[excerpts from] **A terminology for talking about privacy by data minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management**

Excerpts for COMP 2109*, extracted from: v0.34 doc (10-Aug-2010) by Andreas Pfitzmann and Marit Hansen.*
source: http://dud.inf.tu-dresden.de/Anon_Terminology.shtml

[... Papers] about *privacy by data minimization* ... deal with *anonymity*, *unlinkability*, *unobservability*, and *pseudonymity*.

"**Privacy** is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."

Data minimization means: 1) the possibility to collect personal data about others should be minimized; then 2) collecting personal data should be minimized; then 3) the time how long collected personal data is stored should be minimized.

[Our setting involves] *entities* (*subjects* and *objects*) and *actions*. [S]ubjects execute actions on objects ... subjects called *senders* send objects called *messages* to subjects called *recipients* using a *communication network*, i.e., *stations* send and receive messages using *communication lines*.

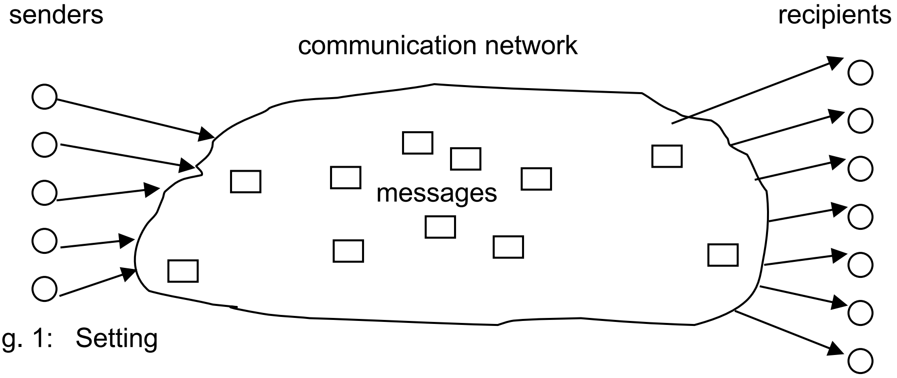


Fig. 1: Setting

All statements are made from the perspective of an *attacker* who may be interested in monitoring what communication is occurring, what patterns of communication exist, or even in manipulating the communication.

The attacker [... has] *items of interest* (IOIs), e.g., who did send or receive which messages.

Attributes (and their values) are related to IOIs because these attribute values may be of interest themselves or their observation may give information on IOIs: An *attribute* is a quality or characteristic of an entity or action. ... Mainly we are interested in attributes of subjects [such as] "sending a message" or "receiving a message".

... we assume that the attacker is not able to get information on the sender or recipient from the message content. Therefore, we do not mention the message content in these sections.

Anonymity **of a subject means that the subject is not identifiable within a set of subjects, the** ***anonymity set*****.**

The *anonymity set* is the set of all possible subjects. [To avoid the implication of] anonymity as a binary property [...] a slightly more complicated definition is:

Anonymity **of a subject** from an attacker's perspective **means that the attacker cannot** *sufficiently* **identify the subject within a set of subjects, the** *anonymity set*.

[Here] "sufficiently" underlines both that there is a possibility to quantify anonymity and that for some applications, there might be a need to define a threshold where anonymity begins.
[For simplicity], we will mainly discuss the quantity of anonymity [...using the term] "strength of anonymity".

Unlinkability **of 2 or more items of interest (IOIs, e.g., subjects, messages, actions, ...) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not.**

Linkability **of 2 or more items of interest (IOIs, e.g., subjects, messages, actions, ...) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker can sufficiently distinguish whether these IOIs are related or not.** Linkability is the negation of unlinkability.

[For ideal unlinkability: an attacker's ability to relate items doesn't increase by observing the system or interacting with it.]

To describe anonymity in terms of unlinkability, we ... augment the definitions... by making explicit the [relevant attributes]:

A sender *s* is anonymous w.r.t. sending, iff *s* is anonymous within the set of potential senders [= sender anonymity set].
[... analogous definitions hold for recipients of messages].

Now [to] describe anonymity in terms of unlinkability ... we consider sending and receiving of messages as attributes; the items of interest (IOIs) are “who has sent or received which message”. Then, *anonymity* of a subject w.r.t. an attribute may be defined as unlinkability of this subject and this attribute ... So we have:

Sender anonymity of a subject means that to this potentially sending subject, each message is unlinkable.

A *pseudonym* [or nym] **is an identifier of a subject other than one of the subject’s real names.**
“Real name” is the antonym to “pseudonym”.

The subject which the pseudonym refers to is the *holder* of the pseudonym.

A subject is *pseudonymous* if a pseudonym is used as identifier instead of one of its real names.

***Pseudonymity* is the use of pseudonyms as identifiers.**

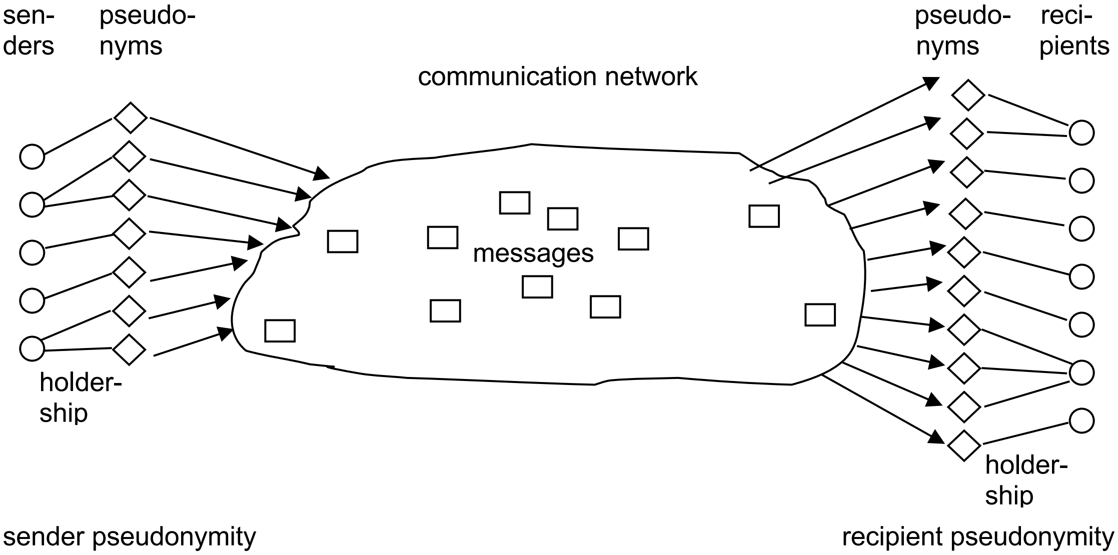


Fig. 2: Pseudonymity

A *digital pseudonym* is a bit string which, to be meaningful in a certain context, is unique as identifier (with very high probability) and suitable to be used to authenticate the holder’s IOIs relatively to his/her digital pseudonym, e.g., to authenticate his/her messages sent.

Ongoing use of the same pseudonym allows the holder to establish or consolidate a **reputation**.

A ***public key certificate*** bears a digital signature of a so-called *certification authority* and provides some assurance to the binding of a public key to another pseudonym, usually held by the same subject. In case that pseudonym is the civil identity (the real name) of a subject, such a certificate is called an *identity certificate*.

Identity management. Setting. To address privacy-enhancing identity management, extend our setting [by dropping the assumption] that an attacker [cannot] get information on [message senders or recipients] from message content and/or the sending or receiving context (time, location information, etc.) of the message. [Assume] the attacker [can] use these attributes for linking messages and, correspondingly, the pseudonyms used with them.

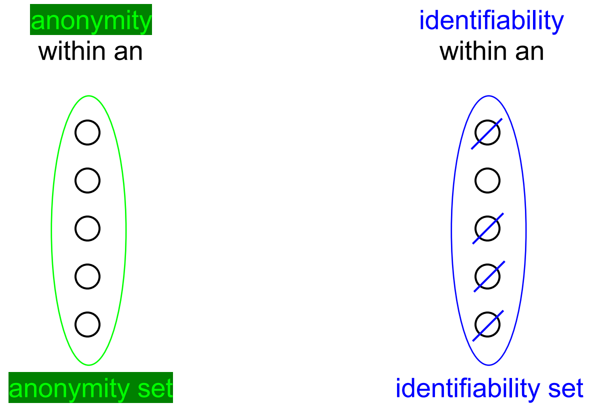
Identity can be explained as an exclusive perception of life, integration into a social group, and continuity, which is bound to a body and – at least to some degree – shaped by society.

[Separate the] concept of identity [into] “I” and “Me” ... “I” is the instance accessible only by the individual self, perceived as an instance of liberty and initiative. “Me” is supposed to stand for the social attributes, defining a human identity that is accessible by communications [...here our interest is “Me”, i.e,] identity as communicated to others and seen by them.

... *identity* can be explained and defined as a property of an entity [as *opposites* of] *anonymity* and *unlinkability*.

***Identifiability* of a subject from an attacker’s perspective means that the attacker can sufficiently identify the subject within a set of subjects, the *identifiability set*.**

Fig.9 contrasts anonymity set and identifiability set.



An *identity* is any subset of attribute values of an individual person which sufficiently identifies this individual person within any set of persons. So usually there is no such thing as "the identity", but several of them.

A **partial identity** is a subset of attribute values of a complete identity, where a *complete identity* is the union of all attribute values of all identities of this person. [Both identities and partial identities] may comprise particular attribute values like names, identifiers, digital pseudonyms, and addresses – but they don't have to.

A ***pseudonym*** might be an identifier for a partial identity. [A pseudonym that's] a *digital pseudonym* [allows] the possibility to authenticate w.r.t. the partial identity, which [can prevent "identity theft", i.e., others taking over the partial identity].

Digital identity denotes attribution of attribute values to an individual person, which are immediately operationally accessible by technical means [e.g., the] identifier of a digital partial identity can be a simple e-mail address.

Identity management means managing various partial identities (usually denoted by pseudonyms) of an individual person, i.e., administration of identity attributes including the development and choice of the partial identity and pseudonym to be (re-)used in a specific context or role.

Establishment of ***reputation*** is possible when the individual person re-uses partial identities.

... identity management is called ***privacy-enhancing*** if it sufficiently preserves unlinkability (as seen by an attacker) between the partial identities of an individual person required by the applications. *This definition focuses on the main property of* **Privacy Enhancing Technologies** (**PETs**), namely: **data minimization** [see p.1 for more context].

Identity management is called ***user-controlled*** if the flow of [a] user's identity attribute values is explicit to the user and the user is in control of this flow.

A Privacy-Enhancing identity managment system [PE-IMS] is an IMS that [for a restricted set of applications] sufficiently preserves unlinkability (as seen by an attacker) between partial identities and corresponding pseudonyms of individuals.

A user-controlled identity management system is an IMS that makes the flow of this user's identity attribute values explicit to the user and gives its user control of this flow. **The guiding principle is "notice and choice".**

Combining user-controlled IMS with PE-IMS means **user-controlled linkability of personal data**, i.e., achieving user-control based on thorough data minimization. [Seems to me: the opposite of what happens in most web sites today.]

Summary of main definitions [left] **and their opposites** [right]

Anonymity of a subject from an attacker's perspective means that the attacker cannot sufficiently identify the subject within a set of subjects, the *anonymity set*.

Identifiability of a subject from ... means ... can sufficiently identify the subject within ... the *identifiability set*.

Unlinkability of 2 or more items of interest (IOIs, e.g., subjects, messages, actions, ...) from an attacker's perspective means that within the system (comprising these and possibly other items), the attacker cannot sufficiently distinguish whether these IOIs are related or not.

Linkability of 2 or more [IOIs] ... means ... can sufficiently distinguish whether these IOIs are related or not

Undetectability of an IoI (item of interest), from an attacker's perspective, means that the attacker cannot sufficiently distinguish whether [the IoI] exists or not.

Detectability of ... means ... can sufficiently distinguish ...