PRIVACY ENGINEERING

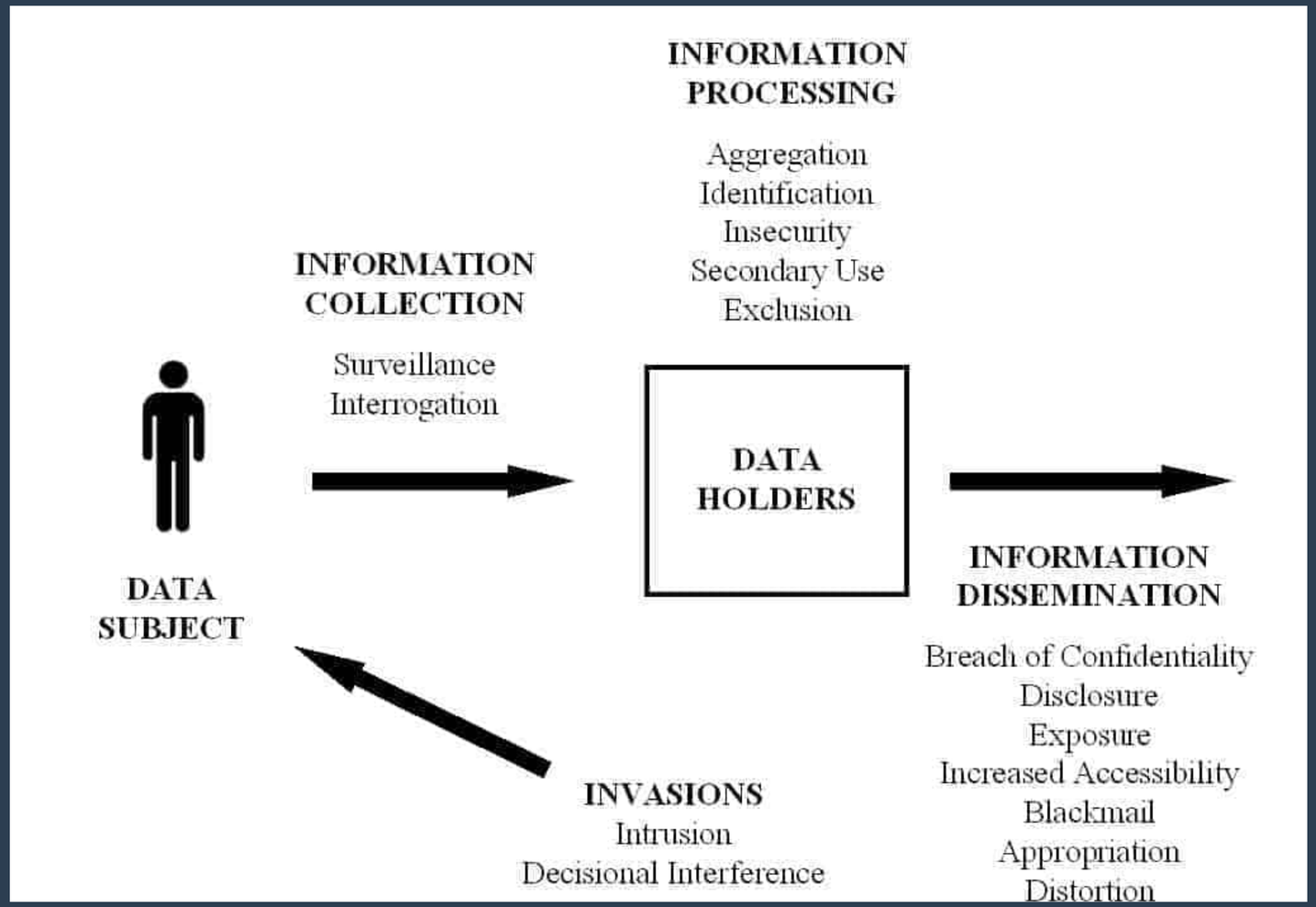
2024-01-24

JENNI REUBEN

PRIVACY

- menti.com Access code: 3645 6077

- Learn a structured way to think about privacy violations and how to go about solving the privacy problem that concerns your use case in hand.



A taxonomy of Privacy [Daniel06]

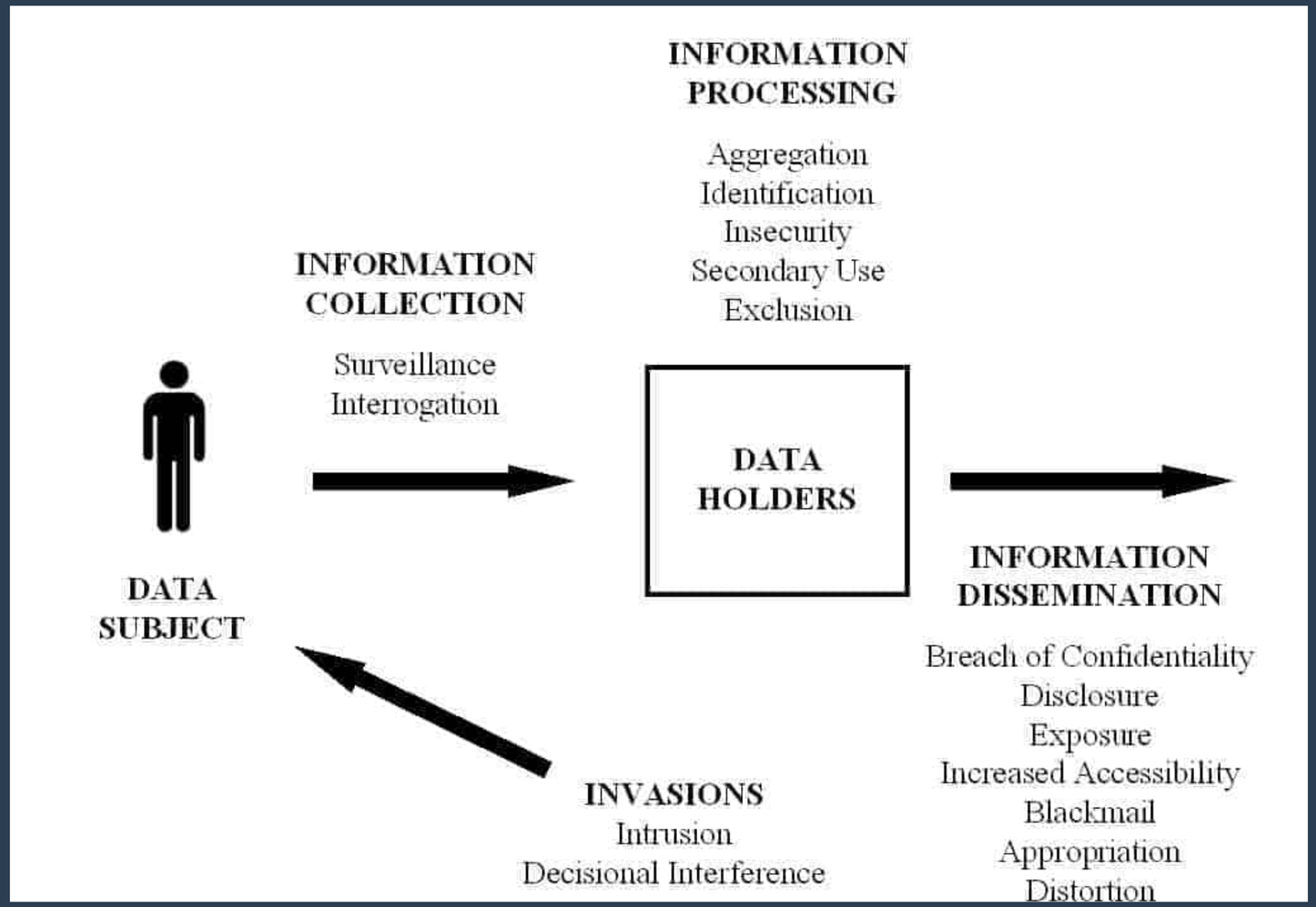
PRIVACY SOLUTIONS

1. With the advancements in computational systems, communications and storage systems, the line between our physical self and digital has never been this blurry before.

2. The amount of data we produce by our actions is rapidly increasing. Individual human actions generate a vast amount of data on a daily basis.

3. With the advancements in information and communication technologies come the increase in tools and applications for data collection, processing and sense making from data (big data analysis) that enable surveillance practices of governments, profiling and tracking by powerful organizations for cashing in our data, and worse lack of accountability and transparency in the collection and usage purposes.

PRIVACY



A taxonomy of Privacy [Daniel06]

WHY PRIVACY ENHANCING TECHNOLOGIES

1. The ideal situation for an end-user is to have enough guarantee about the design of the systems such that it is not necessary for them to trust any third party. But in reality some trust has to placed in other stakeholders, at the best technology can do is to minimize the number of stakeholders that have to be trusted or at least provide ways to challenge or check that the trust is not misused.

2. Privacy Enhancing Technologies (PETs) are technologies that reduce/minimize the perimeter of trust or at least provide accountability and transparency measures to challenge that trust;

3. According to the trust structure in the information and technological solutions there are two categories of PETs

   - Hard Privacy
   - Soft Privacy

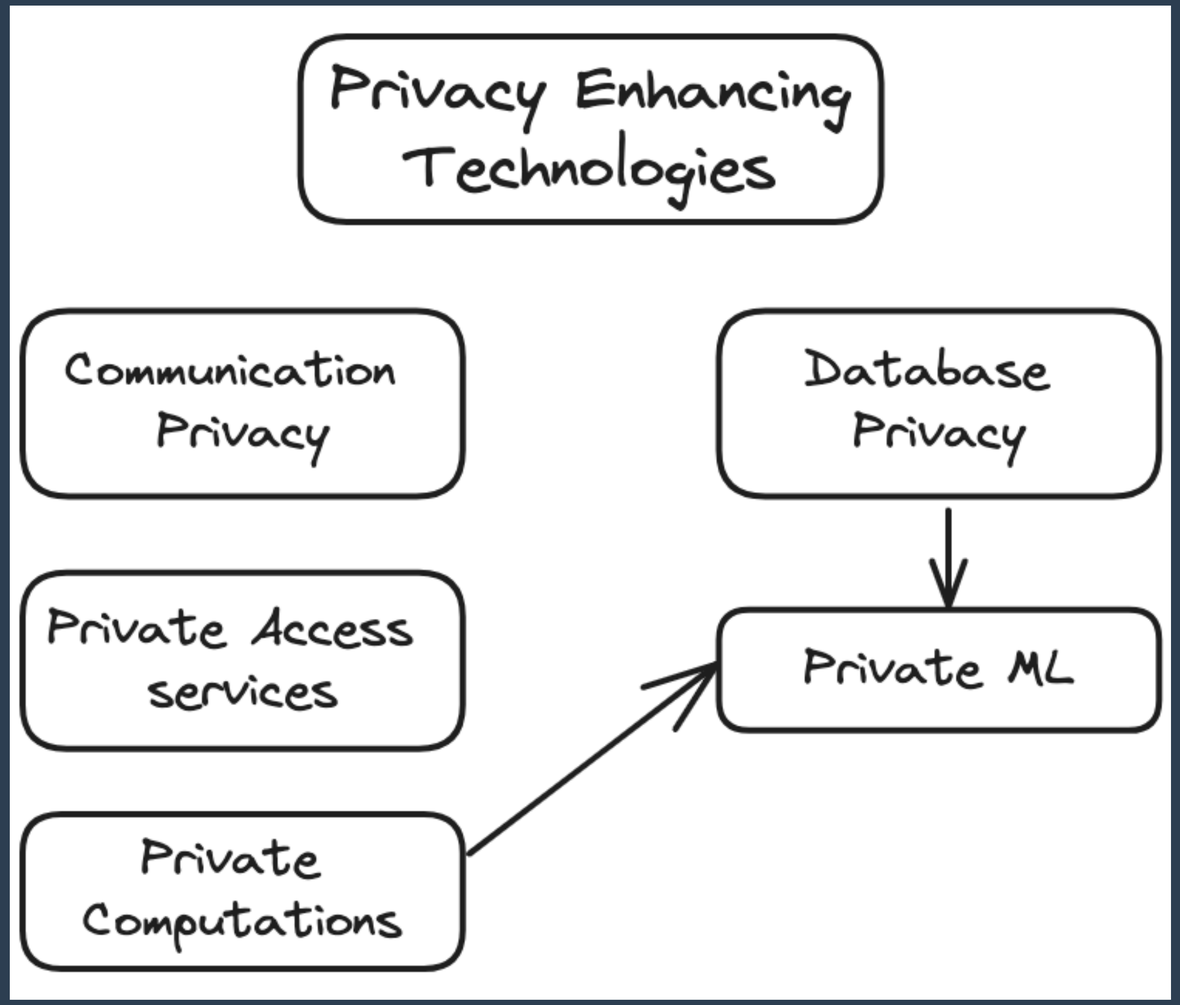
PRIVACY ENHANCING TECHNOLOGIES

1. Hard Privacy

- Technologies that avoid or reduce as much as possible the disclosure of personal data, (i.e enabling ways to avoid placing trust in any stakeholders of a service or system)

2. Soft Privacy

- Technologies for enforcing the rights of the end-users after their personal data is disclosed or processed. The trust structure is based on the assumption that the end-users will eventually lose control over their data when they pursue their need to communicate, access, etc.,and therefore have to place their trust on the service provider. The focus of the technologies is to provide accountability in order to challenge that trust.



Sign Post to Day I and Day II topics

PRIVACY IN COMMUNICATIONS SERVICES

1. The basic expectation for a user in a communication solutions is that her communications are not eavesdropped, not tampered during transmission. This is achieved by means of establishing secure channels and authentication measures in the case of client-server communication scenarios.

2. Public key cryptography technologies helps to establish a secure channel with a service or a server without sharing any prior secrets with it e.g., TLS and SSH protocols. In the case of user to user communications such as VoIP, instant messaging, sending/receiving emails, encryption of the communication is important and it is provided by protocols such as Signal.

3. However, these technologies requires the user to trust a CA for example or a third party. These techniques protect the content of the communications and they don't guarantee me that my identity (ip, role, time) cannot be inferred from my communication patterns e.g., by a passive adversary that observe the network

PRIVACY IN COMMUNICATIONS SERVICES

1. Observing someone persistently browsing for information relating to cancer may reveal their health status, inferring the fact a journalist is talking to a humanitarian agency puts her or someone's life in danger in some authoritarian regimes.

2. Technologies that provide guarantees against such disclosure of communications patterns (meta data) are called anonymous communications.

3. The idea is to introduce intermediate nodes in the network to resend/re-route the messages in a way that it is difficult for a passive or in some cases local active adversaries to infer the relationship between a sender and the receiver (the communication between the sender and receiver is untraceable). 3rd party anonymity is the basic form of protection that is common among the anonymous communication systems.

ANONYMITY

- Opposite of identifiability
- To enable anonymity of the subject, there **always** has to be a **set** of "similar" subjects

> "Anonymity of a subject from a adversary's perspective means
> that the adversary cannot sufficiently identify the subject
> within a set of subjects, the anonymity set.” [Pitzmann17]

- In the case of subjects that are actors, the anonymity set consists of subjects who might have caused an action such as the senders of an email or initiator of a network session.
- In the case of subjects that are actees, the anonymity set consists of subjects who might be acted upon such as the recipients of an email.

ASSUMPTIONS

- We assume that the adversary uses all the information available to him to infer (probabilities of) his/her's items of interest (IOIs),
  - e.g., IOI = who did send or receive which messages.
- The adversary does not forget, the knowledge he/she possess do not decrease, this is good for when you are quantifying the anonymity, or unlinkability properties of your system that you are designing
- we assume that the adversary is not able to get information on the sender or recipient from the message content
  - e.g., in the case when a malicious actor is the receiver of a message m from m's content he/she cannot infer the sender of the message.

TYPES OF ANONYMITY GUARANTEES IN COMMUNICATIONS SERVICES

- 3rd party anonymity – When two known parties communicate with each other, we say that the anonymous communication system provides 3rd party anonymity property if a network adversary (global or partial) cannot infer from my communication pattern whom I am communicating with (sender–receiver relationship).
- Sender anonymity – The recipient receives the message addressed to him/her from a sender but not the sender's identifying information, good to have property when accessing a service but cannot be useful property to have especially when we expect a response, then we need receiver anonymity as well.
- Receiver anonymity – We expect the communication system to guarantee receiver anonymity to enable others in the network to contact me without knowing my identity.

PSEUDONYMITY IN COMMUNICATIONS SYSTEMS

> "A pseudonym is an identifier of a subject other than one of the subject's real names" [Pfitzmann17]

- A subject is pseudonymous if a pseudonym is used as identifier instead of one of its real names [Pfitzmann17] for the purpose of communicating with others or a server.
- It is common to assume that a person's pseudonym won't change over time.
- A group pseudonym refers to a set of holders, i.e., it may refer to multiple holders thus inducing an anonymity set. A transferable pseudonym can be transferred from one holder to another. [Pfitzmann17]

UNLINKABILITY

- What more can we expect when we communicate with a server or another user. Unlinkability!
- Opposite of linkability and w.r.t items of interest

> “Unlinkability of two or more items of interests (IOI)s, (e.g., senders, receivers and actions, messages …) from an adversary’s perspective means that within the system, the adversary cannot sufficiently distinguish whether these IOIs are related or not.” [Pitzmann17]

- Example: Imagine a scenario, where there are 2 senders and 2 messages, the adversary cannot with sufficient probability link the messages to its respective senders is an example of unlinkability property provided by a system to its subjects.

ANONYMITY IN TERMS OF UNLINKABILITY

"A sender s sends a message m anonymously, iff s is anonymous within the set of potential set of senders of m, the sender anonymity set of m" [Pitzmann17]

"A message m is said to be sent anonymously, iff m can have been sent by each potential sender within the sender anonymity set of M" [Pitzmann17]

- Then, anonymity in terms of unlinkability is defined as:

"Anonymity of a subject w.r.t an attribute then maybe defined as unlinkability of this subject and this attribute" [Pitzmann17]

PSEUDONYMITY IN TERMS OF UNLINKABILITY

- With respect to the degree of linkability, various kinds of pseudonyms may be desirable depending on the context for their usage,

1. Person pseudonym: A person pseudonym is a substitute for the holder's name which is equivalent to the holder's civil identity, e.g., personnummer, mobile phone number

2. Role pseudonym: The use of role pseudonyms is limited to specific roles, e.g., a pseudonym such as "customers".

3. Transaction pseudonym: for each transaction, a transaction pseudonym is unlinkable to any other transaction pseudonyms, e.g., randomly generated transaction numbers for online-banking. Therefore, transaction pseudonyms can be used to realize strong anonymity in communications and authentication systems. [Pfitzmann17]

UNDETECTABILITY

- With contrast to anonymity and unlinkability, where the IOIs per say are not protected but rather the relationship between the attributes related to IOIs or relationship between two or more IOIs. Undetectability is protecting the meta-data as such not just links to the subjects

> "Undetectability of an IOIs from an adversary's perspective means that the attacker cannot sufficiently distinguish whether or not the IOI exists" [Pitzmann17]

- Take for example, if we take messages as IOIs, messages are not sufficiently distinguishable from random noise.

- Undetectability of an IOI is not possible if the adversary is involved in the IOI, so the guarantee only applies to IOIs where the adversary is not involved in the IOI.

UNOBSERVABILITY

- We expect from a communication system an assurance of indistinguishability of messages against adversaries not involved in the IOIs. However, if an adversary is involved in an IOI then we expect anonymity of other involved subjects. We say then then the system provides unobservability of the subjects and the IOIs.

- Unobservability of an item of interest means,

  Undetectability of the IOI against all the subjects not involved in it, and

  Anonymity of the subjects involved in the IOI even against the other subjects involved in that IOI [Pfitzmann17]

RELATIONSHIPS BETWEEN THE PROPERTIES

- W.r.t the same adversary, unobservability reveals always only a subset of information that anonymity reveals. Then we can say

$$\text{Unobservability} \Longrightarrow \text{Anonymity}$$

$$\text{Sender unobservability} \Longrightarrow \text{Sender anonymity}$$

$$\text{Receiver unobservability} \Longrightarrow \text{Receiver anonymity}$$

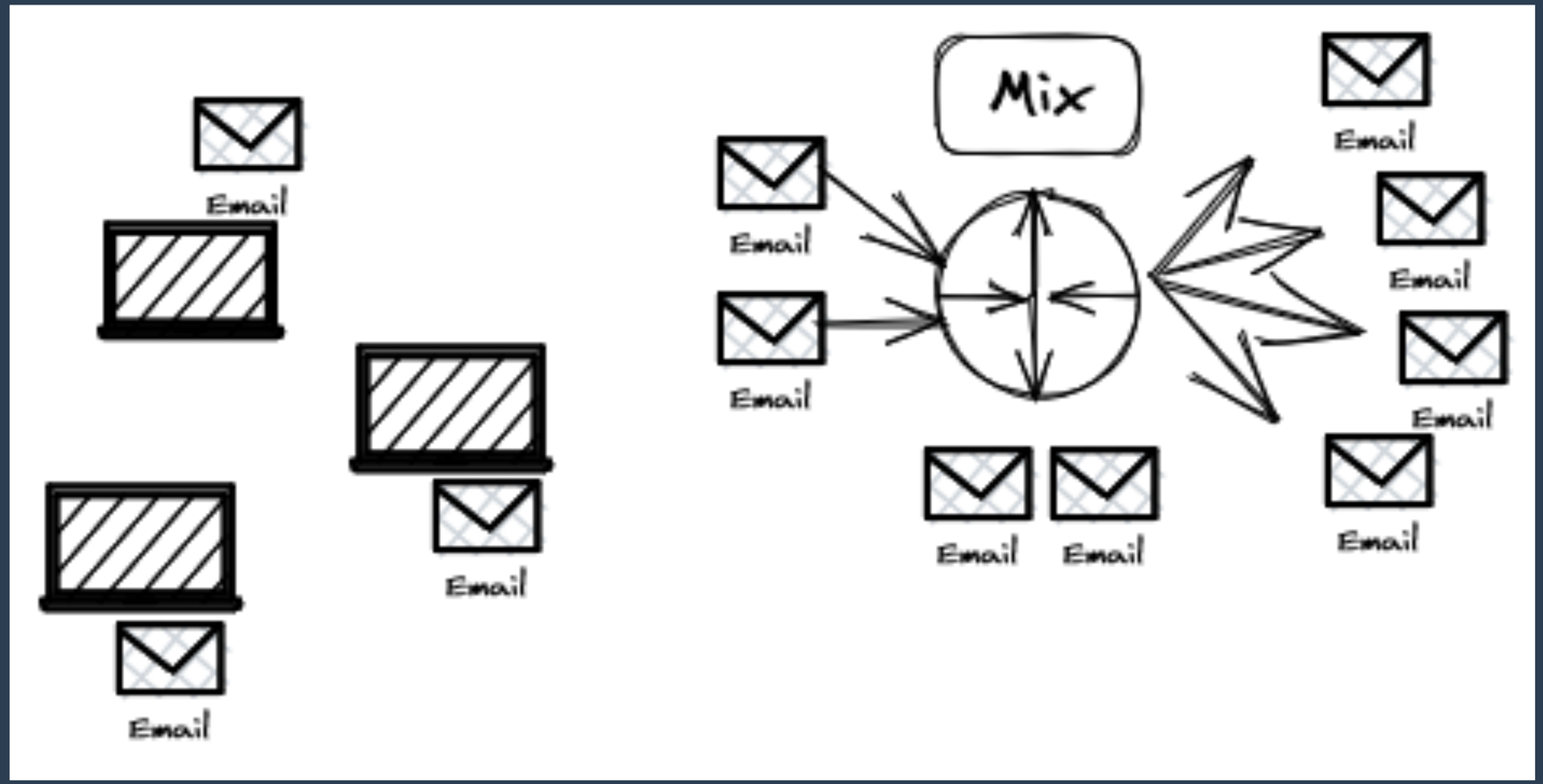
- W.r.t the same adversary, unobservability reveals always only a subset of information that undetectability reveals.

$$\text{Unobservability} \Longrightarrow \text{Undetectability}$$

[Pfitzmann17]

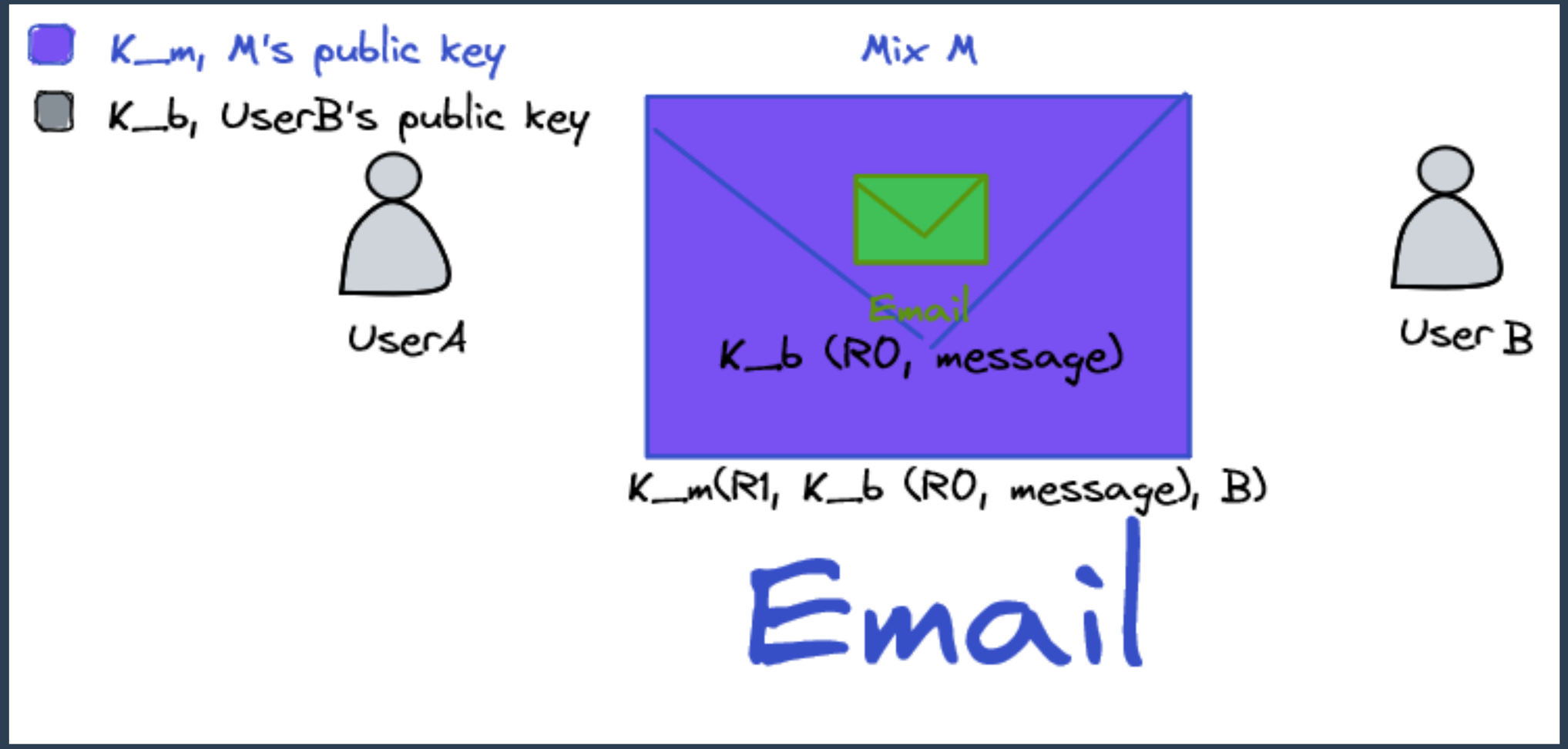
MIXNET

- David Chaum in his seminal paper "untraceable electronic mail, return addresses and pseudonyms" introduce mixnet,
- Mixnet a anonymous communication system for exchanging emails among the users of the system that ensures sender, receiver and 3rd party anonymity.
- The messages in the mix network are bounced around a bunch of mixes before arriving to its destination thus making it hard to trace the sender and receiver relationship.



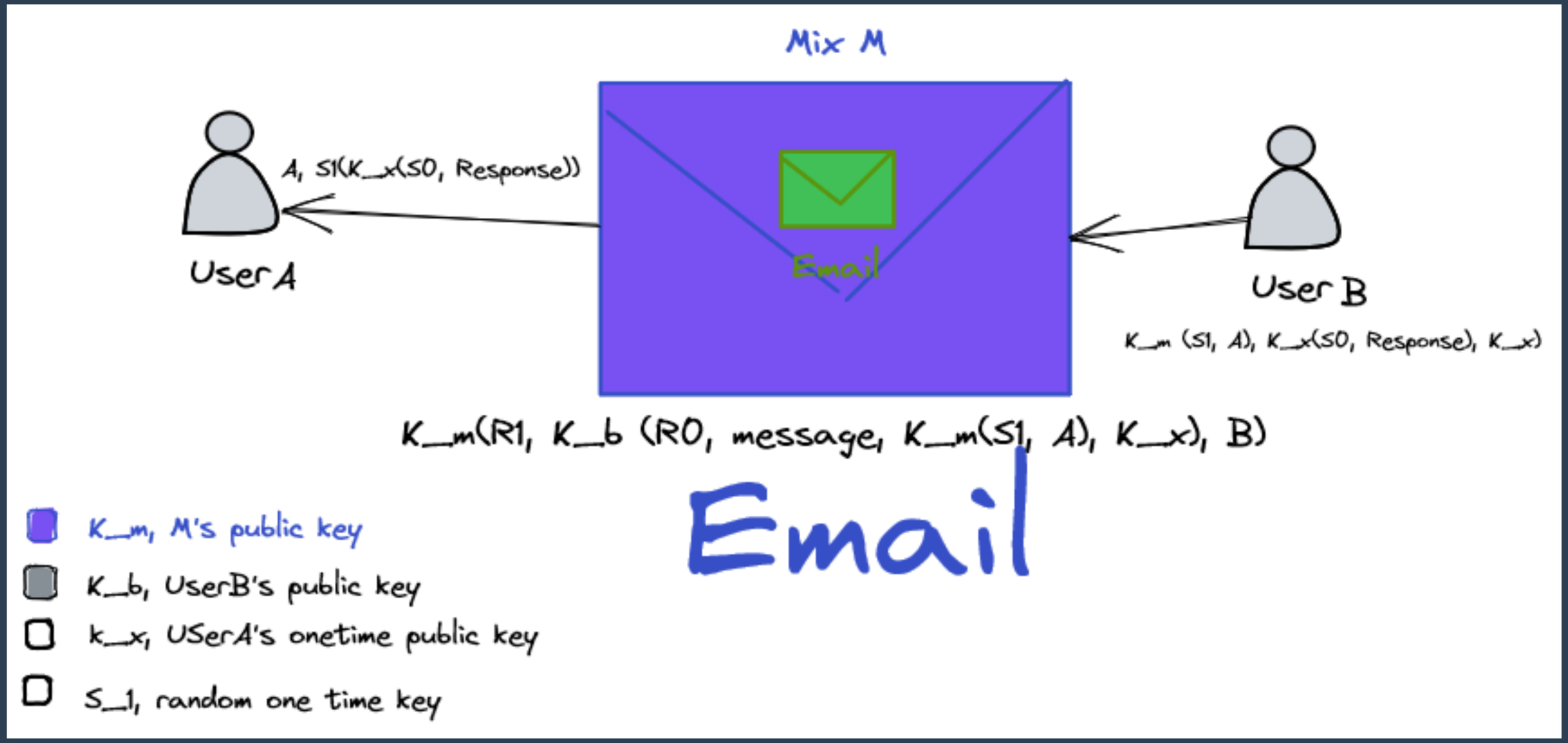
Mixnet

MIXNET - HOW IT WORKS



A sending message to B using mix M

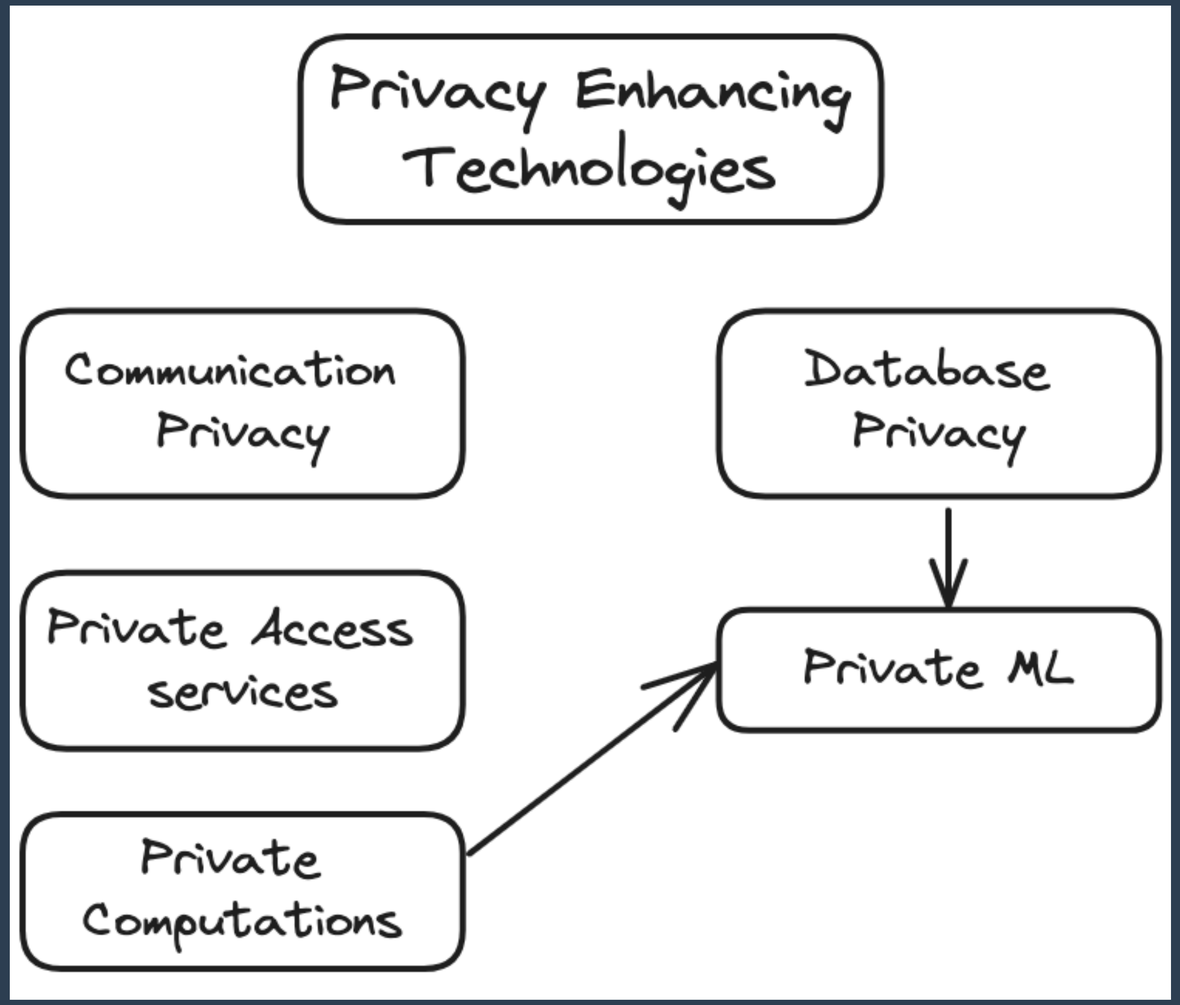
MIXNET - HOW IT WORKS



B sending response to B

MIXNET

- Limitations
- Because all clients uses the same route, the latency of the network is high, further messages are queued thus delayed at very component mix for shuffling
- Smaller anonymity set, the shuffling strategy at the component mix is threshold mixing, the mixes queues the messages for shuffling until a threshold number of messages are queued. So there are 25% chance of an adversary to link the incoming packet to the outgoing packet.
- Vulnerable to active attacks such adversaries inject flow with unique signature and identifying this packets.



Sign Post to Day I and Day II topics

PRIVACY IN ACCESS SERVICES

- Users of a networked information system is linked to devices or entities that they have access in order to retrieve confidential information and execute privileged actions. Once this link is established that the user and the other resources know each other identities and can be subjected to access control policies. Several standards and frameworks exists for identity management such as federated identity management systems.

- Federated identity managements, in such systems apart from the user and the service providers there exists an identity provider. The users register with a identity provider that he/she wants to use a service with certain identity. Then the identity provider authenticates the user using this identity for using a third party service A.

- Example of federate identity management is single-sign-on (SSO) services provided by major service providers such as google, Apple, Microsoft with varying "privacy" guarantees. Shibboleth identity management system on the other hand allows users to provide only a subset of their attributes to a service provider after a successful authentication. Shibboleth is an identity management system used by a network of major United States educational institutions.

PRIVACY IN ACCESS SERVICES

1. In Shibboleth identity management system, the user can decide whether or not to reveal even his/her user-id to a service provider depending on whether or not the service provider absolutely needs her user-id for providing the service

2. However, the privacy protection offered by identity management systems are only robust against observations by third party services.

3. The privacy guarantee for the user is that the identity provider is trustworthy hence does not misuse the user information for other purposes. However, users cannot blindly trust the identity provider.

4. The identity providers can observe every authentication session that a user par-takes in and the services that they are using in order to construct a profile. Worse, the identity provider and the service provider can collide, which will lead to complete disclosure of user activities or information to each other.

ATTRIBUTE BASED CREDENTIALS

1. Different from the federate identity management, central to Attribute based credentials systems is the concept called attributes. Examples of attributes; name, age, hair colour, eye colour, date of birth, grades, diplomas, etc.

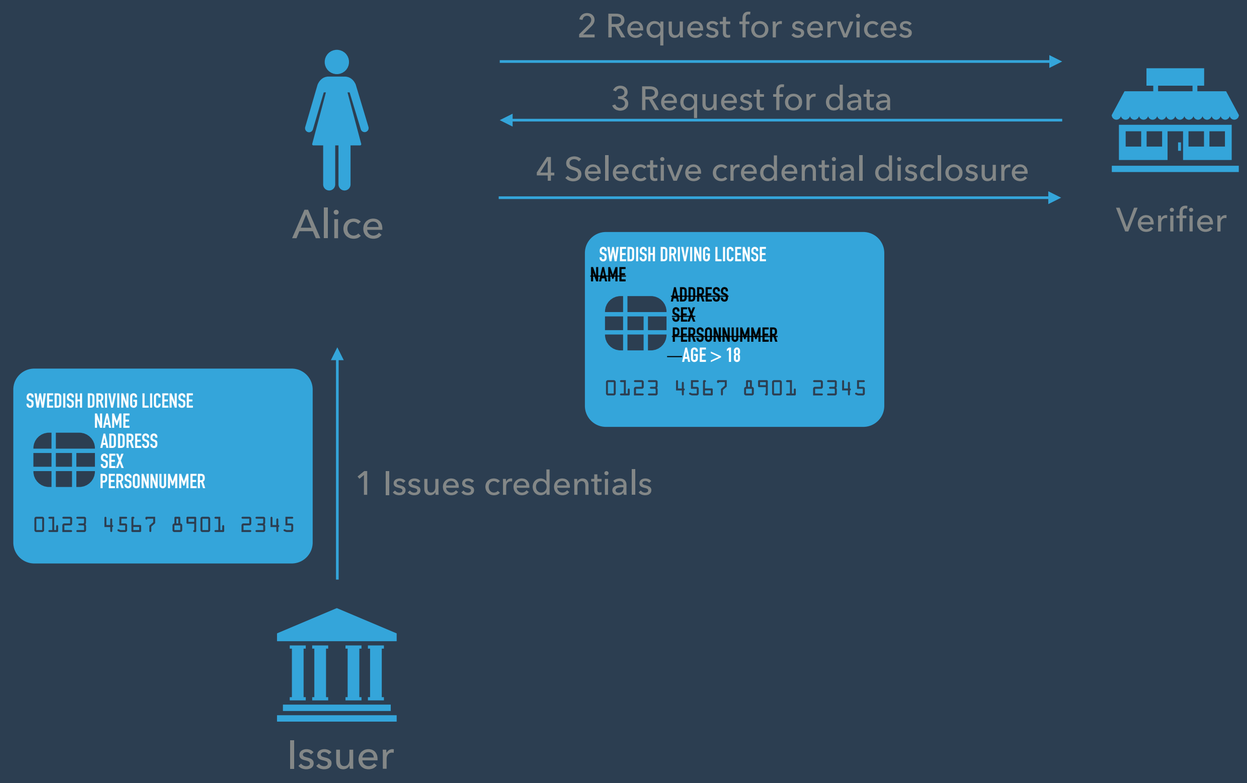
2. Attributed based credentials enables a user (the prover) to securely and privately prove ownership of an attribute or a set of attributes to a service provider (the verifier) without revealing the attributes. The attributes are stored in a secure capsule or magnetic chip called **credentials** very much like certificates.

3. The credentials are issued by a trusted credential issuer who provides accurate values to the attributes. For example, skatteverket is the trusted credential issuer for your date of birth, government is the trusted source for your nationality and so on. Credentials are linked to the private key of the user (the prover).

4. Different from certificates, credentials are not be shown as it is thus not resulting in disclosure of all the attributes and their values and make it possible to trace the prover. Instead, credentials systems implements a selective disclosure protocol that allow the prover to select a subset of attributes from her/ his credentials to be disclosed to the verifier. The other attributes in the credentials are hidden, and certain attribute based credentials even allow the user to only disclose a function computed on the attributes.

ANONYMOUS AUTHENTICATION PROTOCOL EXAMPLE

- Goal: Authenticate Alice to use a video renting service.
- General Idea: No need to reveal all of Alice's information to a verifier or would want Alice’s certificate issuer to track all of her transactions
- Example:
  - Alice wants to prove she is over 18 without revealing her DOB and other attributes [Sim17]



ANONYMOUS CREDENTIALS

- There are two deployed anonymous credential systems and they offer slightly different privacy guarantees. Again what do we expect from such a systems is unlinkability, whenever the credential is used it should not be possible for anyone,

- to link the credential to when it was issued.
- to link the credential to its previous case where it was used before, and
- multiple uses of a credential to the same provider should not linked

SINGLE CREDENTIALS AUTHENTICATION SYSTEMS

1. credentials are used only once. Provers must request for fresh credentials with the same set of attributes from the credential issuer every time they want to disclose an attribute from it.

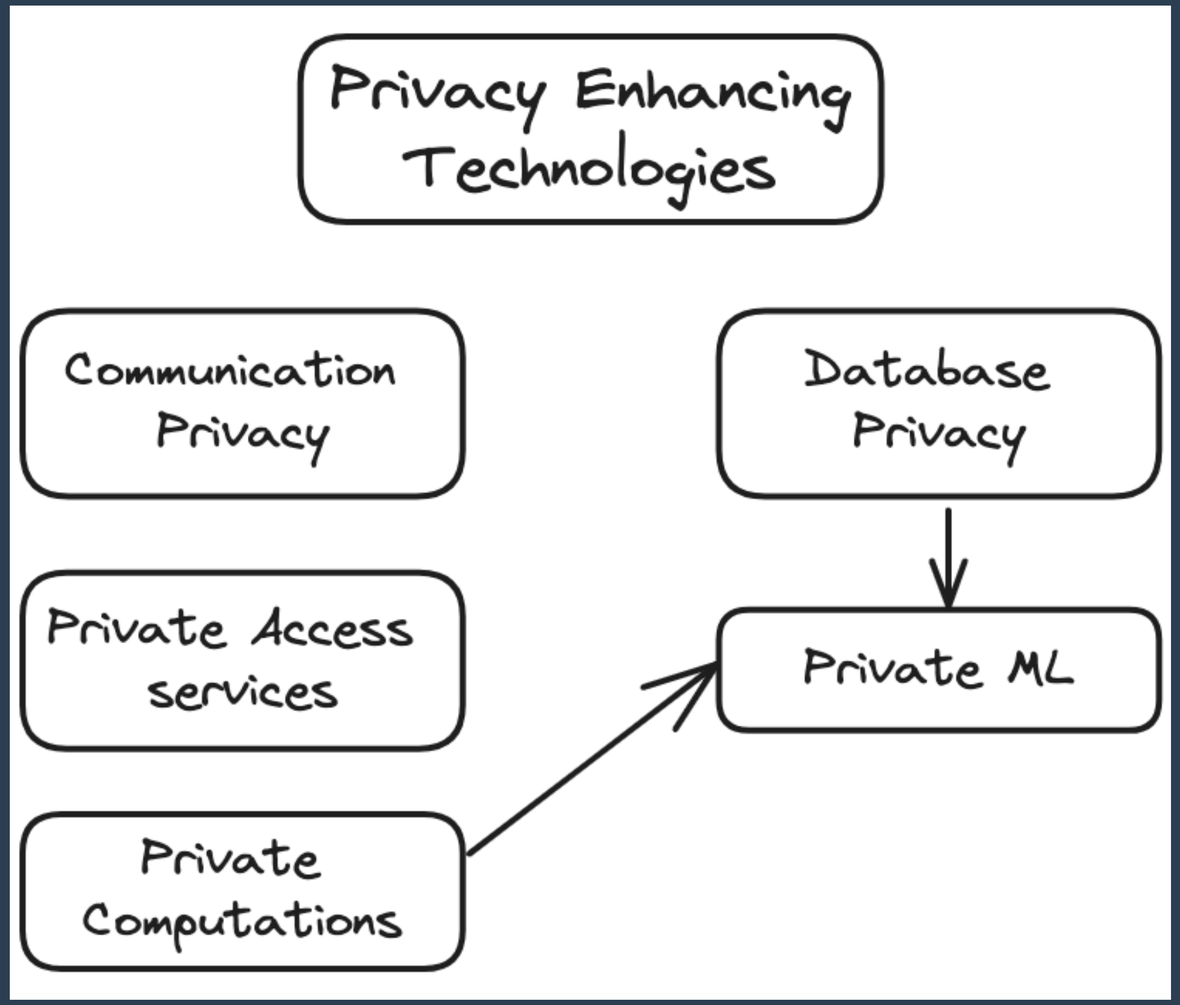
2. Unlinkability between multiple use of selective disclosure.

3. Uses blind signature protocols to ensure unlinkability of a credential between its issuing and the later disclosing of the attributes in it. Blind signatures hides the credential from its issuer.

4. U-Prove by Microsoft is a deployed system that is based on single use credentials, it is easy to setup and efficient. But the disadvantage is that the prover have to be online to obtain fresh credentials every there is a need for authentication.

MULTIPLE CREDENTIALS AUTHENTICATION SYSTEMS

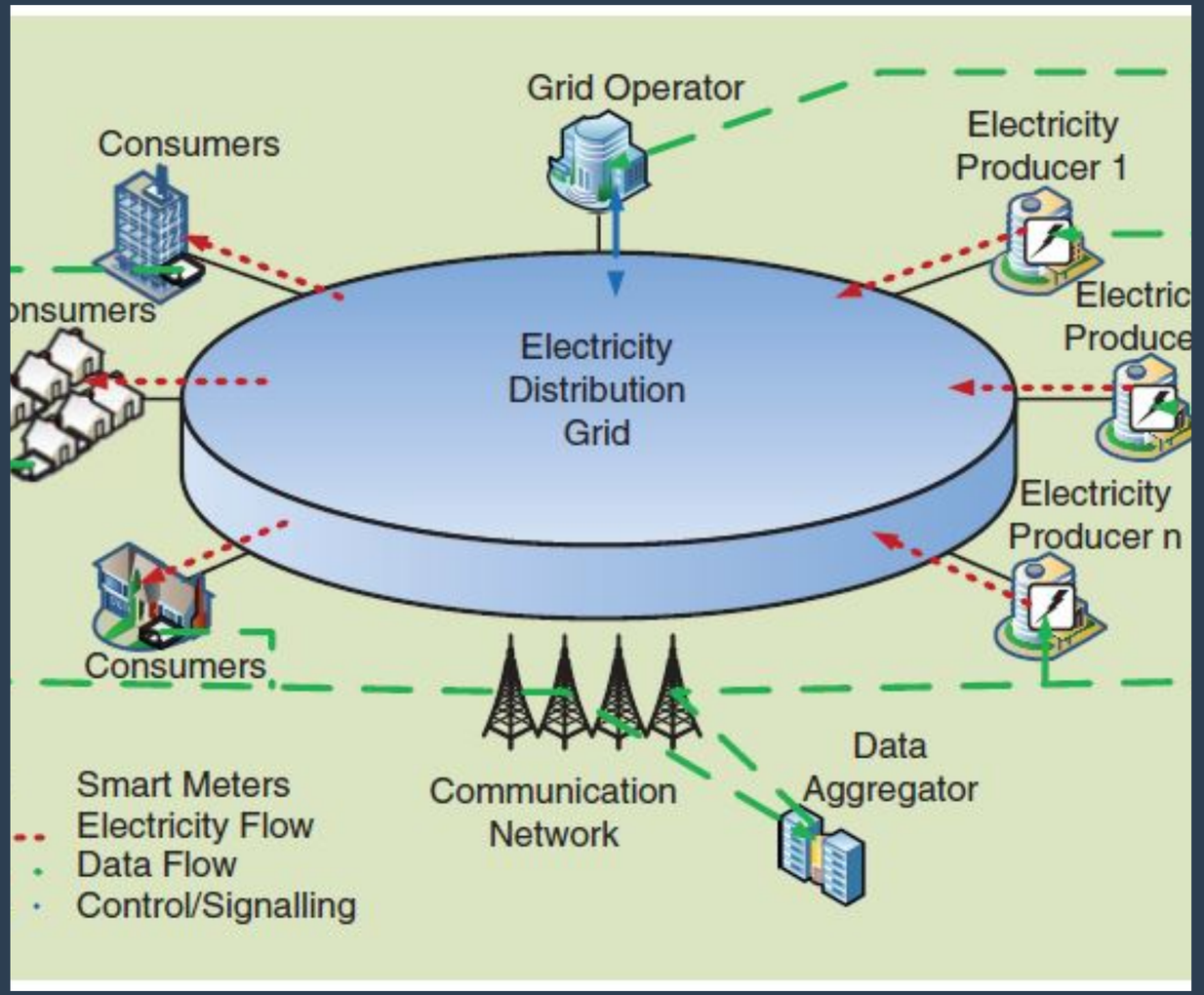
- The alternative strategy is to enable the prover to use the credentials multiple times.
- Unlinkability can be ensured by hiding the credential from the verifier and the issuer.
- This is achieved by a sophisticated crypto protocol called zero-knowledge proof. Zero-knowledge proof enables a user to prove the possession of a secret without revealing the actual secret.
- Idemix is an anonymous credential scheme owned by IBM and is based on the multiple use credentials strategy.
- In Idemix, the provers prove using zero-knowledge proof that they own the credentials that is signed by a valid issuer and that the credentials contains the disclosed attributes.
- Idemix is more complex hence less efficient when it comes to implementation and deployment.



Sign Post to Day I and Day II topics

PRIVATE COMPUTATIONS

- Example: Smart grids
- Consists of three segments;
- Power generation systems,
- Transmission-distribution network and
- smart meters (IoT devices) - to remotely read the measurements for cost calculation



Centralized Smart Grid Architecture

PRIVATE COMPUTATIONS

- Centralized Setup
  - the smart meters sends measurements of short slot intervals to a central data storage that acts as a hub and communicates with each smart meter.
  - The aggregator database is then used for consumption calculation, load balancing calculation and billing. The users get access to the stored data to get information about her consumption.
- De-centralized Setup
  - the smart meters play the role of an aggregator, the calculations such as total consumption, load balancing and billing on the metered data are distributed among consumers.
  - The meters perform a partial data aggregation themselves, calculating the total energy consumption for each billing period and communicate to the energy suppliers.
  - Grid management and load balancing are performed collaboratively by the users.

PRIVATE COMPUTATIONS

- What kind of trust questions is involved in this type of applications?
- fine granular measurements are privacy invasive,
  - simulated attacks have shown to detect from the smart meter data, the presence/absence of residents in a household.
- The utility provider wants to perform analysis for grid management and billing,
  - This is achieved by secure-multi party computations and homomorphic encryption

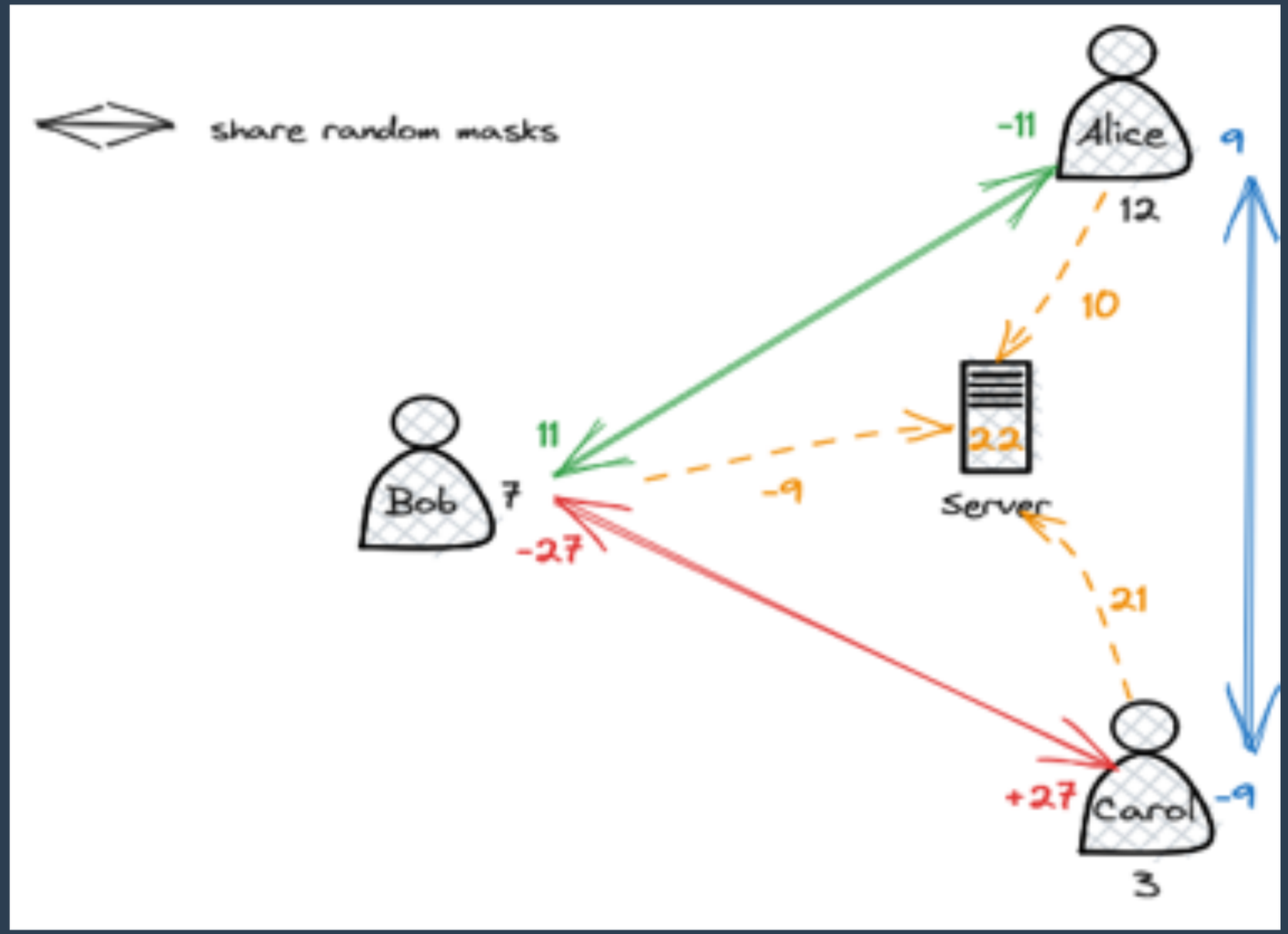
PRIVATE COMPUTATION TECHNOLOGIES

- Homomorphic encryption
  - It is a type of encryption that allows the receiver of the cipher text (the encrypted smart meter data measurements) to compute an operation on these encrypted values like adding the daily fees without having to decrypt them.
- Secure multi-party computations
  - It is a protocol that allows several parties to perform a common computation on their individual values without disclosing their respective values to the others involved in the protocol.

PRIVATE COMPUTATIONS

Privacy preserving computation of total computation of a cell with three users Alice, Bob, Carol along with a utility company.

- Each user in the protocol shares a separate zero-sum mask with every other user in the federation.
- before anything is sent out to the server, each parties scramble their shares by adding it to the shared masks.
- When the server adds up all the inputs the masks cancel out and the server gets the sum of the users' inputs.



REFERENCES FOR FURTHER READING

[Pfitzmann17] – A Terminology for Talking about Privacy by Data Minimization, 2017

[Chaum81] – D. L. Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms", Communications of the ACM Volume 24 Issue, 2 Feb, 1981, pp 84-90, https://doi.org/10.1145/358549.358563

[Camenisch02] – J. Camenisch and E. Van Herreweghen, “Design and implementation of the idemix anonymous credential system,” in Proceedings of the 9th ACM conference on Computer and communications security, New York, NY, USA, Nov. 2002, pp. 21-30. doi: 10.1145/586110.586114.

[Danezis10] – G. Danezis and S. Gürses, “A critical review of 10 years of Privacy Technology,” 2010.

[Shen11] – Shen, Y., & Pearson, S. (2011). Privacy enhancing technologies: A review. Hewlet Packard Development Company. Disponible en https://bit. ly/3cfpAKz.

[Daniel06] – D. Solove, “A Taxonomy of Privacy”, University of Pennsylvania Law Review, 2006.

[Nissenbaum09] – H. Nissenbaum, Privacy in Context: Technology, Policy and the Integrity of Social Life. Stanford University Press, Palo Alto, CA, 2009.