

Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology

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
Draft v0.6	Nov. 26, 2000	Changed definitions, unlinkable pseudonym 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

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*, and *pseudonymity* (*pseudonyms* and *digital pseudonyms*, and their attributes).

1 Introduction

Early papers from the 80ies 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.

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. 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 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 the sequel.

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

The *anonymity set* is the set of all possible subjects³. 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

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

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 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. Due to linkability, cf. below, the anonymity set can only decrease.

⁵ 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.

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.⁶

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).⁷

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)⁸, *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.⁹

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

⁶ 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.

⁷ Normally, the attacker’s knowledge cannot decrease (analogously to Shannon’s definition of “perfect secrecy”, see above). 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.

⁸ 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.

⁹ Unlinkability is a sufficient condition of anonymity, 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.

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.¹⁰

Unobservability is the state of IOIs being indistinguishable from any IOI at all.¹¹

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.¹²

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.¹³ 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

¹⁰Unobservability 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.

¹¹ 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.

¹² 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.

¹³ [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, receiver, 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 \Rightarrow relationship anonymity
recipient anonymity \Rightarrow relationship anonymity

sender unobservability \Rightarrow relationship unobservability
recipient unobservability \Rightarrow 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].

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 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, 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 pseudonym¹⁴.

Pseudonymity is the use of pseudonyms as IDs.¹⁵

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.

¹⁴ 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.

¹⁵ 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 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.

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 linkability¹⁶

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. Using the same pseudonym more than once, the holder may establish a reputation. Some kinds of pseudonyms enable dealing with claims in case of abuse of unlinkability to holders: Firstly, third parties may have the possibility to reveal the identity of the holder in order to provide means for investigation or prosecution. Secondly, third parties may act as liability brokers of the holder to clear a debt or settle a claim.

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:

11.1 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 cannot decrease. Typical kinds of 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 identity of the holder in case of abuse.
- c) *initially unlinkable pseudonym*:
The linking between an initially unlinkable pseudonym and its holder 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 the holders.

Public pseudonyms and initially unlinkable pseudonyms can be seen as extremes 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

¹⁶ 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.

strength of anonymity depends on the knowledge of certain parties about the linking relative to the chosen attacker model.

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:

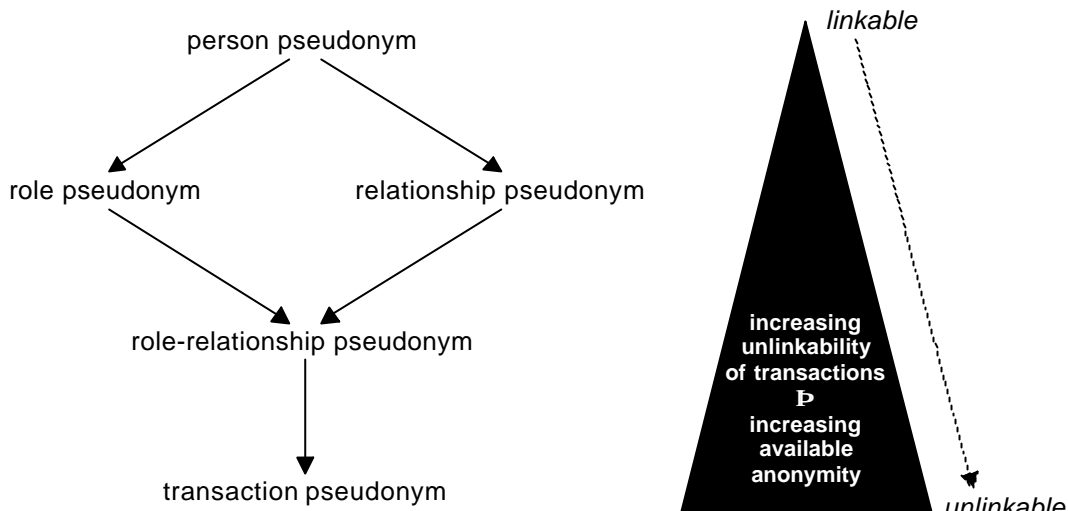
- a) *person pseudonym*:
A person pseudonym is a substitute for the holder's name which is regarded as representation for the holder's civil identity. It may be used in all contexts, e.g., a nickname, the pseudonym of an actor, or a phone number.
- b) *role pseudonym*:
The use of role pseudonyms is limited 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.
- e) *transaction pseudonym*¹⁷:
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.¹⁸

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".¹⁹

¹⁷ Apart from "transaction pseudonym" some employ the term "one-time-use pseudonym", taking the naming from "one-time pad".

¹⁸ 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.

¹⁹ " \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 often and the less context-spanning pseudonyms are used and therefore the less data about the holder can be linked.
- ... the more often pseudonyms are changed over time.

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 PGP²⁰.

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 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 subject²¹ [Chau81, Chau85, Chau90],
- guaranteed uniqueness²² [Chau81, StSy00],
- transferability to other subjects,

²⁰ 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.

²¹ 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.

²² E.g., "globally unique pseudonyms".

- convertability, i.e. transferability of attributes of one pseudonym to another²³ [Chau85, Chau90],
- 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 can spot the currently valid one among different certificates which are inconsistent with one another.

13 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.

14 References

- 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.
- ISO99 ISO IS 15408, 1999, <http://csrc.nist.gov/cc/>.
- 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.

²³ This is a property of convertible credentials.

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 C. 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.