes.



Security and Privacy Governance in Organizations

- Data Gathering/Collection
- Data Analysis
- Data Cleansing
- Data Storage and Retrieval
- Data Disposal

- Data Analytics
- Data Insight Management
- Data Visualization

- Software Versioning
- Configuration Management
- Data Lifecycle
- Data Up-to-dateness
- Data Standard

- Security Logging
- Anomaly Detection
- Security & Privacy Auditing

- Data Encryption
- Data Access Control
- Network Security
- Application Security
- Infrastructure Security

- Personal Data Protection
- Privacy Policies
- Privacy Framework

Privacy Law and Regulatory Compliance

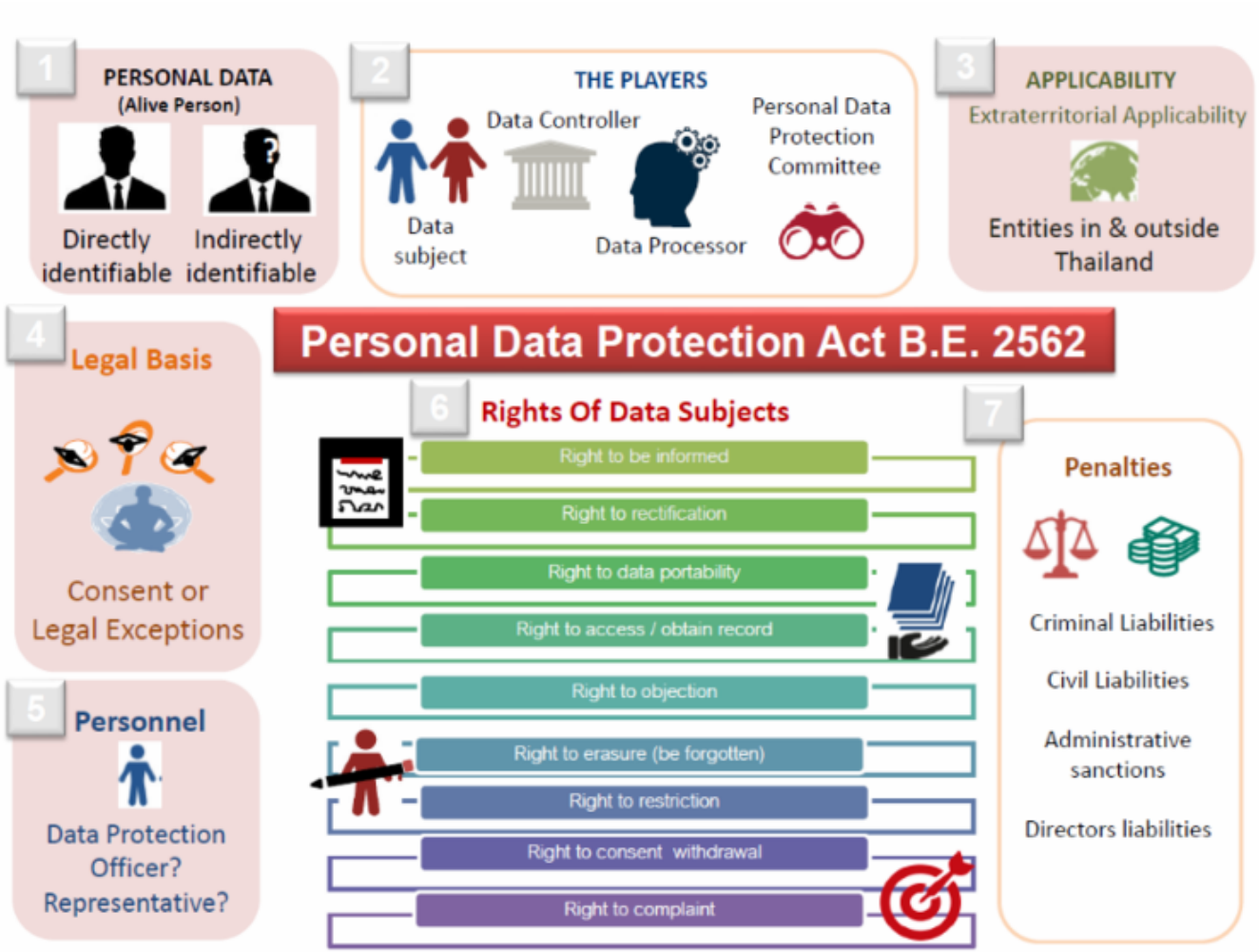
- There are many security/privacy laws and regulations enforced all over the world:
  - **EU General Data Protection Regulation (GDPR)**
  - **California Comsumer Protection Act (CCPA)**
  - **Health Insurance Portability and Accountability Act (HIPAA)**
  - **Thailand's Personal Data Protection Act B.E. 2562 (PDPA)**
  - **Thailand's Cybersecurity Act B.E. 2652 (CSA)**
  - **Thailand's Electronic Transaction Act B.E. 2544 (ETA)**

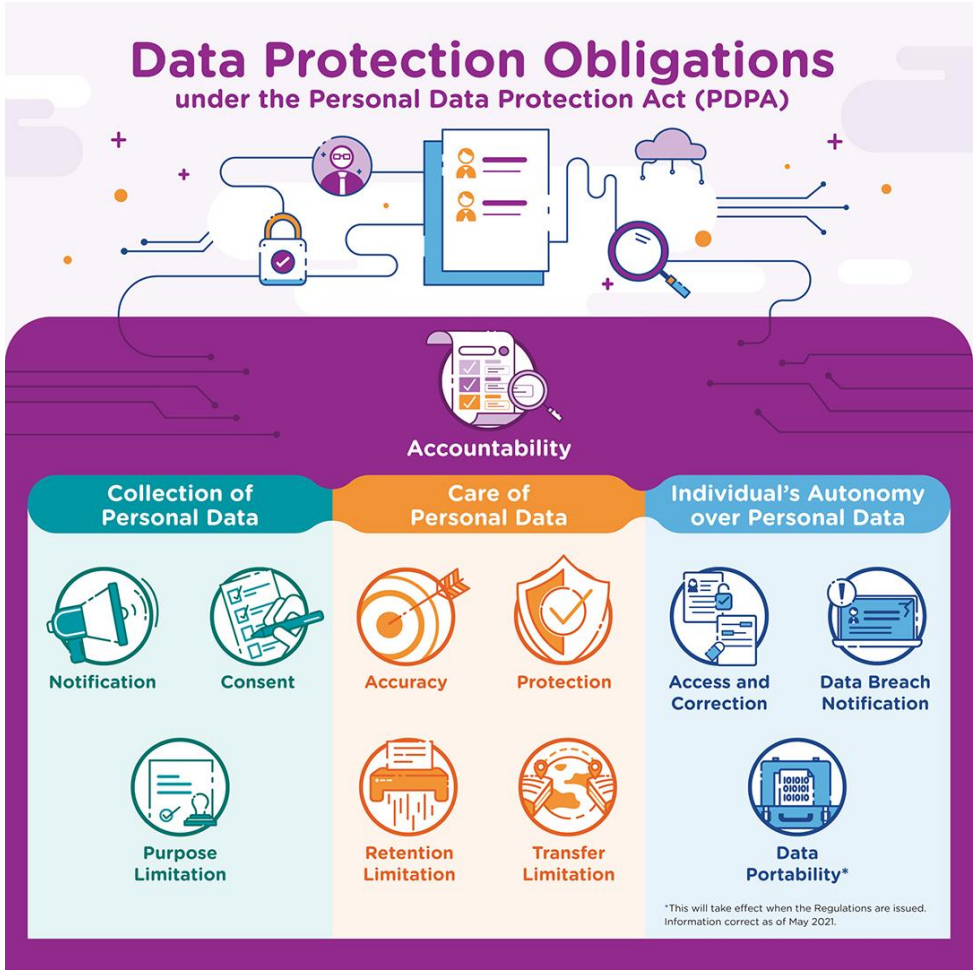
Privacy Law and Regulatory Compliance

- **EU General Data Protection Regulation (GDPR)**

Privacy Law and Regulatory Compliance

- Thailand's Personal Data Protection Act B.E. 2562 (PDPA)





Introducing Data Privacy Preservation

- **Data privacy preservation** is a discipline focused on defining who has access to personal or sensitive data.

Policies to define data protection **strategies** and **processes**

- Does the organization put forward to protect personal data?
- Does the management level personnel recognize the important of personal data protection?
- Does the organization have clear processes with data privacy protection awareness included?

Users have **full control** over their data

- Do users know what personal data they provided to the organization?
- Do users know about the purpose of using or manipulating their data?
- Can users reject or refuse the use of personal data when they are no longer have business with the organization?

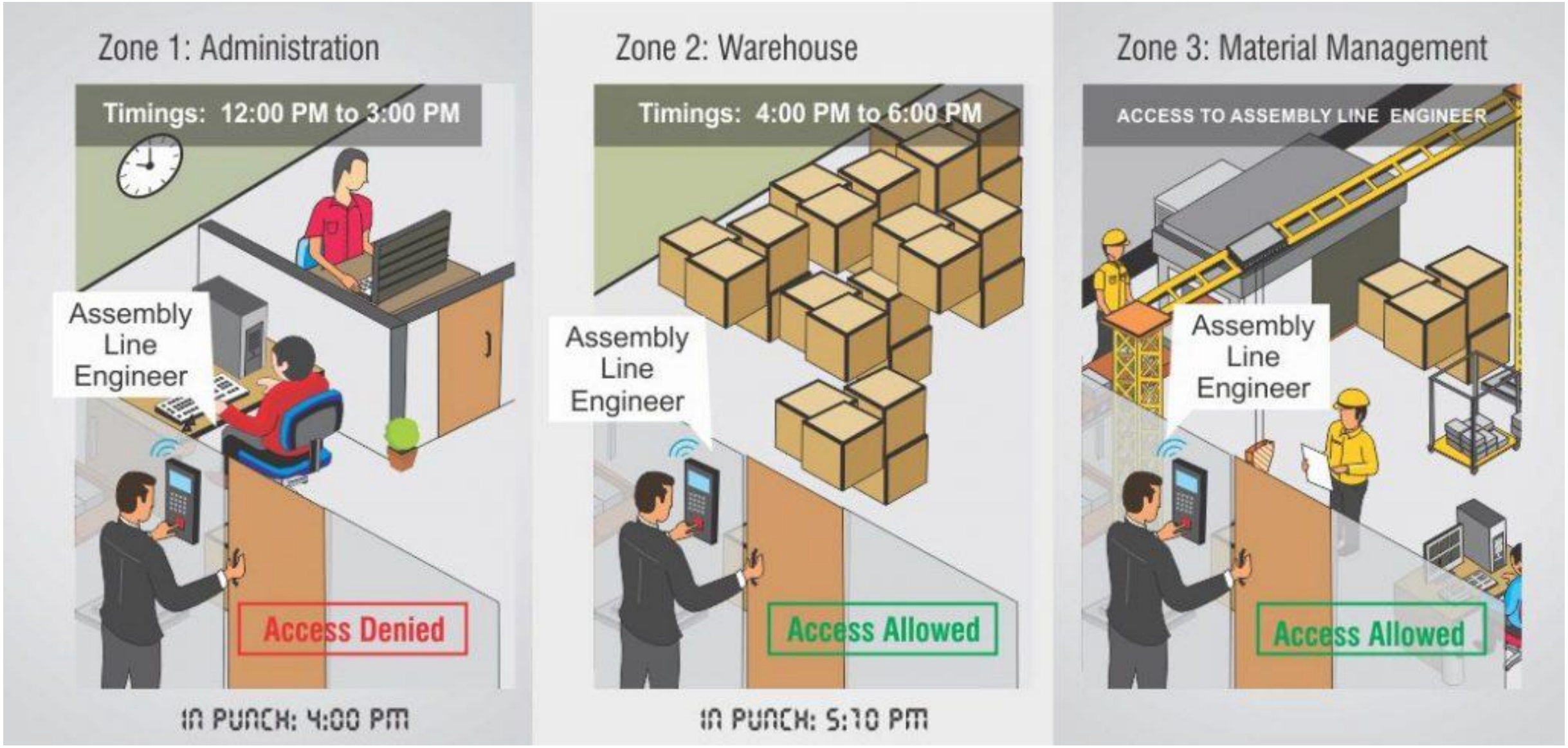
Tools & Techniques to preserve data privacy must be in place

- Do IT or cyber systems in the organization employ data protection tools or techniques properly?
- With the tools/techniques employed, does the organization still get insight from the data?

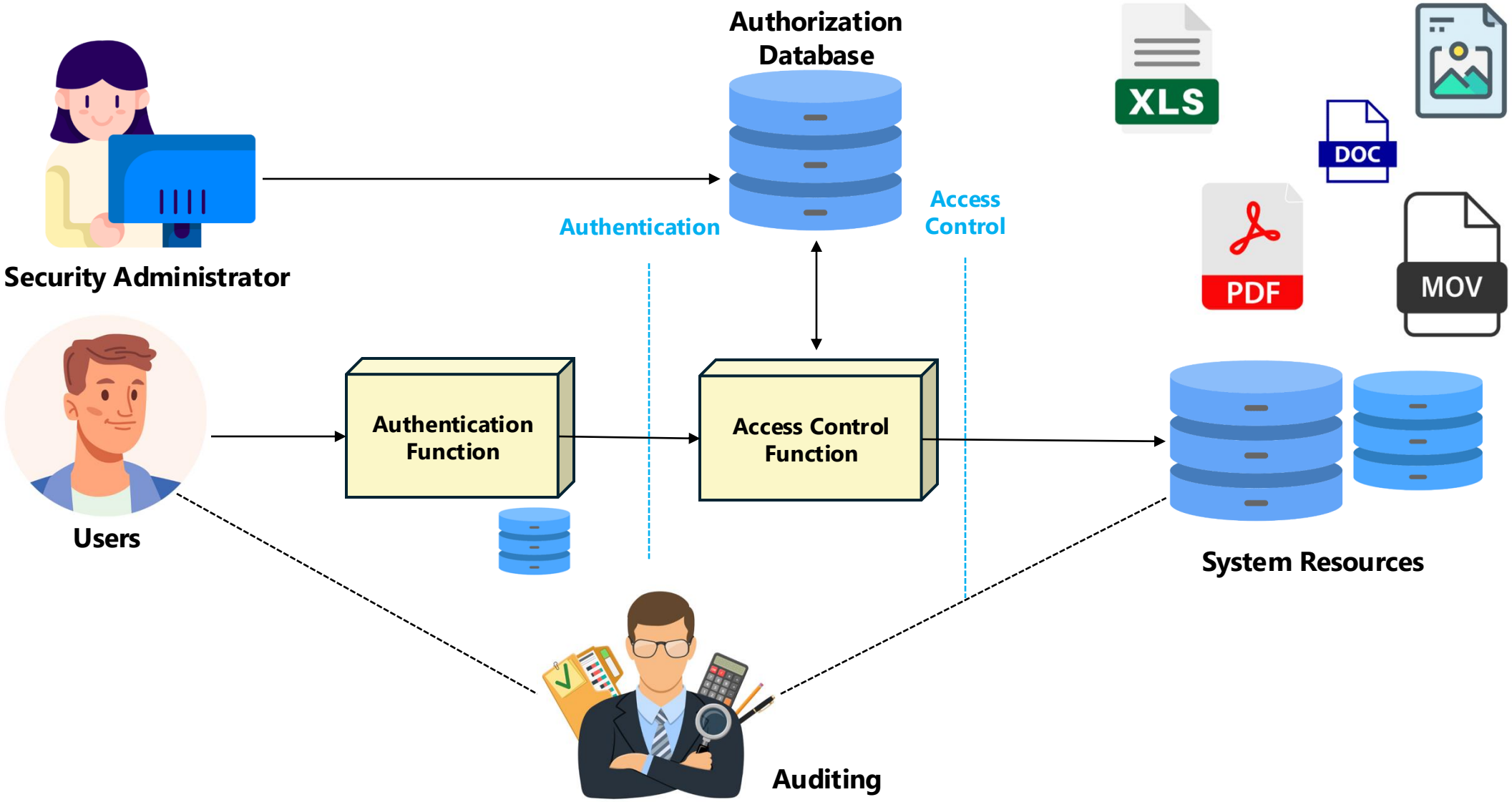
What is Access Control?

- **Social Network:** You can control who can access your personal information, e.g., you can access to all your personal profile, but your friends can access some parts of the profile.
- **Web Browsers:** Access only to a website (same origin policy).
- **Operating Systems:** One user cannot arbitrarily access/kill another user's files/processes.
- **Memory Protection:** Code in one region (e.g., Ring 3), cannot access the data in another more privileged region (e.g., Ring 0).
- **Firewalls:** If a packet matches with certain conditions, it will be dropped.

What is Access Control?



What is Access Control?



Authentication Procedure

- The most common authentication procedure is as follows:

1. **An individual arrives at a checkpoint**

   (login dialogues, doors, security gates, …)

2. **The individual claims an identity**

   (username, smart cards, tokens, ….)

3. **The individual present the item need to prove the identity**

   (password, PIN, biometric, passport, driving license, …)

"On the Internet, nobody knows you're a dog."

User Authentication Can Use

Something You Know
(Password, PIN, …)

Something You Have
(Keys, Badges, Tokens, Cards, …)

Something You Are
(Biometric, Retina Patterns, …)

Something You Do
(Handwriting, …)

Where You Are

Where you are?

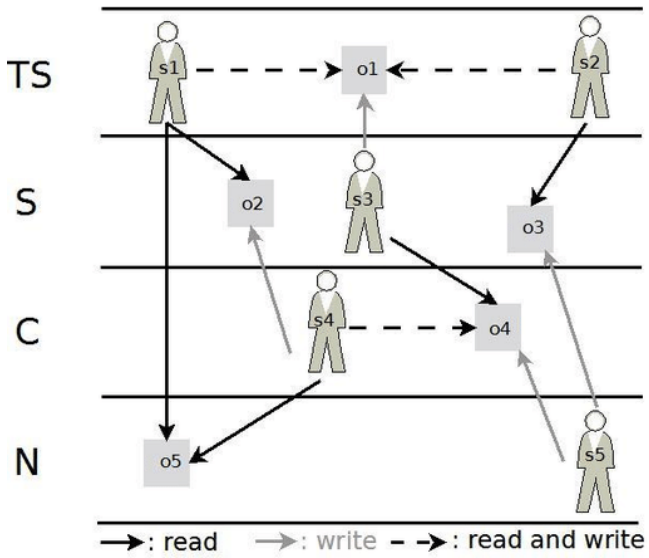
- ❑ does not verify identity, ***unless only one person can enter that location.***
- ❑ could ***reduce*** the number of possible identities.
- ❑ but should rather be thought of as ***access restriction***, than an authentication mechanism

Access Control Policy Models

- **Discretionary Access Control (DAC)**
  - **Definition:** *An individual user can set access control mechanisms to allow or deny access to an object he/she owns.*
  - Relies on the **object owner** to control access.
  - DAC is widely implemented in **operating systems**, such as Windows, UNIX, etc.
  - Its **flexibility** is the key strength of DAC, so users can dynamically control access to their own objects.
- **Mandatory Access Control (MAC)**
  - **Definition:** *A system-wide policy decrees who is allowed to have accesss and individual user cannot alter that access policy.*
  - Relies on the **system** to control access.
  - **Example:** The law allows a court to access driving records without the owners' permission.
  - Traditional MAC mechanisms have been tighly coupled to **a few security models**.
  - Recently, systems supporting flexible security models start to appear (e.g., SELinux, Trusted Solaris, TrustedBSD, etc.)

Access Control Policy Models

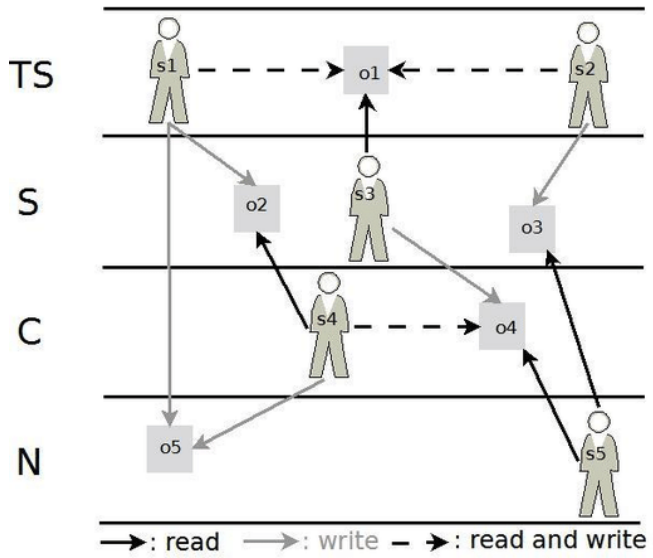
- **Mandatory Access Control (MAC)**



	o1	o2	o3	o4	o5
s1	read write	read			read
s2	read write		read		
s3	write			read	
s4		write		read write	read
s5			write	write	

Bell LaPadula Model

Confidentiality
Rule 1 (Simple Confidentiality): No Read Up
Rule 2 (Star Confidentiality): No Write Down
Rule 3 (Strong Star Confidentiality): No Read Up Write Down



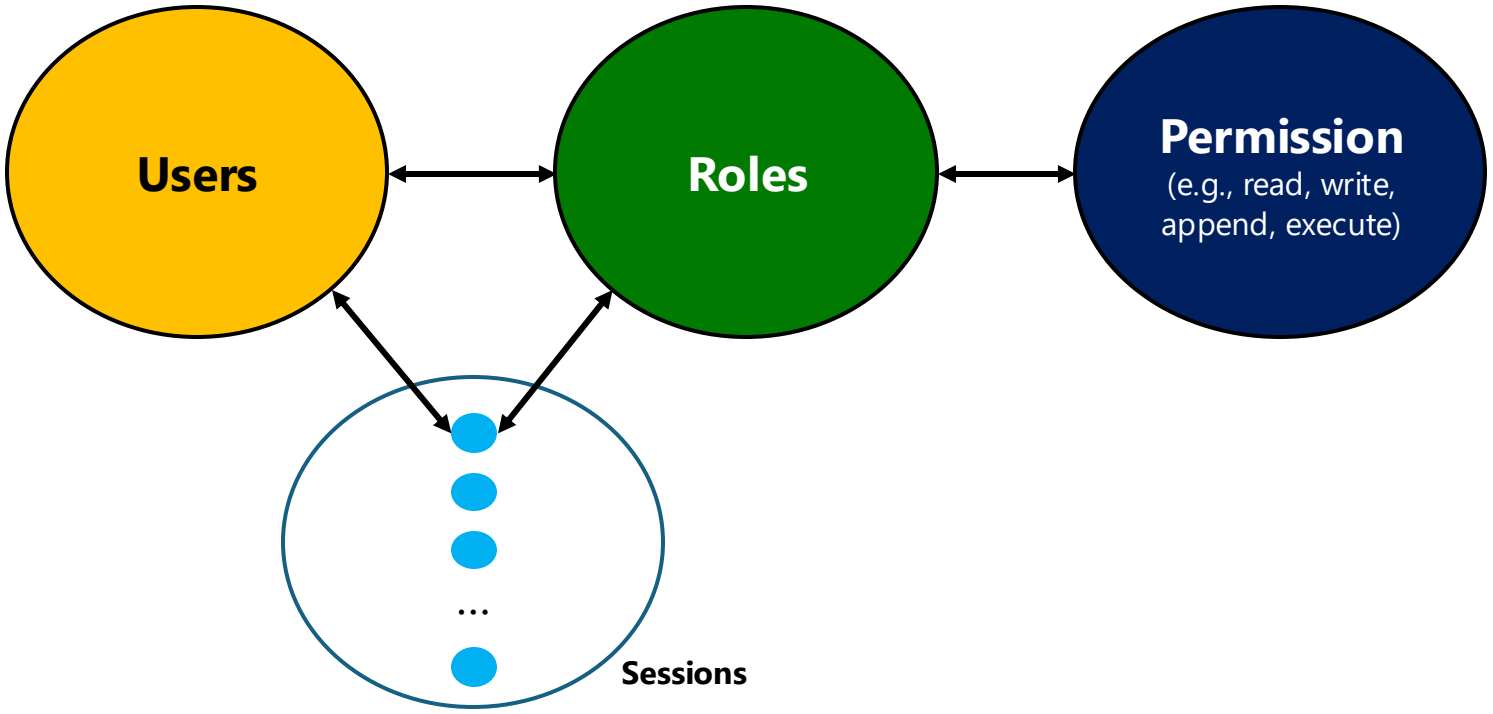
	o1	o2	o3	o4	o5
s1	read write	write			write
s2	read write		write		
s3	read			write	
s4		read		read write	write
s5			read	read	

Biba Model

Integrity
Rule 1 (Simple Integrity): No Write Up
Rule 2 (Star Integrity): No Read Down
Rule 3 (Strong Star Integrity): No Write Up Read Down

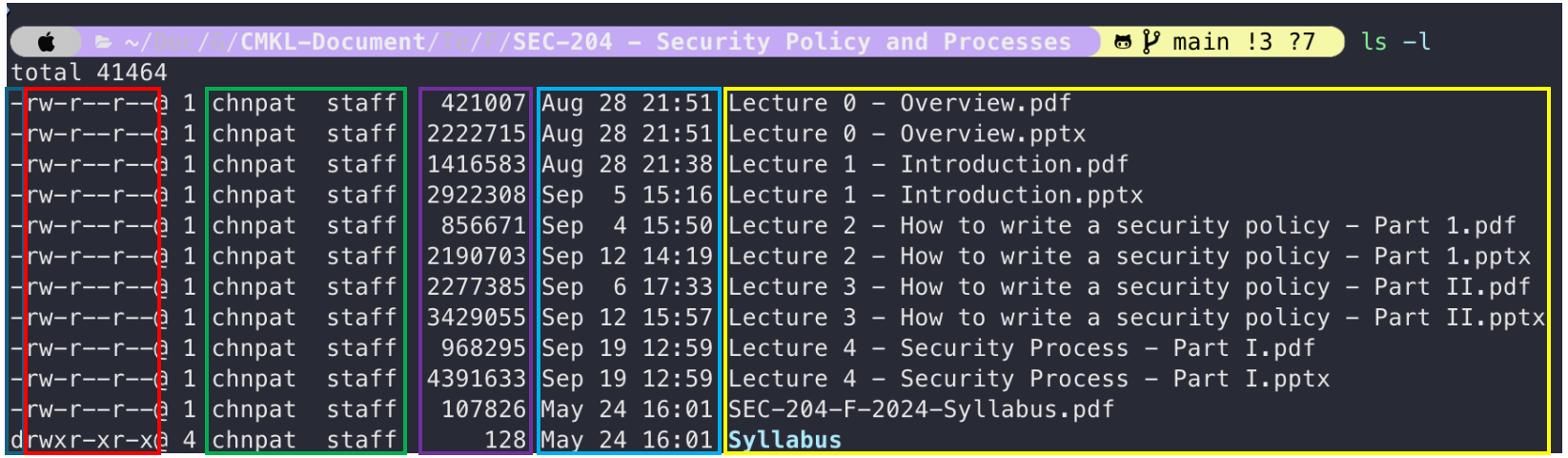
Access Control Policy Models

- **Role-Based Access Control (RBAC)**
  - RBAC can be configured to do MAC, i.e., roles of the entire system.
  - RBAC can be configured to do DAC, i.e., roles based on the identities.
  - RBAC changes the underlying **subject-object** model.
    - Access policies are relations between **roles**, **objects**, and **permissions**.
  - Subjects are now assigned to roles – **Role Assignments**
  - Roles can be structured in **a hierarchy**.



An Example of Access Control

- An example of real-world access control tools: **UNIX File Access Control**.



File Type
D = directory
- = file

Permission List
R = Read
W = Write
X = Execute
--- | --- | ---
Owner | Group | Others

Owner/Group

Last Edited Date and Time

Filename

File size in bytes

Introducing Data Privacy Preservation

- **Data privacy preservation** is a discipline focused on defining who has access to personal or sensitive data.

Policies to define data protection **strategies** and **processes**

- Does the organization put forward to protect personal data?
- Does the management level personnel recognize the important of personal data protection?
- Does the organization have clear processes with data privacy protection awareness included?

Users have **full control** over their data

- Do users know what personal data they provided to the organization?
- Do users know about the purpose of using or manipulating their data?
- Can users reject or refuse the use of personal data when they are no longer have business with the organization?

Tools & Techniques to preserve data privacy must be in place

- Do IT or cyber systems in the organization employ data protection tools or techniques properly?
- With the tools/techniques employed, does the organization still get insight from the data?

Data Privacy Preservation Techniques and Tools

- **Data Anonymization**
- Syntactic Anonymization
  - K-Anonymity, L-Diversity, T-Closeness
- Differential Privacy
- **Zero-Knowledge Proofs**
- Trusted Execution Environment (TEE)
- Secure Multiparty Computation (SMPC)
- Homomorphic Encryption
- Decentralization
- Federated Learning

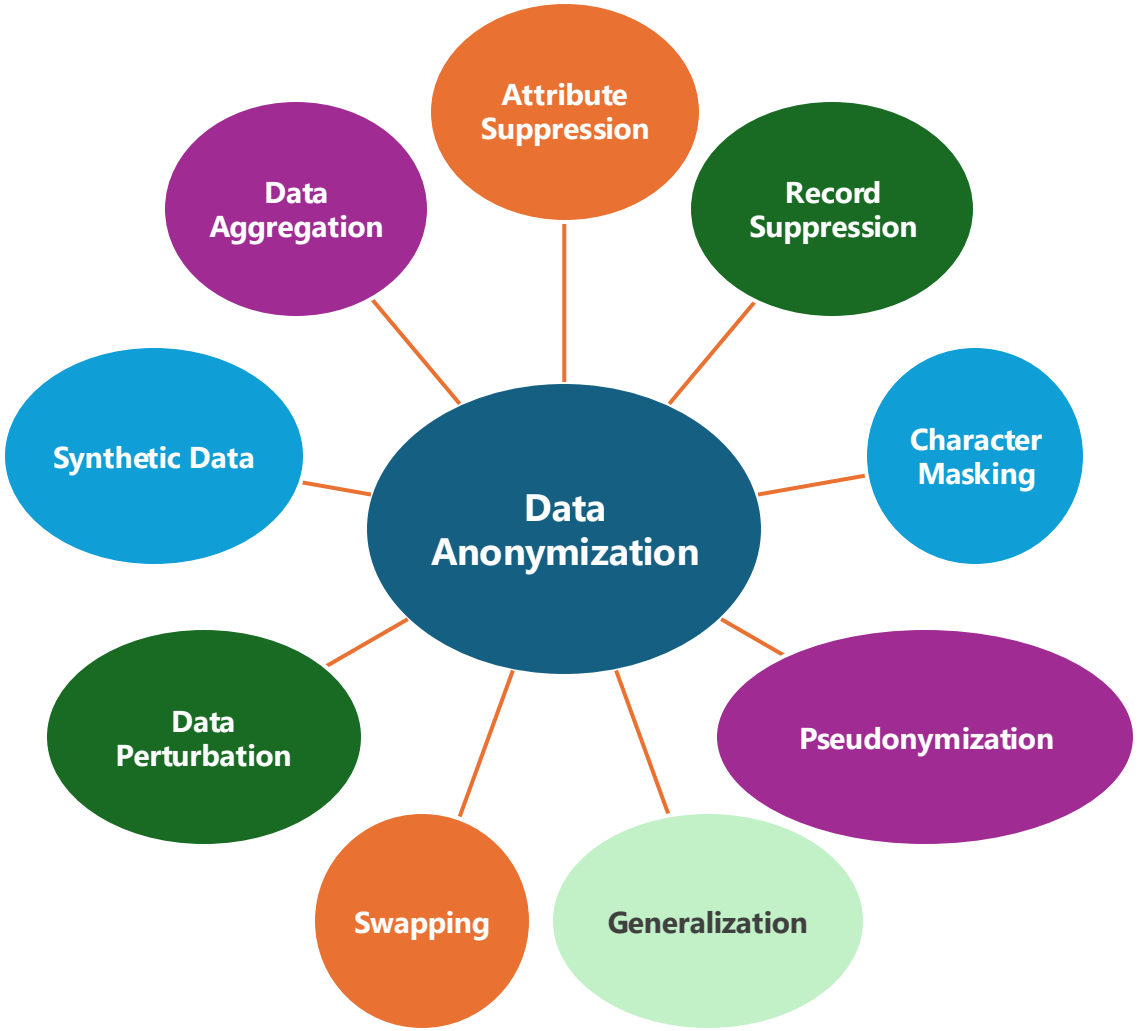
Data Anonymization

- **Data anonymization** is a privacy-preserving technique that transforms:

$$Personal\ Data \rightarrow Anonymized\ Data$$

- Anonymized data must be **irreversible**, meaning it should not be possible to reconstruct the original data.
- Anonymized data is expected to **remain useful**, preserving key characteristics such as:
  - Similar data distribution
  - Consistent frequency representation
  - Retention of essential data characteristics

Data Anonymization Techniques



1. Attribute Suppression

Student	Trainer	Test Score
John	Tina	87
Yong	Tina	56
Ming	Tina	92
Poh	Huang	83
Linnie	Huang	45
Jake	Huang	67

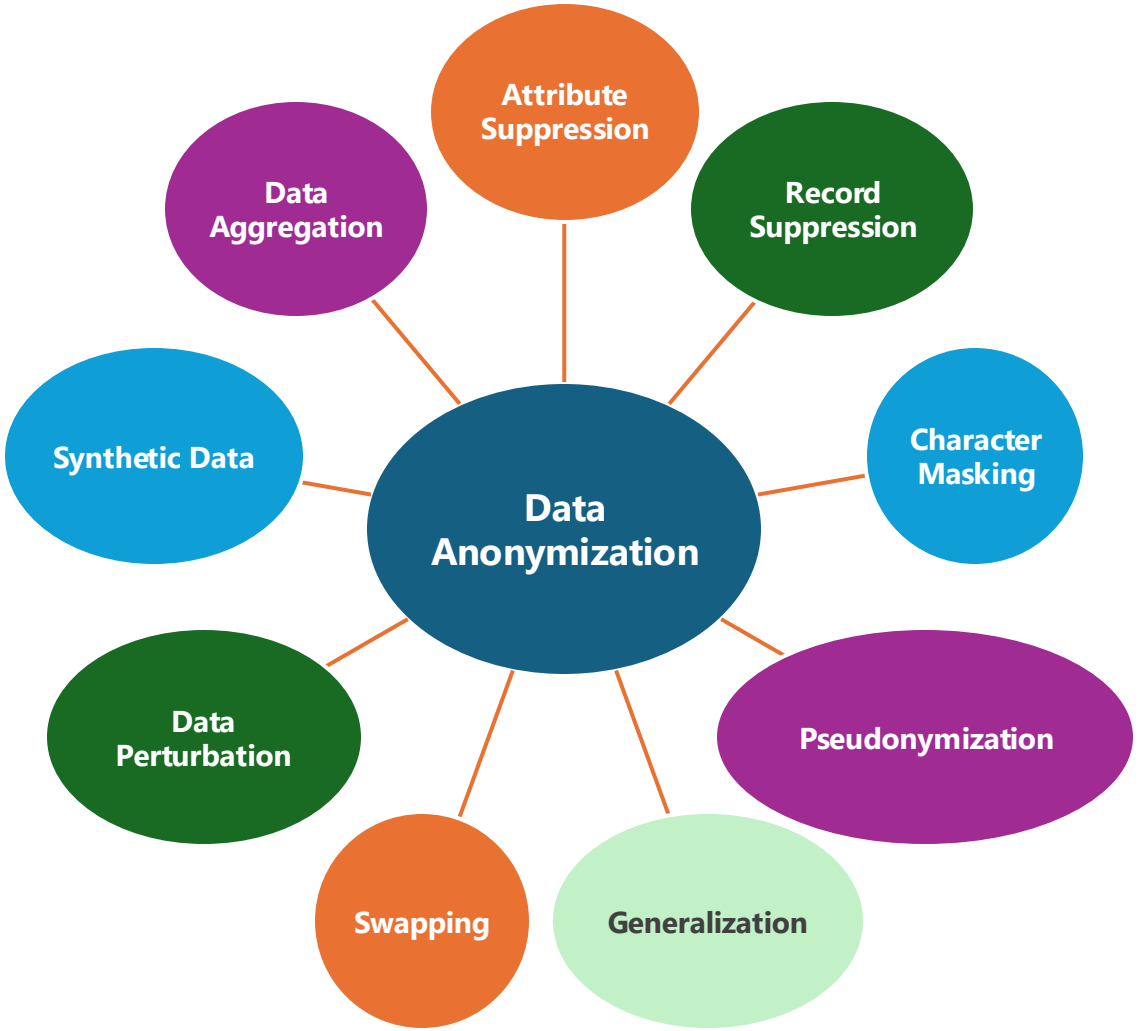
Trainer	Test Score
Tina	87
Tina	56
Tina	92
Huang	83
Huang	45
Huang	67

2. Record Suppression

Student	Trainer	Test Score
John	Tina	87
Yong	Tina	56
Ming	Tina	92
Poh	Huang	83

Student	Trainer	Test Score
John	Tina	87
Yong	Tina	56
Ming	Tina	92

Data Anonymization Techniques



3. Data Masking

Postal Code	Favorite Slot	Avg. # orders
100111	8 pm to 9 pm	2
200222	11 am to 12 noon	8
300333	2pm to 3pm	1

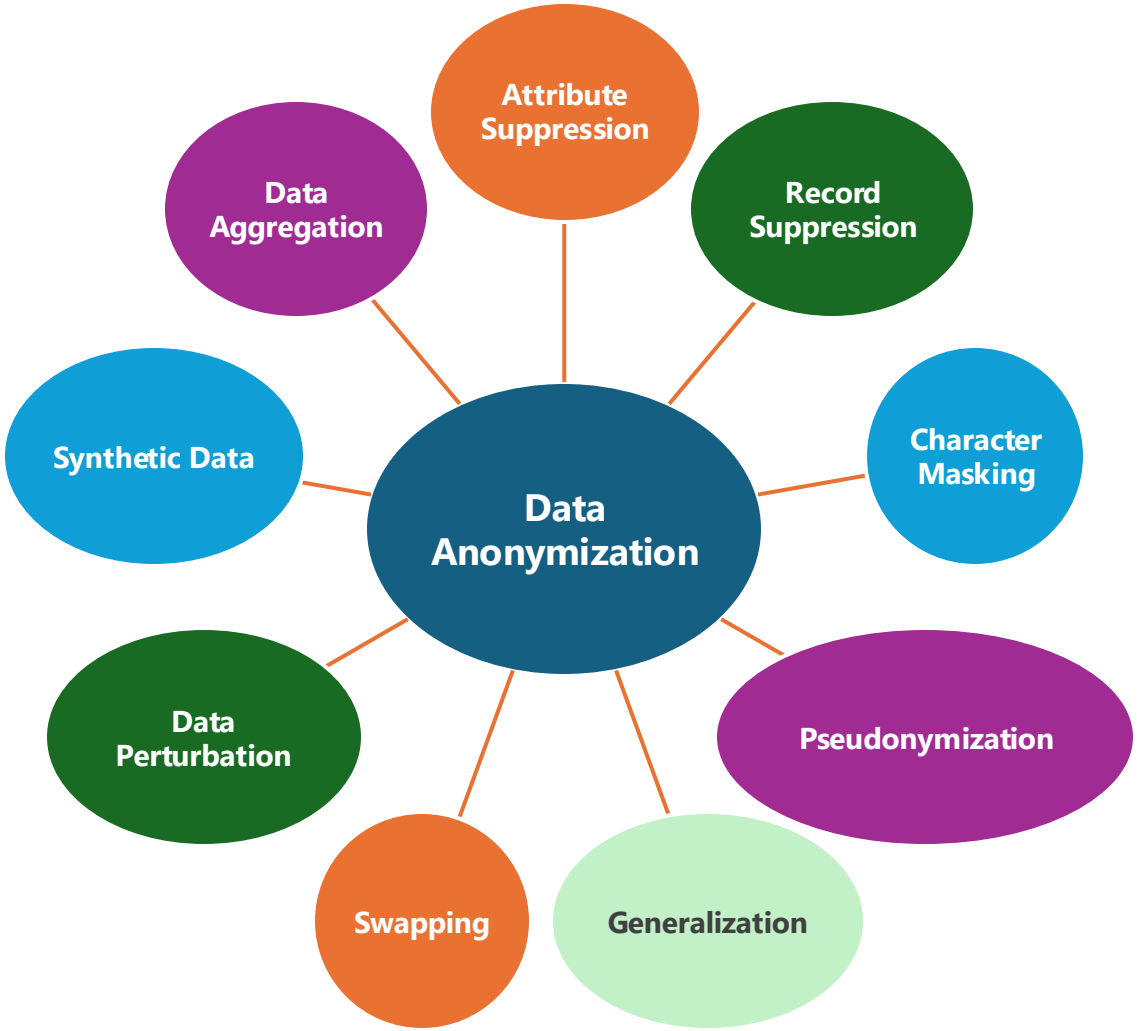
Postal Code	Favorite Slot	Avg. # orders
10xxxx	8 pm to 9 pm	2
20xxxx	11 am to 12 noon	8
30xxxx	2pm to 3pm	1

4. Pseudonymization

Person	Grade	Hours used
Joe Phang	A	20
Zack Lim	B	26
Eu Cheng San	C	30

Person	Grade	Hours used
416765	A	20
562396	B	26
964825	C	30

Data Anonymization Techniques



5. Generalization

S/N	Person	Age	Address
1	357703	24	1 Chalongkrung 1 Ladkrabang Bangkok
2	233121	44	25 Ramkhamhang 148 Saphan Sung Bangkok
3	888948	75	77 Bang Khunnon Bangkok Noi Bangkok

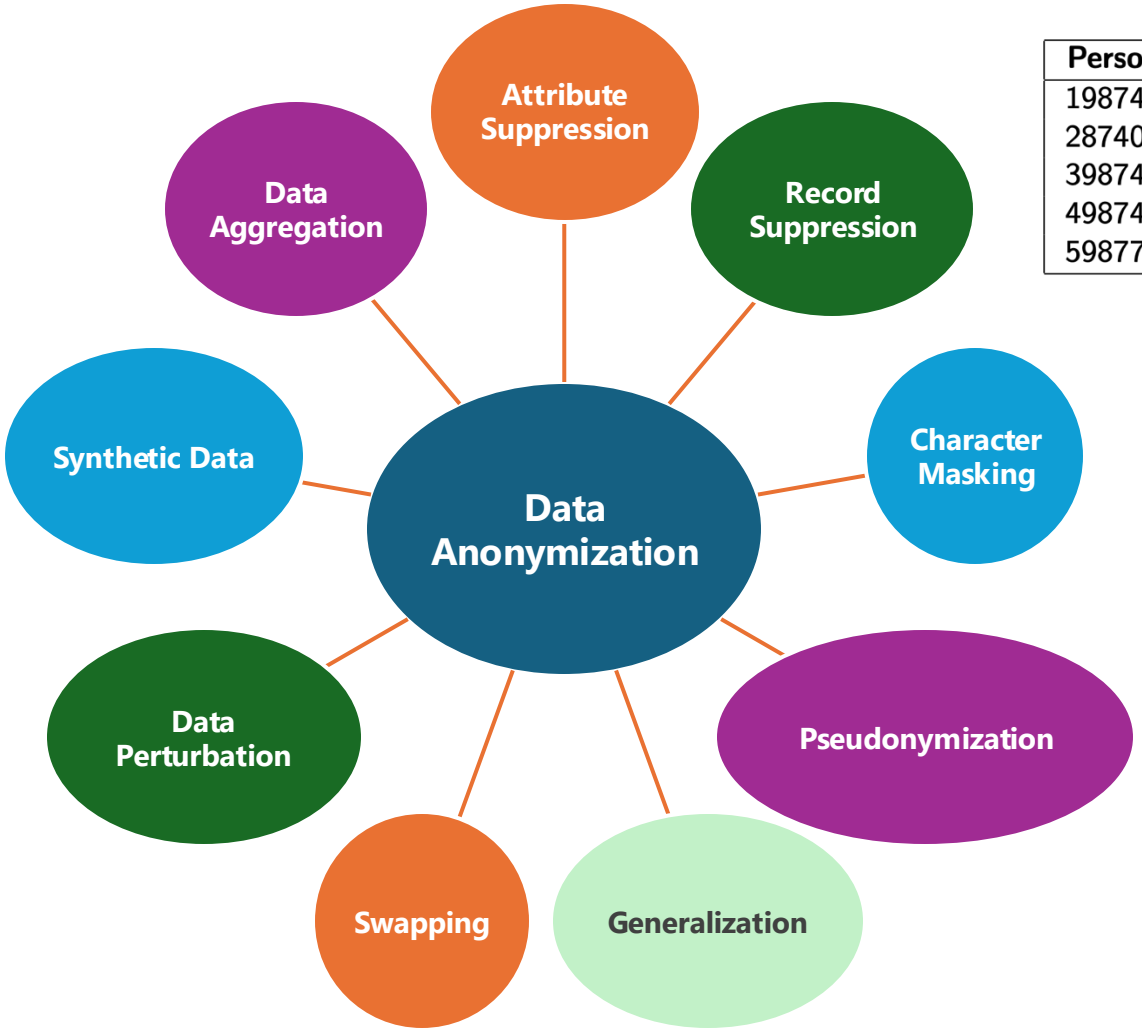
S/N	Person	Age	Address
1	357703	21-30	Ladkrabang Bangkok
2	233121	41-50	Saphan Sung Bangkok
3	888948	>60	Bangkok Noi Bangkok

6. Swapping

Person	Job Title	DoB	Type	#Visits
A	Dean	03-01-1970	Sliver	0
B	Salesman	05-02-1972	Platinum	5
C	Lawyer	07-03-1985	Gold	2

Person	Job Title	DoB	Type	#Visits
A	Dean	05-02-1972	Gold	5
B	Salesman	07-03-1985	Sliver	0
C	Lawyer	03-01-1970	Platinum	2

Data Anonymization Techniques



7. Data Perturbation

Person	Height	Weight	Age	Smoke?
198740	160	50	30	No
287402	177	70	36	No
398747	158	46	20	Ye
498742	173	75	22	N
598772	168	82	44	Ye

Person	Height	Weight	Age	Smoke?
198740	160	41	30	No
287402	175	69	36	No
398747	160	45	21	Yes
498742	175	75	21	No
598772	170	81	45	Yes

8. Data Aggregation

Donor	Monthly Income	Amount Donated
Donor A	$ 4,000	$ 210
Donor B	$ 4,900	$ 420
Donor C	$ 2,200	$ 150
Donor D	$ 4,200	$ 110
Donor E	$ 5,500	

Monthly Income	# of Donation	Donated
1,000–1,999	4	$ 1,470
2,000–2,999	5	$ 1,220
3,000–3,999	3	$ 290
4,000–4,999	5	$ 1,520
5,000–5,999	3	$ 870
Grand Total	**20**	**$5,370**

Data anonymization enables the preservation of data privacy without compromising its utility.

However, there are still situations where sharing and accessing actual personal data is necessary.

How Could We Share Information Without Knowledge?

- The key requirement for discussing personal data or information without revealing it is:
  - The data recipients have **no knowledge of the actual data**, yet they are able to verify certain insights.

- This is the foundational idea behind the invention of **Proof Systems**.

Foundations of Prover Systems

- We will have **a prover** and **a verifier**.
- A prover knows the information, such as a magic word, and doesn't want to disclose it.
- A verifier wants to ensure that the prover knows the information.

- **Goal:** The prover can convince the verifier that something is true without revealing anything that would allow the verifier to learn from it.

Introducing Zero-Knowledge Proofs

- At a high level, a Zero-Knowledge Proof (ZKP) works by allowing the verifier to challenge the prover with a series of tasks that can only be successfully completed **if the prover truly knows the underlying information**.
- If the prover is merely guessing, they will **eventually fail** the verifier's challenge with high probability.

- The **three fundamental properties** that define a Zero-Knowledge Proof are:
  - **Completeness:** If the statement is true, an honest verifier will be convinced by an honest prover that the prover knows the correct input.
  - **Soundness:** If the statement is false, no dishonest prover can convince an honest verifier that they know the correct input.
  - **Zero-Knowledge:** If the statement is true, the verifier learns nothing beyond the fact that the statement is indeed true.

Types of Zero-Knowledge Proofs

Interactive ZK Proofs

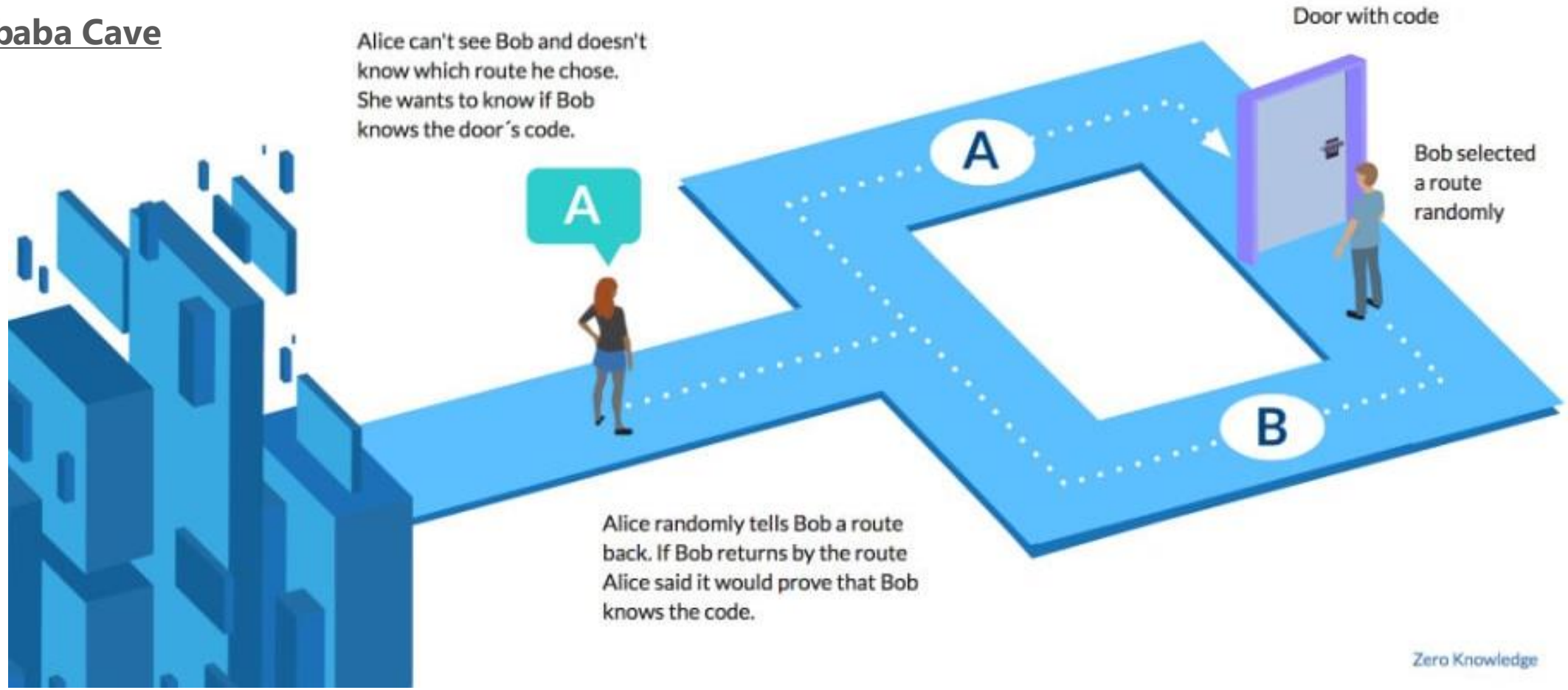
Multiple back-and-forth communications

Non-Interactive ZK Proofs

A single round of communication

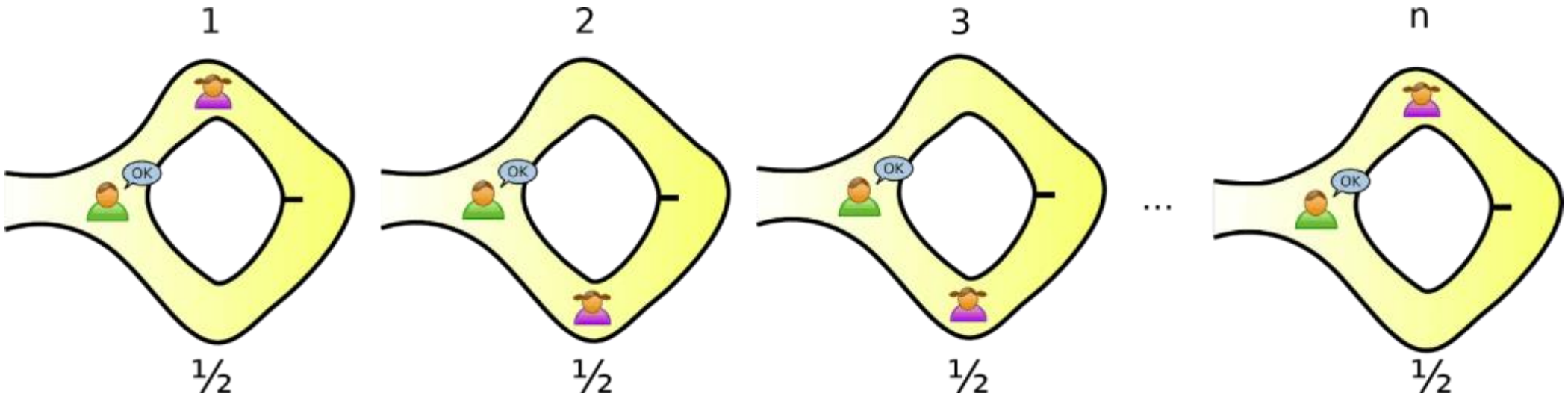
Examples of Interactive ZKP

- **Alibaba Cave**



Examples of Interactive ZKP

- Alibaba Cave



Alice will have only 1 in 1024 chances to guess it right after 10 trials.

Examples of Interactive ZKP

- **Pokemon Cards**

- You have **two Pokemon cards**: **One is a fake card**, while **another is a real card**. These two cards are **very similar** and they are very hard to distinguish between real and fake.
- Your friend bluffs on you that he/she knows which one is fake.
- How could you make your friend prove the claim without telling how to identify the real and fake ones.

Examples of Non-Interative ZKPs

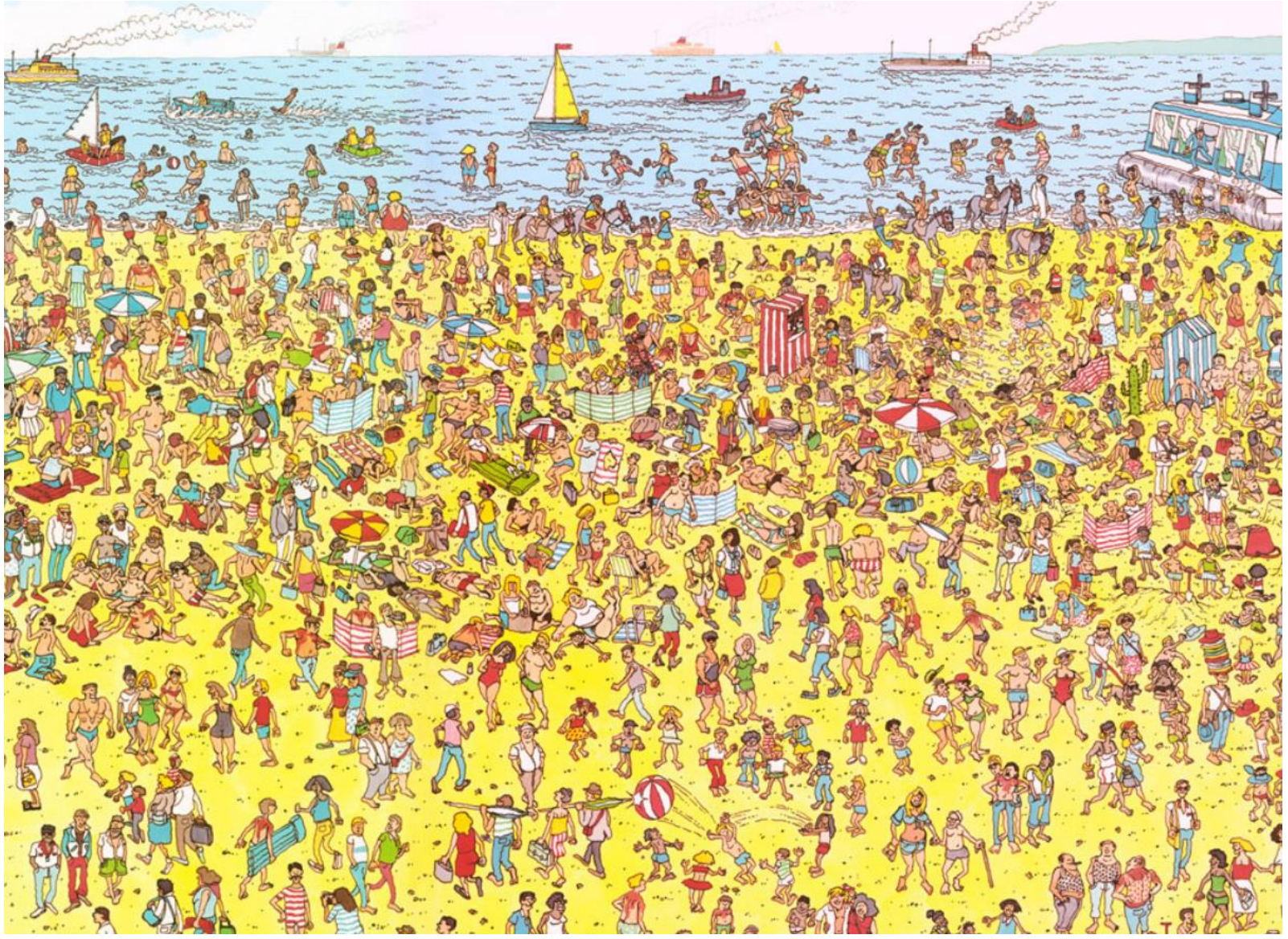
- **Club Safe Combination**

- There is a lock safe that only members of the activity club know the code combination to open it.
- Suppose you meet someone you don't know, yet he/she claims to also be a member of the club you are in.

Examples of Non-Interactive ZKPs

<- This is Wally.

- **Where is Wally?**



- How can we tell everyone in this room that you know the location where wally is **by not telling the actual location**?
- You need to just prove that you know where is Wally to everyone.

End of the Lecture

Please don’t hesitate to raise your hand and ask questions if you're curious about anything!

Key Takeaways of This Competency

- You have learned about security, privacy, and integrity threats in modern information systems.
- You have also learned how to implement various techniques and tools to protect data security, data integrity, and data privacy.
- Throughout the course, you have developed greater awareness of these aspects and should now be able to design and implement secure information systems.
- Additionally, you now know where to explore further from the topics discussed during the lectures.

Don’t forget that we will have a presentation for your assessment group project next week!!