Anonymity, Unobservability, Pseudonymity, and Identity Management – A Proposal for Terminology

Archives of this Document
http://dud.inf.tu-dresden.de/Literatur_V1.shtml (v0.15 and all succeeding versions)

Change History

Draft v0.1	July 28, 2000	Andreas Pfitzmann, pfitza@inf.tu-dresden.de
Draft v0.2	Aug. 25, 2000	Marit Köhntopp, marit@koehntopp.de
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, adam@zeroknowledge.com, Andreas Pfitzmann, Marit Köhntopp Changed definitions, unlinkable pseudonym
Draft v0.6	Nov. 26, 2000	Andreas Pfitzmann, Marit Köhntopp Changed order, role-relationship pseudonym, references
Draft v0.7	Dec. 07, 2000	Marit Köhntopp, Andreas Pfitzmann
Draft v0.8	Dec. 10, 2000	Relationship to Information Hiding Terminology
Draft v0.9	April 1, 2001	Andreas Pfitzmann, Marit Köhntopp IHW review comments
Draft v0.10	April 9, 2001	Andreas Pfitzmann, Marit Köhntopp Clarifying remarks
Draft v0.11	May 18, 2001	Marit Köhntopp, Andreas Pfitzmann
Draft v0.12	June 17, 2001	Annotations from IHW discussion
Draft v0.13	Oct. 21, 2002	Andreas Pfitzmann Some footnotes added in response to comments By David-Olivier Jaquet-Chiffelle, jld@hta-bi.bfh.ch
Draft v0.14	May 27, 2003	Marit Hansen, Andreas Pfitzmann Minor corrections and clarifying remarks
Draft v0.15	June 3, 2004	Andreas Pfitzmann, Marit Hansen Incorporation of comments by Claudia Diaz; Extension of title and addition of identity management terminology
Draft v0.16	June 23, 2004	Andreas Pfitzmann, Marit Hansen Incorporation of lots of comments by Giles Hogben, Thomas Kriegelstein, David-Olivier Jaquet-Chiffelle, and Wim Schreurs; relation between anonymity sets and identifiability sets clarified
Draft v0.17	July 15, 2004	Andreas Pfitzmann, Marit Hansen Triggered by questions of Giles Hogben, some footnotes added concerning quantification of terms; Sandra Steinbrecher caused a clarification in defining pseudonymity

Abstract

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

1 Introduction

Early papers from the 1980ies already deal with anonymity, unlinkability, unobservability, and pseudonymity and introduce these terms within the respective context of proposed measures. We show relationships between these terms and thereby develop a consistent terminology. Then we contrast these definitions with newer approaches, e.g., from ISO IS 15408. Finally, we extend this terminology to identity management.

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.

This document is organized as follows: First the setting used is described. Then definitions of anonymity, unlinkability, and unobservability are given and the relationships between the respective terms are outlined. Afterwards, known mechanisms to achieve anonymity and unobservability are listed. The next sections deal with pseudonymity, i.e. pseudonyms, the corresponding mechanisms, and their properties. Thereafter, this is applied to privacy enhancing identity management. Finally, concluding remarks are given.

2 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. We do not assume the attacker to be an outsider tapping communication lines only, but to be an insider able to participate in normal communications and controlling at least some stations.

Throughout the Sections 3 to 12 we assume that the attacker is not able to get information on the sender or recipient from the message content.[1] Therefore, we do not mention the message content in these sections. For most applications it is unreasonable to assume that the attacker forgets something. Thus, normally the knowledge of the attacker only increases.

3 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*.**[2]

[1] In real life, this cannot easily be achieved as information cannot be removed from messages. Of course, encryption of messages provides protection of the content against attackers observing the communication lines.

[2] From [ISO99]: “[Anonymity] ensures that a user may use a resource or service without disclosing the user’s identity. The requirements for anonymity provide protection of the user

The *anonymity set* is the set of all possible subjects[3]. With respect to actors, the anonymity set consists of the subjects who might cause an action. With respect to addressees, the anonymity set consists of the subjects who might be addressed. 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. The anonymity sets may vary over time.[4]

All other things being equal, 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.[5,6]

4 Unlinkability

Unlinkability only has a meaning after the system of which we want to describe anonymity, unobservability, or pseudonymity properties has been defined. Then:

***Unlinkability* of two or more items (e.g., subjects, messages, events, actions, ...) means that within this system, these items are no more and no less related than they are related concerning the a-priori knowledge.**[7,8]

---

identity. Anonymity is not intended to protect the subject identity. [...] Anonymity requires that other users or subjects are unable to determine the identity of a user bound to a subject or operation." Compared with this explanation, our definition is more general as it is not restricted to identifying users, but any subjects.

[3] I.e. the "usual suspects" :-) The set of possible subjects depends on the knowledge of the attacker. Thus, anonymity is relative with respect to the attacker.

[4] Since we assume that the attacker does not forget anything he knows, the anonymity set cannot increase. Especially subjects joining the system in a later stage, do not belong to the anonymity set from the point of view of an attacker observing the system since an earlier stage. Due to linkability, cf. below, the anonymity set normally can only decrease.

[5] The entropy of a message source as defined by Claude E. Shannon might be an appropriate measure to quantify anonymity – just take who is the sender/recipient as the "message" in Shannon's definition.

[6] 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 [Shan49], whose definition takes into account that no security mechanism whatsoever can take away knowledge from the attacker which he already has.

[7] From [ISO99]: "[Unlinkability] ensures that a user may make multiple uses of resources or services without others being able to link these uses together. [...] Unlinkability requires that users and/or subjects are unable to determine whether the same user caused certain specific operations in the system." In contrast to this definition, the meaning of unlinkability in this text is less focused on the user, but deals with unlinkability of "items" and therefore is a general approach. Note that we chose a relative definition of unlinkability, referring to a-priori knowledge and its possible change. We may differentiate between "absolute unlinkability" (as in [ISO99]; i.e. "no determination of a link between uses") and "relative unlinkability" (i.e. "no change of knowledge about a link between uses").

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).[9]

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.

5 Anonymity in terms of unlinkability

If we consider sending and receiving of messages as the items of interest (IOIs)[10], *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.[11]

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.

Relationship anonymity means that it is untraceable who communicates with whom. In other words, sender and recipient (or recipients in case of multicast) are unlinkable. Thus, relationship anonymity is a weaker property than each of sender anonymity and recipient anonymity, as it may be traceable who sends which messages and it may also be possible to trace who receives which messages, as long as the relationship between sender and recipient is not known.

6 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.[12]

---

[8] As the entropy of a message source might be an appopriate measure to quantify anonymity (and thereafter "anonymity" might be used as a quantity), we may use definitions to quantify unlinkability (and thereafter "unlinkability" might be used as a quantity as well). Quantifications of unlinkability can be either probabilities or entropies, or whatever is useful in a particular context.

[9] Normally, the attacker's knowledge cannot decrease (analogously to Shannon's definition of "perfect secrecy", see above). An exception of this rule is the scenario where the use of misinformation leads to a growing uncertainty of the attacker which information is correct. In the special case where it is known before that some items are related, of course the probability of these items being related stays the same. Even in this "degenerated" case it makes sense to use the term unlinkability because there is no *additional* information.

[10] In this context, the term IOI is used for events like sending or receiving messages. Nevertheless, the general term IOI is chosen in order to be able to more easily extend the meaning in later versions, e.g., including communication relationships.

[11] Unlinkability is a sufficient condition of anonymity (since we defined anonymity in absolute terms, i.e. not relative to the a-priori knowledge of an attacker, but unlinkability only relative to the a-priori knowledge of the attacker, this is not exactly true, but it would be if we either made the definition of unlinkability stronger or the definition of anonymity weaker), but it is not a necessary condition. Thus, failing unlinkability does not necessarily eliminate anonymity as defined in Section 3; in specific cases even the degree of anonymity may not be affected.

[12] Unobservability can be regarded as a possible and desirable property of steganographic systems (see "Known mechanisms"). Therefore it matches the information hiding terminology [Pfit96, ZFKP98]. In contrast, anonymity, describing the relationship to *IDs*, does not directly fit into that terminology, but independently represents a different dimension of properties.

***Unobservability* is the state of IOIs being indistinguishable from any IOI at all.**[13],[14]

This means that messages are not discernible from "random noise".

As we had anonymity sets of subjects with respect to anonymity, we have *unobservability sets* of subjects with respect to unobservability.[15]

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.

7 Relationships between terms

With respect to the same attacker, unobservability reveals always only a true subset of the information anonymity reveals.[16] We might use the shorthand notation

unobservability ⇒ anonymity

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

sender unobservability ⇒ sender anonymity
recipient unobservability ⇒ recipient anonymity
relationship unobservability ⇒ relationship anonymity

[13] From [ISO99]: "[Unobservability] ensures that a user may use a resource or service without others, especially third parties, being able to observe that the resource or service is being used. [...] Unobservability requires that users and/or subjects cannot determine whether an operation is being performed." As seen before, our approach is less user-focused and insofar more general. With the communication setting and the attacker model chosen in this text, our definition of unobservability shows the method how to achieve it: preventing distinguishability of IOIs. Thus, the ISO definition might be applied to a different setting where attackers are prevented from observation by other means, e.g., by encapsulating the area of interest against third parties.

[14] In some applications (e.g. steganography), it might be useful to quantify unobservability to have some measure how much uncertainty about an IOI remains after the attacker's observations. Again, we may use probabilities or entropy, or whatever is useful in a particular context.

[15] Actually, unobservability deals with events instead of subjects. Though, like anonymity sets, unobservability sets consist of the subjects who might possibly send and/or receive.

[16] [ReRu98] propose a continuum for describing the degree of anonymity with the following states named: "absolute privacy" (the attacker cannot perceive the presence of communication, i.e. unobservability) – "beyond suspicion" – "probable innocence" – "possible innocence" – "exposed" – "provably exposed" (the attacker can prove the sender, recipient, or their relationship to others). Although we think that the terms "privacy" and "innocence" are misleading, the spectrum is quite useful.

As noted above, we have

sender anonymity ⇒ relationship anonymity
recipient anonymity ⇒ relationship anonymity

sender unobservability ⇒ relationship unobservability
recipient unobservability ⇒ relationship unobservability

8 Known mechanisms for anonymity and unobservability

DC-net [Chau85, Chau88] and MIX-net [Chau81] 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 [PfPW91].[17]

Broadcast [Chau85, PfWa86, Waid90] and private information retrieval [CoBi95] are mechanisms to achieve recipient anonymity against strong attackers. If we add dummy traffic, both provide for recipient unobservability.

Of course, dummy traffic[18] alone can be used to make the number and/or length of sent messages unobservable by everybody except for the recipients; respectively, dummy traffic can be used to make the number and/or length of received messages unobservable by everybody except for the senders. As a side remark, we mention steganography and spread spectrum as two other well-known unobservability mechanisms.

9 Pseudonymity

Pseudonyms are identifiers of subjects[19], in our setting of sender and recipient. (If we would like to, we could generalize pseudonyms to be identifiers of *sets* of subjects, but we do not need this in our setting.[20]) The subject which the pseudonym refers to is the *holder* of the pseudonym[21].

Being *pseudonymous* is the state of using a pseudonym as ID.[22]

[17] If dummy traffic is used to pad sending and/or receiving on the sender's and/or recipient's line to a constant rate traffic, MIX-nets can even provide sender and/or recipient anonymity and unobservability.

[18] Misinformation may be regarded as semantic dummy traffic, i.e. communication from which an attacker cannot decide which are real requests with real data or which are fake ones. Assuming the authenticity of misinformation may lead to privacy problems for (innocent) bystanders.

[19] "Pseudonym" comes from Greek "pseudonumon" meaning "falsely named" (pseudo: false; onuma: name). Thus, it means a name other than the "real name". As the "real name" (written in ID papers issued by the State) is somewhat arbitrary (it even can be changed during one's lifetime), we will extend the term "pseudonym" to all identifiers, including all names or other bit strings. You may think of a pseudonym as a mapping of the identifier "real name" into another name. The "real name" may be understood as a pseudonym resulted from the neutral mapping.

[20] Such *group pseudonyms* may act as an anonymity set, i.e. using the information provided by the pseudonym only, an attacker cannot decide whether an action was performed by a specific person within the group.

[21] 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. The holder may be a natural person (in this case we have the usual meaning and all data protection regulations apply), a legal person, or even only a computer.

[22] Please note that despite the terms "anonymous" and "pseudonymous" are sharing most of their letters, their semantics is quite different: Anonymous says something about the state of a subject

Seeing using pseudonyms more as a process comprising preparing for the use of pseudonyms e.g. by establishing certain rules how to identify holders of pseudonyms by so-called identity brokers or to prevent uncovered claims by so-called liability brokers (cf. Section 11), leads to the more general notion of pseudonymity:

***Pseudonymity* is the use of pseudonyms as IDs.**[23,24]

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.

10 Pseudonymity with respect to accountability

A *digital pseudonym* is a bit string which is
- unique as ID and
- suitable to be used to authenticate the holder and his/her IOIs, e.g., messages sent.

Using digital pseudonyms, accountability can be realized with pseudonyms.

11 Pseudonymity with respect to linkability[25]

Whereas anonymity and accountability are the extremes with respect to linkability to subjects, pseudonymity is the entire field between and including these extremes. Thus, pseudonymity comprises all degrees of linkability to a subject. Ongoing use of the same pseudonym allows the holder to establish or consolidate a reputation. Some kinds of pseudonyms enable dealing with claims in case of abuse of unlinkability to holders: Firstly, third parties (identity brokers) may have the possibility to reveal the civil identity of the holder[26] in order to provide means for investigation

---

with respect to identifiability, pseudonymous only says something about employing a mechanism, i.e. using pseudonyms. Whether this mechanism helps in a particular setting to achieve something close to anonymity, is a completely different question.

[23] From [ISO99]: "[Pseudonymity] ensures that a user may use a resource or service without disclosing its user identity, but can still be accountable for that use. [...] Pseudonymity requires that a set of users and/or subjects are unable to determine the identity of a user bound to a subject or operation, but that this user is still accountable for its actions." This view on pseudonymity covers only the use of digital pseudonyms. Therefore, our definition of pseudonymity is much broader as it is does not necessarily require disclosure of the user's identity and accountability. Pseudonymity alone – as it is used in the real world and in technological contexts – does not tell anything about the degrees of anonymity or accountability; these degrees depend on several properties, cf. below.

[24] Quantifying pseudonymity would primarily mean quantifying the state of using a pseudonym according to its different dimensions (cf. the next two sections 10 and 11), i.e. quantifying the accountability gained and quantifying the anonymity left over (e.g. using entropy as the measure). Roughly speaking, well-employed pseudonymity would mean appropriately fine-grained accountability to counter identity theft or to prevent uncovered claims in e-commerce using e.g. the techniques described in [BüPf90], combined with much anonymity retained. Poorly employed pseudonymity would mean giving away anonymity without preventing uncovered claims.

[25] Linkability is the negation of unlinkability, i.e. items are either more or are either less related than they are related concerning the a-priori knowledge.

[26] If the holder of the pseudonym is a natural person or a legal person, civil identity has the usual meaning. If the holder is, e.g., a computer, it has to be defined what "civil identity" shall mean. It could mean, for example, exact type and serial number of the computer (or essential components of it) or even include the natural person or legal person responsible for its operation.

or prosecution [Chau81]. Secondly, third parties may act as liability brokers of the holder to clear a debt or settle a claim [BüPf90].

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 aspects and give some typical examples:

11.1 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, for non-transferable pseudonyms the knowledge of the linking cannot decrease.[27] Typical kinds of such pseudonyms are:

a) *public pseudonym*:
   The linking between a public pseudonym and its holder 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 its holder 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 civil identity of the holder in case of abuse.
c) *initially unlinked pseudonym*:
   The linking between an initially unlinked pseudonym and its holder is – at least initially – not known to anybody with the possible exception of the holder himself/herself. Examples for unlinked pseudonyms are (non-public) biometrics like DNA information unless stored in databases including the linking to the holders.

Public pseudonyms and initially unlinked pseudonyms can be seen as extremes of the described pseudonym aspect 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 and that the pseudonym cannot be transferred to other subjects, a public pseudonym never can become an unlinked 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.

If the pseudonym is transferable, the linking to its holder can change. Considering an unobserved transfer of a pseudonym to another subject, a formerly public pseudonym can become non-public again.

[27] With the exception of misinformation which may blur the attacker's knowledge (see above).

11.2 Linkability due to the use of a pseudonym in different contexts

With respect to the degree of linkability, various kinds of pseudonyms may be distinguished according to the kind of context for their usage:

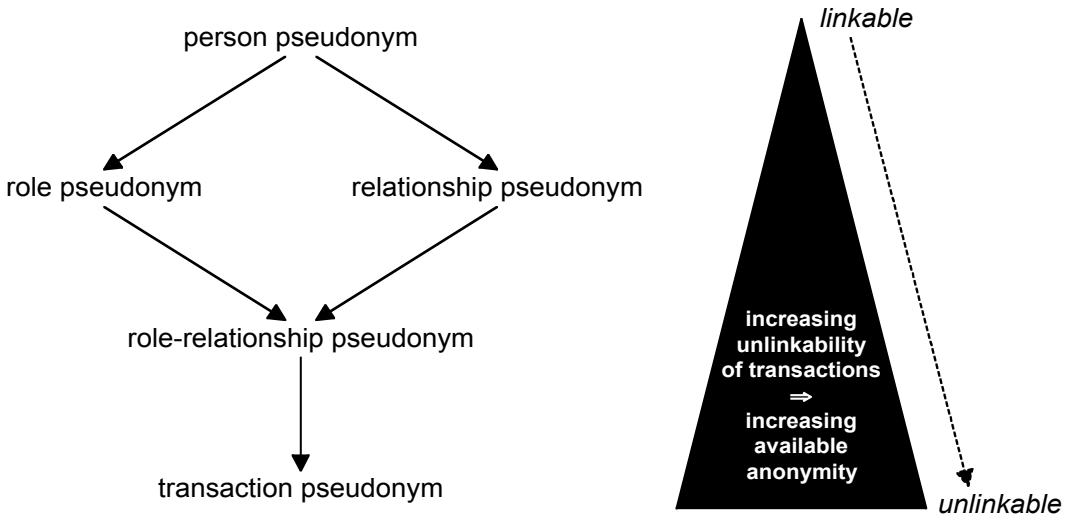
a) *person pseudonym*:
   A person pseudonym is a substitute for the holder's name which is regarded as representation for the holder's civil identity. It may be used in all contexts, e.g., a number of an identity card, the social security number, a nickname, the pseudonym of an actor, or a phone number.
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-relationship pseudonym*:
   For each role and for each communication partner, a different role-relationship pseudonym is used. This means that the communication partner does not necessarily know, whether two pseudonyms used in different roles belong to the same holder. On the other hand, two different communication partners who interact with a user in the same role, do not know from the pseudonym alone whether it is the same user.
e) *transaction pseudonym*[28]:
   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.[29]

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".[30]

[28] Apart from "transaction pseudonym" some employ the term "one-time-use pseudonym", taking the naming from "one-time pad".

[29] 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. Another possibility to achieve strong anonymity is to prove the holdership of the pseudonym or specific properties (e.g., with zero-knowledge proofs) without revealing the information about the pseudonym or properties itself. Then, no identifiable or linkable information is disclosed.

[30] "$\rightarrow$" is not the same as "$\Rightarrow$" of Section 7, which stands for the implication concerning anonymity and unobservability.



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-relationship pseudonyms, the use of which is restricted to both the same role and the same relationship. Ultimate strength of anonymity is obtained with transaction pseudonyms.

Anonymity is the stronger, ...
- ... the less personal data of the pseudonym holder can be linked to the pseudonym;
- ... the less often and the less context-spanning pseudonyms are used and therefore the less data about the holder can be linked;
- ... the more often independently chosen, i.e. from an observer's perspective unlinkable, pseudonyms are used for new actions.

The degree of information of linked data can be reduced by different subjects using the same pseudonym (e.g. one after the other when pseudonyms are transferred or simultaneously with specifically created group pseudonyms[31]) or by misinformation.

12 Known mechanisms and other properties of pseudonyms

A digital pseudonym 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 [Chau81]. The most prominent example for digital pseudonyms are public keys generated by the user himself/herself, e.g., using PGP[32].

[31] The group of pseudonym holders acts as an inner anonymity set within a, depending on context information, potentially even larger outer anonymity set.

[32] In using PGP, each user may create an unlimited number of key pairs by himself/herself (at this moment, such a key pair is an initially unlinked pseudonym), 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.

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. There are several other properties of pseudonyms within the system of their use which shall 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[33] [Chau81, Chau85, Chau90],
- guaranteed uniqueness[34] [Chau81, StSy00],
- transferability to other subjects,
- authenticity of the linking between a pseudonym and its holder (possibilities of verification/falsification or indication/repudiation),
- convertability, i.e. transferability of attributes of one pseudonym to another[35] [Chau85, Chau90],
- possibility and frequency of pseudonym changeover,
- re-usability and, possibly, a limitation in number of uses,
- validity (e.g., guaranteed durability and/or expiry date, 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 civil 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.

13 Identity Management

To adequately address identity management, we have to extend our setting:

- It is not realistic to assume that an attacker might not get information on the sender or recipient of messages from the message content and/or the sending or receiving context (time, location information, etc.) of the message. We have to consider that the attacker is able to use these properties for linking messages and, correspondingly, the pseudonyms used with them.
- In addition, it is not just human beings, legal persons, or simply computers sending messages and using pseudonyms at their discretion as they like at the moment, but they use application programs, which strongly influence the sending and receiving of messages and may even strongly determine the usage of pseudonyms.

Identity can be explained as an exclusive perception of life, integration into a social group, and continuity, which is bound to a body and shaped by society. This concept of identity distinguishes between "I" and "Me" [Mead34]: "I" is the instance that is accessible only by the individual self, perceived as an instance of liberty and initiative. "Me" is supposed to stand for the social

[33] 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.
[34] E.g., “globally unique pseudonyms”.
[35] This is a property of convertible credentials.

attributes, defining a human identity that is accessible by communications and that is an inner instance of control and consistency.[36]

Corresponding to the anonymity set introduced in the beginning of this text, we can work with an "identifiability set"[37] to define "identifiability" and "identity" [Hild03][38]:

***Identifiability* is the possibility of being individualized within a set of subjects, the *identifiability set*.**

An *identity* is any subset of attributes of an individual which uniquely characterizes this individual within any set of individuals. So usually there is no such thing as "the identity", but several of them.

Role:
In sociology, a "role" or "social role" is a set of connected behaviors, as conceptualized by actors in a social situation (i.e., situation-dependent identity attributes and properties). It is mostly defined as an expected behavior in a given individual social context.

Partial identity:
Each identity of a person comprises many partial identities of which each represents the person in a specific context or role. Partial identities are subsets of attributes of a complete identity. On a technical level, these attributes are data.
Thus, a *pseudonym* might be an identifier for a partial identity.
Whereas we assume that an "identity" uniquely characterizes an individual (without limitation to particular identifiability sets), a partial identity may not do, thereby enabling different degrees of anonymity. But we may find for each partial identity appropriately small identifiability sets[39], where the partial identity uniquely characterizes an individual.[40]

Digital identity:
Digital identity denotes attribution of properties to a person, which are immediately operatively accessible by technical means. More to the point, the identifier of a digital partial identity[41] can be a simple e-mail address in a news group or a mailing list. Its owner will attain a certain reputation. More generally we might consider the whole identity as a combination from "I" and "Me" where the "Me" can be divided into an implicit and an explicit part: Digital identity is the digital part from

---

[36] For more information see [ICPP03].
[37] The *identifiability set* is a set of possible subjects.
[38] This definition is compatible with the definitions given in: Giles Hogben, Marc Wilikens, Ioannis Vakalis: On the ontology of Digital Identification; Intern Conf. on Ontologies, Databases and Applications of Semantics, 3-7 Nov. 2003, Catania, Sicily (paper to be published in LNCS, Springer); and it is very close to that given by David-Olivier Jaquet-Chiffelle in http://www.calt.insead.edu/fidis/workshop/workshop-wp2-december2003/presentation/VIP/vip_id_def2_files/frame.htm: "An identity is any subset of attributes of a person which uniquely characterizes this person within a community."
[39] For identifiability sets of cardenality 1, this is trivial, but it may hold for "interesting" identifiability sets of larger cardenality as well.
[40] The relation between *anonymity set* and *identifiabilityset* can be seen in two ways:
1. Within an a-priori anonymity set, we can consider a-posteriori identifiability sets as subsets of the anonymity set. Then the largest identifiability sets allowing identification characterize the a-posteriori anonymity, which i.e. is zero iff the largest identifiability set allowing identification equals the a-priori anonymity set.
2. Within an a-priori identifiability set, its subsets which are the a-posteriori anonymity sets characterize the a-posteriori anonymity. It is zero, iff all a-posteriori anonymity sets have cardenality 1.

[41] A *digital partial identity* is the same as a *partial digital identity*. In the sequel, we skip "partial" if the meaning is clear from the context.

the explicated "Me". Digital identity should denote all those personally related data that can be stored and automatically interlinked by a computer-based application.

Virtual identity:
Virtual identity is sometimes used in the same meaning as digital identity or digital partial identity, but because of the connotation with “unreal, non-existent, seeming” the term is mainly applied to characters in a MUD (Multi User Dungeon), MMORPG (Massively Multiplayer Online Role Playing Games) or to avatars.

Identity management:
Identity management means managing the various partial identities, i.e., their valuation as "applicable to one self" (role taking) or forming them (role making). A prerequisite to choose the appropriate partial identity is to recognize the situation the person is acting in.

Privacy enhancing identity management:
Given the restrictions of an application, identity management is called *perfectly privacy enhancing* if by choosing the pseudonyms carefully, it does not provide more linkability between partial identities to an attacker than giving the attacker all the data with all pseudonyms omitted.
The identity management is called *privacy enhancing* if it does not provide essentially more linkability between the partial identities.[42]

Privacy enhancing application design:
An application is designed in a privacy enhancing way if neither the imposed pattern of sending/receiving messages nor the attributes given to entities (i.e. humans, organizations, computers) imply more linkability than is strictly necessary to achieve the purposes of the application.

Identity management system (IMS):
Technology-based identity management in its broadest sense refers to administration and design of identity attributes.
We can distinguish between identity management system and identity management application: The term "identity management system" is seen as an infrastructure, in which "identity management applications" as components are co-ordinated. Identity management applications are tools for individuals to manage their socially relevant communications, which can be installed, configured and operated at the user's and/or a server's side.
A technically supported identity management has to empower the user to recognise different kinds of communication or social situations and to assess them with regards to their relevance, functionality and their security and privacy risk in order to make and take an roles adequately.
In general the identity management application should help the user in managing one’s partial identities, meaning that different pseudonyms with associated data sets can be used according to different roles the user is acting in and according to different communication partners.

Privacy enhancing identity management system (PE-IMS):
A Privacy Enhancing IMS makes the flow of personal data explicit and gives its user a larger degree of control [CPHH02]. The guiding principle is “notice and choice”, based on a high level of data minimization: This means user-controlled linkage of personal data.[43]

[42] Note that due to our setting, this definition focuses on the main property of Privacy Enhancing Technologies, namely data minimization: This property means to ensure as much unlinkability as possible. We are aware of the limitation of this definition: In the real world it is not always desired to achieve utmost unlinkability. We believe that the user as the data subject should be empowered to decide on the degree of linkage of his or her personal data within the boundaries of legal regulations, i.e., in an advanced setting the privacy enhancing application design should also take into account the support of “user-controlled linkage”.

According to respective situation and context, such a system supports the user in making an informed choice of pseudonyms, representing his or her partial identities. A PE-IMS supports the user in managing his or her partial identities, i.e., in particular the processes of role taking and role making. It acts as a central gateway for all communication between different applications, like browsing the web, buying in Internet shops, or carrying out administrative tasks with governmental authorities [HBCC04].

14 Concluding remark

This text is a first proposal for terminology in the field “anonymity, unobservability, pseudonymity, and identity management”. 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.

15 References

BüPf90 Holger Bürk, Andreas Pfitzmann: Value Exchange Systems Enabling Security and Unobservability; Computers & Security 9/8 (1990) 715-721.

Chau81 David Chaum: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms; Communications of the ACM 24/2 (1981) 84-88.

Chau85 David Chaum: Security without Identification: Transaction Systems to make Big Brother Obsolete; Communications of the ACM 28/10 (1985) 1030-1044.

Chau88 David Chaum: The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability; Journal of Cryptology 1/1 (1988) 65-75.

Chau90 David Chaum: Showing credentials without identification: Transferring signatures between unconditionally unlinkable pseudonyms; Auscrypt '90, LNCS 453, Springer-Verlag, Berlin 1990, 246-264.

CoBi95 David A. Cooper, Kenneth P. Birman: Preserving Privacy in a Network of Mobile Computers; 1995 IEEE Symposium on Research in Security and Privacy, IEEE Computer Society Press, Los Alamitos 1995, 26-38.

CPHH02 Sebastian Clauß, Andreas Pfitzmann, Marit Hansen, Els Van Herreweghen: Privacy-Enhancing Identity Management; The IPTS Report 67 (September 2002) 8-16.

HBCC04 Marit Hansen, Peter Berlich, Jan Camenisch, Sebastian Clauß, Andreas Pfitzmann, Michael Waidner: Privacy-Enhancing Identity Management; Information Security Technical Report (ISTR) Volume 9, Issue 1 (2004), Elsevier, UK, 35-44, http://dx.doi.org/10.1016/S1363-4127(04)00014-7.

Hild03 Mireille Hildebrandt (Vrije Universiteit Brussel): presentation at the FIDIS workshop 2nd December, 2003; slides: http://www.calt.insead.edu/fidis/workshop/workshop-wp2-december2003/presentation/VUB/VUB_fidis_wp2_workshop_dec2003.ppt.

ICPP03 Independent Centre for Privacy Protection & Studio Notarile Genghini: Identity Management Systems (IMS): Identification and Comparison Study; commissioned by the Joint

[43] And by default unlinkability of different user actions so that communication partners involved in different actions by the same user cannot combine the personal data disseminated during these actions.

Research Centre Seville, Spain, September 2003,
http://www.datenschutzzentrum.de/projekte/idmanage/study.htm.

ISO99 ISO IS 15408, 1999, http://www.commoncriteria.org/.

Mead34 George H. Mead: Mind, Self and Society, Chicago Press 1934.

Pfit96 Birgit Pfitzmann (collected by): Information Hiding Terminology -- Results of an informal plenary meeting and additional proposals; Information Hiding, LNCS 1174, Springer-Verlag, Berlin 1996, 347-350.

PfPW91 Andreas Pfitzmann, Birgit Pfitzmann, Michael Waidner: ISDN-MIXes -- Untraceable Communication with Very Small Bandwidth Overhead; 7th IFIP International Conference on Information Security (IFIP/Sec '91), Elsevier, Amsterdam 1991, 245-258.

PfWa86 Andreas Pfitzmann, Michael Waidner: Networks without user observability -- design options; Eurocrypt '85, LNCS 219, Springer-Verlag, Berlin 1986, 245-253; revised and extended version in: Computers & Security 6/2 (1987) 158-166.

ReRu98 Michael K. Reiter, Aviel D. Rubin: Crowds: Anonymity for Web Transactions, ACM Transactions on Information and System Security 1(1), November 1998, 66-92.

Shan49 Claude E. Shannon: Communication Theory of Secrecy Systems; The Bell System Technical Journal 28/4 (1949) 656-715.

StSy00 Stuart Stubblebine, Paul Syverson: Authentic Attributes with Fine-Grained Anonymity Protection; Financial Cryptography 2000, LNCS Series, Springer-Verlag, Berlin 2000.

Waid90 Michael Waidner: Unconditional Sender and Recipient Untraceability in spite of Active Attacks; Eurocrypt '89, LNCS 434, Springer-Verlag, Berlin 1990, 302-319.

ZFKP98 J. Zöllner, H. Federrath, H. Klimant, A. Pfitzmann, R. Piotraschke, A. Westfeld, G. Wicke, G. Wolf: Modeling the security of steganographic systems; 2nd Workshop on Information Hiding, LNCS 1525, Springer-Verlag, Berlin 1998, 345-355.