

# Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology

## Change History

Draft v0.1	July 28, 2000	Andreas Pfitzmann, <a href="mailto:pfitza@inf.tu-dresden.de">pfitza@inf.tu-dresden.de</a>
Draft v0.2	Aug. 25, 2000	Marit Köhntopp, <a href="mailto:marit@koehtopp.de">marit@koehtopp.de</a>
Draft v0.3	Aug. 26-Sep. 01, 2000	Andreas Pfitzmann, Marit Köhntopp
Draft v0.4	Sep. 13, 2000	Andreas Pfitzmann, Marit Köhntopp Changes in sections Anonymity, Unobservability, Pseudonymity
Draft v0.5	Oct. 03, 2000	Adam Shostack, <a href="mailto:adam@zeroknowledge.com">adam@zeroknowledge.com</a> , Andreas Pfitzmann, Marit Köhntopp Changed definitions, unlinkable pseudonym

## Abstract

Based on the nomenclature of the early papers in the field, we propose a set of terminology which is both expressive and precise. More particularly, we define *anonymity*, *unlinkability*, *unobservability*, and *pseudonymity* (*pseudonyms* and *digital pseudonyms*, and their attributes).

We hope that the adoption of this terminology might help to achieve better progress in the field by avoiding that each researcher invents a language of his/her own from scratch. Of course, each paper will need additional vocabulary, which might be added consistently to the terms defined here.

## Setting

We develop this terminology in the usual setting that *senders* send *messages* to *recipients* using a communication network. For other settings, e.g. users querying a database, customers shopping in an e-commerce shop, the same terminology can be derived by abstracting away the special names “sender”, “recipient”, and “message”. But for ease of explanation, we use the specific setting here.

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.

## Anonymity

To enable anonymity of a subject, there always has to be an appropriate set of subjects with potentially the same attributes.

***Anonymity is the state of being not identifiable within a set of subjects, the *anonymity set*.***

The *anonymity set* is the set of all possible subjects who might cause an action. Therefore, a sender may be anonymous only within a set of potential senders, his/her *sender anonymity set*, which itself may be a subset of all subjects worldwide who may send messages from time to time. The same is true for the recipient, who may be anonymous within a set of potential recipients, which form his/her *recipient anonymity set*. Both anonymity sets may be disjoint, be the same, or they may overlap.

Anonymity is the stronger, the larger the respective anonymity set is and the more evenly distributed the sending or receiving, respectively, of the subjects within that set is.<sup>1</sup>

### **Unlinkability**

**With respect to the system of which we want to describe anonymity, unobservability, or pseudonymity properties, *unlinkability* of two or more items means that within this system, these items are no more and no less related than they are related concerning the a-priori knowledge.**

This means that the probability of those items being related stays the same before (a-priori knowledge) and after the run within the system (a-posteriori knowledge of the attacker).<sup>2</sup>

E.g. two messages are unlinkable if the probability that they are sent by the same sender and/or received by the same recipient is the same as those imposed by the a-priori knowledge.

### **Anonymity in terms of unlinkability**

If we consider the sending and receiving of messages as the items of interest (IOIs), *anonymity* may be defined as unlinkability of an IOI and an identifier of a subject (ID). More specifically, we can describe the anonymity of an IOI such that it is not linkable to any ID, and the anonymity of an ID as not being linkable to any IOI.

So we have *sender anonymity* as the properties that a particular message is not linkable to any sender and that to a particular sender, no message is linkable.

The same is true concerning *recipient anonymity*, which signifies that a particular message cannot be linked to any recipient and that to a particular recipient, no message is linkable.

A weaker property than each of sender anonymity and recipient anonymity is *relationship anonymity*, i.e. it may be traceable who sends which messages and it may also be possible to trace who receives which messages, but it is untraceable who communicates to whom. In other words, sender and recipient (or recipients in case of multicast) are unlinkable.

### **Unobservability**

In contrast to anonymity and unlinkability, where not the IOI, but only its relationship to IDs or other IOIs is protected, for unobservability, the IOIs are protected as such.

***Unobservability* is the state of IOIs being indistinguishable from any IOI at all.**

---

<sup>1</sup> One might differentiate between the term anonymity and the term indistinguishability, which is the state of being indistinguishable from other elements of a set. Indistinguishability is stronger than anonymity as defined in this text. Even against outside attackers, indistinguishability does not seem to be achievable without dummy traffic. Against recipients of messages, it does not seem to be achievable at all. Therefore, the authors see a greater practical relevance in defining anonymity independent of indistinguishability. The definition of anonymity is an analog to the definition of "perfect secrecy" by Claude E. Shannon, whose definition takes into account that no security mechanism whatsoever can take away knowledge from the attacker which he already has.

<sup>2</sup> Normally, the attacker's knowledge can only increase (analogously to Shannon's definition of "perfect secrecy", see above).

This means that e.g. sending messages is not discernible from random noise. In the digital world, this might be achieved by sending dummy traffic.

As we had anonymity sets with respect to anonymity, we have *unobservability sets* with respect to unobservability.

*Sender unobservability* then means that it is not noticeable whether any sender within the unobservability set sends.

*Recipient unobservability* then means that it is not noticeable whether any recipient within the unobservability set receives.

*Relationship unobservability* then means that it is not noticeable whether anything is sent out of a set of could-be senders to a set of could-be recipients.

### **Relationships between terms**

With respect to the same attacker, unobservability reveals always only a true subset of the information anonymity reveals. We might use the shorthand notation

unobservability  $\Rightarrow$  anonymity

for that. Using the same argument and notation, we have

sender unobservability  $\Rightarrow$  sender anonymity  
recipient unobservability  $\Rightarrow$  recipient anonymity  
relationship unobservability  $\Rightarrow$  relationship anonymity

As noted above, we have

sender anonymity  $\Rightarrow$  relationship anonymity  
recipient anonymity  $\Rightarrow$  relationship anonymity

sender unobservability  $\Rightarrow$  relationship unobservability  
recipient unobservability  $\Rightarrow$  relationship unobservability

### **Known mechanisms for anonymity and unobservability**

DC-net and MIX-net are mechanisms to achieve sender anonymity and relationship anonymity, respectively, both against strong attackers. If we add dummy traffic, both provide for the corresponding unobservability.

Broadcast and anonymous information retrieval are mechanisms to achieve recipient anonymity against strong attackers. If we add dummy traffic, both provide for recipient unobservability.

As a side remark, we mention steganography and spread spectrum as two other well-known unobservability mechanisms.

## Pseudonymity

*Pseudonyms* are identifiers of subjects, in our setting of sender and recipient. (If we would like to, we could easily generalize pseudonyms to be identifiers of sets of subjects, but we do not need this in our setting.) The subject that may be identified by the pseudonym is the *holder* of the pseudonym<sup>3</sup>.

### **Pseudonymity is the use of pseudonyms as IDs.**

So *sender pseudonymity* is defined by the sender's use of a pseudonym, *recipient pseudonymity* is defined by the recipient's use of a pseudonym. In contrast to anonymity, pseudonymity allows the holder of a pseudonym to establish a reputation while using the same pseudonym. Moreover, accountability can be realized with pseudonyms. Pseudonyms enable dealing with claims in case of abuse of anonymity: Firstly, third parties may have the possibility to reveal the identity of a subject in order to provide means for investigation or prosecution. Secondly, third parties may act as liability brokers of the subject to clear a debt or settle a claim.

A *digital pseudonym* is a bit string which is unique as ID and which can be used to authenticate the holder and his/her IOIs, e.g. messages. It could be realized as a public key to test digital signatures where the holder of the pseudonym can prove holdership by forming a digital signature which is created using the corresponding private key. The most prominent example for digital pseudonyms are public keys generated by the user himself/herself, e.g. using PGP<sup>4</sup>.

A *public key certificate* bears a digital signature of a so-called *certification authority* and pertains to the binding of a public key to a subject. An *attribute certificate* is a digital certificate which contains further information (*attributes*) and clearly refers to a specific public key certificate. Independent of certificates, attributes may be used as identifiers of sets of subjects as well. Normally, attributes refer to sets of subjects (i.e. the anonymity set), not to one specific subject.

## Two dimensions with respect to anonymity of pseudonyms

There are many properties of pseudonyms which may be of importance in specific application contexts. In order to describe the properties of pseudonyms with respect to anonymity, we limit our view to two dimensions and give some typical examples:

1. Initial knowledge of the linking between the pseudonym and its holder  
The knowledge of the linking may not be a constant but change over time for some or even all people. Normally, the knowledge of the linking only increases.
  - a) *public pseudonym*:  
The linking between a public pseudonym and a subject may be publicly known even from the very beginning. E.g. the linking could be listed in public directories such as the entry of a phone number in combination with its owner.
  - b) *initially non-public pseudonym*:  
The linking between an initially non-public pseudonym and a subject may be known by certain parties, but is not public at least initially. E.g. a bank account where the bank can look up the linking may serve as a non-public pseudonym. For some specific non-public pseudonyms, certification authorities could reveal the identity of the holder in case of abuse.

---

<sup>3</sup> We prefer the term "holder" over "owner" of a pseudonym because it seems to make no sense to "own" IDs, e.g. bit strings. Furthermore, the term "holder" sounds more neutral than the term "owner", which is associated with an assumed autonomy of the subject's will.

<sup>4</sup> In using PGP, each user may create an unlimited number of key pairs by himself/herself, bind each of them to an e-mail address, self-certify each public key by using his/her digital signature or asking another introducer to do so, and circulate it.

c) *initially unlinkable pseudonym*:

The linking between an initially unlinkable pseudonym and a subject is – at least initially – not known to anybody with the possible exception of the holder himself/herself. Examples for unlinkable pseudonyms are (non-public) biometrics like DNA information unless stored in databases including the linking to a subject.

Public pseudonyms and initially unlinkable pseudonyms can be seen as endpoints of the described pseudonym dimension whereas initially non-public pseudonyms characterize the continuum in between.

Anonymity is the stronger, the less is known about the linking to a subject. The strength of anonymity decreases with increasing knowledge of the pseudonym linking. In particular, under the assumption that no gained knowledge on the linking of a pseudonym will be forgotten, a public pseudonym never can become an unlinkable pseudonym. In each specific case, the strength of anonymity depends on the knowledge of certain parties about the linking relative to the chosen attacker model.

2. Linkability due to the use of a pseudonym in different contexts

a) *person pseudonym*:

A person pseudonym is a substitute for the subject's name which is regarded as representation for the subject's civil identity. It may be used in all contexts, e.g. a nickname, the pseudonym of an actor, or a phone number.

b) *role pseudonym*:

The use of role pseudonyms is limited to specific roles, e.g. a customer pseudonym or an Internet account used for many instantiations of the same role "Internet user". The same role pseudonym may be used with different communication partners. Roles might be assigned by other parties, e.g. a company, but they might be chosen by the subject himself/herself as well.

c) *relationship pseudonym*:

For each communication partner, a different relationship pseudonym is used. The same relationship pseudonym may be used in different roles for communicating with the same partner. Examples are distinct nicknames for each communication partner.

d) *role instantiation pseudonym*:

For each role and for each communication partner, a different role instantiation pseudonym is used. This means that the communication partner need not be aware that two pseudonyms used in different roles belong to the same holder.

e) *transaction pseudonym*<sup>5</sup>:

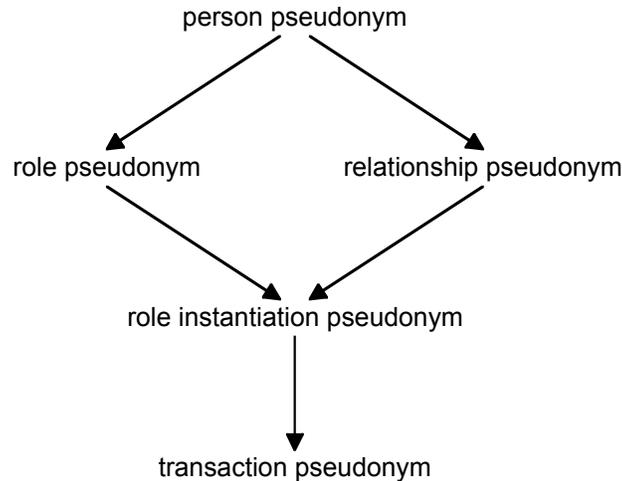
For each transaction, a different transaction pseudonym is used, e.g. randomly generated transaction numbers for online-banking. Thus, there is at least no possibility to link different transactions by equality of pseudonyms. Therefore, transaction pseudonyms can be used to realize as strong anonymity as possible.<sup>6</sup>

The strength of the anonymity of these pseudonyms can be represented as the lattice that is illustrated in the following diagram. The arrows point in direction of increasing anonymity, i.e.  $A \rightarrow B$  stands for "B enables stronger anonymity than A".

---

<sup>5</sup> Apart from "transaction pseudonym" some employ the term "one-time-use pseudonym", taking the pattern from "one-time pad".

<sup>6</sup> In fact, the strongest anonymity ("transaction anonymity") is given when there is no identifying information at all, i.e. information that would allow linking of anonymous entities, thus transforming the anonymous transaction into a pseudonymous one. If the transaction pseudonym is used exactly once, we have the same degree of anonymity as if no pseudonym is used at all.



In general, anonymity of both role pseudonyms and relationship pseudonyms is stronger than anonymity of person pseudonyms. The strength of anonymity increases with the application of role instantiation pseudonyms, the use of which is restricted to both role and relationship. Ultimate strength of anonymity is obtained with transaction pseudonyms.

Anonymity is the stronger, ...

- ... the less often and the less context-spanning pseudonyms are used and therefore the less data about the subject can be gathered.
- ... the more often pseudonyms are changed over time.

### **Other properties of pseudonyms and known mechanisms**

There are several other properties of pseudonyms which should only be shortly mentioned but not discussed in detail in this text. They comprise different degrees of, e.g.,

- limitation to a fixed number of pseudonyms per subject<sup>7</sup>,
- guaranteed uniqueness<sup>8</sup>,
- transferability to other subjects,
- convertability, i.e. transferability of attributes of one pseudonym to another<sup>9</sup>,
- possibility and frequency of pseudonym changeover,
- limitation in number of uses,
- validity (e.g. time limit, restriction to a specific application),
- possibility of revocation or blocking, or
- participation of users or other parties in forming the pseudonyms.

In addition, there may be some properties for specific applications (e.g. addressable pseudonyms serve as a communication address) or due to the participation of third parties (e.g. in order to circulate the pseudonyms, to reveal identities in case of abuse, or to cover claims).

Some of the properties can easily be realized by extending a digital pseudonym by attributes of some kind, e.g. a communication address, and specifying the appropriate semantics. The binding of attributes to a pseudonym can be documented in an attribute certificate produced either by the holder himself/herself or by a certification authority. If the attribute certificate is dated, anybody

---

<sup>7</sup> For pseudonyms issued by an agency that guarantees the limitation of at most one pseudonym per individual, the term “is-a-person pseudonym” is used.

<sup>8</sup> E.g. “globally unique pseudonyms”.

<sup>9</sup> This is a property of convertible credentials.

can spot the currently valid one among different certificates which are inconsistent with one another.

**Concluding remark**

This text is a first proposal for terminology in the field “anonymity, unobservability, and pseudonymity”. The authors hope to get feedback to improve this text and to come to a more precise terminology. Everybody is invited to participate in the process of defining an essential set of terms.