FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

A Survey and Taxonomy on Privacy Enhancing Technologies

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Wirtschaftsinformatik

eingereicht von

Peter Zimmermann, BSc
Matrikelnummer 0828244

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: Univ.-Prof. Dipl.-Ing. Dr. techn. A Min Tjoa
Mitwirkung: Mag.rer.soc.oec. Johannes Heurix

Wien, 20.11.2014

(Unterschrift Verfasser) (Unterschrift Betreuung)

FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

A Survey and Taxonomy on Privacy Enhancing Technologies

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Peter Zimmermann, BSc
Registration Number 0828244

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: Univ.-Prof. Dipl.-Ing. Dr. techn. A Min Tjoa
Assistance: Mag.rer.soc.oec. Johannes Heurix

Vienna, 20.11.2014

(Signature of Author) (Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Peter Zimmermann, BSc
Lainzer Straße 157/3/8, 1130 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum) (Unterschrift Verfasser)

Acknowledgements

I want to thank my family and my friends for their support during my study - especially my parents who always supported me in any possible way. I also want to thank all my colleagues for a lot of interesting discussions, insights and a broad variety of opinions.

Abstract

Due to the continuous increase of data storage e.g. by Google and Facebook an increasing need for data protection is created. One basic requirement for implementing data protection is an overview of the technological state of the art. In the fast moving IT industry this is especially difficult, as technological developments, changed habits, and as a consequence the risk exposure may change fast.

For everyone dealing with the implementation of systems for data protection it is thus especially important to know recent developments and possibilities. This is particularly important for persons dealing with this topic for the first time. Unfortunately, there exists few literature providing an overview about this topic which allows to get started fast - especially if the importance of up to date information is considered.

For this reason, this thesis offers an overview about recent developments in the field of Privacy Enhancing Technologies. To provide a better structure a taxonomy is desired - unfortunately, there does not exist an appropriate one for this purpose. Thus an own taxonomy for Privacy Enhancing Technologies is created. The techniques are evaluated by means of this taxonomy. Finally, an evaluation of possible combinations of the clusters of techniques as well as exemplary combinations of specific techniques is provided.

Kurzfassung

Aufgrund der kontinuierlich steigenden Datensammlung unter anderem durch Google und Facebook entsteht ein stärker werdendes Bedürfnis nach Datenschutz. Eine Grundvoraussetzung, um Datenschutz implementieren zu können, ist einen Überblick über den Stand der Technik zu haben. Das ist in der schnelllebigen IT Industrie besonders wichtig, da sich hier technologische Entwicklungen, geänderte Verhaltensweisen und damit einhergehend auch das Gefährdungspotential rasch ändern können.

Für alle, die sich mit der Implementierung von Systemen zum Datenschutz befassen, ist es daher besonders wichtig, aktuelle Entwicklungen und Möglichkeiten zu kennen. Ganz besonders gilt das für Personen, die neu in dieses Thema einsteigen wollen. Leider gibt es jedoch wenig Überblicksliteratur, die einen raschen und schnellen Start ermöglicht - speziell, wenn die Wichtigkeit von aktuellen Informationen berücksichtigt wird.

In dieser Arbeit wird deshalb ein Überblick über aktuelle Entwicklungen im Bereich der Privacy Enhancing Technologies gegeben. Zur besseren Strukturierung ist eine Taxonomie wünschenswert - leider gibt es jedoch noch keine für diesen Zweck geeignete. Daher wird eine entsprechende Taxonomie für Privacy Enhancing Technologies erstellt. Anhand dieser Taxonomie werden die untersuchten Techniken evaluiert. Abschließend erfolgt eine Auswertung über Kombinationsmöglichkeiten der gebildeten Gruppen von Techniken sowie beispielhaft von mehreren konkreten Techniken.

Contents

CHAPTER 1

Introduction

Privacy becomes more and more important - not only for experts, but also in public perception. Examples for the growing importance of privacy are discussions related to data collection by Google (e.g. resistance against street view) and Facebook [34] (automatic facial recognition, timeline, the initiative Europe vs. Facebook [111]), discussions related to data retention (in German „Vorratsdatenspeicherung“), passports with biometrical information included, and the growing concentration of video monitoring.

Every time when handling sensitive data, considerations about privacy have to be taken into account. Except for enforcement by laws, it is also an important consideration from the business point of view as a lack of privacy may harm/prevent business. On the other hand, one may be able to attract customers from a competitor if privacy is implemented well (and of course this may be - depending on the business domain - an important point for the customers).

Various ways of enhancing privacy with specific pros and cons and areas of application exist, and not all of them are equally good. Therefore, they are described in detail and then compared to each other to provide a solid base for choosing the correct technique for the correct application scenario. However, it is hard to compare characteristics without a taxonomy for means of classification.

For this reason, a taxonomy (cf. page 9) is highly desirable to support discussions by the use of a standardised naming convention. In the field of security, there already exist some important taxonomies: [3] introduce a taxonomy on secure and dependable computing which, due to its general focus, is (at least partially) applicable on security in general. Unfortunately, its focus is too broad for being able to deal with privacy. On the other hand, [1] introduce a taxonomy which deals with privacy from the policy/requirements point of view. Although it seems to be somehow close to what is needed in this case, it takes privacy from a different point of view into account; it enables to classify policies e.g. on websites but not Privacy Enhancing Technologies. The taxonomy presented in [99] is related to privacy too. However, it is not suitable to classify Privacy Enhancing Technologies as the provided dimensions do not fit the requirements for a taxonomy like it is proposed here. [5] provide a taxonomy that has got some interesting aspects like the purpose for that data may be used. Unfortunately, it is not sufficient

either to classify Privacy Enhancing Technologies properly. [59] introduces a taxonomy related to privacy too, but it is too limited due to its narrow focus on the threats (rather than possibilities to deal with these threats) and „(...)mobile computing and communications technologies used for personal-health monitoring“, [59, p1]. [110] focus on possibilities to deal with threats, but focus on information security in general causing the resulting taxonomy being too general and with too many irrelevant aspects. [91] focus on the information security for service centric systems, but is again too generic.

While these taxonomies all deal with information security in general, they are inadequate to precisely cover the privacy-specific properties of PETs. Therefore, in this thesis, a taxonomy specialised on Privacy Enhancing Technologies is proposed.

The underlying research question regarding the taxonomy is: *How can Privacy Enhancing Technologies be classified in a standardised way?* Besides providing a classification scheme it should ease the evaluation of techniques in this field by providing a base for structured discussions on this topic. In order to establish the taxonomy, literature research is conducted to collect information about Privacy Enhancing Technologies. This information is analysed and used to derive a proper classification.

1.1 Motivation

In general, there are several reasons for the need for privacy. Three of the most common reasons are enforcement through law or the market (competitors), the customers' need for privacy and privacy as a requirement to reach own goals. The first two reasons are easy to understand: if one wants to stay in a market and it is requested by the customers (e.g. sales of products with sexual relation) or even law (e.g. elections), one has to conform with it. However, it may also be in one's own interest to respect privacy to ensure a base of trust and openness that maybe would not be possible otherwise. For example, complaints might be published anonymously to prevent potential negative consequences.

But why is privacy so important? One common argument against the need of privacy (especially when privacy infringement is enforced by law, society, or an organisation) is that if one has not done anything wrong, one need not have concerns about privacy and therefore no need for privacy exists (if you have nothing to hide, you have got nothing to fear). However, this cannot be seen as an argument. You alone should decide how your data is used. If you do not want it to be used it should not be necessary to tell reasons for it. For example, if a stranger just walks around asking, what one has eaten at the last meal or how long it is ago that one has taken a shower, it would probably not harm anyone to answer, but it seems not to be likely that this stranger would get a lot of answers (because it is not his/her business). Another example is if a stranger (or even a familiar person) does not stop staring and observing one's actions - this may easily cause people to feel uncomfortable even if they have not done anything wrong (and therefore have nothing to hide).

Of course privacy may also be exploited to harm others, the society, or state misuse should be prevented. However, a balance must be found between the freedom of the individual and the needs of the society to prevent misuse. Anyway, one has to always take care of existing laws

as well as basic concepts of democracy, and the constitutional state („Grundsatz der Verhältnismäßigkeit“).

1.2 Problem Statement and Aim of the Work

Unfortunately, currently available literature related to privacy or Privacy Enhancing Technologies is often specialised and has got a quite narrow focus - concrete problems or techniques are addressed and often a specific domain like Radio Frequency Identification (RFID) or health care like in [66] and [105] are addressed. This is sufficient for people already involved in this field but in turn makes it hard to get started with this topic (as basic knowledge is missing as well as knowledge about the state of the art). Unfortunately, this implies a limited number of publications which address people, who want to get started with privacy or Privacy Enhancing Technologies. As in the IT industry changes occur in a short period of time, up-to-date literature is very important. This is because this progress may have impacts like new approaches for protecting/threatening privacy, adopted habits related to the usage of IT and, last but not least, the evolution of existing technologies. As a consequence, literature which has been published several years ago, may not be up to date, rendering out-dated publications and their results obsolete.

Therefore, in this master thesis the state of the art knowledge related to Privacy Enhancing Technologies is summed up, in order to provide an overview over recent advances in the area of PETs. Furthermore, a taxonomy on PETs is introduced to allow a structured analysis and discussion of technologies regarding this topic, followed by an examination of recent developments concerning PETs. Additionally, basic techniques are also taken into account which have not changed recently but are relevant as the basis for other more advanced techniques. Then the techniques are evaluated based on the taxonomy in order to show which parts are covered by which technique. As each technique has got strengths and weaknesses, some possibilities for combinations of techniques are suggested.

The most important topics that are examined in this thesis are

- How could Privacy Enhancing Technologies be compared?
- Which (recent) approaches exist to enhance privacy?
- What are the pros and cons and in which contexts may these approaches be applied?
- Comparison of the different approaches based on the taxonomy.
- How may different approaches be combined to enhance privacy for different scenarios and needs?

1.3 Methodological Approach & Structure of the Work

This work consists of 4 main parts (namely a taxonomy to classify Privacy Enhancing Technologies, approaches how to achieve privacy, an evaluation of the investigated approaches based on

the taxonomy introduced in this thesis, and finally suggestions for combinations of techniques), which require different methodologies.

In the first part (cf. chapter Taxonomy at page 15) a taxonomy for classifying Privacy Enhancing Technologies is described. To the best of my knowledge no appropriate taxonomy covering all needed aspects exists - for this reason, a custom taxonomy for the classification is introduced within this thesis. This taxonomy consists of two (sub)-taxonomies: one from a technical point of view and one for the different perspectives. In order to find out if an appropriate taxonomy exists, literature research has been conducted. The resulting taxonomy was developed using knowledge from the investigated taxonomies as well as knowledge from the literature research of the second part (approaches for Privacy Enhancing Technologies) by using abstraction and clustering to find out common and differentiating attributes of these approaches.

For the second part (cf. chapters Basics of Privacy at page 5 and Techniques at page 33) recent developments in the field of Privacy Enhancing Technologies are presented which are organised in 3 sub-parts depending on what aspect is considered: the identity, the content, or the behaviour. Literature research is required for this part because currently available information is spread over a great number of publications and often with a very narrow scope. To ensure a broad scope of application (which is one aim of this thesis), it is important that the approaches discussed are generic and applicable in various scenarios/contexts and domains. For example, shielding in order to block electromagnetic waves may be an good idea for enhancing privacy in the field of RFID but is obviously less useful for protecting data stored in a database. Therefore, approaches which are quite limited will only be discussed in special cases and at least the majority of approaches are applicable in a broad scope.

In the third part (cf. chapter Evaluation at page 57), each investigated technique is evaluated with respect to the taxonomy. Afterwards, a summary is presented so that a quick overview about the techniques and their evaluation is possible. Therefore, the knowledge gathered in the first two parts is accumulated. The investigated approaches for enhancing privacy are evaluated regarding the introduced taxonomy in order to find clusters of related/similar techniques, check if the taxonomy can be used for classification, and to provide a better overview of the investigated techniques. This is done based on the knowledge gathered in the second part and comparing the described characteristics of the approaches to the characteristics defined for the taxonomy.

In the last part (cf. chapter Synergy Effects of PET Groups at page 105), combinations are investigated. First techniques are clustered into groups which are then combined afterwards. Finally a selection of specific techniques is combined in order to allow more precise judgement about the combination. These combinations of techniques intent to cover more aspects of privacy than one single approach. For this chapter, combining and applying all knowledge gathered in the previous parts is required.

CHAPTER 2

Basics of Privacy

This section covers the most important terms, concepts, and considerations related to the topic of privacy in general as well as Privacy Enhancing Technologies. First, a set of definitions for important terms is introduced, followed by some considerations regarding the topic.

2.1 Definitions

The definitions provided below should give an introduction to the reader to form a base to get started with this topic and enable him to follow the considerations, discussions, and evaluations later on in this thesis. Furthermore, these definitions are required to build up a base of common vocabulary so that misunderstandings due to unclear or ambiguous terms are avoided. If not stated otherwise, every time a term is used within this thesis, it is used referring to these definitions.

Privacy

„Privacy is the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about themselves is communicated to others“, [119] quoted in [72, p124] is a good way to start defining privacy. It sums up that everyone is able to determine which parts of one's data are communicated to others and how this takes place. However - especially in context of modern IT systems and data processing it seems necessary to add these issues to the definition too. Therefore, a new definition is introduced which builds up on the first one:

Privacy is the claim of individuals or institutions to determine for themselves when, how and to what extend data about themselves is communicated to whom and which information may be processed by whom including the purpose of processing.

Important laws in the context of privacy and data protection are the Austrian DSG 2000 (Data Protection Law, in German Datenschutzgesetz) [25], the German BDSG (Federal Data Protection Law, in German Bundesdatenschutzgesetz) [16], the Data Protection Directive 95/46/EG

of the European Union [31], and the US Privacy Act [81]. However, these laws will not be investigated in detail as this thesis has a technical and not a legal focus.

Anonymity and Anonymisation

„Anonymity of a subject (...) means that the attacker cannot sufficiently identify the subject within a set of subjects (...)“, [79, p10]. This set of subjects is called anonymity set and has the property that the items are not characterised in a unique way by their attributes (within the set). Therefore, one can measure the anonymity by measuring how many items cannot be distinguished, which corresponds to the probability of knowing the real identity of an item.

Of course the anonymity is the stronger, the more items cannot be distinguished, which corresponds to a lower probability. Anyway, unprotected or unconsidered attributes which allow the identification may exist. This may, for example, be the combination of date and time, the handwriting, or the content of the message itself. If, for example, an employee or customer always complains about something (and no one else complains about it) and then a message related to this appears in the complaint box, it will assumed to be related to this person (however, this may be exploited in sense of spoofing).

The anonymity for every item may differ. Therefore, in a set with usually high anonymity, items with low anonymity or even uniquely characterised items may also be contained (cf. [79, p8-11]). The process of establishing anonymity is called anonymisation.

Pseudonymity and Pseudonymisation

„A pseudonym is an identifier of a subject other than one of the subject's real names.“, [79, p21]. Pseudonym is composed of the Greek words pseudo (false) and onuma (name) which means false name (or as stated in the definition an other name than the real one). The use of a pseudonym by a group (group pseudonymity) is possible too, as well as the transfer of the pseudonym (transferable pseudonym). For a group pseudonym it is however not sufficient that a pseudonym has more than one holder but it has really to be the same pseudo identity (however, the same name may be used several times by different persons or in different systems). In case of the group pseudonymity, privacy may be restricted if the members that hold the pseudonym are known (cf. [79, p21ff]). If the members and the usage of a group pseudonym are not known (and one single person is assumed to hold the pseudonym), privacy may even be increased as it gets more difficult to analyse the data related to the pseudonym. The use of pseudonyms instead of the real identity is called pseudonymisation: „A subject is pseudonymous if a pseudonym is used as identifier instead of one of its real names.“, [79, p21].

Distinction between Anonymity and Pseudonymity

Both terms seem to be close, which is to a certain extend true if the pseudonyms are applied carefully. However, some differences exist: pseudonymisation only ensures that not the real identity is used (cf. [79, p22]) - however, the used pseudonym may be weak and therefore the real identity obvious. „PeterZimmermann“, „PeterZimmerm.“, and „PeterZimmermann1234“ are all pseudonyms as they do not correspond to my real identity. However, they won't help to improve

my privacy as it is easy to guess that they are used instead of my identity. Even if the pseudonyms are used carefully, another point which distinguishes anonymity from pseudonymity exists: anonymity defines that the real identity is not known. With pseudonymity, it is supposed that the real identity is not known (for public). However, persons may exist who may know it (this may also be used for ensuring accountability - this is explained later on).

Trusted Third Party

„A *trusted third party* (TTP) is an entity that facilitates the interactions between two parties who both trust the third party. In TTP models, the relying parties use this trust to secure their own interactions. However, in real life a TTP can become untrusted or malicious.“, [120, p2]. The use of trust in this sense suggests that it is a boolean attribute stating that a Third Party may either be trusted or not. While this may be true in some cases, the question comes up if trust should be considered as something in between absolute trust and absolute distrust. However, this issue should only indicate possible need for further research and is out of scope for this thesis. Therefore, it will not be discussed in more detail and will not be considered beyond its common meaning within this thesis.

Privacy Enhancing Technology

„Privacy-enhancing technologies (PET) have been developed (...) to meet the demand of technological controls providing protection of personal information and online privacy. Various technologies are available (...) from encryption to anonymisation and to full identity management systems. (...) PETs are referred to a variety of technologies that safeguard personal privacy by minimizing, masking, or eliminating the collection of identifiable information (...) to protect: the identities of users by providing anonymity, pseudonymity, unlinkability, unobservability of users(...) and the confidentiality and integrity of personal data.“ [102, p3]. This definition covers only the identity protection. However, the content as well as the access itself are protected by PET too. Therefore, for this thesis the above definition gets extended and it is defined that PET protect these aspects too.

Security

A popular definition of security is as follows: „Security is a composite of the attributes of confidentiality, integrity, and availability, requiring the concurrent existence of 1) availability for authorized actions only, 2) confidentiality, and 3) integrity with 'improper' meaning 'unauthorized.'“, cf. [3, p13].

Computational Security

„A scheme is said to be computationally secure if the security depends on (...) carrying out some computation that in principle is possible but in which all of the known methods of execution require an infeasible amount of computation.“, [96]

Information Theoretic Security

„A system is called (...) information-theoretically (or unconditionally) secure if the ciphertext provides no information about the plaintext, regardless of the adversary's computational power.“, cf. [26, p200].

Although the definition above targets cipher systems, it can be applied to privacy too. The definition below will be used for Information Theoretic Security in this thesis. It is based on the definition above and takes the contrary of Computational Security into account:

In contrast to Computational Security schemes providing Information Theoretic Security cannot be compromised even with unlimited time and computing resources. The only thing an adversary may know is a probability that he or she is right. However, this probability cannot be increased to 100%.

Deniability

For the scope of this thesis it is defined that „(...) a deniable scheme is also non-committing (...) and secure under selective opening attacks on whichever parties can equivocate.“, [76, p525].

Indistinguishability, Unlinkability

Indistinguishability and *Unlinkability* refer to similar concepts that are applied to the *Protected Aspects Identity* (*Indistinguishability*) respectively *Behaviour* or *Content* (both *Unlinkability*). They are used within this thesis as defined below.

Unlinkability is defined as „the infeasibility of an adversary to correlate two transactions initiated by the same user“, [57].

„Two (..) ensembles (...) are said to be indistinguishable if there exists no witness algorithm (...)“, [128, p83]. Or in other words: „Two ensembles (...) are called statistically indistinguishable if their statistical difference is negligible“, [30, p256]. Although both definitions refer to ensembles, they are applicable within the scope of this thesis too. Therefore, these definitions are joined and form the base for a new definition for *Indistinguishability*:

Two identities are said to be indistinguishable if there exists no witness algorithm and their statistical difference is negligible.

Both definitions state that it is not possible to differentiate between several instances of the *Protected Aspect* (*Identity* respectively *Behaviour* or *Content*). Both concepts (Indistinguishability and Unlinkability) are part of the taxonomy introduced in the chapter Taxonomy starting at page 15.

Data Minimisation

„Data minimization means that first of all, the possibility to collect personal data about others should be minimized. Next within the remaining possibilities, collecting personal data should be minimized. Finally, the time how long collected personal data is stored should be minimized.“, [79, p6].

For example, for buying goods online with age restriction it is not necessary to reveal your identity - the only thing of interest is your age. Therefore, a third party may check your identity

(for example, with your passport number) against your age and then only transmit the information to the shop if you are allowed to buy the good or not.

In reality of course other measures have to be found so that no link to the identity is possible because the example mentioned is somehow a contradiction: it would be difficult to argue, why this sensitive data is given to a third party (even if it would only contain the mapping between passport number and age) just to enhance privacy. Additionally, the passport number alone is of course not enough to identify a person over the internet as persons from the same household (as under aged kids for example) could use it. Therefore, an additional PIN or password is necessary. Data Minimisation may be seen as a general way of improving privacy and the handling of data. Additionally, it may also improve the usability as one has to enter less data (maybe without decreasing the quality of service).

Explicit Identifier

The Explicit Identifier is the „set of attributes, such as name and social security number (SSN), containing information that explicitly identifies record owners“, [40, p14:4]. In terms of Relational Data Base Management Systems (RDBMS), this corresponds to the primary key and potentially also to the candidate keys (as they too distinguish all records in a relation). However, some candidate keys may not be explicit identifiers as they either may only be Quasi Identifiers (e.g. there may occur non-unique values, although no records with duplicate values are present) or for some reason (e.g. relation with at some point in time few records) common combinations.

Quasi Identifier

The Quasi Identifier is the „set of attributes that could potentially identify record owners“, [40, p14:4]. In this work, common abbreviation QID is used. In RDBMS this may be a subset of the candidate keys (as explained above) as well as combinations of attributes, which are not considered as key candidates (e.g. in a table of 1 million entries, a QID that occurs only twice must not be considered as a key candidate but nearly identifies one record).

Taxonomy

For the scope of this thesis, the definition of a taxonomy is as stated here: „A taxonomy consists of the core part which is a schema (or more schemas) and linkbases. The schema is the part that contains definitions of elements whereas linkbases provide relationships between them. The definition of elements and their attributes are the basic of the taxonomy.“, cf. [114, p11].

The main purpose of taxonomies is to „allow for a complex subject involving many items (...) to be simplified into a limited number of major categories of related items, to make them more accessible for identification and use. They *[the taxonomies, author's comment]* aid in the understanding of a group of items by allowing for a standardized and heuristic approach to defining and describing the items that make up the subject.“ [22] cited in [60, p2].

Furthermore, taxonomies can be used to analyse, compare, and categorise existing as well as future schemes, are able to show needs for further research (if, for example, one classifica-

tion is not covered sufficiently) (cf. [67, p385], [106, p1]) and enable classification and, as a consequence, (structured) discussions become easier.

Metadata

„Metadata is data linked to some data item, i.e., metadata is data about data. The metadata of a data item specifies how the data item was created, in which context it can be used, how its was transformed, or how it can be interpreted or processed.“, [54].

Confidentiality

Confidentiality is „the absence of unauthorized disclosure of information“, cf. [3, p13] or in other words: Confidentiality ensures that only authorised information disclosure is possible.

2.2 General Considerations

After getting known to the most important definitions, some general considerations are made which are important for the topic of privacy (enhancement) and data protection. In addition to the definitions above, this is important to get started with this topic and to correct some possibly wrong assumptions (e.g. regarding misuse).

Considerations to Anonymity - Misuse

Anonymity, as well as pseudonymity, may increase the risk of misuse. This is because people may feel more confident about their increased privacy and therefore do things they usually would not do, especially because they may have less fear of reprisals. One may expect a noticeable amount of misuse of anonymisers, but an evaluation of the JAP Project [82] however showed a completely different result: in a timespan of more than three years (July 2001 to December 2004), about four to five requests per month have been made from law enforcement or private persons related to abuse (the absolute peak was ten) (cf. [36, p6f]). Especially when considering the amount of data anonymised (three to four terabytes per month), this is only a small proportion of misuse. Another interesting fact is that although the number of users increased, the number of inquiries stayed constant. Two possible explanations for this are (i) a low number of people, who exploit the system for misuse, or (ii) the fact that authorities got used to the fact that they will not receive data from the project (cf. [79, p6f]). Although this data is already several years old, it shows interesting relations and especially only a small portion of misuse.

However, this is not an effect that could only be observed in former times when anonymisation services have not been as popular as they are now, or which is limited to one specific service. A recent research on various anonymisation services (TOR [83], I2P [74], and the commercial GoTrusted System [44]) from October 2010 shows an increased proportion of misuse. However, this too can be considered as only a small overall percentage. Two big problems of misuse were investigated: Spam and illegal filesharing. Only a small percentage of the overall misuse was conducted by using anonymisation services. Depending on the investigation method,

this means 0,67-4,42% (Spam) and below 0,05% (filesharing) (cf. [63]). This shows that illegal activities take place anyway, independently if there exist systems for anonymisation or not. Persons conducting illegal activities are somehow attracted to anonymisation services, but the broad majority of these activities is conducted without the usage of anonymisation services. As a consequence, neither the observation nor the ban of anonymisation services would improve the situation significantly. Therefore, they should be seen as a way how honest people may improve their privacy, and for fighting illegal activities, other countermeasures have to be taken.

So the essence is that misuse of anonymisation services occurs (as with any other service too) but it is not as common as one may expect, and the majority of users is honest and only interesting in protecting their privacy. In contrast to pseudonymity, it is more difficult to deal with misuse in case of anonymity as the real identity usually cannot be reconstructed (one exception to this are some techniques building on Anonymous Credentials, cf. page 53). Depending if one takes identity protection for the Data Consumer or the Record Owners into account, undesired activities have to be handled differently. In case of the Data Consumer, one way to deal with misuse may be the revocation of users, which may be possible anonymously. In case of the Record Owner, things get more complicated but one may preserve a copy of the original data so that users could be re-identified in predefined scenarios. One may consider this as pseudonymity as the published data may be seen as a pseudonym for the real (full) data, but it will still be considered to be anonymous for the scope of this thesis. The reason for this is that one wants to protect sensitive data from Data Consumers. This happens even if a copy of the original data is kept. Furthermore, no new identifier (pseudonym) to the published data is assigned and, last but not least, it is usually not possible to identify one single person but only a group of persons because most techniques protecting the Record Owner ensure that there exists a minimum number of records per group. For these reasons, keeping a copy of the original data does not exclude anonymity - at least from the view of the external „attacker“.

Considerations to Pseudonymity

Generation and Assignment

The generation of the pseudonyms and their assignment to the users is a crucial task, on which a big part of the protection of the real identity depends. If, for example, the user is able to choose his or her pseudonym, it may allow to draw conclusions about the real identity. Additionally, the date, time, or place of creation (or all of these), as well as the place where the real person usually stays may be revealed (cf. [79, p21]). All these things probably lower the protection of the real identity and, as a consequence, privacy. To ensure adequate privacy, one may decide to refuse user defined pseudonyms and only allow the use of automatically generated pseudonyms which fit all necessary criteria.

Misuse and Accountability

Two ways of preventing or at least limiting the consequences of misuse are the identity broker [79, p22] and funds attached to a pseudonym as deposit. The identity broker is a possibility to manage the relationship between pseudonym and real identity. When a pseudonym is created,

the identity broker stores information about the real identity. Under well defined conditions (like e.g. criminal prosecution or abuse), the real identity may get revealed (at least to a limited group of persons). The deposit is an easy way of preventing damage from a system: under the assumption that everyone wants the deposit back, one will probably behave in a way which does not exclude this. Anyway, if someone behaves in an undesired way, the deposit may be used to cover the expenses caused by the damage. Of course a hybrid system combining those two approaches is also possible (cf. [79, p24]). This provides a higher security to the providers of the system as well as affected partners, but on the other hand, it may lower the attractiveness (even for users who do not intent to abuse it) a lot. If you are about to build up a system (especially with a not well known organisation like for example, a start-up), it is probable that a broad majority will not be willing to pay a deposit or to give you information which makes them identifiable (which can be considered as really sensitive data) to you. In this case, a balance between attractiveness and security for the provider has to be found.

Requirements of Privacy Enhancing Systems

The most important requirements of privacy enhancing systems include that one must not trust the network operator nor a centralised node (this may be either a server, natural person, organisation, or any other kind of involved subject). Other important requirements are usability, performance (for example, no one would use a world wide web anonymiser that needs 5 minutes for every request), and reliability (cf. [36, p2ff]). A survey of the JAP Project [82] shows that a broad majority (more than 70%) of the users of the anonymiser are heavy users (more than three times per week), 44% are intensive users (more than 5 hours per day online), and 98% use it at least partly for business purposes (cf. [100]). Taking this into account shows how important the reliability - especially for commercial products one has to pay for - is.

Market

A survey of the JAP Project [82] showed, that only 60% of the users are willing to pay for this service and that only 10% are willing to pay more than 5 euro per month. The intensity of the usage of the service has however no significant influence on the willingness to pay. From the billing model, a flatrate with a monthly payment is the most preferred option for the heavy users (cf. [100]). This shows that it is possible to earn money with privacy services; however, the willingness to pay is limited. Therefore, the question if an economically viable operation is possible has to be investigated carefully. The broad majority (44%) of the requested content of the JAP Project [82] is related to entertainment (the biggest subcategory is sexual content with 33%), followed by services like route planning, and search engines (18%) (cf. [79, p5].

Data Protection in the EU and US

The approach for data protection in the US is completely different than in the EU. In the EU, data has to be considered as private unless there exists another regulation, which means that the usage of data is only allowed for specified purposes. However, in the US, data protection works just the other way round: one can do anything with data, except it is forbidden. Additionally, the review

and control of access to sensitive data by authorities like, for example, the FBI (Federal Bureau of Investigation) is not possible in the US, as access to the data is not documented (cf. [109]). All this raises a lot of questions related to misuse by authorities, the principles of data protection, how much and which kind of data may be used by the state for law enforcement and counter terrorism etc. As this would exceed the limits of this thesis, these questions will not be answered in detail but it seems like in this field further considerations are necessary. Important points to be considered may be the necessity and effectiveness, proportionality, and control/review of access (cf. [109]) as well as secondary use of data.

CHAPTER 3

Taxonomy

As mentioned in chapter Introduction at page 1, a taxonomy for Privacy Enhancing Technologies is required. As no appropriate taxonomy exists yet, a custom taxonomy is introduced in this chapter [51]. The evaluation of the investigated techniques (cf. page 57) is based on this taxonomy. However, this evaluation is enriched with suggestions for combinations of different technologies in order to make them more powerful and combine their pros (cf. chapter Synergy Effects of PET Groups at page 105). Note that an underlined term in the figures indicates a link between a generic and a more fine grained figure.

This taxonomy is based on two different viewpoints concerning privacy (cf. figure 3.1 at page 15):

- The *Technology* used to ensure privacy with a focus on general approaches than on specialised and specific techniques.
- The *Perspective* on privacy dealing with different *Views*/stakeholders that one can take into account regarding privacy.

Each of these aspects which is explained in more detail in a separate section is the base of a separate (sub)taxonomy. These *Views* on privacy may overlap in certain aspects. Note that these taxonomies were inspired by and are (partially) based on the taxonomies mentioned in chapter 1.

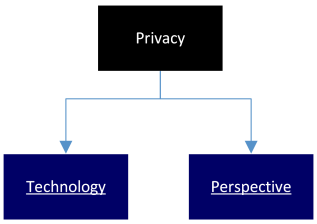


Figure 3.1: 2 Points of view on the taxonomy for Privacy Enhancing Technologies

3.1 Technology

The taxonomy aspect Technology (cf. figure 3.2 at page 16) considers privacy as the root of the taxonomy and takes various technological considerations on and characteristics of privacy into account. From a technological point of view, Privacy Enhancing Technologies can be classified into the following groups of attributes: *Protected Aspect*, *Means of Protection*, *Affected Data Type*, *Strength of Privacy*, and *Base of Security* (cf. figure 3.2 at page 16). Each of these groups is explained in more detail in a separate section.

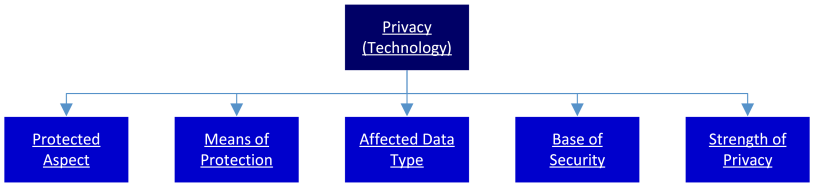


Figure 3.2: Taxonomy from the technological point of view

Protected Aspect

The *Protected Aspect* states what should be protected. This may be the *Content*, the *Identity*, or the *Behaviour* (cf. figures 3.3, 3.4). The *Protected Aspect* forms the base structure for the techniques presented in chapter Techniques at page 33.

Content

The *Content* is what communication or interaction is about. It may be any kind of *Content* as specified in the *Affected Data Type* criteria (cf. page 19).

- One part of the *Content* is the *Metadata* which may either be *added* afterwards (e.g. tags, keywords etc.) or is *generated* during the usual process (like e.g. communication partners, date, and time etc.). This distinction is important because usually *generated Metadata* is related to the process itself and for this reason, the possibility to include it in privacy protection exists. Adding *Metadata* like tags, however, cannot be prevented as this would mean to prevent the adversary from having control of *any* system (including his or her own ones), which is obviously not possible.

- The second part is the actual *Data* related to the process.

Identity

Figure 3.4 at page 18 shows the different instances of the *Protected Aspect Identity*: *Anonymity* and *Pseudonymity*. As they may be a bit difficult to distinguish sometimes, section Distinction between Anonymity and Pseudonymity at page 6 offers some considerations regarding this topic.

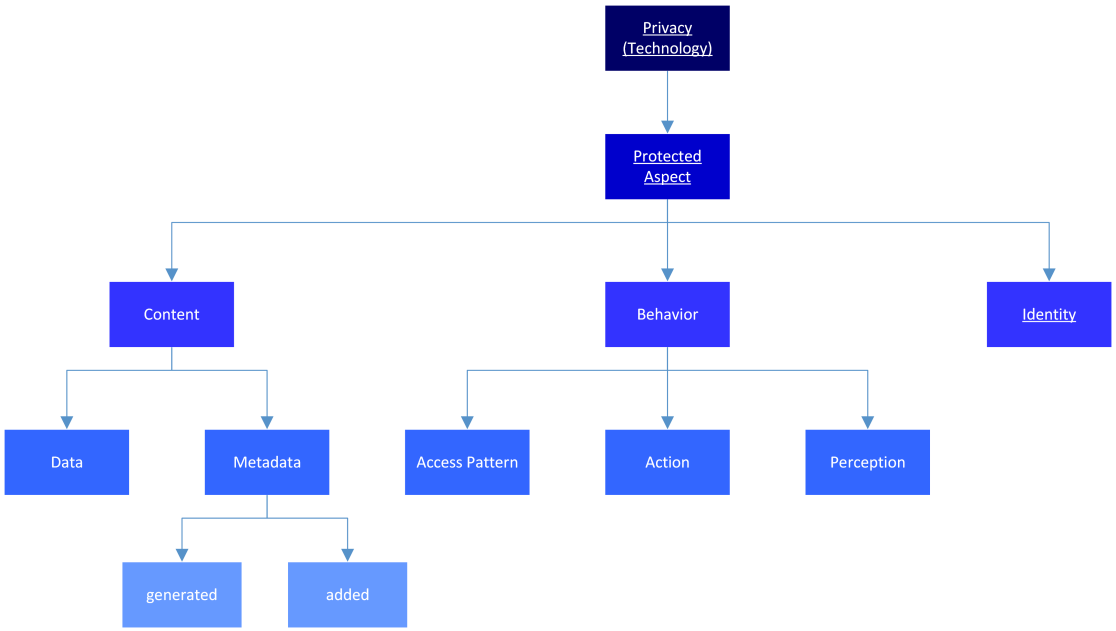


Figure 3.3: Taxonomy from the technological point of view: Protected Aspect

Anonymity and *Pseudonymity* are characterised by an important property: the *Directionality*. It states whether *Anonymity*/*Pseudonymity* is guaranteed only in one direction/for one party (*One-Sided*) or for both/all parties (*Two-Sided*).

In addition to *Anonymity*, *Pseudonymity* has two more important properties:

- The *Holder* defines by whom the Pseudonym is held. This may either be an *Individual*, so the Pseudonym is only used by one entity or person, or a *Group* that shares a common Pseudonym.

- The *Cardinality* states if the number of pseudonyms a *Holder* can create is *limited* (independent to which number) or if he or she has no restrictions and could create an *unlimited* number of pseudonyms. Depending on the number of allowed pseudonyms, *limited Cardinality* may very well have similar properties as *unlimited Cardinality*. As the maximum *Cardinality* is a property that may be configured for a specific system and therefore, can be changed as desired and adopted to the specific needs of the system/users, this variety is assumed to be from administrative nature. Therefore, it will not be considered in more detail within this taxonomy.

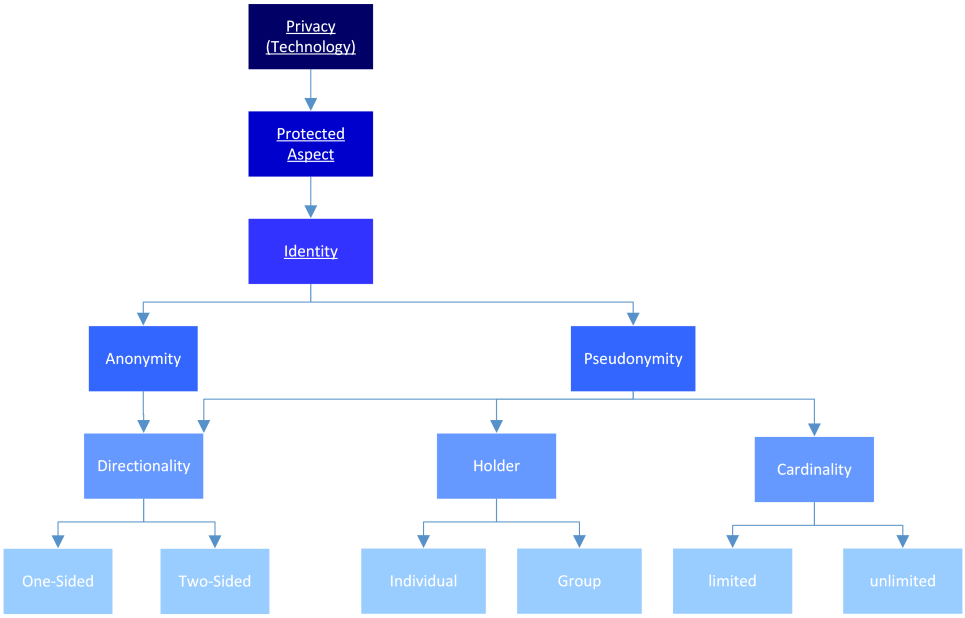


Figure 3.4: Taxonomy from the technological point of view: Identity

Behaviour

As the name already suggests, the *Protected Aspect Behaviour* covers all criteria related to how and when which parts of a service are used including reactions to interactions from the system. Although the *Behaviour* itself is usually not an identifier, it may be used in order to identify a user and to generate profiles, so it is an important aspect that needs to be protected from a privacy point of view. It can further be classified into:

- *Access Patterns* characterised, for example, by the timestamp (date and time), accessed records, or the location from which the access takes place (e.g. country, region, or range of IP addresses). When comparing this description to the one of (generated) *Metadata* (cf. page 16), it becomes obvious that some similarities and common attributes exist and that *Metadata* could be used in order to find out about *Access Patterns*.
- *Actions* taken by the user like, for example, the used functions, their parameters, and any other characteristic describing how the user acts.
- *Perception* refers to the physical perception (e.g. movement of eyes) as well as the mental perception (e.g. how some content is perceived).

Means of Protection

The *Means of Protection* states by which means privacy should be protected. This may either be *Deniability*, *Indistinguishability*, *Unlinkability*, or *Confidentiality* (cf. figure 3.5). An

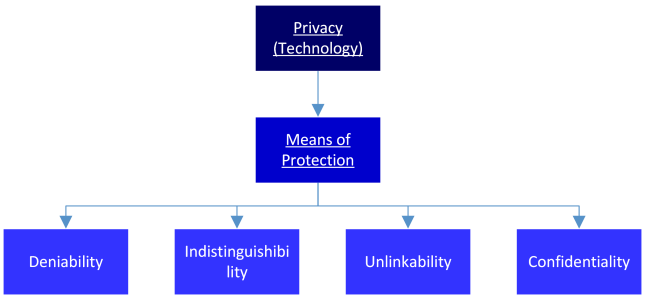


Figure 3.5: Taxonomy from the technological point of view: Means of Protection

important technical property of all elements within the *Means of Protection* attribute is the possibility of Enforced Reversal stating if it is possible to reverse the protection (i.e. reverse the privacy-enforcing operations) without the consent/*Cooperation* of the Data Owner or not using the technique as specified (cf. Reversibility at page 22).

Indistinguishability & Unlinkability

Indistinguishability and *Unlinkability* refer to similar concepts that are applied to the *Identity* (*Indistinguishability*) and *Behaviour* or *Content* (both *Unlinkability*). For more details cf. the respective definitions and explanations at page 8.

Deniability

Deniability (cf. definition at page 8) states that an entity is able to deny having anything to do with some *Data* or *Action*. It is the contrary of accountability. It may be important to privacy as it may allow actions which would usually be punished. On the other hand, one has to take care of misuse. For considerations on misuse cf. section Considerations to Anonymity - Misuse at page 10.

Confidentiality

Confidentiality (cf. definition at page 10) states that access is only possible for authorised persons.

Affected Data Type

The *Affected Data Type* states which kind of information (in a technical way) is treated. This may be: *Stored*, *Transmitted*, *Processed*, or *Combined* (cf. figure 3.6). *Stored* means that only stored data is considered. A typical example are online storage services. *Transmitted* data excludes storage and any kind of processing. It states that the counterpart just receives (and maybe forwards/(re)distributes) data. A typical example is a proxy server which receives requests and then just forwards the request and the respective reply without altering or processing it in any

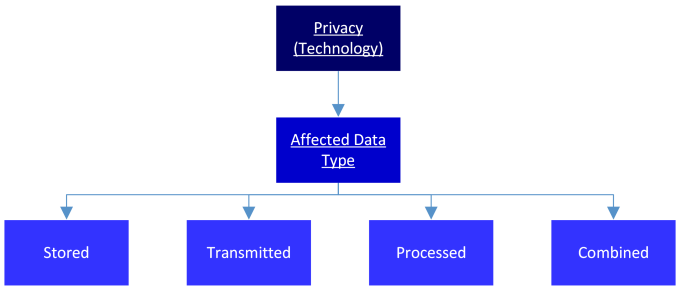


Figure 3.6: Taxonomy from the technological point of view: Affected Data Type

way nor (under the assumption that it does not have log files) storing it. *Processed* data states that the counterpart receives data, then conducts operations on it and optionally returns the result. It must not store data nor transmit it to any third party. After all operations are completed, no evidence exists that it has ever dealt with the data. A typical example are online calculators e.g. for exchange rates of currencies (again under the assumption that they do not keep log files). As in practice a lot of systems do not only deal with one of these data types, the *Affected Data Type Combined* is required to model combinations of the three elementary types.

Strength of Privacy

The *Strength of Privacy* defines how strongly and robustly privacy is protected. It has the following attributes: *Enforceability*, *Strength of Security*, *Trusted Third Party Participation*, and *Reversibility* (cf. figure 3.7). Each of them is described in more detail in a separate section below.

Enforceability

Enforceability is the ability of one party (this may either be the *Server* or the *Client*) to force the counterpart to participate in a protocol in the specified manner. One simple example is ordinary cryptography: if one party starts sending only encrypted messages, the counterpart has to participate in this new protocol (or one is not able to participate at all).

Privacy enhancement may be enforced (with increasing security) by: no one (*None*), the *Client*, the *Server*, or *Both* (*Client* and *Server*).

Note that *Enforceability*: *None* need not mean that the technique is not worth anything and that privacy may be compromised easily. Although cooperation may be required for a technique it may only improve another technique or approach which has higher *Enforceability* but e.g. lower performance. Furthermore, it does not mean that privacy is violated - it only states that both parties could prevent this specific technique from being used.

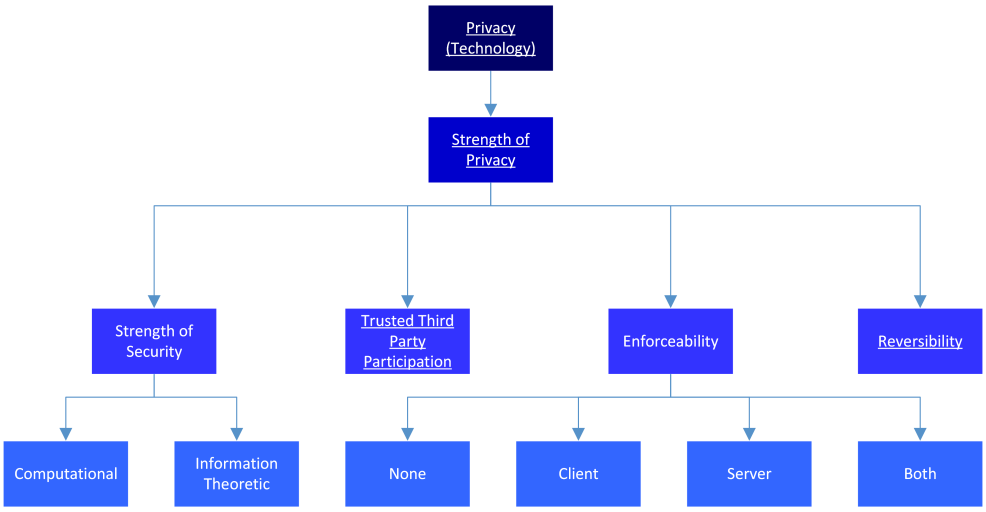


Figure 3.7: Taxonomy from the technological point of view: Strength of Privacy

Strength of Security

The *Strength of Security* defines the security model of the technology. It can be distinguished between:

- *Computational Security* (cf. definition at page 7) which is based on the assumption that an attacker has limited computing resources. Therefore, he or she is assumed of not being able to solve the computational hard problems privacy is based on within a reasonable amount of time. Cryptographic problems, are - in general - solvable by anyone. A predefined algorithm exists how the underlying problem could be solved which is usually known publicly (Kerckhoffs' Principle: „only secrecy of the key provides security" respectively Shannon's assumption that „the adversary knows the system", cf. [70] cited in [9]). The protection comes from the secret (key) that authorised persons know. This speeds up decryption a lot. If, for example, the used algorithm is RSA, the underlying problem is prime factorisation (cf. [15, p165-175]). This can in general be performed by any arbitrary person, but the algorithm is considered secure as long as the required resources to execute the algorithm without secret information are infeasibly high (which means that the algorithm cannot be broken by an attacker with polynomial or any other form of limited resources in reasonable time). However, the term limited resources is relative as the advances in computational power of computer systems have shown limitations of existing cryptographic algorithms. With increased computational power, simple attacks such as brute forcing become feasible, effectively circumventing the algorithms' hard problems. One exception that cannot be broken (even with unlimited resources) is, encryption based on a one-time pad that has equal length as the plaintext (cf. [92], [15, p99-102]) which is actually an information theoretically secure approach (see below).

- In contrast to this *Information Theoretic Security* (cf. definition at page 8) means that even with infinite computational power it is not possible to solve the problems, on which privacy is based. An example for this is the trivial solution for Private Information Retrieval, where one downloads the whole database in order to hide the accessed records. Of course, it is impossible to find out which records are accessed - even with infinite computer power.

Trusted Third Party Participation

The *Trusted Third Party Participation* deals with the question if/how a Trusted Third Party is involved in some technique (for an illustration cf. figure 3.8). It has three attributes:

- The *Frequency* states how often the TTP participates in the protocol. This may either be *Always*, only in *Specific Scenarios* (such as e.g. dissent or problems) or *Never*.

- The *Phase* states when the TTP (cf. page 7) participates in the protocol. This may either be in the *Setup Phase* only, during the whole *Operation Phase* or *None* in case that no Trusted Third Party is required.

- The *Background* is a property defining whether a TTP is involved for performing *Operations* or *Checks*, or if *No TTP participation* at all is desired in the protocol. In case of *Operations*, parts of one host's workload may be delegated to the TTP while in case of *Checks*, it is only verified if some specified condition is (not) satisfied. Note that *Operations* and *Checks* involving a TTP need not necessarily cause a decrease of privacy like in case when Homomorphic Encryption is used. The *Background* of the *Trusted Third Party Participation* should be seen as an agreement between Data Owner and service provider on how trust and checks are handled within the system.

Reversibility

The *Reversibility* states if and how much of the original information may be restored and if some kind of cooperation of the data owner is required (for an illustration cf. figure 3.9). It therefore consists of two sub-criteria:

- The *Degree* which may be
  - *Fully Reversible*: all information can be restored
  - *Partially Reversible*: parts of the information can be restored
  - *Not Reversible*: no information at all can be restored
  - *Deniable Reversible*: the Data Owner is able to reverse either the original or an alternative information. The latter may be used if the Data Owner is forced to reveal information. It is not possible to identify if the information revealed is the original or the alternative information.

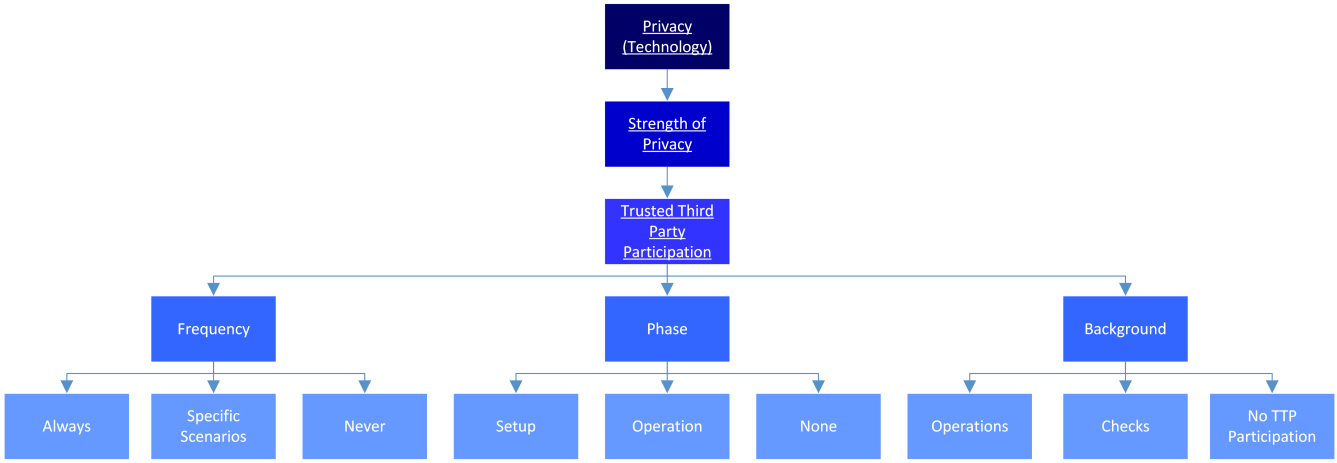


Figure 3.8: Taxonomy from the technological point of view: TTP Participation

These classifications are ordered by increasing security. Note that the strength of *Not Reversible* and *Deniable Reversible* may, depending on the arguments, be reversed. However, *Deniable Reversible* is assumed as the strongest *Degree of Reversibility* in this taxonomy, as it offers the most possibilities for the Data Owner.

- The *Cooperation* states if cooperation from the Data Owner is required to reverse the process of protection or not. However, the basic assumption within this thesis is that the technique is used as specified i.e. attacks or other action beyong the specification is neglected.

Base of Security

The *Base of Security* describes how privacy is guaranteed. Basically three different approaches exist (cf. figure 3.10):

- The approach based on *Statistics* intents to prevent e.g. the identification of a person by lowering the probability of correctly guessing him or her. An example for this is k-Anonymity where privacy is based on the principle that an identifying attribute is shared by at least k records. It therefore makes it statistically hard to unambiguously identify that person within this group (this is only possible with some probability). Usually techniques following this approach cannot be reversed (increasing the probability of guessing an attribute to 100% again). As they are not based on number-theoretic assumptions, they offer *Information Theoretic Security* as the *Strength of Privacy* but only under specific circumstances (i.e. without access to additional information): statistical methods/attacks

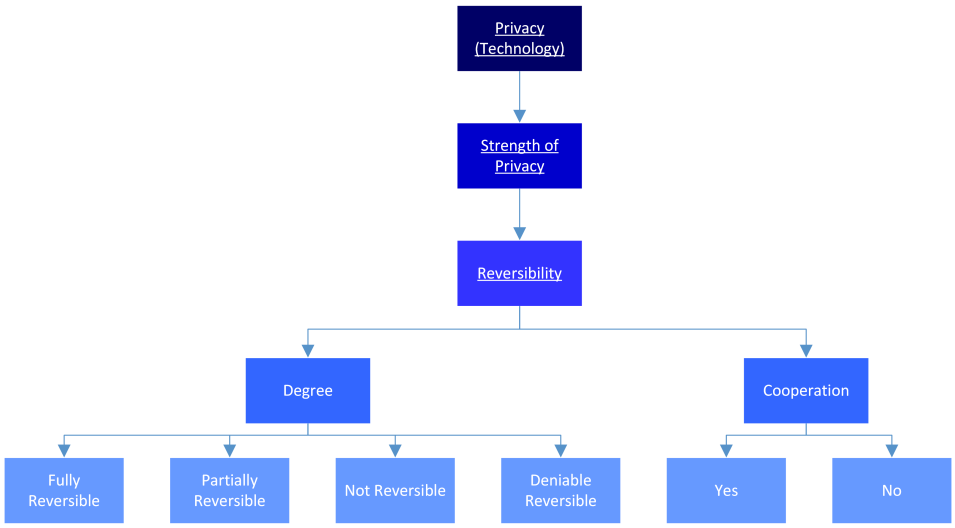


Figure 3.9: Taxonomy from the technological point of view: Reversibility

exist that could help to increase the probability again - in case of poor chosen data, even up to 100% (resulting in re-identification).

- In contrast to this, the second *Base of Security* builds on computational hardness of *Cryptography*. This may further be distinguished into *symmetric* encryption, where the same key is used for encryption and decryption, and *asymmetric* encryption, where different keys are used for encryption and decryption (cf. [96]), like it is the case in the RSA algorithm (cf. [87]). Furthermore, an even more fine grained distinction of cryptography would be possible - however, this is out of the scope of this thesis. *Cryptography* relies on the secrecy of certain information that helps in conducting some kind of computation. This implies that with unlimited computing resources one will usually be able to solve the problem and compromise privacy. Therefore, *Cryptography* as the *Base of Security* usually offers only *Computational Security* as the *Strength of Privacy*. With unlimited resources, cryptography can be broken or circumvented such that the problem of deciphering a ciphertext can be solved without the secret decryption key.

- The third *Base of Security* are *Mathematical Problems* like e.g. the Quadratic Residuosity Problem which may be used in cryptography (cf. e.g. [43]) as well as for Private Information Retrieval (cf. [129, p14ff]). In case of usage for Private Information Retrieval, this problem has no cryptographic background but still protects privacy as it obfuscates which bit of a database is of interest to the user (cf. [129, p14ff]). In contrast to *Statistics*, it is (with unlimited resources) possible to reverse this process, break the protection and therefore gain again 100% certainty. However, *Mathematical Problems* may have a broad variety of hardness. For example, the problem to calculate the sum of two numbers could be considered as a mathematical problem too - however, it is not a hard problem

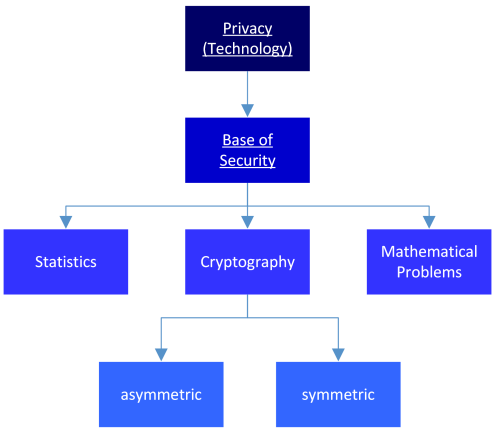


Figure 3.10: Taxonomy from the technological point of view: Base of Security

at least with the computational power and mathematical understanding of today. The discrete logarithm problem (cf. e.g. [69, 95], and [15, p177-192]) on the other hand, is a hard problem. In order to classify the hardness of these problems, the concept of computational complexity has been introduced. Popular complexity classes are P (solvable in polynomial time), NP (Nondeterministic Polynomial), and EXP (solvable in exponential time), cf. [77, p260-265].

The comparison above demonstrates that *Cryptography* and *Mathematical Problems* have a strong relation to each other. It is important to note that *Cryptography* is based often on *Mathematical Problems*. However, it is possible too to use *Cryptography* without *Mathematical Problems* (e.g. one-time pad) or to use *Mathematical Problems* for other approaches than *Cryptography* (e.g. the Quadratic Residuosity Problem for Private Information Retrieval). For an illustrative picture cf. figure 3.11. Most current *asymmetric* encryption schemes are based on *Mathematical Problems* while most *symmetric* ones are based on other concepts like S-boxes. For this reason, *Cryptography* (especially *asymmetric Cryptography*) and *Mathematical Problems* are often but not always correlated and are treated and evaluated independently. Note that although *symmetric Cryptography* could be based on *Mathematical Problems*, this is usually not the case (for current schemes).

Cryptography does not only consist of encryption techniques. For example, key exchange protocols like the Diffie Hellman Key Exchange Protocol (cf. [28]) which is based on the discrete logarithm problem and digital signature schemes are part of cryptography too.

3.2 Perspectives

The taxonomy aspect *Perspective* (cf. figure 3.12 at page 27) considers privacy from a perspective oriented approach as the root of the taxonomy and takes the different *Perspectives* on privacy

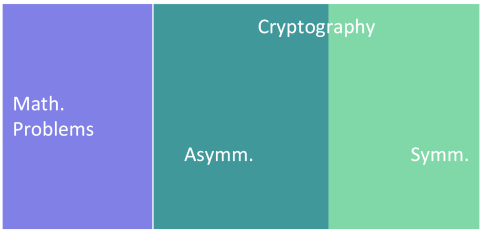


Figure 3.11: Relation of Mathematical Problems and Cryptography (symmetric and asymmetric separated by the white vertical line)

into account. All distinctions are based on the stakeholder. This may be either the Data Provider or the Data Consumer (user) which results in:

- The *Users View* which can further be differentiated into the *User Perspective* protecting his or her *Identity* and the *Profile Perspective* preventing the generation of profiles about the user. The *Profile Perspective* can further be distinguished into the *Content Perspective* protecting the actual *Content* and the *Access Perspective* protecting the privacy of the accessed records by hiding what the user is interested in. Note that the *User Perspective* and the *Data Perspective* may seem somehow similar. For a discussion and some considerations on this please, cf. section Similarities between the User Perspective and the Data Perspective at page 31. For the *Users View*, the adversary is another user or a Data Provider the user is communicating with. The user wants to protect himself or herself by hiding his or her *Identity* (*User Perspective*) or preventing the generation of profiles (*Profile Perspective*).

- The *Data Providers View* which does not get further divided into more *Perspectives* consists only of the *Data Perspective*. In this case, the adversary is the Data Consumer rather than the Data Provider. The Data Provider intents to protect sensitive information by hiding the user's *Identity* in order to be allowed to publish this sensitive data. Data Consumers (as well as other Data Providers) may, however, have an interest in reconstructing the Data Owner's *Identity* e.g. by combining information from various sources or performing statistical analysis. These Data Providers and Data Consumers, trying to reconstruct the *Identity*, are considered as the adversary for techniques of the *Data Perspective*.

In order to provide a better overview and allow more structured discussions, the *User Perspective* and the *Profile Perspective* are summed up to the *Users View*. The corresponding counterpart is the *Data Source View* which consists of the *Data Perspective*. For explanations and additional information on this topic, cf. section Two Views and Four Perspectives on Privacy at page 27.

Users View

The *Users View* consists of the *User Perspective* and the *Profile Perspective* which is composed of the *Content Perspective* and the *Access Perspective*.

User Perspective

The *User Perspective* intents to protect the user's *Identity* which is similar to the *Data Perspective*. These two approaches have similar ideas as a common base but are associated to a different stakeholder are assigned to different *Perspectives*. However, they are not identical similarities. For more details on the similarities and differences between these *Perspectives* cf. section Similarities between the User Perspective and the Data Perspective at page 31.

Profile Perspective

The second *Perspective* associated to the *Users View* is the *Profile Perspective* which protects the user from the generation of profiles. It consists of the *Content Perspective* (*Protected Aspect Content*) and the *Access Perspective*, taking the *Protected Aspect Behaviour* into account.

Data Source View

The only instance of the *Data Source View* is the *Data Perspective* which intents to protect the user's *Identity* e.g. in published data.

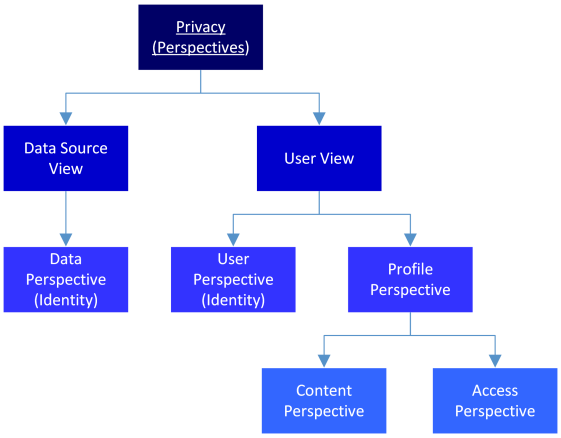


Figure 3.12: Taxonomy from the Perspective point of view

Two Views and Four Perspectives on Privacy

In order to provide a better overview and an easier understanding, the *Perspectives* are illustrated (cf. figures 3.13, 3.14, 3.15, 3.16). Note that the dotted lines illustrate what is visible to the server/user, and the solid lines illustrate the information flow.

In case of the *Access Perspective*, the Data Consumer requests data without the Data Provider knowing which data he or she is interested in. For this reason, the Data Provider receives a request for some data and processes it (however, this processing is a blackbox to it symbolised by the cloud symbol). The Data Consumer, however, knows exactly which data he or she is interested in (cf. figure 3.13 at page 29). An example for the *Access Perspective* is the trivial solution for Private Information Retrieval (cf. page 44): if the Data Consumer requests the whole database, although he or she is only interested in one record, the Data Provider has no possibility to find out which record actually was interesting for the Data Consumer.

In contrast to the *Access Perspective*, the *Content Perspective* has no intention of hiding which record the user is interested in. The Data Provider knows exactly which record it is expected to deliver - however, it has no idea what this record is about as the record is a blackbox to the Data Provider, which is indicated by the cloud symbol (cf. figure 3.14 at page 29). A typical example for the *Content Perspective* is encryption preventing unauthorised access to confidential data.

The *Data Perspective*, on the other hand, hides the relation between a record and the user this record is about. Therefore, the Data Consumer is able to request some specific record but he or she does not know who is affected by this record (often the Data Provider does not know it neither to enforce stronger privacy and security). This is illustrated in figure 3.15 at page 30. Techniques like k-Anonymity (cf. page 47), that allow *Anonymisation* of records, are able to protect privacy from the view of the *Data Perspective*.

Last but not least, the *User Perspective* hides from the Data Provider who is requesting data but not which data is requested. Therefore, the Data Consumers are separated from the Data Provider by a blackbox (again illustrated with the cloud symbol) so that the Data Provider finally knows what records it should deliver but neither who receives which record nor which Data Consumers are involved in the current requests at all (cf. figure 3.16 at page 30). A simple example for a technique from the *User Perspective* is a proxy server preventing that the counterpart knows about one's identity.

One may ask why the distinction into the *Views* is necessary when a further distinction into the *Perspectives* with a strong similarity is introduced. Although it would actually be possible to relinquish the *Views* and just use the four *Perspectives*, it seems to be more intuitive to have the stakeholder as a criterion too. Furthermore, it allows a quick categorisation (and selection) of Privacy Enhancing Technologies appropriate for a specific scenario and stakeholder as well as a more intuitive way to enable research for a combination of several approaches to deal with various different needs. For these reasons, this thesis sticks to the use of the two *Views* as well as the four *Perspectives*.

In this thesis, the *Data Source View* will usually be referred to as the *Data Perspective* while on the other hand, the term *Users View* is used to refer to both the *Users Perspective* and the *Profile Perspective*. Although this classification seems to be strict, techniques exist that cover both *Views* (*Users View* and *Data Source View*) or more than one of the four introduced *Perspectives* (*Data Perspective*, *User Perspective*, *Content Perspective*, and *Access Perspective*).

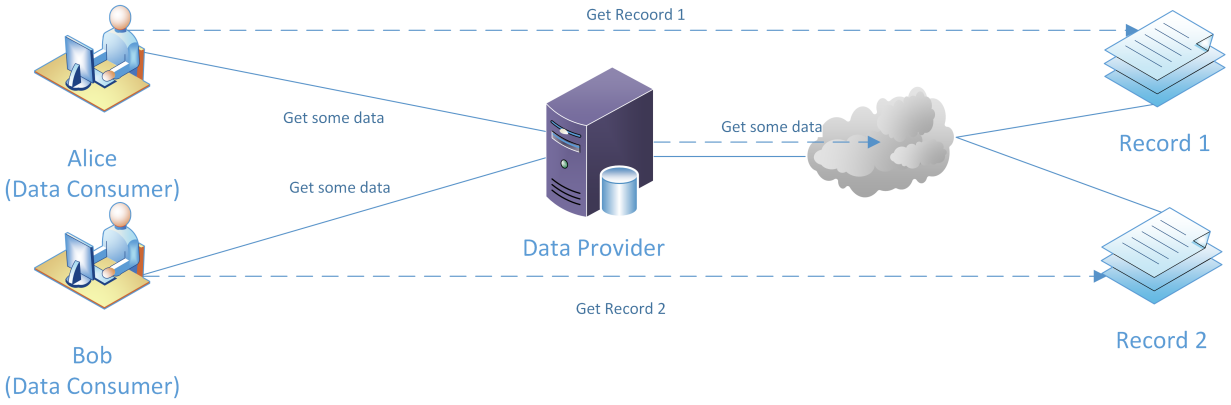


Figure 3.13: Access Perspective

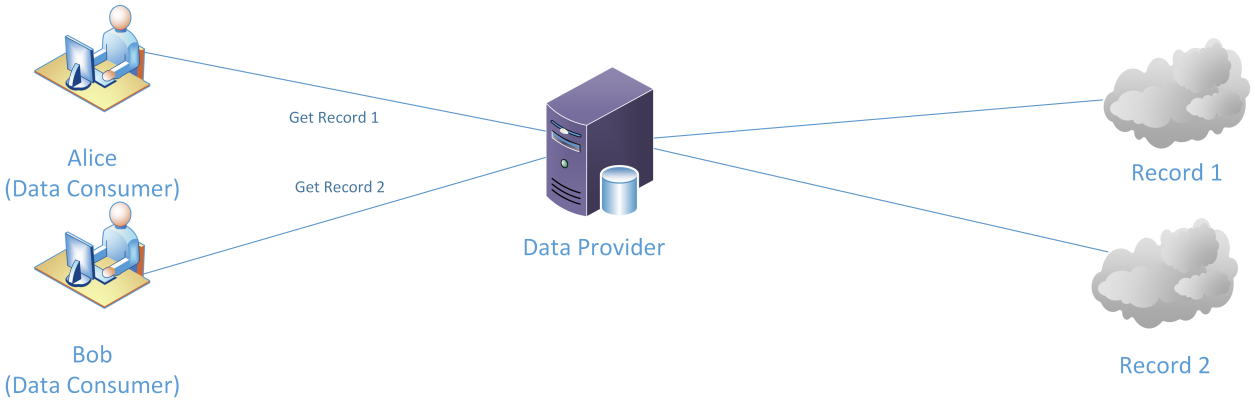


Figure 3.14: Content Perspective

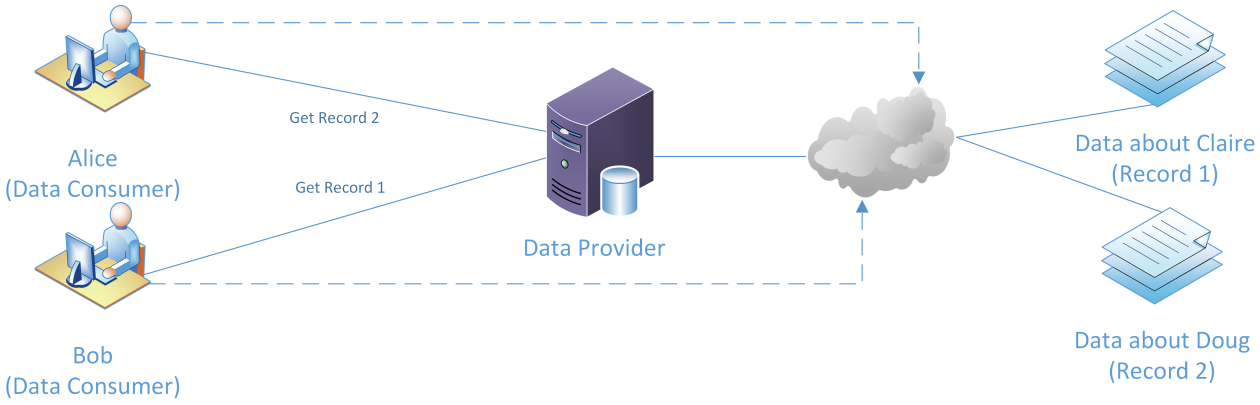


Figure 3.15: Data Perspective

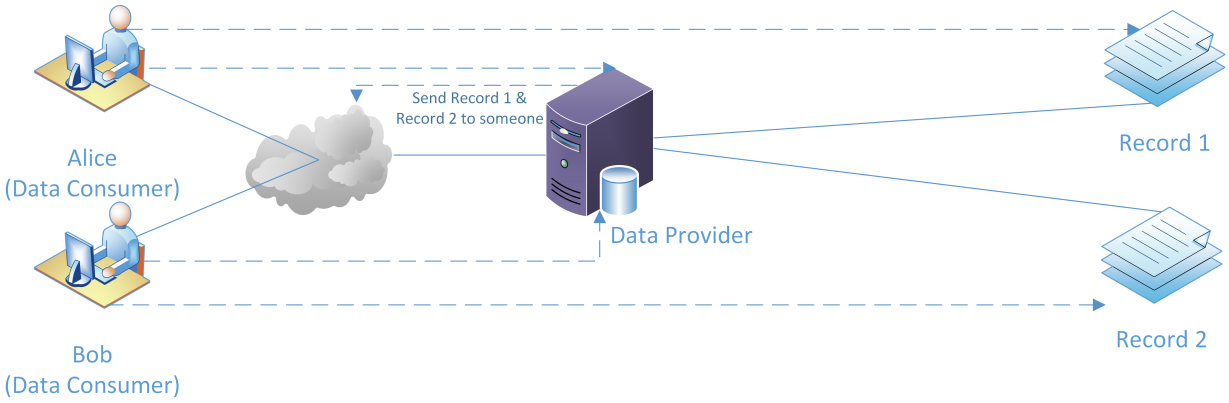


Figure 3.16: User Perspective

Similarities between the User Perspective and the Data Perspective

The *User Perspective* and the *Data Perspective* take care of similar affairs. Both want to protect the user's *Identity* and intend to prevent others from learning about it. Therefore, they basically tackle the same problems from a different points of view and within slightly different scenarios: the Data Provider in the *Data Perspective* may be seen as a merchant offering data. However, he or she is only allowed to offer this data under certain conditions like e.g. that it is not possible to draw one's conclusions about the person's *Identity* when taking a look at the provided data when and the Data Provider does not trust the Data Consumers. It is assumed that pooled data already exists. On the other hand, the *User Perspective* assumes that data is not pooled at the beginning but that data is collected and then may be assembled to profiles. Of course, data may be pooled too, but the difference is that this has to be done by the counterpart and is an ongoing process, whereas at the *Data Perspective* the pooled data already exists.

CHAPTER 4

Techniques

In this chapter Privacy Enhancing Technologies are presented. The majority of these techniques reflects recent developments starting from the year 2009. However, there exist some exceptions where older publications are considered because there exist important developments or the technique itself is important to demonstrate basic concepts or forms the base of other techniques.

4.1 Proxy Re-Encryption

A common problem with encrypting data is that one may decide to share it with another person. In traditional encryption schemes, one would decrypt it and then share it encrypted under a new key. The well studied problem of key exchange over an insecure channel is not very hard to solve anymore: either one uses algorithms for key exchange or uses Public Key Encryption. However, in recent time a tendency that more and more data is stored remotely instead of locally exists. This means that it has to be retrieved from a server, then re-encrypted, uploaded again to the server, and the recipient has to download it again so the data has to be transferred 3 times in order to share it. While this may be nasty but feasible with small files, this may cause problems if big files have to be shared (although in a local environment it may still be possible). Things get even worse if the server is not located within the same network, but data is transferred over the internet (e.g. using cloud services or servers without local connection) or even mobile devices are involved. To solve this kind of problem, Proxy Re-Encryption (PRE) has been invented. It builds on Public Key Encryption and enables a semi-trusted third party to transform a ciphertext (encrypted under some public key) into a ciphertext encrypted under some different public key without actually decrypting it (or in other words without learning anything about the plaintext). If, for example, Alice (*delegator*) decides to share some encrypted data with Bob (*delegatee*), she does not need to download it but advises the server (*proxy*) to re-encrypt it so that Bob can decrypt it with his own Private Key. The plaintext never gets exposed to the server. Furthermore, no need for sharing any keys exists (neither her own Secret Key nor any symmetric key which would have to be managed). An application for PRE is sharing any kind of data without decrypting it as e.g. when forwarding an encrypted email.

The two main fields of security in PRE tackle Message Security and Collusion Resistance. Message Security defines that only authorised users can know the plaintext. It can be classified into the security levels Chosen Plaintext Secure (CPA-secure), Replayable Chosen Ciphertext Secure (RCCA-secure) and Chosen Ciphertext Secure (CCA-secure) where CPA-secure provides the weakest and CCA-secure the strongest security. Collusion Resistance defines that proxy and delegatee cooperate in order to - as supposed - calculate the Re-Encryption Key (weak Private Key), but they are not able to calculate the delegator's Private Key. The delegatee then uses the Re-Encryption Key to decrypt the ciphertext (cf. [93, p1]). In general there exist some fundamental differences in implementing PRE. It may be uni-/bidirectional and it may be transferable/non-transferable. Unidirectional PRE is considered to be stronger as it could be used twice so that bidirectional PRE is created. Non-transferable states that a delegatee is not able to re-delegate his or her decryption rights. However, this may be hard to prevent as the delegatee always may redirect the decrypted plaintext (cf. [49, p1]).

In recent time a lot of new possibilities to deal with weaknesses when applying PRE have been developed, improving existing schemes, or adding new functionality. Some of the progress refers to different security models whereas other parts offer new/enhanced functionality. However, this enhanced functionality may require additional restrictions. What sounds a bit strange at the beginning may actually be useful. For example, one may want to limit the messages that may be re-encrypted using Conditional PRE (C-PRE) (cf. [117]) which was improved in [118] or Type-based PRE (cf. [103]). C-PRE uses conditions to restrict the re-encryption rights whereas in Type-based PRE there are types associated to data and the decryption capabilities are restricted to specific types. K-times PRE (cf. [116]) has been introduced to bound the number of times a ciphertext can be re-encrypted. If the proxy would re-encrypt data too often, the proxy's Private Key would get revealed (however, this implies that the proxy has got its own Private Key and Public Key). Another restriction is unidirectional PRE which deals with situations where trust relations are not symmetric (e.g. in a hierarchy) (cf. [64]). In [93] this got extended to a Multi-Use Unidirectional PRE (stating that the number of re-encryptions is not limited as in K-times PRE). If the number of re-encryptions should be bounded or unbounded depends of course on the aim of the system. Therefore, a flexible way to switch between these schemes would be desirable.

The next four developments are all based on C-PRE and the first and fourth again enforce (desired) restrictions. However, the third approach also increases functionality. The first one is Attribute-based C-PRE and enables the delegator a more fine grained control of the decryption rights by „(...) setting conditions in the form of access structure and attribute set.“, [135, p1]. It is based on key-policy attribute-based encryption and token-based encryption (cf. [135]). The second development is Identity-based C-PRE (cf. [94]) and the main difference to C-PRE is that not a Public Key but the recipient's identity is used to encrypt data. This seems to be an interesting approach when C-PRE should be introduced with avoiding big structural changes for the end users (such as e.g. in Email Systems). The third approach combines the ideas of Searchable Encryption (cf. page 36) and C-PRE into one scheme (cf. [136]). The fourth development deals with Anonymous C-PRE stating that no information about the condition has to be revealed to the proxy (cf. [35]).

Another development regards Non-Transferability of the ciphertexts. The delegator's Private

Key is needed to construct the Re-encryption Key which intents to solve two Problems in PRE: the PKG despotism problem and the key escrow problem. Non-Transferability is insured by revealing one's secrets instead of adding tracing information in order to find out who actually has distributed data in a non legitimate way (cf. [49]). Another important advance is the introduction of PRE with an invisible proxy (satisfying CCA-Security). The invisible proxy property refers to the fact that the „(...) adversary cannot distinguish a regular ciphertext from a re-encryption of a type-based ciphertext, even if the adversary can obtain all the keys.“, [55, p1]. As stated in the definition above, the proposed scheme builds on Type-based PRE (cf. [55]). Furthermore, an efficient Identity-base PRE scheme which achieves master secret security, was proposed (cf. [115]).

Another interesting proposal is Accountable Anonymity (cf. [123]) which protects the sender's privacy as long the sender behaves well. However, it enables a victim, the Registration Database and the Key Generator to reveal the senders identity if they cooperate. In theory this sounds like an interesting approach that intents to deal with the problem of misuse in privacy enhancing systems. However, there are some weak points in the system. According to the definitions of *Anonymity* (cf. page 6) and *Pseudonymity* (cf. page 6) the system does not provide *Anonymity* but *Pseudonymity*. In this case the pseudonym is the whole chain of information that is used to reveal the identity as described in [123, p113]. Actually it is a system similar to the IP addresses: as long as no complaint occurs, one uses both of them as a pseudonym and on order of a court the mapping between Pseudonym and real identity gets revealed. Furthermore, it only enables accountability when the receiver of some message is a victim. If the receiver is not a victim but a perpetrator too, the receiver has no interest to report the message and therefore the system will not work. This may be the case, for example, with illegal filesharing as well as reputational damage by rumours, consuming illegal services, and arrangements for criminal activities and terrorism. As argued above, lots of illegal activities could be conducted without fear of getting caught. Therefore, the idea was interesting, but it seems like Accountable Anonymity is still not achieved with this system. It is arguably impossible to achieve it with reference to the definition of *Anonymity*. However, this approach offers an interesting possibility to fight some specific type of misuse like blackmailing or spamming.

4.2 Homomorphic Encryption

With ordinary encryption one faces the problem that data has to be decrypted before operations can be conducted on it (and naturally has to be re-encrypted after the operation). However, this may be impossible, for example, when data is stored on a third party server which must not be able to decrypt stored data. Homomorphic Encryption deals with this problem as it allows operations on data without the need to decrypt it (the result of the operation is the same as if one would decrypt the value first and re-encrypt it afterwards). This implies that an untrusted server in a data centre could perform these operations as it is not able to learn anything about the encrypted values.

In former times, Homomorphic Encryption suffered from a variety of drackbacks like e.g. that it was inefficient or could only perform additions or multiplications but not both of them (cf. [107, p5f]). However, in 2009, a thesis was published that presented the first approach

for Fully Homomorphic Encryption that could be conducted efficiently on arbitrary operations which led to compact ciphertext [42] by Craig Gentry. Although the constructions based on Gentry's Blueprint suffered from several weaknesses like limited efficiency (concerning performance as well as the length of keys and ciphertext) and non standard cryptographic assumptions, this was a big step for further research. Although there has been made progress in improving the original scheme (cf. [107]), it seems like further research is necessary in order to provide the possibility for usage under real world conditions. As a simplification Somewhat Fully Homomorphic Encryption (Somewhat FHE) has been introduced. It has the weakness that (especially) the number of multiplications that may be conducted without fostering noise, is limited. This noise may lead to errors preventing decryption.

Although it seems that further research is necessary, recent developments seem to be promising for the practical application of Fully Homomorphic Encryption. These developments include the extension from bitwise encryption as proposed in Gentry's Blueprint (cf. [42]) to integers (cf. [108]) which was further improved with respect to its complexity [84, 125]). Furthermore, there has been developed an extension to messages of arbitrary length that is based on matrix calculations where the results are transformed into a vector, which implies easy calculations (cf. [56]). Last but not least, an algorithm has been developed that enables privacy preserving outsourcing of calculations. It allows to protect the whole equation (including exponents and not only coefficients) that is the base for the calculations while still enabling the calculation of the result. Naturally, this implies that the involved server is not able to learn anything from the calculations (neither the form of the equation nor the values and obviously not the result) so it is a non-interactive way of outsourcing calculations (cf. [20]).

Under the assumption that Homomorphic Encryption works in a practical way, rights management may be needed so that only authorised persons can perform operations on encrypted data e.g. in order to prevent problems with integrity/digital signatures or (illegal) manipulation of encrypted data. Another example are restrictions on which kind of operation may be conducted by whom. In other scenarios, it may be interesting to retrieve the result of a computation (decrypted) while it is not possible to get any information about the base of the calculations. Last but not least, a big improvement may be to base Fully Homomorphic Encryption not on lattice problems but number theoretical problems like factorisation or logarithmic functions (cf. [107, p13f]). Anyway, there have already been promising developments in Homomorphic Encryption but besides technical issues (like e.g. complexity, key size, and the maximum number of multiplications) one crucial issue seems to be the involved complexity for understanding the techniques and algorithms as well as implementing them.

4.3 Searchable Encryption

Searchable Encryption deals with the problem that with ordinary encryption it is not possible to perform search requests on encrypted data. One exception that may be realised easily are search queries with an exact match. This may be done by encrypting the keyword that one wants to look up and then looking for the encrypted keyword. However, strong limitations are the consequence: first it may only be applied to deterministic encryption schemes where the same plaintext is always encrypted to the same ciphertext (when using the same key). Next it is only

possible if data gets encrypted word by word (from the moment on, where 2 words are encrypted together this technique will not work anymore) meaning that fulltext search is excluded. Last but not least, a lot of well developed search techniques in the plaintext domain cannot be performed, such as wildcard search, error tolerant search, or multi-keyword search. In order to deal with these problems, a lot of techniques have been developed. As this thesis cannot deal with all of them, the most important recent developments are summed up below.

Recent developments can be categorised in two different groups: enhanced functionality/usability and improved performance. Below the most important developments according to those two groups are presented:

Enhanced functionality/usability

In this group many advances were made leading to important developments. Although the performance is a critical point, it is scalable (at least to a certain amount). Therefore, usability and functionality are probably the most important things (which cannot be modified by the used infrastructure). Additionally, increased functionality is often related to increased performance too, as it does not need to be implemented in an ineffective and complicated way. One example for this suboptimal implementation is the trivial solution to the fulltext search problem: one has „just“ to retrieve the whole set of documents, decrypt them on the client and then perform fulltext search over the plaintext. However, it is obvious that this is inefficient (bandwidth, computation effort, local storage and time) nor does it make sense to outsource documents and then retrieve all of them with every search request. Another example is the problem that one faces if one of the formerly authorised users get revoked (cf. [127, p1]). In case of symmetric re-encryption, one had to decrypt the whole database in order to prevent unauthorised access (especially if one does not consider an access control system as sufficient for dealing with this task).

Fortunately, there exist solutions for these problems like e.g. [127] which allow Multi User Handling for Searchable Encryption in this case. In this paper, a Searchable Encryption Scheme was developed that deals with accountability (including unforgeablity of the queries) as well as user management (adding/revoking users). A drawback of this approach is that only one user has got write access to the database (cf. [126]). Anyway, those approaches need to be combined in order to meet all needs of outsourced databases. Even if they get combined, traditional (unencrypted) data has got more sophisticated mechanisms for access control. Nevertheless, these approaches are big steps towards the practical relevance of Searchable Encryption. Additionally, the Separation of Duty (restriction of a user's query abilities) is proposed as an extension of the multi-user Searchable Encryption which may be one of those more sophisticated approaches (cf. [127, p270]).

Furthermore, a number of more useful functions for Searchable Encryption are suggested: Conjunctive Keyword Search which combines more than one keyword into a search query (cf. [90], [127, p270f]), its extension the Phrase Search (where a match has to contain the whole phrase rather than only all of the keywords, cf. [104]), various approaches for Fuzzy Search (cf. [21, 53], [127, p271]), or Error Tolerance which are based on (non-cryptographic) Hash Functions and Bloom Filters (cf. [14]). Another interesting approach (cf. [113]) works similar to a search engine by not only enabling the user to search within encrypted data but also rank the

results. This approach builds on Order Preserving Symmetric Encryption (OPSE) and intents to not only find documents containing some keyword but also taking the keywords' relevance (or in other words frequency) into account. Another big pro of this approach is that it is able to deal with updates without rebuilding the whole index. The possibility of authenticating the (ranking of the) search results by using Hash Chains is also an interesting extension which prevents the loss or wrong ordering of results.

However, some of the listed techniques may need further development. For example, the approach introduced by [53] requires the participation of a third party and also obfuscates the index by padding all entries to the same size as the longest entry. This is done in order to prevent the server from gaining information that may be used in form of an dictionary attack. It is obvious that this may cause incredible overhead in the index size.

All these developments (namely Multi User Support, Separation of Duty, Multiple Keywords, Fuzzy Search/Error Tolerance, and Ranked Search) offer a lot of possibilities for search within encrypted data. One big drawback is that with most techniques only boolean combinations of several keywords are possible while more complex actions are still not possible. Anyway, if all the mentioned approaches could be combined into one encryption scheme, this would support operations which are beyond the „usual“ (e.g. Fuzzy Search/Error Tolerance is not offered in a lot of applications) or at least at the same scope of search in unencrypted data. However, a crucial point beside combining the approaches is of course the performance.

Improved performance

A variety of possibilities to improve the performance of Searchable Encryption were developed. Some of them tackle improvements in means of the storage overhead (cf. [90]), others intent to reduce computation/communication effort (cf. [29]) or improve the performance of existing extensions like e.g. [21] where incremental updates are developed in order to prevent rebuilding the whole index structure when a new file is added. However, again further improvements seem to be necessary in order to compete with operations on unencrypted data in a feasible way.

4.4 Order Preserving Encryption

Order Preserving Encryption (OPE) is an operation that deals with encrypting values while still preserving their order. This may also be seen as Homomorphic Encryption (in this case the operation that is conducted is ordering, cf. [80]). Of course this weakens the security of the encryption as an adversary is able to deduce the order of the values which already may be critical (like e.g. in databases containing information about the salary). However, it allows a lot of operations such as ordering, range queries (by only encrypting the upper and lower bound and then querying for these values), and min/max queries to be conducted at the same performance as in case of unencrypted data while still granting more privacy. Additionally, only the encryption but not the underlying application/database system itself has to be changed. Therefore, it is important to be aware of the fact that some information gets revealed by OPE and that it is

really important to protect other attributes well (especially the Explicit- and Quasi Identifiers). Anyway, if this type of queries is needed, OPE is still more preferable to no encryption at all.

The aim of OPE is to pretend randomness for the values but also for the distance between them (cf. [68]). As for OPE strict monotonic increasing functions are needed, one faces the risk of overflows resulting in wrong order as well as loss of data (overflows would result in inability of decryption). The selection of the key (respectively encryption function) is crucial. One straight forward approach is to take the logarithm of this function, but this may result in floating point numbers, where one faces the problem of rounding errors (cf. [68]).

There are also several publications dealing with the security of OPE. In [68] two approaches for analysing the encryption function's quality are proposed which are based on constructing a sequence to be analysed as a signal and the difference between an attackers approximation and the real function respectively. [121] propose a method to analyse OPE Security by calculating the average min-entropy.

In [10], the first proved secure OPE was introduced. However, in [121] it was proved that the theoretical approach of looking for an OPE Object that is indistinguishable from the ideal OPE Object may not meet the expected security properties. On the other hand, there also exists a suggestion for increased performance of the scheme proposed in [10] which builds on the use of probabilistic middle range gap instead of the Euclidean middle range gap (cf. [131]). Later mutable Order Preserving Encoding (mOPE) was proposed which really achieves optimal security and a strongly improved performance. The approach in general is simple: the values are ordered and then mapped to integer numbers in increasing order. Therefore, the order is the same for every plaintext combination and the distance between the encrypted values is always 1. In order to enable quick decryption this approach uses a B* Tree. It is called mutable OPE because if there are inserted new values, the data structure is able to change (mutate) in order to keep security at an optimal level. Additionally, it offers, besides the popular honest but curious scenario, also a scenario for dealing with malicious servers that try to retrieve more information. An adoption of this algorithm is also able to deal with this attack, and even this implementation is a lot faster than the previous ones (cf. [80]).

Another recently presented approach (cf. [132]) deals with additive OPE. However, it is not clear how effective it is because if additions on encrypted data are needed, then it is probable that other operations are also needed, which directly leads to Homomorphic Encryption (cf. page 35). On the other hand, one main scenario mentioned in the paper is the ranking of documents, but again another technique seems to be more appropriate for dealing with this: Searchable Encryption (cf. page 36).

Last but not least, a Multi User approach for OPE was developed that also enables strong security measures where even the cooperation of the database server and a key agent will reveal few information. The two suggested approaches are digest based OPE (DOPE) and its extension Oblivious Encryption-DOPE (OE-DOPE) (cf. [122]). As the article promises applicability with every OPE algorithm, a combination of the approaches described in [80] and [122] seems to be interesting and promising.

4.5 Deniable Encryption

In ordinary encryption an encrypted text can only be decrypted to one plaintext (otherwise decryption would not be possible). This property is exploited, for example, with the digital signature in order to ensure that a message is sent by some specific person and has not been faked nor altered. However, for some scenarios this may be an undesirable circumstance like, for example, when the existence of encrypted data is known and the owner can be forced to decrypt it. Note that this does not only refer to law enforcement in a democracy but also law enforcement by a dictator. In this scenario, Plausibly Deniable Encryption (PDE) can be used to decrypt the ciphertext to another (innocuous) plaintext (cf. [8, p125f]). Sometimes just erasing data may be sufficient, but in some settings (as observed transmissions) data cannot be deleted or must not be deleted (e.g. for legal reasons) (cf. [76, p529f]).

There exist several properties by which PDE can be classified. First it can be distinguished by who is able to deny the plaintext (namely sender, receiver, or both) resulting in Sender-/Receiver-/Bi-Deniability. However, the purpose for Bi-Deniability is not as broad as it seems: in order to avoid inconsistencies, sender and receiver have to agree on a common fake plaintext that may be revealed instead of the real plaintext. Under the assumption that they need a secure channel, the question raises why they do not exchange the real plaintext over this channel instead of the fake plaintext (in this case there would be no need for deniability) (cf. [8, p126]). However, they may invent the fake plaintext for several transmissions in advance. In an extreme example, they may invent enough plausible fake plaintexts for the transfers of several years and just have to exchange them once. Therefore, the risk that exactly this one transmission is observed is low in comparison to the risk that further transmissions may be observed. Furthermore, another big risk gets eliminated: subsequent examination.

Other important properties of PDE are the way how the ciphertext is produced and interactivity. The ciphertext may be produced by the same algorithm as in honest encryption (fully deniable) or by a fake algorithm with the possibility to decrypt it to the real or a fake plaintext (multi distributional). Multi distributional PDEs have got the drawback that they can be revealed by checking if it is possible to generate the observed ciphertext with the fake plaintext or not (which is not possible) (cf. [8, p126]). Therefore, fully deniable PDEs are highly desirable, but full deniability is only possible in an interactive way or only for the sender (cf. [8, p129]). Interactivity depends on whether the system is based on Public Key Encryption (non-interactive) or involves communication (interactive) (cf. [8, p127]).

In recent time there have been several publications related to PDE, but some of them seem to be already outdated due to new developments. For example, [52] offer only basic functionality which has been improved and extended (e.g. by [76]). Others like [112] offer a scheme based on alternative mathematical assumptions (in this case a lattice based algorithm). An important development is [76] which offered multi-distributional bi-deniability for the first time. Although fully bi-deniability implies sender deniability as well as receiver deniability, this is not possible for multi-distributional schemes. Therefore, the bi-deniability may be considered as sufficient to construct any combination of sender and/or receiver deniability in the fully deniability setting. However, for the multi-distributional setting bi-deniability only implies sender deniability but *not* receiver deniability (cf. [8, p133] and [76, p533f]). [8] propose a way to measure the security

of PDE. According to this proposal, the bi-deniable scheme proposed by [76] achieves negligible security stating that the probability of detecting the deniability is negligible (and therefore stronger security than polynomial security). It shows that negligible security is possible for all schemes that are not fully deniable and non-interactive (cf. [8, p129]). [41] proposed the first deniable information hiding encryption that is secure under the chosen ciphertext attack. A proposal for the first fully deniable non-interactive, sender deniable encryption scheme (cf. [33]) was published as well but it was successfully attacked by Chris Peikert and Brent Waters.

As mobile devices become increasingly popular and have got different requirements for PDE, this topic needs to be investigated besides the „normal“ PDE research. One initial article dealing with it is [98] which transfers knowledge about PDE and data leakage into a mobile environment (treating e.g. characteristics of mobile OS/filesystems, storage media etc.). It also evaluates an implementation for Android, but there are still other mobile OS (like e.g. iOS) left. The approach for the hidden volume within another encrypted volume is interesting: the offset of the hidden volume is calculated dependently on the chosen password. This means that it does not need to be stored at any particular place, which makes it more difficult to prove its existence. On the other hand, it increases the probability of discovering the hidden volume (if the usage of this algorithm is known) as it is known that the hidden volume is located at the end of the encrypted volume and consumes between 25% and 50% of the overall volume size. Furthermore, it limits the flexibility with respect to the size of the hidden volume. An alternative approach for enhancing the flexibility in volume size is to hash the password several times in order to achieve different volume sizes. Although this is an easy way to increase flexibility, it does not solve the basic problems.

4.6 Oblivious Transfer

In Oblivious Transfer (OT) in its most basic version, the receiver initialises the transfer of exactly one out of two possible bits of his or her choice without the sender knowing which bit has actually been transferred. OT is used for secure computing without trusted majority. Two possibilities to construct OT are Homomorphic Encryption (cf. page 35) and enhanced trapdoor permutations (cf. [65, p520f]). More generalised variants of OT are 1-out-of-n (based on k-Anonymity, cf. page 47) and k-out-of-n OT where k items are retrieved out of a pool of n items (similar to the idea of (X,Y)-Privacy, at page 47) (cf. [73, p1]). [73] propose 2 interesting OT protocols which try to prevent the retrieval of more than the specified k items: Protocol II intents to daunt unfaithful receivers with a high probability of getting caught whereas Protocol III prevents the retrieval of any item in case of unfaithful behaviour. Protocol I has got less interesting properties but is used as a base to construct the other two protocols.

Priced Oblivious Transfer (POT) deals with buyer-seller relations related to digital goods (such as commercial downloads of software or movies) where it has to be kept secret what actually was bought. However, a big weakness of POT is that fairness is usually not guaranteed, which means that buyer and seller could cheat on each other. The seller could refuse to deliver the goods after receiving the payment. On the other hand, the buyer could accuse the seller spuriously of just this behaviour and it cannot be proved that this is wrong. With the introduction of Optimistic Fair POT these problems are solved. This is achieved by an adjudicator who gets

involved only in case of disagreement (therefore, the protocol is optimistic) and clarifies the situation in case of problems (even in this case neither the vendor nor the adjudicator know about what has been purchased). Regardless of this, the user's privacy still is assured. Furthermore, it is possible to extend every secure POT to a Optimistic Fair POT (cf. [86]).

Another approach that has got interesting features for e-commerce purposes is [17] which treats Unlinkable POT with the possibility to recharge the digital wallet. It extends POT so that not only the retrieved good (*Access Perspective*) but also the price and the user's identity (*User Perspective*) are kept secret, and even different transactions cannot be linked to a user. Therefore, real *Anonymity* and not only *Pseudonymity* is provided (for the distinction between these paradigms cf. page 6). Furthermore, two from the economical point of view important features are added: overdrawing the account is prevented without revealing the actual account balance and the digital wallet can be recharged. The latter prevents problems when the account balance is too low to buy goods anymore: no new account has to be created and the remaining money does not get lost.

In contrast to the scenario above where basically everyone (who pays) is able to access data/goods, there may exist scenarios where this behaviour is not desired and the access should be restricted to authorised users while still granting user privacy. This leads to the deployment of Access Control to OT (Access Control OT, AC-OT). In AC-OT it is required that a (possibly outsourced) server learns nothing about the credential nor the retrieved record, which is why AC-OT ensures Access Privacy (retrieved record) as well as User Privacy (credential). In the following the results of three interesting recent developments in the field of AC-OT are summed up: The first one provides strong privacy by hiding most of the sensitive data from the database as well as from the user (cf. [18]). The second one allows disjunction without the need to duplicate records (cf. [134]). The third one deals with Threshold AC-OT where a specified number of attributes possessed by a user has to be exceeded to conform the access policy (cf. [124]).

The first AC-OT scheme provides a powerful attempt to enforce privacy which is based on Anonymous Credentials (cf. page 53). It hides sensitive information from the users as well as from a (possibly outsourced) database: The database does not know anything about the access policy and the result of the request (allowed or refused). The user does not retrieve any information about the access policy nor does the user get any information about the database (structure) or access policy except if his or her request was rejected or not. The user's access is limited to 1 record per request and of course the user can only access records for which the user satisfies the specified constraints (roles, attributes or rights). Last but not least, user revocation is possible (cf. [18]).

The second AC-OT scheme provides the possibility of disjunctions in the access control policy without duplicating signatures in OT that get incorporated into ciphertext-policy attribute-based-encryption. Due to the omission of duplication, it saves resources for storage, computation (e.g. indexing, search), and enhances data quality by preventing inconsistencies. Depending on the underlying cryptography, this may furthermore result in decreased security as attacks on the encryption and the identification of duplicated records may be facilitated (cf. [134]). This scheme seems of high importance for practical realisations as disjunctions are part of most access policies e.g. in form of various groups that may access a resource (however, in most cases a disjunction instead of an intersection of these groups is sufficient to obtain access to a resource).

The third AC-OT scheme provides the possibility to define a threshold for the number of attributes which a user has to possess, in order to get access. It is based on the combination of fuzzy identity-based encryption and a credential signature scheme. [18] is a special case of this construction for the case that the threshold is set to the total number of attributes. This scheme provides an interesting base to construct more complex access control mechanisms expressed by Linear Secret Sharing Schemes or a composition of „threshold“, „and“, and „or“ operators if a proper attribute based encryption is selected (cf. [124]). Unfortunately, this scheme results in inaccurate access policies as only the number of the possessed attributes but not their importance is taken into account which strongly reduces practical applicability. For this reason, further research is necessary to build up more advanced schemes that are able to provide accurate access control. Nevertheless, it is an important step and a good base for further development.

4.7 Direct Anonymous Attestation

Direct Anonymous Attestation (DAA) makes remote authentification of a Trusted Platform Module possible while still protecting privacy (the *Identity* stays hidden and actions are unlinkable). The technical background for DAA is public key cryptography where each TPM has got a private key that is used to create a signature and a shared group key is used to prove membership of a group. For revocation usually the private key has to be published to the verifiers, or relaxed requirements for unlinkability are necessary (cf. [12]). Both possibilities clearly weaken the privacy protection. On the other hand, revocation is a legitimate and important operation. For combining both positions - the desire for privacy as well as the possibility for revocation - Enhanced Privacy ID (EPID) has been introduced. It enables user revocation without relaxations on the unlinkability or publication of the private key (cf. [12]). In contrast to most other DAA schemes, it is also applicable in scenarios beyond the authentification of Trusted Platform Modules: it could be used in driving licenses or id cards in order to proof the validity of the document and the possession of certain attributes (such as legal age) without revealing the person's identity (cf. [12, 346f]). As this approach limits the information disclosure to the minimal possible amount of data, it is a highly desirable development from a privacy point of view while the counterparts are still able to meet their obligations. For example, the users can check that they do not sell alcohol to under-aged persons without knowing the actual ages - an approach conforming to Data Minimisation (cf. page 8). The basic intention of the scheme presented in [12] is similar to the idea of Anonymous Credentials (cf. page 53).

Another interesting extension of DAA is forward security which prevents an attacker from forging signatures pertaining to the past. This is highly desirable because in case of the leakage of the private key, the consequences are somehow limited to the future while old signatures are still valid and need not be changed/resigned. Note that in the suggested scheme the issuer is not able to open the signature (cf. [37]). Forwarding Security may be especially important in scenarios where a leaked/stolen private key needs some time to be revoked. In this case an attacker would not be able to backdate a signature, and the victim therefore clearly is not responsible for any action after it announced that the key has been stolen.

The second scheme that extends DAA beyond its original scope builds up a cloud service architecture based on DAA. The aim of this architecture is to enable users to specify the amount

of data shared between several of their accounts and proving platform integrity while still staying anonymous. Furthermore, the cloud provider is able to assess user account activities and therefore enforce security measures/policies (cf. [45]).

Unfortunately, not all recent developments are of positive nature. An attack on DAA has been discovered by treating a Trusted Platform Module as a static Diffie Hellman oracle, which has not been taken into account in a lot of security proofs. Therefore, the security is weakened considerably (in case of the TMP 1.2, specified by the Trusted Computing Group, to 70 instead of 104 bits). This means that corrupted hosts could launch attacks to extract the secret key. However, DAA is believed not to be broken due to this development, although its security is weakened. Fortunately, two mitigation strategies were also proposed. The first one is a quick fix for systems already deployed to avoid firmware updates and simply adopts the change of the prime numbers. The second one which of course provides more security, is to avoid the static Diffie Hellman Assumption (cf. [11]).

Another interesting development is Improved DAA, an adoption of DAA which is optimised with respect to the computational complexity. Due to their limited computing power, this is especially important for embedded devices such as used in cell phones or machine-to-machine communication (internet of things) (cf. [50]).

4.8 Private Information Retrieval

Private Information Retrieval (PIR) is based on the assumption that not the *Identity* of the user but the accessed resources should be hidden from a Data Provider. There exist 2 key differentiators in PIR: security level (computational or information theoretic) and the number of servers involved (single server or multi server). For the security level information theoretic security states that none of the involved servers can find out what data the user requested (even with unlimited resources). Computational PIR however is (like cryptography) based on computationally hard *Mathematical Problems* that have to be solved. If they are solved, a Data Provider will know which data has been accessed.

For the single server scenario, a trivial solution exists that assures information theoretic security: to download the whole database. It is obvious that this causes enormous communication overhead and therefore research concentrated on two more effective scenarios: computational security with a single server and information theoretic security with multiple servers (note that this is based on the assumption that at least a subset of the servers does not collude) (cf. [129]). [97] concluded in 2007 that (regarding the computational practicality) no better solution than the trivial one exists in the single server scenario. Unfortunately, Quantum Computing is not able to improve this (cf. [6]). However, these results were extended by [75] stating that multi-server information-theoretic and single-server lattice-based PIR schemes have 1-3 orders of magnitude lower end-to-end response times than the trivial solution for realistic scenarios. Every time when multiple servers are involved, one faces the problem that some of them may return wrong results (either by fault or by intention), which refers to the problem of Byzantine fault toleration. In [27] Byzantine robustness for PIR was improved to the maximum possible value and additionally the performance is improved too.

As always it seems like there does not exist one optimal technique - one has to find a trade off between computational and network overhead and take the scenario into account too. For example, high network overheads due to the trivial solution may not be a problem if the database gets queried a lot of times. However, this will become a problem on mobile devices or if data changes frequently like in case of stock values (cf. [75, p170]).

Homomorphic Encryption (cf. page 35) may be used for PIR easily: one just has to use a function dependent on the index using Homomorphic Encryption to receive data without the server knowing what has been retrieved (cf. [61, p4ff]). In [61] such a scheme is described in more detail.

Another approach is Extended PIR which enables a user to privately evaluate a (publicly known) function with two parameters, a block from the database and a string. To add more flexibility, this notion has been further extended e.g. to be capable of evaluating polynomial functions (cf. [13]).

For some types of data perturbation (optionally with combining requests of several users to recover accuracy), it may be another possibility to hide what one is actually looking for. If zero mean noise gets added for all users, they need not even trust the other users (cf. [85, p4631f]). However, this will only work where accurate data is not essential like e.g. in location-based services (for an evaluation on this topic cf. [23]). Another interesting approach is to bundle queries of several users in order to hide not only the origin (user) or the target (accessed data) from the server, but the association who accessed what.

One of the biggest drawbacks of PIR (besides its high complexity) is that it relies on the cooperation of the data provider. However, it is doubtful if the data provider has an interest in cooperating in this way: if this type of data would not be of interest for it, the provider will not trace it anyway. But if it is, why should it relinquish it (cf. [85, p4632])? One countermeasure is the usage of fake queries to hide the real ones. However, one has to take care of timings and the distribution of keywords/value to prevent information leakage that could lead to the identification of the real query (cf. [85] for more details). Another drawback is that the user has to know the location (index) of data the user is interested in and it has to be modelled as a vector. This is not possible in some of the applications where it would be most desirable like e.g. search engines. [32] face all these problems by introducing h(k)-private PIR. It basically invents k-1 additional keywords for fake queries in order to satisfy k-Anonymity (cf. page 47).

4.9 Blind Signature

Blind Signature enables digital signature without the signer learning any critical information about the content of the signed message (blindness), but the user cannot create more signatures than the number of interactions with the signer (unforgeability). In order to prevent fraud, two variations of Blind Signatures exist: Fair Blind Signatures which allow a Trusted Third Party to recover the *Identity* in case of dispute and Partially Blind Signatures which require that the signer has some control of the message (e.g. the user should not be able to control the expiration date of a voucher). Fair Blind Signature however has to be considered as providing *Pseudonymity* instead of *Anonymity* as *Anonymity* cannot be reversed (cf. Distinction between Anonymity and Pseudonymity at page 6). Possible applications of Blind Signatures include but are not limited

to e-cash and e-voting (cf. [89, p34]). Fair Blind Signatures can be based on Oblivious Transfer (cf. [101]) or on other *Mathematical Problems* like the discrete logarithm problem. Partially Blind Signatures are usually based on public key cryptography (cf. [133, p546]).

In [39] the first Fair Blind Signature Scheme in the standard model security was proposed which exhibits the highest security level that has been achieved by a Fair Blind Signature Scheme so far. This scheme prevents framing of innocent users/signatures. Another interesting development deals with Fair Partially Blind Signatures which is the combination of Fair Blind Signatures with Partially Blind Signatures and is secure in the standard model too (cf. [89]). Note that these developments are independent from each other although some of the results seem to be similar.

Most of the time, only the optimal case where goes well and the protocol finishes without any problems is considered. Another assumption is that if a user starts an action, it will be finished. Although this may be true often, the possibility that an action fails or gets aborted by the user has to be taken into account too in a realistic environment. The same applies for the signer/server. Fortunately, there exists a way to convert every secure Blind Signature scheme into a Selective-Failure Blind Signature scheme. That is a scheme where blindness is ensured even if the signer learns about the abortion of some of the executions. This improvement requires only one more computation and therefore the computational overhead is low. Every transformed 3-move blind signature scheme remains unforgeable in case of abortion but uniqueness (one signature per key/message pair) is not guaranteed after the transformation (cf. [38]).

Abortions are not the only problem that may affect a cryptographic protocol. Usually cryptography gets weakened considerably if parts of the key are leaked. However, in [88] a protocol is presented that is able to deal with the leakage of a 1 - o(1) fraction of the private key. This scheme is lattice-based and with respect to the performance promising for the future as it offers quasi linear complexity in the key size (cf. [88]). Under the assumption that the key size is direct proportional to available computing power, this means that the performance will stay nearly equal in contrast to schemes with higher complexity in the key size. Furthermore, the scheme offers statistical blindness as well as unforgeability and is able to deal with abortion (cf. [88]).

Ordinary (non blind) signature schemes can be built upon arbitrary one way functions. For Blind Signature schemes this possibility as well as the construction from the random permutation oracle have been proven to be impossible even in the weakest scenario by [58].

Usually Blind Signature Schemes are not collective, so the partners need not sign both to cause any action. This is sufficient for many cases but when a contract should be signed, it is required that all parties sign it simultaneously. For this reason, a Blind Collective Signature Protocol that is based on the Discrete Logarithm Problem has been introduced. This scheme is the privacy-preserving version of Collective Digital Signatures. Furthermore, another scheme is proposed that enables several contracts to be signed by different sets of signers simultaneously (cf. [71]). This leads to interesting possibilities for complex contractual situations where several contracts depend on each other as well as on a set of parties.

A hybrid approach combining Blind Signature with Oblivious Transfer (cf. page 41) is explained in the section k-Anonymity on page 47.

4.10 k-Anonymity

The general concept of k-Anonymity is a simple way to prevent record linkage using the QID. It defines that every record has to be indistinguishable from k-1 other records with respect to the QID (which means that every QID has to occur at least k times). Therefore, the probability of revealing the real identity is bounded to at maximum 1/k. However, if more than one QID exists, this is not a problem neither as k-Anonymity may be applied to both of them. Therefore, this also secures data against de-anonymisation and does not require further mechanisms. One has to take into account that the result of combining two QIDs to an overall QID or treating them as two different QIDs is not the same. If they are treated separately the data quality is better as less attributes which have to be distorted for achieving k-Anonymity exist. On the other hand, if they are treated separately, one may face serious privacy problems if an attacker is able to find a data source containing the conjunction of both QIDs (which may even allow a unique identification). One weakness of the k-Anonymity method is that it assumes that every record owner has only got 1 entry - if more entries are available, *Anonymity* can be broken (cf. [40, p14:8f]) (although the attacker has to know that several records belong to one person). To solve this and other issues, extensions of k-Anonymity like (X,Y)-Privacy and MultiR k-Anonymity have been developed.

In [78] three schemes are proposed which are based on the idea of k-Anonymity. The first one is based on Blind Signature (cf. page 45), the second one on Oblivious Transfer (cf. page 41), and the third one combines those two in order to hide the user requesting data (User Perspective) as well as the accessed data (Access Perspective) (cf. [78]). Although the protocol based on Oblivious Transfer is based on the idea of k-Anonymity, this does not seem to apply for the protocol based on blind signature. However, the combination of both protocols in order to protect the *Access-Perspective* as well as the *User-Perspective* is an interesting idea.

4.11 (X,Y)-Privacy

(X,Y)-Privacy is an extension of k-Anonymity which basically defines that for each value on X k distinct values on Y have to exist. For example, X may be considered as a QID and Y as a sensitive value or an explicit identifier. This means that each group of QIDs is linked to at least k different sensitive values or explicit identifiers respectively (cf. [40, p14:10]). The difference to k-Anonymity is that it does not only take the QID into account and ensures that k indistinguishable values are contained in the published data but additionally specifies with respect to which attribute the k indistinguishable values have to exist.

4.12 MultiR k-Anonymity

MultiR k-Anonymity (also known as Multi Relational k-Anonymity) is an extension of k-Anonymity too. It assumes that data is distributed over several relations as it is common in Relational Data Base Management Systems (RDBMS) and that 1 table exists containing the explicit identifier (and probably other data) and various tables containing other data (like sensitive values or QID) but not the explicit identifier. The method of MultiR k-Anonymity ensures that after joining all tables (with an inner join) k-Anonymity is satisfied for the result or, in other words that for

each explicit identifier the QID is shared with at least k other records. Important is that this is granted not only on record- but on record owner basis and therefore does not suffer from privacy problems when record owners may have multiple records. This also may be seen as special case of (X,Y)-Anonymity where X is the QID and Y the explicit identifier (cf. [40, p14:10]), but it is not exactly the same as it explicitly considers the record owner level and an unlimited number of relations that are joined. This means that in joint tables the records have to be indistinguishable from k-1 other records.

4.13 Distinct l-Diversity

Distinct l-Diversity (as well as Entropy l-Diversity, and (c,l)-Diversity) belongs to the family of l-diversity which in general defines that every QID group contains l or more well represented values; however, it is not that easy to define what exactly well represented means. This is the reason why several different methods within this family exist. The easiest is Distinct l-Diversity (also known as p-sensitive k-Anonymity). It simply defines that each QID group must contain l or more values. This is technically the same method as k-Anonymity, but from another point of view as it tries to bound the probability of guessing the parameter directly.

However, this does not prevent probabilistic inference attacks which take the distribution of the values into account (some values will appear more frequently than others) (cf. [40, p14:11f]). One big drawback of the family of l-Diversity methods is that it assumes more or less equally distributed values for the sensitive attributes. If this condition is violated, it may cause a high information loss, as the QID groups are defined so that in each of them a sensitive attribute occurs at least once. The number of occurrences of the least frequent sensitive value also defines the minimum number of groups (as it has to occur in each group at least once) (cf. [40, p14:12]). Decreasing the number of groups of course implies increasing group size (with the total number of records being constant) and therefore the accuracy of the dataset is decreased drastically. However, it may be possible to apply the skewness attack to l-Diversity.

4.14 Entropy l-Diversity

As mentioned Entropy l-Diversity also belongs to the group of l-Diversity methods. As a base it takes the entropy (measure for the amount of contained information) and defines it as a constraint:

$$-\sum_{s\in S}\frac{s}{GE}\log\left(\frac{s}{GE}\right)\geq\log(l) \tag{4.1}$$

where S is the sensitive attribute and GE the respective QID Group Extend. The value of the left hand side of the expression is the higher, the more evenly the sensitive values are distributed (which is shown in the example below). The right hand side is used as a threshold which defines the minimum level of confidence that has to be achieved (cf. [40, p14:12]).

An adversary could try to join at least two tables (of which at least one is anonymised) with QIDs. As one can see in table 4.1 no information exists which allows the disclosure of the real identity of the record owners. However, if one also considers table 4.2 and assumes that every person from the patient table (table 4.1) is also contained in the external table (table 4.2), some

real identities may be revealed. For example, two male lawyers are present, but only one of them is 38 years old. Therefore, one knows that the 38 year old lawyer in the patient table is Doug and that he has got HIV (cf. [40, p14:8]). With the same technique, one can also find out that Bob and Fred have got Hepatitis. To tackle this problem, methods like k-Anonymity, (MultiR) k-Anonymity, and (X,Y)-Privacy have been introduced in order to build up indistinguishable groups. This attack is known as Record Linkage (cf. [40, p14:7f]).

Although Record Linkage may not be possible, one could also take another scenario into account. Again one assumes an anonymised (table 4.1) and an external (table 4.2) table and again concentrates on the linkage of the QID. However, one does not need to find a clear mapping between the two tables; it is sufficient to have a high probability of assigning the sensitive attribute(s) to one person. Again it is assumed that the attacker knows that a record in table 4.1 has got a mapping to table 4.2.

However, one need not get certainty for infringing privacy. If one considers both tables, it turns out that both dancers have got HIV and both engineers have got Hepatitis (with 100% probability) (cf. [40, p14:11]). In both cases, not even the age is needed because the disease has got only one attribute for the mentioned jobs. Another example (again with only 1 attribute) is the gender. If one knows only the gender of the person, one can assume with a probability of 75% that a female has got HIV in this database. This shows, how dangerous and unexpected side effects a table may have as not even age or profession but only the gender needs to be known (which one can find out easily). Even if the age of every female record would be different or unknown, this would not change anything. Implementing 3-anonymity (cf. table 4.3) does not change this problem at all. This attack is referred to as Attribute Linkage.

The difference to Record linkage (as explained above) is that one only tries to find out if attributes automatically infer (with high probability) that another sensitive attribute has a specific value instead of linking the QIDs in order to reveal the identity. In case of Attribute Linkage, the identity is never linked directly. However, one can link the sensitive value to an identity anyway. To prevent this attack, methods like l-Diversity, Entropy l-Diversity, Confidence Bounding etc. have been developed.

If one applies Entropy l-Diversity to table 4.3, this means that for the first group (Professional, Male, 35-40 years) the attribute Hepatitis occurs in 2 (out of 3) cases and HIV in 1 (out of 3) case(s) which leads to

$$-\frac{2}{3}\log\left(\frac{2}{3}\right) \qquad and \qquad -\frac{1}{3}\log\left(\frac{1}{3}\right) \tag{4.2}$$

respectively which are then (according to the equation 4.1) summed up:

$$-\frac{2}{3}\log\left(\frac{2}{3}\right)-\frac{1}{3}\log\left(\frac{1}{3}\right)=\log(1{,}89) \tag{4.3}$$

For the group Artist, Female, 30-35 years this leads to

$$-\frac{3}{4}\log\left(\frac{3}{4}\right)-\frac{1}{4}\log\left(\frac{1}{4}\right)=\log(1{,}75) \tag{4.4}$$

Therefore, the overall table only satisfies l-Diversity if l<1,75. Please note that the results in the equations 4.3 and 4.4 are rounded. However, the biggest problem with Entropy l-Diversity is that it does not measure the confidence level in an intuitive way (cf. [40, p14:11f]).

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV
Dancer	Female	30	HIV
Dancer	Female	30	HIV

Table 4.1: Patient table, [40, p14:8]

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30
Fred	Engineer	Male	38
Gladys	Dancer	Female	30
Henry	Lawyer	Male	39
Irene	Dancer	Female	32

Table 4.2: external table, [40, p14:8]

Job	Sex	Age	Disease
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	Hepatitis
Professional	Male	[35-40)	HIV
Artist	Female	[30-35)	Flu
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV
Artist	Female	[30-35)	HIV

Table 4.3: 3-anonymised patient table, [40, p14:8]

4.15 (c,l)-Diversity

(c,l)-Diversity is also a member of the family of l-Diversity methods. Basically it defines that the frequency of values is not too high nor too low. The intention behind this is that an attacker should not be able to infer privacy although the attacker is excluding some possibilities by applying background knowledge. With $f_i, i \in 1...n$ as the frequency of the i^{th} most common value, (c,l)-Diversity is defined as following: f_1 has to be lower than the totalised frequencies of the (n - l + 1) least common values multiplied by the user defined factor c or written as a formula:

$$f_1 < c \sum_{i=l}^{n} f_i \tag{4.5}$$

This intents to balance the frequency of the most common value and the frequencies of the least common values. If this applies to all of its QIDs groups, a table has recursive (c,l)-diversity. However, big drawbacks of this system exist: the risk is shown in a non-intuitive way, it focuses too much on the most frequent and most infrequent value(s), and it does not allow the definition of different thresholds for different values of a sensitive attribute (only for different sensitive attributes). Anyway, recursive (c,l)-Diversity helps to prevent table linkage in a better way than Confidence Bounding which intents to resolve the drawbacks of it (cf. [40, p14:12f]).

4.16 Confidence Bounding

Confidence Bounding is a simple concept which, as its name already suggests, intents to limit the probability with which one is able to guess that a sensitive value is related to a record. It defines that for every group within a QID, the confidence of assuming a certain value for a sensitive attribute, has to be below a defined threshold. It is important to note that the maximum probability of a value within the QID is important. For table 4.3 this means that every threshold below 75% is violated for the sensitive attribute Disease with the value HIV. This is because one knows that a member of the group Professional, Male, 35-40 years has got 33,33% and the group Artist, Female, 30-35 years 75% probability of having HIV (cf. [40, p14:13]). As the highest of these values is 75%, it cannot be guaranteed that the probability of guessing a sensitive attribute may be lower than 75%, which therefore limits the confidence level. Naturally, all higher confidence levels (which means lower probabilities) would be violated too.

4.17 (α, k)-Anonymity

(α, k)-Anonymity is an approach which combines the methods of k-Anonymity (k indistinguishable records per QID group) with Confidence bounding (confidence with which a sensitive attribute value for a QID group may be guessed is limited) (cf. [40, p14:8,13]). It therefore combines the advantages of both methods but also increases the data loss.

4.18 (k,e)-Anonymity

With applying k-Anonymity logics to numerical values, one would face the problem that a wide range of numbers that occur exists. In case of a sensitive attribute that measures a percentage as integer for example, 100 different values exist that have to be covered. So statistically 200 values are needed in order to apply 2-Anonymity. In reality however, less would be needed as the numbers of the percentage probably will not be equally distributed. But if one takes floating point numbers into consideration, it becomes obvious that the problem is very unlikely to be solved without excluding a lot of records. Therefore, (k,e)-Anonymity has been proposed. It is basically k-Anonymity applied to numeric values. The additional parameter e defines a minimum range of the values, and the parameter k defines minimum number of indistinguishable records (as it is the case in the method of k-Anonymity too). Anyway, this method has its weaknesses too: if the majority of records has got a similar but distinct value and one value that is different, the similar values can be guessed with high probability, although the variables e and k indicate a stronger anonymity (cf. [40, p14:13]). As an illustrative example, the explanation of this problem in [40, p14:13] is used as a base: a table with 10 records for some QID exists, and for the sensitive attribute 7 different values within this group exist. From the 10 values, 9 are in the range from 30 to 35 and the last one is 80. This allows to guess that the sensitive value of a record owner is with 90% probability in the range 30 to 35, although (k,e)-anonymity is satisfied with k=7 and e = 50. This is because 7 different values (=k) exist and the range of the sensitive value within this group is 80-30=50 (=e) (cf. [40, p14:13]).

4.19 (ϵ, m)-Anonymity

(ϵ, m)-Anonymity is a method that simply limits the probability that one can guess parameters in the range s$\pm\epsilon$ to $\frac{1}{m}$ where s is any sensitive value within a relation. It is also effective when statistical outliers exist (cf. [40, p14:13]).

4.20 Personalized Privacy

Personalized Privacy builds on a principle similar to the one used in Confidence Bounding: one tries to limit the probability with which an attacker can assign a sensitive attribute to a record owner. The difference however is that the record owner can specify the level of confidence the record owner wants to have for his or her record. Therefore, a tree of the information is necessary and as one is moving from the root to the leaves one gets more detailed information with every layer. Personalized Privacy sets a limit for a record so that one cannot extract more information than to a specified layer. Again, taking the tables 4.1 and 4.2 into account, Cathy may specify that an attacker may know that she is present in table 4.1 (and therefore that she has got some kind of infectious disease), but she does not want anyone to know which disease she is suffering from. Bob however may not mind if anyone knows about his disease so he can specify a lower protection level which leads to less information loss (cf. [40, p14:14]). A reason for this is that Cathy knows that any infectious disease ranging from a cold to HIV is stored, so the significance of her being in the database is limited anyway.

With this approach it is possible to decrease the information loss as only the minimum level of privacy is satisfied, but problems with the record owners are avoided anyway, because they are fine with the protection level of their record. However, as they will probably have a lack of knowledge about the rest of the records which should be published, they may tend to overprotect their privacy and also do not allow any knowledge about common attribute values (cf. [40, p14:14]). Anyway, it seems to be possible to distribute generalised and aggregated data (as for example, the count percentage of the occurrence of their sensitive values in other records) in order to enable them to judge this situation better without a considerable overhead.

4.21 t-Closeness

t-Closeness tries to prevent the skewness attack and therefore intents to conserve the distribution of the original data as good as possible in the published data. Therefore, the Earth Mover Distance (EMD) is used to compare the distributions. The limit of the closeness must not be lower than specified by the parameter t. Anyway, some drawbacks with this method exist too: like a lot of other methods (except of Personalized Privacy), it does not allow any distinction between the attribute values. Additionally, it must not be used for numerical values as the EMD function may allow Attribute Linkage. Last but not least, it may cause a huge data loss as the sensitive values have to have the same values (with the same distribution) in all QID groups. As a consequence, the correlation between QID and sensitive attributes would be harmed in a serious way (cf. [40, p14:14]).

4.22 Anonymous Credentials

Although it affects only a small percentage of the overall users, some of them may misuse anonymisation services (cf. page 10). Anonymous Credentials however offer *Anonymity* while still being able to enforce authentification/accountability and to deal with misbehaving users.

This approach can either be based on Blind Signatures or on Zero Knowledge proofs. It also takes Data Minimisation (cf. page 8) into account (cf. [62, p1f]) and is thus a privacy friendly technique for achieving the goals of all involved parties: users get the desired privacy and furthermore are able to decide what information gets revealed, while service providers still are able to fulfil legal restrictions (such as e.g. age restrictions) and ensure that, for example, only a specified group of users is able to use their service. However, companies whose business is to collect, analyse, and sell user data will not be happy about enhancing user privacy (cf. [62, p3f]). Nevertheless, this is an important business case for a lot of companies. Anonymous Credentials are also responsible for certain tasks which are better handled by traditional systems (using a combination of username and password) but still need to be managed by them as well. In the following, the most important recent developments in the field of Anonymous Credentials are summarised.

User revocation or blacklisting of misbehaving users seems to be one of the most difficult tasks related to Anonymous Credentials whereas it is much more difficult to be managed in traditional systems because the users' identities or at least *Pseudonyms* need to be known for access blocking. If one assumes that the system is open for everyone to register things get more

complicated as the user may try to re-register, if his or her account is blocked. However, things get a lot more complicated when dealing with Anonymous Credentials as *Anonymity* is ensured. Fortunately, a lot of research related to this topic was done. In [62] user revocation is discussed in depth. Besides user revocation, it also deals with Mobile Anonymous Authentification. [47] offer a possibility to efficiently remove misbehaving users, but the approach relies on cooperation between a Public Authority and the Issuer, and furthermore is argued that „These keys can be later on used to identify malicious users and attackers (...)“, [47, p213]. This concept is extended in [48]. However, this is a conflict with the definition of Anonymity (cf. page 6), and therefore it seems to be necessary to consider the used concept to offer *Pseudonymity* but no *Anonymity*.

Another important publication is [7] which introduces the first scheme that offers efficient delegatable Anonymous Credentials. Although this approach is interesting (it is argued that, without delegation, the Data Provider would completely be overwhelmed by the flood of requests, cf. [7, p108f]) major weaknesses of the system exist: at first, it offers just delegation but no rights management. The lack of rights management in combination with an anonymous system seems to be risky as (group) dynamics may lead to the loss of control who can access the system. More and more users could be invited with the undesirable effect that users may have more privileges than desired (although it can be traced back by whom they were invited). For some systems, this may be less of a problem (e.g. if they want to have users with equal privileges and the access of the first user to a public database is read-only) but, in general, this is indeed an undesirable effect. Secondly, this system relies on a Trusted Third Party (TTP). However, this only shifts the trust requirements away to another party, which is still a sub optimal solution. Privacy should be achieved in general 'by design' (e.g. by using cryptographic means).

With these considerations in mind a system for user revocation which does not need a TTP to be involved was introduced (cf. [2]). It allows servers to automatically apply complex policies to blacklist misbehaving users based on scoring - the user gets blacklisted if he or she exceeds a certain threshold in one or more categories. However, this may (optionally) be improved again by positive ratings. In order to offer more fine grained control, an extension that offers the possibility to take the severity of the misbehaviours into account was developed.

Furthermore, several interesting improvements not related to user revocation were proposed as well. For example, [46] enable the usage of encrypted attributes in Anonymous Credentials in order to hide information from the user. This is due to the fact that the user may not be the only one who wants to hide information: the issuer of the credential may want to hide some confidential information from the user too, which is still required as attribute to deliver the desired service, which again leads to problems. Another extension of Anonymous Credentials, especially important for real world applications, is that the credentials may be issued from several authorities. However, sometimes it may be necessary to combine credentials from several authorities. Therefore, [19] propose a system based on indexed aggregate signatures which is the first one that enables the combination of credentials from different authorities.

[24] deal with the problem of access control in oblivious databases. As the user is not known it may be difficult to enforce access control policies. However, an approach based on directed graphs was proposed which extends Anonymous Credentials to be stateful. It is an interesting scheme as it ensures privacy from the access perspective (no information is revealed about the accessed items) as well as from the user perspective (the users identity is not revealed

neither). Additionally, history-based access is possible with this scheme too. However, it needs to be improved in order to support concurrent access (unfortunately, only sequential access is considered).

As mobile devices continue to become more important, [4] describe a scheme called Anonymous Credentials Light which is based on Blind Signatures (cf. page 45) with Attributes. It is specifically designed to be efficient which is, due to limited computing power, especially interesting for mobile devices.

A hybrid approach combining Blind Signature (cf. page 45) with Oblivious Transfer is explained in the section k-Anonymity on page 47.

CHAPTER 5

Evaluation

In this chapter, the methods presented in chapter Techniques at page 33 are evaluated based on the introduced taxonomy. Due to the broad variety of presented techniques, the evaluation is kept general. This means that the overall approach of the techniques (like e.g. Searchable Encryption or Oblivious Transfer) is evaluated and rated. When possible the techniques are clustered to groups so that a better overview is possible while still considering the individual special properties of the techniques to ensure a detailed analysis.

The evaluation is based on the taxonomies regarding *Technology* and *Perspective* - both is considered for every technique in an own subsection.

5.1 Techniques

k-Anonymity

k-Anonymity (cf. page 47) is a simple but effective technique to enhance privacy.

Technology

k-Anonymity has *Statistics* as the *Base of Security* because it is based on lowering the probability of guessing the respective identity (from initially 100%) to a predefined threshold.

Following the argumentation from above, the *Strength of Security* is *Information Theoretic Security*: it is impossible to calculate the relation between data and the respective identity without further knowledge; one just has to know it (and if not, one could only guess it with some probability below 100%). This cannot be changed with any amount (even unlimited) of computation power. Per definition this results in *Information Theoretic Security*. For the same reason, the *Reversibility* has the *Degree Not Reversible* and *Cooperation* is required to reverse this process (the Data Owner has to reveal his or her *Identity* to reverse the process). Only the *Server* (=Data Provider) has the *Enforceability* because it controls the data before it gets published. Each Data Owner is only able to control his or her own information that is disclosed to the Data

Provider, but not how the data is aggregated at the Server. For this reason, the client is not considered to be capable of enforcing this technique. The *Trusted Third Party Participation* occurs with *Frequency Never* and, therefore, affects the *Phase None*, both resulting in the *Background No TTP Participation* because it is not required nor desired that any Third Party (regardless if trusted or untrusted) participates in the protocol. The protocol only obliges the Data Provider to remove data, which it can do itself. As the strongest scenario possible is considered, it is assumed that no TTP will participate.

The *Affected Data Type* is *Stored* - *Transmitted* and *Processed* data is not relevant for this technique.

As stated above *Protected Aspect* is the *Identity* (to be more precise by establishing *One-Sided Anonymity*). This is because in contrast to *Pseudonymity* (where the *Identity* gets replaced with a Pseudonym), the *Identity* is removed (at least this applies for the published data). One may think of the possibility that the Data Provider could preserve the removed information and that this may lead to Pseudonymity. This is not true as not even the Data Source can restore all information, of course, assuming that the original data is not considered (this would not make sense anyway, as this data already contains all data).

The *Means of Protection* is *Indistinguishability* because the *Identity* is hidden rather than the *Behaviour* or the actual *Content*. *Deniability* is not existent in this case. One could argue that due to *Indistinguishability Deniability* exists. This is (at least to some extend) true because one could deny having anything to do with the published data. However, if the original data is taken into account, it becomes clear that *Deniability* does not apply for k-Anonymity.

Perspective

In a typical scenario for this technique, the Data Source wants to publish some data but is only allowed to do this after applying k-Anonymity. As one can easily see, the Data Source is the main stakeholder. Therefore, the associated *Perspective* is the *Data Perspective*.

(X,Y)-Privacy

(X,Y)-Privacy (cf. page 47) is an extension of k-Anonymity preventing Attribute Linkage Attacks (cf. explanation at page 49).

Technology

The *Base of Security* for (X-Y)-Privacy is *Statistics* due to the fact that the probability of guessing the correct assignment of identities to records is lowered by enforcing collisions within the QIDs as well as the related sensitive attributes. It is not possible anymore to reverse this process and increase it up to 100% again.

The *Strength of Security* (X,Y)-Privacy has clearly to be classified as *Information Theoretic Security* as it is not reversible with any arbitrary amount of computing power without further knowledge. Regarding the *Trusted Third Party Participation*, the *Frequency* with which the TTP participates in the protocol is *Never* and accordingly the *Phase* is *None*, both resulting from the *Background* that *No TTP Participation* is desired nor required for this technique. As this

technique requires an honest Data Provider pre-processing all data before publishing it, it is obvious that the *Server* can enforce it. Unfortunately, it is not possible for a *Client* to enforce it the other way round, which directly leads to the *Enforceability* being only possible by the *Server*. Regarding the *Reversibility*, the *Degree* is *Not Reversible* as the Data Provider would not be able to recover all data and increase the probability of mapping an record owner to a record to 100% again (of course assuming the original data is not available). Therefore, only the record owner himself is able to reveal his or her *Identity* by providing the mapping between sensitive attribute and the QID. For this reason, *Cooperation* is needed to reverse this process.

As this technique affects data that is being published from a database, the *Affected Data Type* is *Stored*.

The *Means of Protection* is based on *Indistinguishability & Unlinkability* (to be more precise on *Indistinguishability*) as the *Protected Aspect* is the *Identity*.

As indicated above, the *Protected Aspect* is the *Identity* which is protected by means of *Anonymisation* because the explicit IDs are removed rather than substituted. The *Directionality* is *One-Sided* because only the *Identity* of the Data Owner is hidden.

Perspective

Examining the use cases of this technique, it becomes clear that the stakeholder is a Data Provider that wants to publish confidential data without harming privacy. As a consequence, this leads to the *Data Perspective* as the associated *Perspective*.

MultiR k-Anonymity

MultiR k-Anonymity (cf. page 47) is an extension of k-Anonymity that considers that data is split into multiple relations. It explicitly takes the record owner level and not only record level into account, which is why it supports multiple records per record owner too.

Technology

Like its base technology k-Anonymity, this technique generalises the data that is published so that the probability that a record can be de-anonymised is lowered to the desired level. Therefore, the *Base of Security* is *Statistics*.

The identifiers are not only substituted by a pseudonym but are completely removed from the data to be published. For this reason, it is clear that the *Base of Identity Protection* is *Anonymisation*.

Statistics as the *Base of Security* implies that this technique is not reversible and, therefore, the *Strength of Security* is *Information Theoretic Security*. For applying this technique, no Third Party is required; all operations can be conducted by the Data Provider itself. Therefore, no *Trusted Third Party Participation* exists, which implies the *Phase* and the *Frequency* are *Never* respectively *None* and the *Background* is *No TTP Participation*. Again, the *Client* has no influence on the way data is stored on the *Server*, which is why the *Enforceability* is only possible by the *Server*.

It becomes clear that the *Reversibility* has the *Degree Not Reversible* and that *Cooperation* of the Data Owner is required to reverse this process because no-one (not even the Data Provider) is able to reverse the *Anonymisation* (under the assumption that the original data is not available).

The *Affected* Data Type is *Stored* data; applying this technique to *Transmitted* or *Processed* data is not possible (although it may influence *Processed* or *Transmitted* data that has been stored applying MultiR k-Anonymity in the first place).

The *Means of Protection* is ensured by means of *Indistinguishability* & *Unlinkability*. To be more precise, it is based on *Indistinguishability* because the *Protected Aspect* is the *Identity* rather than the *Content* or *Behaviour*.

As argued above, the *Protected Aspect* is the *Identity* by applying the concept of *Anonymisation*. The *Directionality* for the identity protection is *One-Sided Anonymity* (where only the Data Owner gets protected) which is not reversible (leading to *Reversibility* with *Degree Not Reversible where Cooperation is requried*).

Perspective

This technique is used to enable a Data Provider to publish data in an anonymous way so that the Data Owners cannot be identified. Therefore, the main stakeholder is the Data Provider, which leads to *Data Perspective* as the associated *Perspective*.

Distinct l-Diversity

Distinct l-Diversity (cf. page 48), which is also called p-sensitive k-Anonymity, is a technique that requires l different values for each QID group. Basically it is based on a similar approach as k-Anonymity but from another point of view as it prevents guessing of parameters.

Technology

Distinct l-Diversity shares some common properties with its base (k-Anonymity). Both techniques remove information in order to lower the probability that the *Identity* of a record owner could be guessed. This security concept states that the probability of guessing the record owner's *Identity* has to be below a specified threshold, which implies *Statistics* as the *Base of Security*.

The *Strength of Security* is *Information Theoretic Security* as removed information may not be restored with any computing resources available. As a consequence, the *Degree* is (for the same reason as in case of k-Anonymity) *Not Reversible* and *Cooperation* is required because no-one, not even the Data Provider, is able to restore the original information that has been removed (under the assumption that the original data is not available). All operations can be conducted by the Data Provider itself, so no need for any *Trusted Third Party Participation* exists. Therefore, it is not desired that any third party (independently if trusted, untrusted, or semi-trusted) participates in this protocol, which is why the *Background of TTP* is *No TTP Participation*. This implies the *Phase* and the *Frequency* of the *Trusted Third Party Participation* are *None* respectively *Never* because it is not needed. The Data Provider is the only participant that is able to enforce the application of this technique leading to the *Enforceability* for the *Server* only. Due to the deletion of the QIDs it is not possible to de-anonymise data, so no *Reversibility* is possible,

which, as a consequence, leads to the *Degree Not Reversible* and requires *Cooperation* of the Data Owner in order to reverse the process.

Regarding the *Affected Data Type*, only *Stored* data is protected by this technique. Neither *Transmitted* nor *Processed* data will have any protection by applying this technique if the stored data has not been anonymised first.

The *Protected Aspect* is *Anonymity* as the QIDs are removed permanently rather than being substituted by a pseudonym. To be more precise, the *Directionality* of *Anonymity* is *One-Sided* because only the Data Owner's *Identity* is protected.

The *Means of Protection* is based on *Indistinguishability & Unlinkability* and, to be more precise, it is *Indistinguishability* because the *Protected Aspect* is the *Identity*.

Perspective

For this technique, the main scenario is that a Data Provider wants to publish data containing confidential information. As this is only permitted when the Data Owner's privacy cannot be compromised, it is in the Data Provider's interest to anonymise data so that publication is permitted. Therefore, the main stakeholder is the Data Provider, which implies that the *Perspective* for this technique is the *Data Perspective*.

Entropy l-Diversity

Entropy l-Diversity (cf. page 48) is a variation of l-Diversity based on the concept of entropy but follows a similar approach. The basic idea of this technique is to limit the possible information gain of the presence of an attribute.

Technology

As it intends to limit the information that could be extracted from the presence of an attribute, it basically just limits the probability that a person can be related to a specific attribute. Therefore, this technique builds on methods from the field of *Statistics* as the *Base of Security* which do not allow to generate knowledge from the data.

The *Base of Identity Protection* is *Anonymisation* because Entropy l-Diversity is a measure to ensure the quality of *Anonymisation*. As the QIDs are removed and are not present anymore in the published data, the *Reversibility* has the *Degree Not Reversible* and *Cooperation* is required for de-anonymisation because when the original data is selected, it is not possible for anyone (including the Data Provider) to restore it again regardless of available computing capabilities. Therefore, the *Strength of Security* is *Information Theoretic Security*. In order to follow this technique's protocol, no Third Party is needed or desired. For this reason, the *Trusted Third Party Participation* occurs in the *Phase None* and with *Frequency Never* due to the *Background No TTP Participation*. The *Client* has no possibility to influence if/how this technique is applied on the data; this can only be done by the *Server*. For this reason, only the *Server* has *Enforceability* for Entropy l-Diversity.

Entropy l-Diversity may be applied to *Stored* data only but has no effect on *Transmitted* or *Processed* data if it has not been anonymised before. Therefore, the *Affected Data Type* is *Stored*.

The *Means of Protection* of privacy is ensured with *Indistinguishability* & *Unlinkability* and because the *Protected Aspect* is the *Identity*, this technique deals with *Indistinguishability*.

The *Protected Aspect* is - as mentioned - the *Identity*. *Identity* protection is based on *One-Sided Anonymisation* because only the *Identity* of the Data Owner is hidden. This leads to the *Directionality One-Sided*.

Perspective

The main use-case for applying this technique is to remove all data that prevents the Data Provider from publishing a privacy-preserving dataset. It is clear that the main stakeholder is the Data Provider, because it wants to publish sensitive data without harming the privacy of the Data Owners. As a consequence, the associated *Perspective* is the *Data Perspective*.

(c,l)-Diversity

(c,l)-Diversity (cf. page 51) is a technique taking the distribution of values into account. It specifies that the frequency of the most common value has to be lower than the totalised frequencies of the (n - l + 1) least common values multiplied by some user defined factor.

Technology

This technique establishes boundaries for the frequency of values within the published data set so that no value occurs too often or too rarely. It is obvious that this has *Statistics* as the *Base of Security*.

Again, the QIDs are removed to ensure privacy protection by means of *Anonymisation* as the *Base of Identity Protection*.

As the QIDs have been removed, the *Reversibility* has the *Degree Not Reversible* and *Cooperation* is required because no-one, not even the Data Provider itself, is able to reconstruct the original data anymore. The amount of available computing resources has no influence on the (lack of) ability to restore the original data. This is why the *Strength of Security* is *Information Theoretic Security*. For this technique, only the Data Provider is required. Involving any other (Trusted) Third Party is not required. As a result the *Background* is *No TTP Participation*, the *Phase* and *Frequency* of the *Trusted Third Party Participation* are *None* respectively *Never*. Again, the Data Consumer has no possibility to influence if/how the Data Provider follows the specified protocol and removes the QIDs before publishing the data. Therefore, the Data Provider has the exclusive power to control and enforce (c,l)-Diversity, which results in the *Enforceability* for the *Server* only.

The *Affected Data Type* of (c,l)-Diversity is *Stored* data. On *Transmitted* or *Processed* data it only has an effect if this technique has been applied before.

The *Means of Protection* is *Indistinguishability* & *Unlinkability*. Because the *Protected Aspect* is the *Identity*, this technique deals with *Indistinguishability*.

As indicated by the removal of the QIDs, the *Protected Aspect* of this technique is the *Identity*. It's protection is based on *One-Sided Anonymisation*. The *Directionality* is *One-Sided* because only the Data Owners' *Identity* is hidden and the Data Provider's *Identity* keeps unprotected.

Perspective

The main stakeholder having an interest in this technique is the Data Provider. After applying (c,l)-Diversity to sensitive data, it can be published, which would not be possible without *Anonymisation*. This implies that the associated *Perspective* is the *Data Perspective*.

Confidence Bounding

Confidence Bounding (cf. page 51) ensures privacy by limiting the probability with which a certain manifestation of an attribute within a QID group may be guessed.

Technology

As the description above suggests, the *Base of Security* for Confidence Bounding is *Statistics*.

This technique is a one way function (removed information cannot be restored again), which results in the *Reversibility* having the *Degree Not Reversible* and requiring *Cooperation* because not even the Data Provider can restore the original information. Due to this, the *Strength of Security* can be considered as *Information Theoretic Security* because this cannot be changed with any amount of computing power. No Trusted Third Party is required to apply Confidence Bounding and, therefore, the *Trusted Third Party Participation* is *Background No TTP Participation* and, due to this, in the *Phase None* and with the *Frequency Never*. The Data Provider is the only one who can have an influence on if/how this technique is applied to data and, therefore, the *Enforceability* is classified as *Server*.

This technique applies to *Stored* data as the *Affected Data Type* - an influence to *Transmitted* or *Processed* data is, however, only possible by storing anonymised data first.

As identifiable information is removed, the *Protected Aspect* is the *Identity*. Identity protection is enforced by applying *One-Sided Anonymisation* because the QIDs are not only replaced by a pseudonym but are definitely removed from the published data. The *Directionality* is *One-Sided* because only the Data Owner's *Identity* is hidden.

The *Means of Protection* is *Indistinguishability* & *Unlinkability*, more specifically by means of *Indistinguishability* as the *Protected Aspect* is the *Identity*.

Perspective

The main scenario for applying Confidence Bounding is that if a Data Provider wants to publish confidential data which is anonymised in order to conform with a privacy policy. Therefore, the main stakeholder is the Data Provider, which, as a consequence, leads to the *Data Perspective* as the associated *Perspective*.

(α, k)-Anonymity

(α, k)-Anonymity (cf. page 51) is formed out of a combination of k-Anonymity and Confidence Bounding and combines the pros of both techniques.

Technology

Based on the two underlying techniques, the evaluation of (α, k)-Anonymity can be considered as a composition of both techniques. As both of them have identical properties regarding this evaluation, it is not a surprise that (α, k)-Anonymity has the same properties too.

The *Base of Security* is *Statistics*, as the probability with which an attacker could guess the real *Identity* is kept at a specified level by removing QIDs.

It is not possible to reverse this process for anybody, which results in *Reversibility* with the *Degree Not Reversible* and required *Cooperation*. Furthermore, *Information Theoretic Security* as the *Strength of Security* is assured. As with both other techniques, no *Trusted Third Party Participation* is needed, resulting in *No TTP Participation* at all as the *Background*, which means that a TTP is consulted in the *Phase None* and with *Frequency Never*. The Data Provider is the only participant that has the power of *Enforceability* resulting in *Enforceability* for the *Server* only.

Again, the *Affected Data Type* is *Stored* data only.

The *Means of Protection* takes place in form of *Indistinguishability* & *Unlinkability*, and because the *Protected Aspect* is the *Identity*, it results in *Indistinguishability*.

The *Protected Aspect* is again the *Identity*. Identity protection is performed by means of *One-Sided Anonymisation* as only the *Identity* of the Data Owner resulting in *One-Sided* as the *Directionality*.

Perspective

This technique supports the Data Provider in privacy preserving publishing of sensitive data. Therefore, the stakeholder is the Data Provider and the associated *Perspective* is the *Data Perspective*.

(k, e)-Anonymity

(k, e)-Anonymity (cf. page 52) is k-Anonymity applied to numerical values having a range of e. However, it does not take the distribution of values into account. The evaluation of (k, e)-Anonymity results in an identical result as for k-Anonymity regarding the investigated criteria. For this reason, only the respective values are provided. For further explanations, cf. the evaluation of k-Anonymity at page 57.

Technology

This technique uses *Statistics* as the *Base of Security*. The *Strength of Privacy* consists of *Enforceability* only by the *Server*. No *Trusted Third Party Participation* is desired resulting in the *Background No TTP Participation*, the *Phase None*, and *Frequency Never*. The *Reversibility*

has a *Degree* that is *Not Reversible* and where *Cooperation* is required. The offered *Strength of Security* is *Information Theoretic Security*.

(k, e)-Anonymity has only *Stored* data as the *Affected Data Type* and the *Server* has exclusive *Enforceability*.

The *Means of Protection* is based on *Indistinguishability* & *Unlinkability* - in more detail it uses *Indistinguishability* because the *Protected Aspect* is the *Identity*. Identity Protection is ensured by *Anonymity* with *One-Sided Directionality*.

Perspective

As explained in the evaluation of k-Anonymity the main stakeholder is the Data Provider and the associated *Perspective* is the *Data Perspective*.

(ϵ, m)-Anonymity

(ϵ, m)-Anonymity (cf. page 52) limits the probability with which one is able to guess values within range of ϵ to $\frac{1}{n}$.

Technology

Considering the description above, it is obvious that this approach makes use of statistical properties that should make it hard to de-anonymise the protected data. For this reason, the *Base of Security* is *Statistics*.

Regarding the *Reversibility*, it has to be considered as *Not Reversible* as the *Degree* and *Cooperating* is needed to recover the original data. As no possibility exists to influence this with any arbitrary amount of computing power, this technique offers *Information Theoretic Security* as the *Strength of Security*. The server is the only one who is able to decide about (ϵ, m)-Anonymity and influence it, so it has *Enforceability* for the *Server* only.

The *Affected Data Type* for this kind of privacy protection is *Stored* data only. For *Transmitted* or *Processed* data it would not be possible to apply this kind of protection (without preprocessing and overwriting stored data).

As the Data Provider is able to operate without any need for cooperation with other systems, the *Background* for the *Trusted Third Party Participation* is *No TTP Participation* and, therefore, in *Phase None* and with *Frequency Never*.

The *Protected Aspect* of this technique is the *Identity* and the *Base of Identity Protection* is *Anonymisation* because identifying information is removed. The *Directionality* is *One-Sided* as only the *Identity* of the Data Owner is hidden.

This results in *Indistinguishability* as the *Means of Protection* as it is the only possibility for the *Protected Aspect* of *Identity*.

Perspective

The main stakeholder is the Data Provider because it can use this technique to anonymise data and ensure privacy friendly data handling. This has to be considered as the reason why it is

permitted to publish confidential data. Therefore, the associated *Perspective* is the *Data Perspective*.

Personalized Privacy

An interesting approach is Personalized Privacy (cf. page 52) because the Data Owner specifies a confidence level (cf. confidence bounding). However, it may be difficult to specify the appropriate level of privacy as one does not know about the underlying distribution of values.

Technology

The *Base of Security* for this technique is *Statistics* because Personalized Privacy uses probabilities that the Data Owner could be identified as base for the specification of the security level.

This technique is flexible and, therefore, makes the evaluation somehow difficult. For example, the *Protected Aspect* may be - depending on the user's choice - either *None* (if no protection at all is desired) or *Identity* (*Anonymity*) (with varying strength) but always with *Directionality One-Sided*.

As a consequence, the *Strength of Security* which could generally be considered as *Information Theoretic Security* may also be *None* at all if the user decides that this is not required. Of course the *Reversibility* depends strongly on the desired privacy too and may have a *Degree* varying from *Not Reversible* with *Cooperation* required to *Fully Reversible* with no *Cooperation*.

However, the base technique is the same for every chosen privacy level and, therefore, no Trusted Third Party is involved in any case resulting in no *Trusted Third Party Participation* with *Phase None*, *Frequency Never*, and, of course, the *Background* is *No TTP Participation*.

The *Enforceability* does not depend on the chosen privacy level neither. It is obvious that the *Server* has the power to enforce Personalized Privacy. However, one may consider a scenario in which the *Client* provides only information at a level of detail that does not compromise his or her desired level of privacy. This may be hard if the underlying distribution of values is unknown. Under the assumption that the distribution itself is not sensitive, it can be used by a user. Thus he or she is able to decide if information is provided or not, resulting in *Enforceability* for *Both - Client* and *Server*.

The *Affected Data Type* is *Stored* data.

The *Means of Protection* is *Indistinguishability* & *Unlinkability*, or more precisely, *Indistinguishability* as the *Protected Aspect* is the *Identity*. Identity Protection works based on *Anonymisation* with *One-Sided Directionality*.

Perspective

As the Data Provider has an interest to use this technique to get the permission to publish/collect data, it has to be considered as the main stakeholder. Due to this description, it is clear that the *Data Perspective* is the associated *Perspective*.

t-Closeness

t-Closeness (cf. page 53) is a technique that intends to conserve the original distribution of data as good as possible in the published (anonymised) dataset.

Technology

As the distribution of data is taken into account, it becomes clear that t-Closeness has to use *Statistics* as the *Base of Security* in order to ensure *Anonymisation* as the *Base of Identity Protection*.

For the *Strength of Privacy* (namely the *Reversibility*, the *Trusted Third Party Participation*, the *Strength of Security*, and the *Enforceability*), it offers similar properties as most techniques based on *Statistics*. The *Degree* of *Reversibility* is *Not Reversible* and *Cooperation* is required to reverse the process because if it is applied and the QIDs are removed, it is not possible to return to the original dataset after applying the one way function of data removal regardless of available computing power (and no further background information), thus the *Strength of Security* level is *Information Theoretic Security*. t-Closeness can be applied to basically all *Stored* data (for which the original distribution is known), which results in the *Affected Data Type Stored*. *No TTP Participation* is required as the *Background* of *Trusted Third Party Participation* in the *Phase None*, resulting in the *Frequency Never*. The *Server* has the control of publishing data. If the user knew about the distribution of the original data, he or she could decide if he or she wants to provide data or not. However, this corresponds to Personalized Privacy and not t-Closeness. Thus, *Enforceability* is only possible for *Server*.

The *Means of Protection* is based on *Indistinguishability* & *Unlinkability* - in more detail it is *Indistinguishability*. This is because this technique is based on *Identity* as the *Protected Aspect*. Data is protected using *Anonymisation* with a *One-Sided Directionality*.

Perspective

The main stakeholder for this technique is the Data Provider because it publishes sensitive data. For this reason, the associated *Perspective* is the *Data Perspective*.

Direct Anonymous Attestation

Direct Anonymous Attestation (cf. page 43) is a technique used for privacy preserving remote authentification and for checking the possession of certain attributes.

Technology

Direct Anonymous Attestation is based on group signatures and protects privacy by preventing that more than the group is revealed. Therefore, the *Base of Security* is *Statistics* for identifying an entity as the probability of guessing the *Identity* decreases with increasing group size. For faking an *Identity* on the other hand, unauthorised creation of a valid signature is required, which is possible with a brute-force attack. Thus the *Base of Security* can either be *Statistics* or *asymmetric Cryptography*.

The *Strength of Security* is *Information Theoretic Security* for the identity protection because the allocation to the groups cannot be reversed if no explicit list is available. On the other hand, it is only *Computational Security* for the authentification (as argued before). The *Trusted Third Party Participation* takes place in the *Setup Phase*, and the *Frequency* is only *In specific scenarios*, namely to generate the group signature and issue the DAA credentials for the TPMs (Trusted Platform Module). The *Background* is not for performing operations and it is obvious that the *Background No TTP Participation* does not apply either. So the last remaining possibility is *Checks* which is accurate as the checks of the signatures are created using information provided by the TTP (note that it is involved in checks only indirectly). This scheme can only be enforced by the *Server*, which results in *Enforceability* for the *Server* only.

In this case *Reversibility* refers to identity protection and not to the authentification process itself, as this technique should be seen as an authentification technique that gets enhanced with identity protection. Thus, *Reversibility* refers to the identity protection but not to the authentification for the evaluation of Direct Anonymous Attestation. Depending on the specific technique used, the *Reversibility* may either have the *Degree Not Reversible* or even *Deniable Reversible*. In both cases, the *Cooperation* of the Data Owner is required to reverse the process.

As Direct Anonymous Attestation is an authentification scheme, it naturally has the *Affected Data Type Transmitted* data (the authentification information) and *Processed* data.

This approach uses *Pseudonymisation* as the *Base of Identity Protection*: the pseudonym is the assigned group.

The *Means of Protection* is based on *Indistinguishability & Unlinkability*. As the *Protected Aspects* are *Behaviour* and the *Identity*, *Indistinguishability* and *Unlinkability* are both satisfied. With respect to the *Behaviour*, the *Access Patterns* are hidden. This is because the actions are not hidden but cannot be assigned to a specific party (Indistinguishability). On the other hand, the identity protection is based on Pseudonyms where the *Holder* is a *Group* and the *Directionality* is *One-Sided* because only the *Identity* of the Data Consumer is hidden. The *Cardinality* is *unlimited* because usually no restriction exists from the technical point of view regarding the groups that could be joined (note that specific techniques/implementations may exist which are exceptions).

Perspective

The main stakeholder for Direct Anonymous Attestation is the user/TPM which is able to authenticate without the need not disclose the *Identity*. Therefore, the Data Provider as a stakeholder is excluded leading to the *User Perspective* as the associated *Perspective*.

Proxy Re-Encryption

Proxy Re-Encryption (cf. page 33) enables a semi-trusted third party to (re)encrypt data for another person (with another key) without getting knowledge about the plaintext.

Technology

As the name already suggests, Proxy Re-Encryption (PRE) is a cryptographic technique. It therefore has *Cryptography* as the *Base of Security* which is usually *asymmetric*. Although the possibility to use PRE with symmetric schemes exists, the preferred variant for high security is still the use of asymmetric schemes. For this reasons, *symmetric Cryptography* will not be considered for PRE (yet).

As a consequence of *Cryptography* as the *Base of Security*, this technique offers only *Computational Security* as the *Strength of Security*. For this protocol, (Semi)Trusted Third Parties may get involved. Usually, this would be a weakness, but in this case this possibility is, explicitly desired as it allows delegation and improved work flows and performance. Nevertheless, the strongest scenario is considered so a TTP is not considered as being required.

The *Trusted Third Party Participation* takes place in the *Phase None* with the *Frequency Never*. The *Background* is *No TTP Participation*.

PRE cannot be enforced by a single party - *Client* and *Server* need to cooperate leading to *Enforceability None*. Although this seems to be a weak *Enforceability* and that the *Server* could refuse to participate in the protocol to comprimise privacy, this is not correct. If the *Server* would not cooperate, then it could prevent the usage of PRE, which corresponds to a Denial of Service - however, privacy is not (directly) affected, as the *Server* does not learn anything about the plaintexts. Furthermore, it only is able to reduce convenience but if the user downloads the file, re-encrypts and afterwards uploads it again, the overall functionality is not changed (however, one faces, of course, degraded performance and convenience).

As supposed for a cryptographic paradigm, the *Reversibility* offers a *Degree* that is *Fully Reversible* because anything else would destroy rather than protect data. The process can be reversed, but this reversal cannot be enforced without the possession of the key (neglecting brute force attacks). Therefore, *Cooperation* from the Data Owner is required.

The *Affected Data Type* of PRE is *Stored* data because data is not specifically created during this operation but existing data converted. As the re-encryption requires operations on the encrypted data, *Processed* and *Combined* have to be added to the *Affected Data Type*.

As typical for encryption schemes, the *Means of Protection* of PRE are based on *Confidentiality*.

The *Protected Aspect* of PRE is the *Content* that cannot be read from unauthorised parties. Unfortunately, the *Metadata* remains unprotected and only the *Data* is protected.

Perspective

As no identity protection is used for this technique and the access is not protected either the only *Perspective* that fits is the *Profile Perspective* (*Content Perspective*). As a consequence, the stakeholder is considered to be the user who is able to outsource confidential data (with the need to process it).

Homomorphic Encryption

Homomorphic Encryption (cf. page 35) is a technique that allows a server to perform operations on encrypted data without learning any knowledge on the data.

Technology

As Homomorphic Encryption is a cryptographic technique, it relies on *Cryptography* (*symmetric* and *asymmetric*) as the *Base of Security*. Most techniques based on Homomorphic Encryption use asymmetric encryption, but variants based on symmetric encryption schemes exist too. For this reason, the *Cryptography* property has been evaluated as *symmetric* and *asymmetric*.

The *Strength of Security* is *Computational Security* as the encryption algorithm may be reversed with (sufficient) computing power.

The reason for implementing Homomorphic Encryption is that one wants to delegate operations to a Third Party without fully trusting it. However, Homomorphic Encryption may be applied to internal systems too in order to protect data even in case access control fails. Thus, *(Trusted) Third Party Participation* is optional. Considering the strongest scenario leads to *Trusted Third Party Participation* in the *Phase* of *None* with *Frequency Never*, and with *Background No TTP Participation*. Involving a TTP would usually be considered as a drawback. Nevertheless, it is (if external systems are used) desired and has definitely to be seen as a pro in this case because otherwise the use of external systems or performing operations on the encrypted data would be impossible.

Neither the *Client* nor the *Server* are able to enforce the usage of this technique, resulting in *Enforceability None* because cooperation between *Client* and *Server* is needed. If, however, the *Server* would not follow the specified protocol, it may prevent the use of Homomorphic Encryption but would not gain any information. Furthermore, in the end it is not possible to prevent the client from conducting the desired operations, as it could download the encrypted data, decrypt it, then perform all operations on the plaintext and encrypt and upload all data again. Although this would result in a loss of performance and convenience, it would not change the end result at all.

In order to allow decryption, *Reversibility* has to be offered (to be more precise the *Degree Fully Reversible*). Note that, however, (depending on the specific technique) the maximum number of operations that can be performed on a dataset may be limited. The reason for this is that the operations (especially multiplications) cause an increasing level of noise. If this limit of operations is exceeded, non-reversible side-effects will occur and will prevent proper decryption. This would result in changing the *Degree* of *Reversibility* to *Not Reversible* as no-one - not even the Data Owner could restore the original information. As this is not relevant if the algorithm is used within its specification (with not more operations than permitted), the evaluation will not take this possibility into account. In any case *Cooperation* is required for decryption.

Homomorphic Encryption may be applied to data that is stored and should be processed in some way, which is why the *Affected Data Type* is *Combined* (*Stored* and *Processed* data).

The *Means of Protection* is *Confidentiality* as this technique prevents unauthorised reading of the protected data.

Like in general for encryption techniques, the *Protected Aspect* of Homomorphic Encryption is the *Content* (to be more precise *Data*) which could not be read from unauthorised entities. *Metadata* like when an operation has been performed, however, is not protected.

Perspective

This technique prevents unauthorised access to data (when a need to perform operations on it exists) . However, it does not protect the user's access but the actual *Content* and, therefore, the associated *Perspective* is the *Profile Perspective* (*Content Perspective*). Due to this, the main stakeholder is the user who can outsource his or her data while still ensuring privacy and *Confidentiality*.

Searchable Encryption

A common problem preventing the use of encryption is that search is not possible anymore. Searchable Encryption (cf. page 36) solves this problem and allows searching within encrypted content without prior decryption.

Technology

As already suggested by the name, Searchable Encryption is a measure relying on *Cryptography* as the *Base of Security*. Schemes are based on *symmetric* and *asymmetric* encryption.

The use of a cryptographic paradigm already indicates the *Strength of Security* is *Computational Security*.

If the third party that stores confidential information would be trusted, then one may not face the obligation to encrypt data but this may indeed be required/recommended for a trusted third party in case of very sensitive information. By applying Searchable Encryption a (Trusted) Third Party is permitted to store sensitive information but one need not worry about privacy because data is protected. Nevertheless encryption can be applied on internal systems too as an additional security measure preventing a single point of failure. As the strongest scenario is considered for this evaluation, it is assumed that no TTP is involved. The resulting *Trusted Third Party Participation* takes place in the *Phase* of *None* and with the *Frequency Never* and the *Background* is *No TTP Participation*.

Unfortunately, this technique cannot be enforced by the *Server* nor the *Client* - they need to cooperate in order to establish this protocol. This leads to the *Enforceability None*; a non-cooperating server will not be provided with the sensitive data in the first place.

Encryption obviously is not intended to be a one way function, but *Reversibility* is required (otherwise the data would be destroyed instead of protected). The resulting *Degree* is *Fully Reversible* and *Cooperation* (decryption by the Data Owner) is required to reverse this process.

The *Affected Data Type* is *Combined* as *Stored* data and *Processed* data is affected. *Stored* data is affected because data that is stored on the server is protected using Searchable Encryption while search capabilities result in *Processed* data.

The *Means of Protection* is *Confidentiality* which is typical for cryptographic schemes.

The *Protected Aspect* of Searchable Encryption is the *Content* (to be more precise the *Data*) because sensitive information is hidden from unauthorised entities.

Perspective

As no kind of identity protection is used and access is not obfuscated, the only possible *Perspective* for this technique is the *Profile Perspective* (*Content Perspective*) - the server is prevented from accessing sensitive data. As a consequence, the main stakeholder is the user because his or her information is hidden on a system he or she does not trust.

Order Preserving Encryption

Order Preserving Encryption (cf. page 38) is a cryptographic paradigm ensuring that encrypted values are sorted in the same order as their plaintext equivalents.

Technology

As the name already suggests, the *Base of Security* for Order Preserving Encryption is *Cryptography* which may either be *symmetric* or *asymmetric*.

The *Strength of Security* level for this technique is only *Computational Security*.

At a high level, Order Preserving Encryption works just like ordinary encryption: the user has some sensitive value that he or she wants to protect and, therefore, he or she encrypts it (without involving any third party). Therefore, the *Phase* and *Frequency* of *Trusted Third Party Participation* are *None* respectively *Never* - no matter if the user actually decides to delegate this work to a TTP, it is not required by this technique. As a result the *Background* is *No TTP Participation*.

Unfortunately, neither the *Client* nor the *Server* is able to enforce this technique because cooperation between both is required leading to *Enforceability None*. Although the *Server* could prevent usage of this technique, privacy cannot be violated but maybe an alternative technique with less performance has to be used.

In order to enable decryption, full *Reversibility* is needed. A secure cryptographic algorithm, however, requires that it is not possible (feasibly) to enforce decryption without the will of the Data Owner. Therefore, the resulting *Degree* is *Fully Reversible* but, of course, *Cooperation* is required for decryption.

The *Means of Protection* is *Confidentiality* as this technique intends to prevent unauthorised access to information (or at least preventing any gain of knowledge from unauthorised access), while still being able to sort data.

The *Protected Aspect* of this technique is the *Content* but the focus is on *Data* only while *Metadata* is not taken into account.

Perspective

As no identity protection is used and the access is unprotected, the only possible *Perspective* left is the *Profile Perspective* (*Content Perspective*). This is correct as unauthorised access to the content is prevented for sensitive information (while still being able to build indexes and

perform range queries). Due to this the main stakeholder of Order Preserving Encryption is the user.

Anonymous Credentials

Anonymous Credentials (cf. page 53) offer the possibility to enforce privacy preserving authentification and accountability. Two basic technologies exist that are based on Blind Signature (cf. the respective evaluation) and zero-knowledge proofs. As these basic technologies have different properties, it is not possible to classify this technique precisely with respect to all properties. For this reason, they are examined independently. Note that in the summary further distinctions exist in order to cover the chain of techniques. If, however, all variants have equal or very similar properties, then no distinction is required.

Anonymous Credentials - Blind Signature

As this variant of Anonymous Credentials builds on Blind Signatures, the properties do not change. Therefore, a separate evaluation is not required, for the results cf. sections Blind Signature - Fair Blind at page 79 and Blind Signature - Partially Blind at page 80.

Note that, however, the *Protected Aspect* is the *Behaviour* (to be more precise the *Access Patterns*) rather than the *Content*. Although Blind Signatures protect the *Content*, this is not the primary aim of Anonymous Credentials. When applying Anonymous Credentials, the protected *Content* is the value of a sensitive attribute - it is only checked if some condition applies (Data Minimisation). In this sense, the *Content* is protected but all further *Content* that is transferred between the parties stays unprotected. For this reason, the *Content* is evaluated as unprotected. On the other hand, the *Protected Aspect Behaviour* (*Access Patterns*) is protected under the condition that *Anonymity* is satisfied so that sessions cannot be linked which implies that it is not possible to observe the *Behaviour*.

Anonymous Credentials - Zero Knowledge

Technology

The *Base of Security* of Anonymous Credentials based on Zero Knowledge is *Statistics*. Basically everyone could be able to deal with the challenge by just sending a random reply. However, it should be designed in a way that the probability of guessing the correct reply is very low and close to 0. This results in *Information Theoretic Security* as the *Strength of Security*.

In the strongest scenario, no Trusted Third Party is required resulting in *TTP Participation* with *Frequency Never*, the *Phase* is *None* and the *Background* is *No TTP Participation*. *Enforceability* is possible for *Both* parties as they could refuse any other form of authentification.

For Anonymous Credentials *Reversibility* refers to identity protection and not to the authentification process itself. The core feature of this technique is to augment the authentification process with privacy protection. Thus, *Reversibility* refers to the identity protection but not to the authentification itself. The *Reversibility* is not possible resulting in a *Degree* of *Not Reversible*. For some techniques *Cooperation* is required to inverse the protection while some techniques

do not require *Cooperation*. Thus, *Cooperation* is evaluated (depending on the technique) as required as well as not required.

As typical for authentification schemes, *Processed* and *Transmitted* data is affected by this technique resulting in *Combined* as the *Affected Data Type*. The *Protected Aspect* is the *Identity* by means of *Anonymisation* with a *One-Sided Directionality* as the other party (e.g. the server) has no identity protection.

As a result, the *Means of Protection* builds on *Indistinguishability* and *Unlinkability* as the *Identity* remains hidden, which prevents linking different sessions of the same user too.

Perspective

The main stakeholder of this technique is the user. As already indicated by the name, the aim is to protect the *Identity* and, therefore, the associated *Perspective* is the *User Perspective*. Furthermore, the *Profile Perspective* (*Access Perspective*) is possible too (additionally). However, this depends on the technique used as the base for Anonymous Credentials. As this evaluation intends to provide a good overview of the possibilities a technique offers, AC is considered to cover both Perspectives (*Access Perspective* and *User Perspective*).

Deniable Encryption

Deniable Encryption (cf. page 40) provides the possibility to reverse the encryption operation in two different ways. The resulting plaintext may either be the original plaintext or another alternative fake plaintext.

Technology

As it is a cryptographic technique, the *Base of Security* for Deniable Encryption is *Cryptography*.

Deniable Encryption shows typical properties for cryptographic schemes: *Trusted Third Party Participation* is not desired causing the *Phase* to be *None* and the *Frequency* to be *Never*. As a result, the *Background* is *No TTP Participation* and the *Base of Security* is *Cryptography* which may be *symmetric* and *asymmetric*. However, the *Strength of Security* is, in contrast to most cryptographic schemes, *Information Theoretic Security*. Although the encryption can be broken using a brute force attack, various plaintexts can be decrypted. One of them is the real plaintext whereas the fake plaintext will be decrypted too with the respective key resulting in at least 2 plaintexts. Thus, it is only possible to get knowledge of the (real) plaintext with some probability which per definition leads to *Information Theoretic Security*.

Deniable Encryption is enforceable only by the sender of the information which is considered to be the *Client*. It can encrypt its message using Deniable Encryption - the receiver has, however, no possibility to decide whether a message he or she gets is encrypted (deniably) or delivered as plaintext. This results in the *Enforceability* for the *Client* only.

In order to be able to decrypt the ciphertext, this technique has to offer *Reversibility*. However, it offers not only the possibility to decrypt the ciphertext to one but to two different plaintexts, which results in *Deniable Reversible* as the *Degree*. However, it depends on the specific

technique if only the sender or receiver or both of them are able to deny the real plaintext. *Cooperation* is required to decrypt the text. Note that, however, one is not able to judge if the real plaintext or the fake plaintext has been decrypted - even if the Data Owner pretends to cooperate. With real cooperation, however, it is not a problem at all to decrypt the real plaintext.

The *Affected Data Type* is *Combined* as this technique may be applied to *Stored* and (especially when considering email or other forms of communication) *Transmitted* data.

The *Means of Protection* is, as already indicated by the name, *Deniability* as the user can deny a plaintext, which is per definition the contrary of accountability (leading to deniability). Additionally, *Confidentiality* applies too as the data is kept secret - even in case of enforced decryption.

The *Protected Aspect* is the *Content* that is encrypted. However, only the *Data* but not the *Metadata* is protected.

Perspective

As for this technique, no identity or access protection exists; the *Profile Perspective* (*Content Perspective*) is the only possible *Perspective*. Because the *Content* is protected, this is appropriate. This implies that the main stakeholder is the user: he or she is able to protect himself or herself by means of decrypting a ciphertext to an alternative plaintext.

Oblivious Transfer

Oblivious Transfer (cf. page 41) is a technique that enables a client to retrieve data from a server while at the same moment preventing the server to learn anything about which data is actually transferred. The various techniques are based on Homomorphic Encryption (cf. page 35) or Anonymous Credentials (cf. page 53), so the evaluation is based on these two techniques too. In the summary, more variants are distinguished in order to cover the chain of techniques when necessary.

Technology

Due to the broad variety of techniques for Oblivious Transfer, the *Base of Security* may either be *Statistics* or *Mathematical Problems* (techniques based on Anonymous Credentials), *asymmetric Cryptography* (techniques based on Anonymous Credentials or Homomorphic Encryption) or *symmetric Cryptography* (Homomorphic Encryption). Accordingly the *Strength of Security* could be *Information Theoretic Security* (Anonymous Credentials) as well as *Computational Security* (Anonymous Credentials and Homomorphic Encryption).

Regarding the *Trusted Third Party Participation*, again two possibilities exist: the *Frequency* may be *Never* and the *Background* may be *no TTP Participation* (Homomorphic Encryption) or the *Frequency* may be *In specific scenarios/Never* and the *Background* may be *Checks* or *No TTP Participation* (Anonymous Credentials). The *Phase* may be *Operations* (Anonymous Credentials) or *None* (Anonymous Credentials and Homomorphic Encryption).

Enforceability is - depending on the used technique and the respective base technique - either *Both* (Anonymous Credentials) or *None* (Homomorphic Encryption).

The *Reversibility* refers to the ability of reversing the privacy protection. In case of Oblivious Transfer this means that the server is able to find out which records have been transferred. *Reversibility* may have a *Degree* of *Partially Reversible* (Anonymous Credentials) or *Fully Reversible* (Anonymous Credentials and Homomorphic Encryption). In case of Anonymous Credentials *Cooperation* is not required while in case of Homomorphic Encryption *Cooperation* is required. Thus, *Cooperation* is evaluated to required as well as not required, depending on the specific technique used.

However, some fixed attributes exist too: the *Affected Data Type* is in any case *Combined* as *Stored* data and *Transmitted* data is affected. Actually it is hard to distinguish these data types in this case because the whole protocol is about the transmission of *Stored* data (without the server knowing which data the user is interested in).

The *Protected Aspect* is the *Behaviour* as the *Access Patterns* are hidden from the server as described above. One may think of the *Content* being a *Protected Aspect* as well but this is not true as the *Content* is not hidden from the sever nor the client. For this reason, only the *Behaviour* is a *Protected Aspect*. In case of Unlinkable POT and AC-OT, the *Identity* is protected too by means of *Pseudonymisation* as the process could generally be reversed in case of dispute. The *Directionality* of *Pseudonymity* is *One-Sided* as the server's real *Identity* is known. The *Holder* is an *Individual* as it is not shared, but it is possible to use an arbitrary number of Pseudonyms (one for each transaction) so the *Cardinality* is *Unlimited*.

The *Means of Protection* is *Indistinguishability* & *Unlinkability* (to be more precise *Unlinkability*) because it is not possible to correlate two transactions and tell if they affect the same data or not (which matches the definition of *Unlinkability*). In most schemes, *Deniability* is covered too: it cannot be told if some data actually has been retrieved. An exception to this is POT. However, usually the possibility of *Deniability* exists so it is considered in this evaluation too.

Perspective

The affected stakeholder is the user, and it is intuitive by the description of this technique that the *Profile Perspective* (*Access Perspective*) is the associated *Perspective* as the accessed data is hidden from the server. For Unlinkable POT and AC-OT, the *User Perspective* is supported too as the *Identity* is hidden from the server.

Private Information Retrieval

Private Information Retrieval (PIR) is a scheme that allows a client to hide what records are retrieved from a server. 3 main concepts for PIR exist: the trivial solution (downloading the whole database), single server computational secure PIR, and multi server information theoretic secure PIR (cf. page 44). All 3 variants are evaluated separately because, due to their fundamental differences, they have a broad variety of attributes. This separation is necessary to give a realistic and complete view on PIR. For a better overview, the differences of all 3 concepts are presented in separate sections, while the attributes that are equal for all of them are evaluated in this general section.

Technology

Trusted Third Party Participation is not required nor desired for PIR and, therefore, the affected *Phase* is *None* and the *Frequency* is *Never*, and, as a consequence, the *Background* is *no TTP Participation*.

The *Affected Data Type* is *Stored*. Although one intuitively may consider the *Affected Data Type Transmitted* too, this is not correct because PIR only hides the selection process of the desired record from the server, but in general does not protect the *Transmitted* data in any special way.

The *Means of Protection* builds on *Unlinkability* as neither the *Identity* is protected (*Indistinguishability*) nor *Deniability* is intended but the accessed *Content* is protected.

The intent of this technique is to hide what has been accessed and not by whom some record has been accessed. As a result, the *Identity* of the user (if not protected otherwise) is disclosed to the server while the *Protected Aspect Behaviour* (to be more precise *Access Patterns*) remains hidden.

Perspective

The *Perspective* of PIR is clearly the *Profile Perspective* (*Access Perspective*), as the stakeholder is the user and the accessed *Content* is hidden while no intent to hide the *Identity* exists.

Private Information Retrieval - Trivial Solution

Technology

The *Base of Security* is *Statistics* as the whole database with n items is downloaded. Therefore, the server knows with probability 1/n in which record the client is interested in (assuming that it is interested in exactly one record).

As indicated above, the *Strength of Security* is *Information Theoretic Security*. As argued above, *Reversibility* refers to the server's ability to find out about the data retrieved by the client. Thus, the *Reversibility* has the *Degree Not Reversible* because no way of finding out which records a user accessed using his or her local copy of the database exists except of asking the user (causing that *Cooperation* is required).

Regarding *Enforceability*, it is obvious that the client is able to enforce trivial PIR by requesting all data (assuming that this is permitted by the server). However, the server is able to enforce trivial PIR too if it does not offer any possibilities for filtering data (in this case all data is sent to the client, which corresponds to trivial PIR). Therefore, *Enforceability* is possible for *Both* (*Client* and *Server*) in case of trivial PIR.

Private Information Retrieval - Single Server with Computational Security Solution

Technology

The *Base of Security* are *Mathematical Problems* like the Quadratic Residuosity Problem that cannot be solved easily by the server, so it cannot identify which records the user is interested in.

Another possibility would be the implementation by means of Homomorphic Encryption, which changes the *Base of Security* to cryptography. For this reason, the *Base of Security* is evaluated as both, *Mathematical Problems* and *Cryptography*.

The resulting *Strength of Security* is *Computational Security*.

Regarding the *Reversibility*, the *Degree* is *Fully Reversible*. *Cooperation* of the client is required to reverse this technique.

The *Enforceability* by the *Server* is obvious: it just provides access to the data in a way specified by a PIR protocol and only clients following this protocol are able to access data. However, the *Client* is able to enforce (e.g. using Homomorphic Encryption) this technique too. Therefore, *Enforceability* is evaluated to *Both* (*Client* and *Server*).

Private Information Retrieval - Multiple Servers with Information Theoretic Security Solution

Technology

PIR involving multiple servers usually builds on the assumption that a maximum number of servers that is below the total number of servers are allowed to collude. Therefore, a threshold is defined stating the maximum number of servers that could collude without harming privacy. It is clear that these assumptions lead to a fragile security model. For this reason, for some of the criteria the alternative scenario (too many servers collude) is taken into account too, in order to provide a realistic evaluation. Note that the final result is based on the version as specified in the techniques (and, assuming that not too many servers collude).

Like in the trivial concept, the *Base of Security* in this concept is *Statistics* too. As indicated above, the *Strength of Security* is *Information Theoretic Security* (in case of colluding servers, it would be *Computational Security*).

Regarding *Enforceability*, this technique is a powerful tool that can be enforced from the *Client* as well as the *Server*. The *Client* can enforce this technique by requesting data from multiple servers. The *Server* can enforce it e.g. by introducing an intermediate proxy that requests data from multiple servers. As a consequence the proxy is the client for the *Servers* but serves as a server for the actual *Client*. The behaviour of the overall system does not change at all - the only difference is that the address of the proxy is used by the *Client* and the *Server*. Thus, the *Enforceability* is *Both* (*Client* and *Server*).

The *Reversibility* is of *Degree Not Reversible* and, if the assumption of non colluding servers holds, not enforceable without the users consent (therefore, *Cooperation* is required).

Blind Signature

Blind Signature (cf. page 45) is a technique that allows the creation of a signature without the signer knowing, what he or she actually signs and, therefore, preventing him or her from acquiring knowledge about the content. Two basic approaches for Blind Signature exist: Fair Blind Signatures and Partially Blind Signatures. As both approaches have different manifestations for some of the characteristics, the evaluation is split up. Common properties that are identical for both approaches are evaluated in this section, while properties that differ are evaluated in sepa-

rate sections for both approaches. As they have few common properties (if partial intersections are neglected), most of them are evaluated in separate sections.

Technology

Only two properties are identical for both approaches: the *Enforceability* and the *Affected Data Type*. The other technological properties are distinct.

The *Enforceability* is in general possible for both participating parties - the *Client* and the *Server* - because it does not matter where the signature is created; it will always be blind but only the creator of the signature has the power to decide how he or she actually does it. However, it is really hard to find a use-case where the *Client* creates a signature but must not know the content. Thus, this possibility is neglected and *Enforceability* is evaluated as only being possible for the *Server*.

The *Affected Data Type* is *Combined* as Blind Signatures (as well as Signatures in general) could be applied to *Stored* and *Transmitted* as well as *Processed* data (depending on the use case).

Perspective

The stakeholder and the *Perspective* are equal for both approaches. Both of them do not handle identity protection nor the obfuscation of the accessed data. For this reason, only the *Profile Perspective* (*Content Perspective*) is possible as the associated *Perspective*, which is appropriate as access to the content is not possible. Therefore, the main stakeholder of this technique is the user.

Blind Signature - Fair Blind

In case of Fair Blind Signatures, a Trusted Third Party exists that is able to recover the *Identity* in case of dispute. For Fair Blind Signatures, a broad variety of different techniques with varying properties exists. Therefore, for a majority of the attributes, more than one manifestation has to be considered. As described, the base technique is Oblivious Transfer so the evaluation will mainly be based on its properties.

Technology

The *Base of Security* may either be *Statistics*, *Mathematical Problems*, or *Cryptography* (*asymmetric* or *symmetric*).

Reversibility refers to the *Protected Aspects Content* and *Identity* stating if one is able to find out about the user requesting a signature and about the content that has been signed. Several possibilities for the *Degrees of Reversibility* exist, ranging from *Fully Reversible* to *Partially Reversible* - depending if a session can be linked to the user's identity. However, no *Cooperation* from the user is required to inverse the process.

Accordingly the *Protected Aspect* is clearly the *Content* (*Data*). Furthermore, it may additionally be the *Identity* by means of *Pseudonymity*, depending on the possibility to link the session to a user. In this case, the *Pseudonymity* has *One-Sided Directionality* and could be used

with *Unlimited Cardinality*. Regarding the *Holder*, no general statement is possible because the pseudonym could be held by *Individuals* as well as by *Groups*.

As mentioned above, the *Trusted Third Party Participation* is a necessity to deal with disputes, which takes place in the *Phase* of *Operation* but with the *Frequency* only *In specific scenarios* (which is dispute). As indicated, the *Background* are *Checks* that are performed in case of disputes only.

No other possibility (regardless of available computing resources) exists to reverse privacy protection than consulting the TTP. Thus the *Strength of Security* is, depending on the technique, *Computational Security* or even *Information Theoretic Security* (if the possibilities of the TTP are not taken into account, which is the default case for well behaving users) can be guaranteed. On the other hand, it is possible for unauthorised users to create a valid signature with a brute-force attack. Thus, the signature itself offers only *Computational Security*.

The *Means of Protection* is based on *Indistinguishability & Unlinkability*, and because the *Protected Aspects* is *Content* and the *Identity* both concepts are applied (note that the sessions of one user cannot be related to each other). The *Identity* is protected by the session and, therefore, the *Holder* of the pseudonym is an *Individual* as identifiable information remains. This may be used by the TTP to uncover a misbehaving user.

Blind Signature - Partially Blind

In case of Partially Blind Signatures, the signer and signee have to agree on the value of certain attributes.

Technology

As the basic technology is public key cryptography, the evaluation shows typical properties for cryptographic measures and the *Base of Security* is *Cryptography* (*asymmetric*).

The *Strength of Security* is only *Computational Security* and as the technique is based on *Cryptography*, it has *Reversibility* resulting in *Degree* that is *Fully Reversible*. In general no *Cooperation* from the user is required to reverse the protection, which could be exploited in case of misuse. No need to involve any third party exists resulting in the *Trusted Third Party Participation* in the *Phase None*, the *Frequency Never*, and the *Background No TTP Participation*.

The *Protected Aspect* is the *Identity* by means of *Anonymisation* with *One-Sided Directionality* and the *Content* (*Data*) because this technique is used to protect the *Identity* of the signee and the signed content.

The *Means of Protection* is based on *Indistinguishability* & *Unlinkability* where both approaches are covered: the *Protected Aspect* of *Identity* (*Indistinguishability*) and the signed content (*Unlinkability*) except of attributes that the signee explicitly does not want to be protected attributes (which is an interesting feature as he or she controls it on his or her own).

5.2 Summary

In this section, the evaluation is summed up in form of tables. As markers *x* is used for indicating that a certain property is present for a technique or the respective cell is left blank otherwise.

However, techniques exist that could be based on more than one approach, so the underlying properties can change too. In this case, the affected properties are marked with *o* indicating that this property may be satisfied but need not be satisfied in every case. In exceptions, text is used as marker to keep the tables simple. For numeric properties, a number or * is used indicating that an arbitrary number is possible.

Technology

Table 5.1: Overview of Techniques - Protected Aspect

Technique	Identity	Identity-Directionality	Identity-Anonymisation	Identity-Pseudonymisation	Identity-Pseudonymisation-Holder	Identity-Pseudonymisation-Cardinality	Behaviour	Behaviour-Access Patterns	Behaviour-Action	Behaviour-Perception	Content	Content-Data	Content-Meta Data
k-Anon	x	one sided	x										
MultiR k-Anon	x	one sided	x										
Distinct l-Div.	x	one sided	x										
Entropy l-Div.	x	one sided	x										
(c,l)-Div.	x	one sided	x										
Conf. Bounding	x	one sided	x										
(α, k)-Anon	x	one sided	x										
(X,Y)-Priv.	x	one sided	x										
(k, e)-Anon	x	one sided	x										

Continued on next page

Table 5.1 – continued from previous page

Technique	Identity	Identity-Directionality	Identity-Anonymisation	Identity-Pseudonymisation	Identity-Pseudonymisation-Holder	Identity-Pseudonymisation-Cardinality	Behaviour	Behaviour-Access Patterns	Behaviour-Action	Behaviour-Perception	Content	Content-Data	Content-Meta Data
(ϵ, m)-Anon	x	one sided	x										
Pers. Priv.	o	one sided	x										
t-Closeness	x	one sided	x										
DAA	x	one sided		x	*	*	x	x					
PRE											x	x	
Hom. Enc.											x	x	
Searchable Enc.											x	x	
OPE											x	x	
AC	x	one sided	o	o	*	*	o	o					
AC 0 Know.	x	one sided	o	o	*	*	o	o					
AC Fair Blind Sig. (OT)	x	one sided		x	1	*	x	x					
AC Fair Blind Sig. (Math. Problems)	x	one sided	o	o	*	*							

Continued on next page

Table 5.1 – continued from previous page

Technique	Identity	Identity-Directionality	Identity-Anonymisation	Identity-Pseudonymisation	Identity-Pseudonymisation-Holder	Identity-Pseudonymisation-Cardinality	Behaviour	Behaviour-Access Patterns	Behaviour-Action	Behaviour-Perception	Content	Content-Data	Content-Meta Data
AC Part. Blind Sig.	x	one sided	x										
Deniable Enc.											x	x	
OT	x	one sided		x	1	*	x	x					
PIR trivial							x	x					
PIR single server							x	x					
PIR multi server							x	x					
Fair Blind Sig.	x	one sided		x	1	*					x	x	
Part. Blind Sig.	x	one sided	x								x	x	

Table 5.2: Overview of Techniques - Means of Protection

Technique	Deniability	Indistinguishability	Unlinkability	Confidentiality
k-Anon		x		
MultiR k-Anon		x		
Distinct l-Div.		x		
Entropy l-Div.		x		
(c,l)-Div.		x		
Conf. Bounding		x		
(α, k)-Anon		x		
(X,Y)-Priv.		x		
(k, e)-Anon		x		
(ϵ, m)-Anon		x		
Pers. Priv.		x		
t-Closeness		x		
DAA		x	x	
PRE				x
Hom. Enc.				x
Searchable Enc.				x
OPE				x
AC		x	x	
Deniable Enc.	x			x
OT			x	
PIR trivial			x	
PIR single server			x	
PIR multi server			x	
Fair Blind Sig.		x	x	
Part. Blind Sig.		x	x	

Table 5.3: Overview of Techniques - Affected Data Type

Technique	Combined	Stored	Transmitted	Processed
k-Anon		x		
MultiR k-Anon		x		
Distinct l-Div.		x		
Entropy l-Div.		x		
(c,l)-Div.		x		
Conf. Bounding		x		
(α, k)-Anon		x		
(X,Y)-Priv.		x		
(k, e)-Anon		x		
(ϵ, m)-Anon		x		
Pers. Priv.		x		
t-Closeness		x		
DAA	x		x	x
PRE	x	x		x
Hom. Enc.	x	x		x
Searchable Enc.	x	x		x
OPE	x	x		x
AC	x	x	x	
Deniable Enc.	x	x	x	
OT	x	x	x	
PIR trivial		x		
PIR single server		x		
PIR multi server		x		
Fair Blind Sig.	x	x	x	
Part. Blind Sig.	x	x	x	

Table 5.4: Overview of Techniques - Base of Security

Technique	Statistics	Cryptography	Cryptography - symmetric	Cryptography - asymmetric	Mathematical Problems
k-Anon	x				
MultiR k-Anon	x				
Distinct l-Div.	x				
Entropy l-Div.	x				
(c,l)-Div.	x				
Conf. Bounding	x				
(α, k)-Anon	x				
(X,Y)-Priv.	x				
(k, e)-Anon	x				
(ϵ, m)-Anon	x				
Pers. Priv.	x				
t-Closeness	x				
DAA	o	o		o	
PRE		x		x	
Hom. Enc.		x	o	o	
Searchable Enc.		x	o	o	
OPE		x	o	o	
AC	o	o		o	o
AC 0 Know.	x				
AC Fair Blind Sig. (OT)		x		x	
AC Fair Blind Sig. (Math. Problems)					x

Continued on next page

Table 5.4 – continued from previous page

Technique	Statistics	Cryptography	Cryptography - symmetric	Cryptography - asymmetric	Mathematical Problems
AC Part. Blind Sig.		x		x	
Deniable Enc.		x	x	x	
OT	o	o	o	o	o
OT (Hom. Enc.)		x	o	o	
OT (AC with Blind Sig.)		o	o	o	o
OT (AC with 0 Know.)	x				
PIR trivial	x				
PIR single server		o	o	o	o
PIR multi server	x				
Fair Blind Sig.	o	o	o	o	o
Fair Blind Sig. (Math. Problems)					x
Fair Blind Sig. (OT based on AC-0 Know.)	x				
Fair Blind Sig. (OT based on Hom. Enc.)		x	o	o	
Part. Blind Sig.		x		x	

Note that for table 5.5 the *Strength of Security* for Personalized Privacy is marked with *o* indicating that it may but need not be *Information Theoretic Security* but may also be not protected at all. For more information on this please cf. section Personalized Privacy at page 66.

Table 5.5: Overview of Techniques - Strength of Privacy (Strength of Security)

Technique	Computational	Information Theoretic
k-Anon		x
MultiR k-Anon		x
Distinct l-Div.		x
Entropy l-Div.		x
(c,l)-Div.		x
Conf. Bounding		x
(α, k)-Anon		x
(X,Y)-Priv.		x
(k, e)-Anon		x
(ϵ, m)-Anon		x
Pers. Priv.		o
t-Closeness		x
DAA	o	o
PRE	x	
Hom. Enc.	x	
Searchable Enc.	x	
OPE	x	
AC	o	o
AC 0 Know.		x
AC Fair Blind Sig. (OT)	x	
AC Fair Blind Sig. (Math. Problems)	x	
AC Part. Blind Sig.	x	
Deniable Enc.	x	
OT	o	o
OT (Hom. Enc.)	x	
OT (AC with Blind Sig.)	x	

Continued on next page

Table 5.5 – continued from previous page

Technique	Computational	Inf. Theoretic
OT (AC with 0 Know.)		x
PIR trivial		x
PIR single server	x	
PIR multi server		x
Fair Blind Sig.	o	o
Fair Blind Sig. (Math. Problems)	x	
Fair Blind Sig. (OT based on AC-0 Know.)		x
Fair Blind Sig. (OT based on Hom. Enc.)	x	
Part. Blind Sig.	x	

Table 5.6: Overview of Techniques - Strength of Privacy (TTP Participation)

Technique	Frequency - Always	Frequency - In specific scenarios	Frequency - Never	Phase - Setup	Phase - Operation	Phase - None	Background - Operations	Background - Checks	Background - No TTP Participation
k-Anon			x			x			x
MultiR k-Anon			x			x			x
Distinct l-Div.			x			x			x
Entropy l-Div.			x			x			x
(c,l)-Div.			x			x			x

Continued on next page

Table 5.6 – continued from previous page

Technique	Frequency - Always	Frequency - In specific scenarios	Frequency - Never	Phase - Setup	Phase - Operation	Phase - None	Background - Operations	Background - Checks	Background - No TTP Participation
Conf. Bounding			x			x			x
(α, k)-Anon			x			x			x
(X,Y)-Priv.			x			x			x
(k, e)-Anon			x			x			x
(ϵ, m)-Anon			x			x			x
Pers. Priv.			x			x			x
t-Closeness			x			x			x
DAA		x		x				x	
PRE			x			x			x
Hom. Enc.			x			x			x
Searchable Enc.			x			x			x
OPE			x			x			x
AC		o	o		o	o		o	o
AC 0 Know.		o	o		o	o		o	o
AC Fair Blind Sig. (OT)		o	o		o	o		o	o
AC Fair Blind Sig. (Math. Problems)		o	o		o	o		o	o
AC Part. Blind Sig.			x			x			x
Deniable Enc.			x			x			x
OT	o	o	o		o	o	o	o	o
OT (Hom. Enc.)			x			x			x
OT (AC with Blind Sig.)		o	o		o	o		o	o
OT (AC with 0 Know.)		o	o		o	o		o	o

Continued on next page

Table 5.6 – continued from previous page

Technique	Frequency - Always	Frequency - In specific scenarios	Frequency - Never	Phase - Setup	Phase - Operation	Phase - None	Background - Operations	Background - Checks	Background - No TTP Participation
PIR trivial			x			x			x
PIR single server			x			x			x
PIR multi server			x			x			x
Fair Blind Sig.		x				x		x	
Part. Blind Sig.			x			x			x

Table 5.7: Overview of Techniques - Strength of Privacy (Enforceability)

Technique	None	Client	Server	Both
k-Anon			x	
MultiR k-Anon			x	
Distinct l-Div.			x	
Entropy l-Div.			x	
(c,l)-Div.			x	
Conf. Bounding			x	
(α, k)-Anon			x	
(X,Y)-Priv.			x	
(k, e)-Anon			x	
(ϵ, m)-Anon			x	
Pers. Priv.				x
t-Closeness				x
DAA				x
PRE	x			
Hom. Enc.	x			
Searchable Enc.	x			
OPE	x			

Continued on next page

Table 5.7 – continued from previous page

Technique	None	Client	Server	Both
AC				x
Deniable Enc.	x			
OT	o			o
OT (Hom. Enc.)	x			
OT (AC with Blind Sig.)				x
OT (AC with 0 Know.)				x
PIR trivial				x
PIR single server				x
PIR multi server				x
Fair Blind Sig.				x
Part. Blind Sig.				x

Table 5.8: Overview of Techniques - Strength of Privacy (Reversibility)

Technique	Degree - Fully Reversible	Degree - Partially Reversible	Degree - Not Reversible	Degree - Deniable Reversible	Cooperation
k-Anon			x		x
MultiR k-Anon			x		x
Distinct l-Div.			x		x
Entropy l-Div.			x		x
(c,l)-Div.			x		x
Conf. Bounding			x		x
(α, k)-Anon			x		x
(X,Y)-Priv.			x		x

Continued on next page

Table 5.8 – continued from previous page

Technique	Degree - Fully Reversible	Degree - Partially Reversible	Degree - Not Reversible	Degree - Deniable Reversible	Cooperation
(k, e)-Anon			x		x
(ϵ, m)-Anon			x		x
Pers. Priv.	o		o		o
t-Closeness			x		x
DAA			o	o	x
PRE	x				x
Hom. Enc.	x				x
Searchable Enc.	x				x
OPE	x				x
AC	o	o			o
AC 0 Know.	o	o			o
AC Fair Blind Sig. (OT)	o	o			o
AC Fair Blind Sig. (Math. Problems)	o	o			
AC Part. Blind Sig.	o	o			
Deniable Enc.				x	x
OT	o	o			o
OT (Hom. Enc.)	x				x
OT (AC with Blind Sig.)	o	o			
OT (AC with 0 Know.)	o	o			o
PIR trivial			x		x
PIR single server	x				x
PIR multi server	x				x
Fair Blind Sig.	o	o			

Continued on next page

Table 5.8 – continued from previous page

Technique	Degree - Fully Reversible	Degree - Partially Reversible	Degree - Not Reversible	Degree - Deniable Reversible	Cooperation
Fair Blind Sig. (Math. Problems)	x				
Fair Blind Sig. (OT based on AC-0 Know.)	o	o			
Fair Blind Sig. (OT based on AC-Hom. Enc.)	x				
Part. Blind Sig.	x				

Perspective

Table 5.9: Overview of Techniques by Perspective

Technique	Data Perspective	User Perspective	Profile Perspective	Content Perspective	Access Perspective
k-Anon	x				
MultiR k-Anon	x				
l-Div.	x				
Entropy l-Div.	x				
(c,l)-Div.	x				
Conf. Bounding	x				
(α, k)-Anon	x				
(X,Y)-Priv.	x				

Continued on next page

Table 5.9 – continued from previous page

Technique	Data Perspective	User Perspective	Profile Perspective	Content Perspective	Access Perspective
(k, e)-Anon	x				
(ϵ, m)-Anon	x				
Pers. Priv.	x				
t-Closeness	x				
DAA		x			
PRE			x	x	
Hom. Enc.			x	x	
Searchable Enc.			x	x	
OPE			x	x	
AC		x	o		o
AC 0 Know.		x	o		o
AC Fair Blind Sig. (OT)		x	o		o
AC Fair Blind Sig. (Math. Problems)		x	o		o
AC Part. Blind Sig.		x	o		o
Deniable Enc.			x	x	
OT		o	x		x
PIR - trivial			x		x
PIR - single server			x		x
PIR - multi server			x		x
Fair Blind Sig.			x	x	
Part. Blind Sig.			x	x	

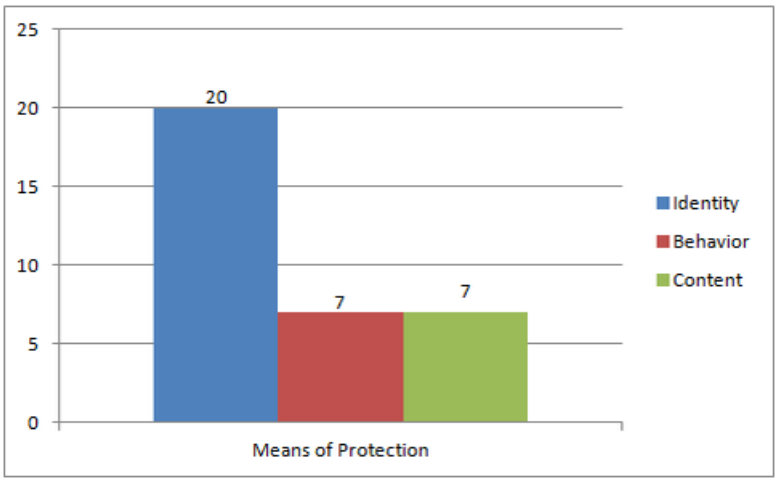


Figure 5.1: Statistics - Protected Aspect

Number of Occurrences per Category

In this section, the number of occurrences of the attributes of the taxonomy is summed up. Note that one technique may have several properties satisfying more than one aspect of the taxonomy. Furthermore, there exist techniques that have varying properties depending on the base technique used. These techniques are split so that a more fine grained evaluation is possible, which is considered in calculating the various sums.

Regarding the *Protected Aspect*, figure 5.1 shows an overview of the evaluation: the *Identity* is protected 20 times while *Behaviour* and *Content* are protected only 7 times (cf. figure 5.1). This is mainly caused by the block of 12 techniques from the field of Privacy Preserving Data Publishing. Examining the *Protected Aspect* in more detail shows that *Anonymisation* (16 times) is overrepresented in comparison to *Pseudonymisation* (6 times). Surprisingly not one single technique exists offering Directionality Two-Sided. This is probably caused by the assumption that the other entity (the Data Consumer in case of the *Data Perspective* or the Data Provider in case of the *User Perspective*) needs no protection. While this may be true and actually makes sense for the *User Perspective* (as the server could then be substituted by an adversary), it would actually make sense in case of the *Data Perspective* so that users can retrieve anonymised data anonymously. For *Pseudonymisation*, the *Holder* (Individual vs. Group) is split equally, but the *Cardinality* is never set to *Limited* (cf. figure 5.4), which is in general good as it is more flexible (cf. figure 5.3). In case of the *Protected Aspect Behaviour* and *Content*, only *Access Patterns* or *Data* are present in the examined techniques. *Metadata*, *Actions* and *Perception* are not protected a single time with these techniques.

For the same reason, as in case of *Identity*, *Indistinguishability* is a bit over represented in the *Means of Protection* (16 times) in comparison to *Unlinkability* (8 times) and *Confidentiality* (5 times) cf. figure 5.5. Note that *Unlinkability* would, however, be over represented anyway as it is covered by identity protection from the *User Perspective* and from the *Data Perspective*.

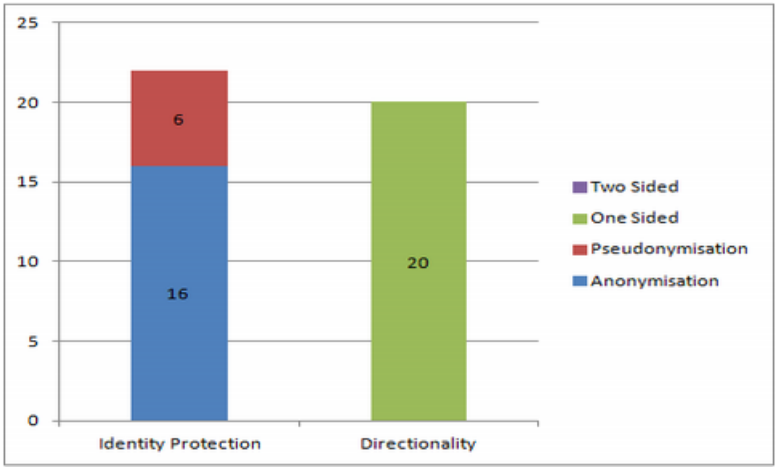


Figure 5.2: Statistics - Identity Protection and Diretionality

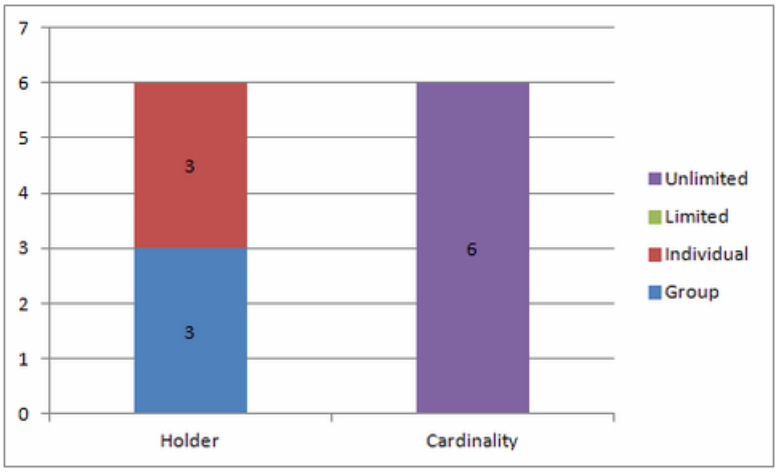


Figure 5.3: Statistics - Pseudonymity

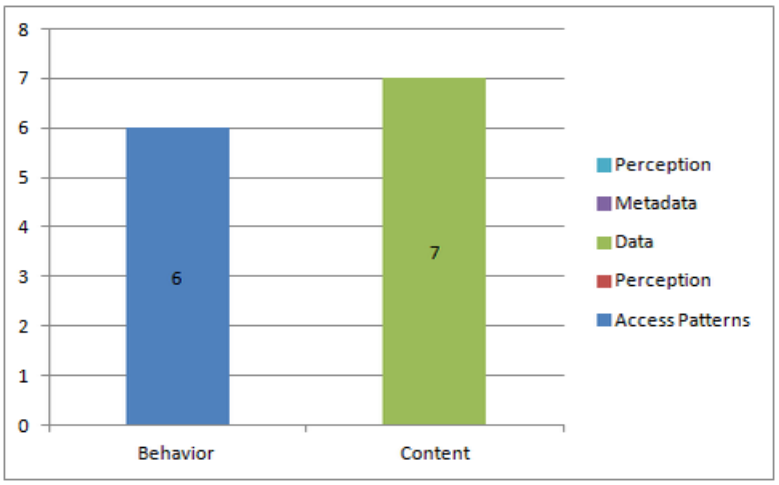


Figure 5.4: Statistics - Behaviour and Content

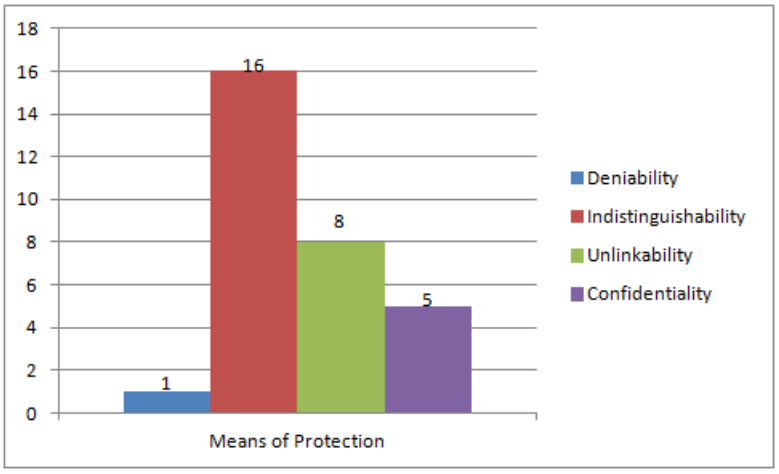


Figure 5.5: Statistics - Means of Protection

Deniability occurs only once: in Deniable Encryption.

Regarding the *Affected Data Type*, *Combined* is used not as frequently (10 times) as expected due to the assumption that in most cases data of more than only one type is treated. This is partially caused by the techniques from the *Data Perspective* which contribute to the over representation of *Stored* data (24 times). *Processed* data is only represented 5 times. In case of *Transmitted* data, the number is low too (6); however, this cannot be seen as representative as a lot of techniques from the field of cryptography exist that cover this area but have not been considered within this thesis as they are already known and discussed intensively. For an overview on the *Affected Data Type* cf. figure 5.6. The *Affected Data Type combined* is composed of the

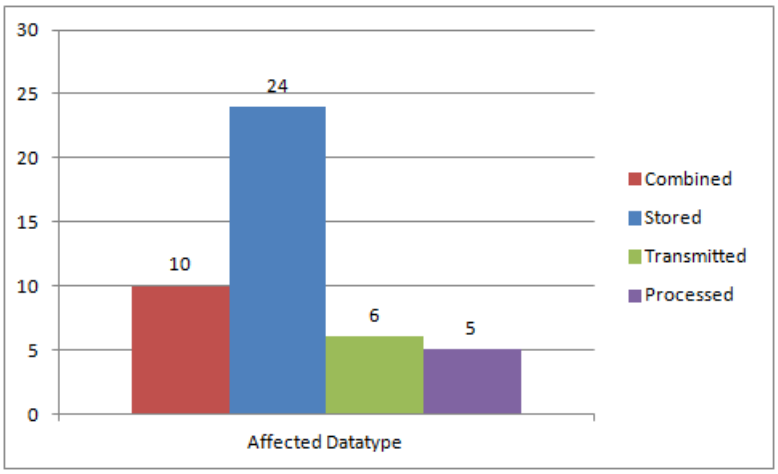


Figure 5.6: Statistics - Affected Data Type

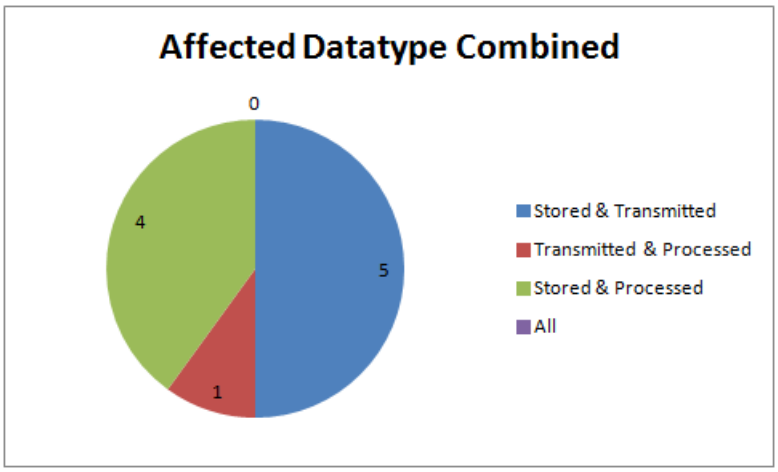


Figure 5.7: Statistics - Affected Data Type Combined (Split)

following combinations: *Stored* & *Transmitted* is covered 5 times, *Stored* & *Processed* 4 times, *Transmitted* & *Processed* once and a combination of all three never (cf. figure 5.7).

The *Base of Security* shows an over representation of *Statistics* (18 out of 30, cf. figure 5.8). This is good on one hand as it offers strong privacy protection but on the other hand, it is not applicable for all scenarios. This over representation is partially caused by the big number of techniques from the field of Privacy Preserving Data Publishing that has been examined. 13 techniques are building on *Cryptography* and 4 building on *Mathematical Problems*. This confirms that the distinction between those two is not easy and that they may overlap, but that it is important to distinguish them and that this is actually relevant not only in theory but in

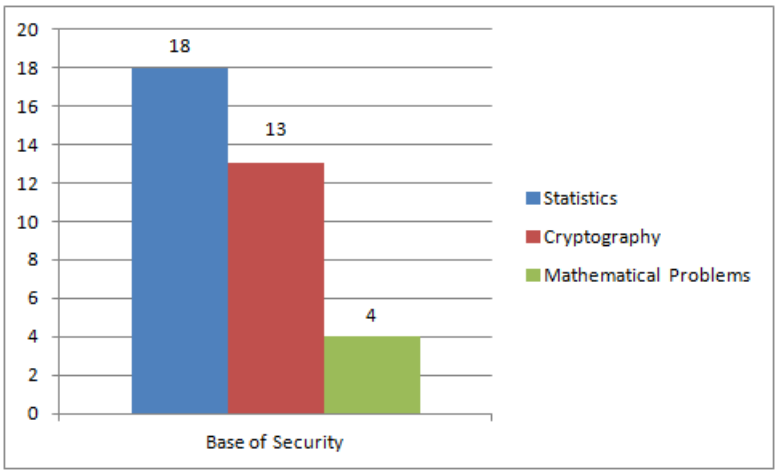


Figure 5.8: Statistics - Base of Security

practice as well. As *Cryptography* can further be distinguished into *symmetric* and *asymmetric Cryptography*, this distinction is taken into account here too: *symmetric Cryptography* was only used as a base 8 times while *asymmetric Cryptography* was used 13 times (cf. figure 5.9). This is, however, not a big surprise as *asymmetric Cryptography* usually is based on *Mathematical Problems* while *symmetric Cryptography* intends to introduce secret states, and one is more flexible to combine math with other techniques than these secret states.

The *Strength of Security* is split into *Information Theoretic Security* 18 times and *Computational Security* 15 times (cf. figure 5.10) which is good because it indicates that the majority of techniques examined are really robust so they cannot be broken with any arbitrary amount of computational power. However, certain aspects of a technique may have different *Strengths of Security* so a detailed study of one specific technique is required before its usage.

Regarding the *Trusted Third Party Participation*, the *Frequency* and the *Phase* and the *Background of TTP Participation* have been examined (cf. figure 5.11). All three dimensions show a strong tendency to avoid Trusted Third Parties at all, in general a favourable approach of relying on privacy by design instead of trust. The *Frequency* is 28 times Never, 7 times only *In Specific Scenarios* and not a single time *Always*; the *Phase* is 29 out of 35 times None, 5 times the *Operation* Phase and only once the *Setup* Phase, and the *Background* is (as a consequence) *No TTP Participation* 28 out of 35 times. The other possibilities (*Checks/Operations* as *Background*) occur 7/0 times.

At first sight, the *Enforceability* has an important drawback: not a single technique can be enforced by the *Client* (cf. figure 5.12) - however, one has to consider that in *Enforceability Both* (11 out of 27 times) the *Client* is included as well, so the client actually has the possibility to enforce nearly half of the techniques. Furthermore, 6 techniques with *Enforceability None* exist. As argued before, this does not mean that they are worthless nor that they cause privacy problems. They may just be improvements with another technique as fall-back so that if the *Server* does not participate, no privacy problems are caused. In 10 cases only the *Server* is able

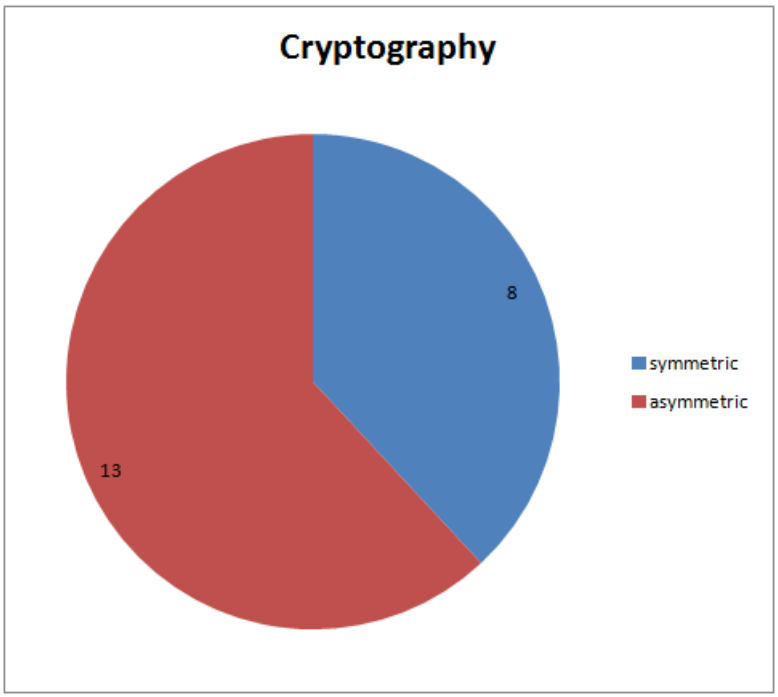


Figure 5.9: Statistics - Base of Security (Cryptography)

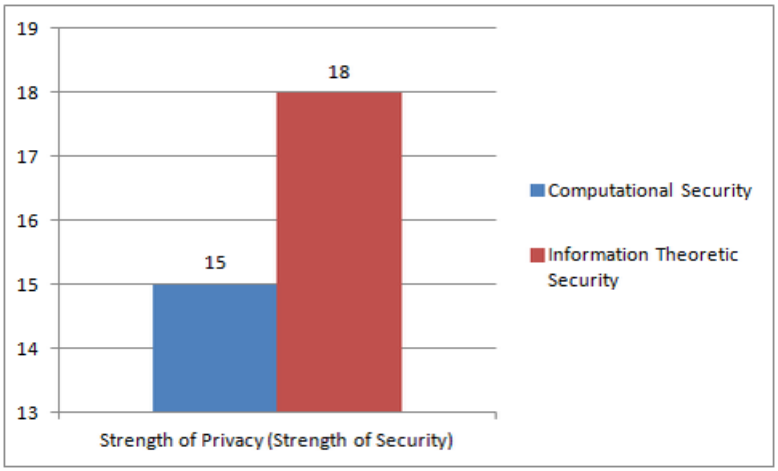


Figure 5.10: Statistics - Strength of Security

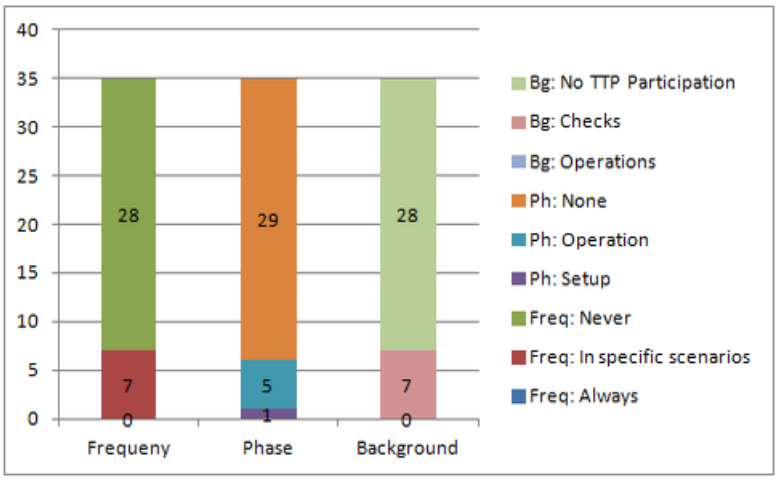


Figure 5.11: Statistics - Trusted Third Party Participation

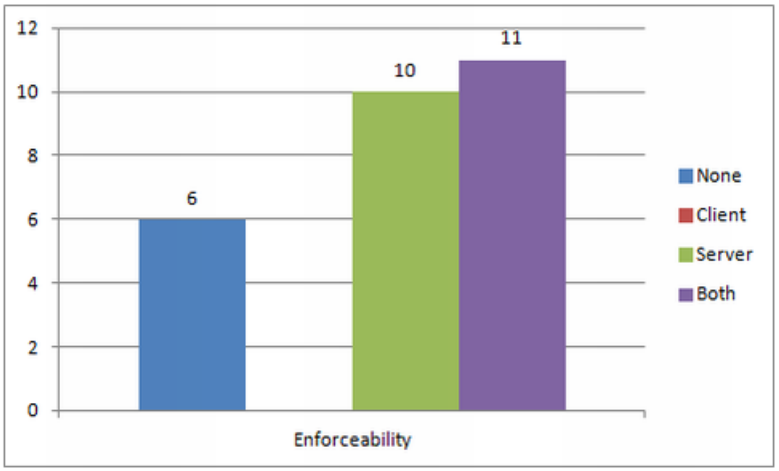


Figure 5.12: Statistics - Enforceability

to enforce the technique. Although this sounds strange, techniques for which this actually makes sense exist and like all examined techniques with *Enforceability* for the *Server* only are from the field of Privacy Preserving Data Publishing this is fine.

Regarding the *Reversibility*, the *Degree* was *Deniable Reversible* twice and *Not Reversible* 14 times (out of 41) which is shows strong privacy protection. 18 techniques are *Fully Reversible* - however, this could be totally in the user's interest too (like, for example, in case of encryption). 7 techniques are *Partially Reversible*. Considering that *Full/Partial Reversibility* could be really desired features this seems to be ok too. *Cooperation is required* for 25 out of 36 techniques.

Taking the *Perspectives* into account, the *Data Perspective* and the *Profile Perspective* are

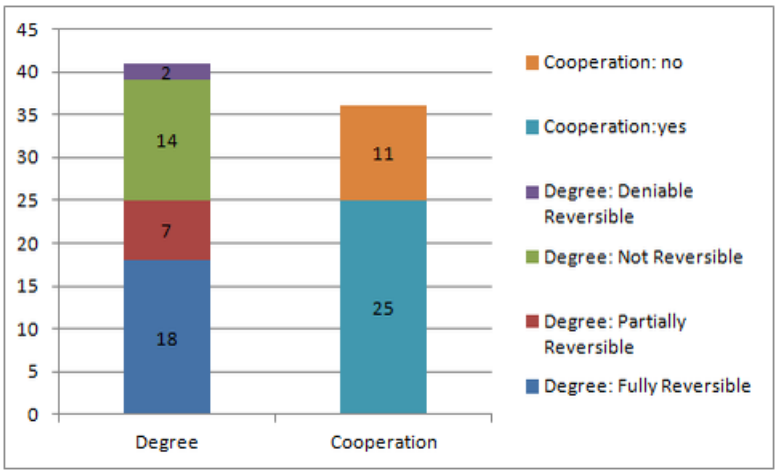


Figure 5.13: Statistics - Reversibility

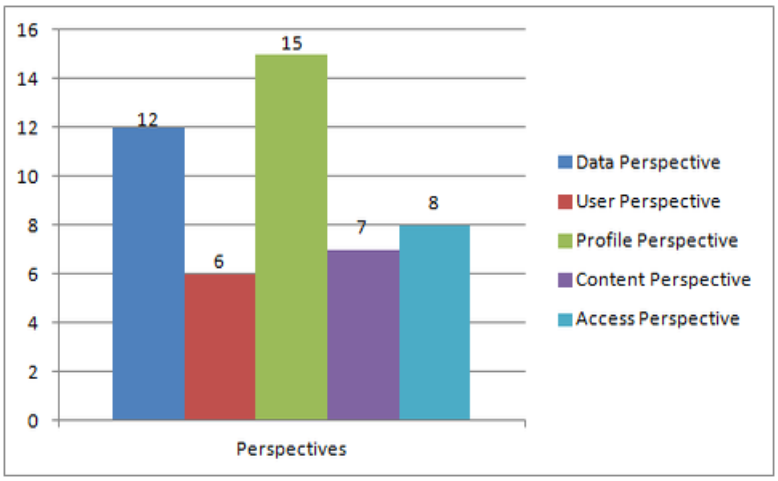


Figure 5.14: Statistics - Perspectives

equally represented (12/15 times). The *User Perspective* is only present in 6 techniques (cf. figure 5.14). However, a lot of mature techniques for covering the *User Perspective* exist that are out of scope of this thesis. Examining the *Profile Perspective* in more detail shows that the *Access Perspective* and the *Content Perspective* occurred 8/7 times.

CHAPTER 6

Synergy Effects of PET Groups

When two approaches are combined, the resulting properties change. Some attributes have got additive characteristics, so adding a second approach covering other parts of privacy improves the overall privacy level. Other attributes, however, are more complicated to combine as the resulting attribute is a mixture rather than a simple combination of the subsets with potentially different properties. This section investigates potential synergy effects when combining techniques of different PET groups in order to achieve a higher privacy level.

In general, this analysis includes both technology and perspective aspects introduced in the taxonomy, but focuses on *Protected Aspect*, *Means of Protection*, and *Affected Data Type* in the technology subtree because the remaining properties are either mutually exclusive (*Base of Security*), are incomparable, or would not provide any significant improvement. The aim is to create groups of PETs with similar properties that could be combined to counter each others' weaknesses or limitations.

This general analysis approach is supplemented by selected more in-depth combination proposals which describe the internal technical interactions as well as the particular advantages of combining two specific PETs in more detail.

6.1 Groups

In this section the investigated techniques are grouped so that techniques with similar properties/aims form a group. As each of the techniques/groups has got strengths and weaknesses, synergy effects between the groups are investigated so that the strengths can be combined while the weaknesses are eliminated or at least reduced. Examples illustrate the combination's relevance.

Privacy Preserving Data Publishing

Privacy Preserving Data Publishing is a group of techniques that allows a server to publish confidential data it would usually not be allowed to publish without compromising an individual's

privacy. This corresponds to the *Data Perspective* (cf. chapter Perspectives at page 25). Privacy Preserving Data Publishing is a cluster consisting of the following techniques that have been evaluated within this thesis:

- k-Anonymity
- (X,Y)-Privacy
- MultiR k-Anonymity
- Distinct l-Diversity
- Entropy l-Diversity
- (c,l)-Diversity
- Confidence Bounding
- (α, k)-Anonymity
- (k, e)-Anonymity
- (ϵ, m)-Anonymity
- Personalized Privacy
- t-Closeness

As they all have got equal properties regarding the presented taxonomy, they are evaluated as a group rather than as single techniques. Note that differences between these techniques exist - however, they are too specific to be covered by the taxonomy.

The group's properties are as follows: the *Protected Aspect* is the *Identity* by means of *One Sided Anonymity*, the *Means of Protection* is *Indistinguishability*, the *Affected Data Type* is *Stored*, and the *Perspective* is the *Data Perspective*.

The *Protected Aspect* of *Identity* has already been fully exploited as *Anonymity* is considered to be stronger than *Pseudonymity* - at least for the *Polarisation One-Sided*. If authentification is required, then the group Privacy Friendly Authentification can be used to ensure privacy for authenticated users too - if this is not required then simple techniques like proxies can be used to hide the *Identity* of the user requesting some data. Therefore, an approach covering the *Content* and the *Behaviour* is desired. The *Content* could be covered easily by a technique from the group Operations on Encryption, while the *Behaviour* can be covered by adding a technique from the group Privacy Friendly Access. The actual selection of the proper technique is highly dependent on the specific data publishing scenario and the corresponding requirements.

When combining the Group Privacy Preserving Data Publishing with the groups Operations on Encryption and Privacy Friendly Access, the *Protected Aspects* of *Identity*, *Behaviour*, and *Content* are covered. Furthermore, the *Means of Protection* of *Indistinguishability*, *Unlinkability*, *Confidentiality*, and (in case of Oblivious Transfer) even *Deniability* are covered, which is equivalent to all possibilities for the *Means of Protection*. The resulting *Affected Data Type*

is *Combined* because *Stored*, *Transmitted*, and *Processed* data is protected, which again corresponds to all possibilities. Last but not least, the covered *Perspectives* are extended too: in addition to the *Data Perspective*, the *Profile Perspective* (*Access Perspective* and *Content Perspective*) is covered and, in case of Oblivious Transfer, the *User Perspective* may be covered too (depending on the specific approach used). Even if it is not covered, it may be included easily e.g. by means of proxy servers. For this reason, the suggested combination is supposed to cover all relevant criteria introduced in this taxonomy.

Additionally protecting already anonymised data with encryption may seem superfluous at first sight, but may be desirable or even necessary depending on the data. Consider the following scenario:

A company may be interested in analysing its business by external business analysts. To ensure its customers' privacy, the data is anonymised before the business analysis are granted access to the records.

In addition the company's policy requires access control so that only authorised employees are able to access this sensitive data. For this reason, encryption is required although the data is already anonymised. If the analysts still need to perform operations on this protected data, then techniques from the group Operations on Encryption are required. The groups of Privacy Friendly Access and Privacy Friendly Authentification are required if the company wants to ensure its employees' privacy (Privacy Friendly Authentification is of course only necessary when authentification is required).

Operations on Encryption

Privacy Preserving Data Publishing is a group of techniques that enables operations on encrypted data (which is not possible with ordinary encryption schemes). Therefore, it fosters the use of encryption in scenarios where it is usually difficult or impossible because the data needs to be processed.

The techniques assigned to this cluster are:

- Homomorphic Encryption
- Searchable Encryption
- Order Preserving Encryption

In case of Homomorphic Encryption arbitrary operations are possible on the data (considering the strongest techniques without limitations to additions/multiplications). In case of Searchable Encryption it is obvious that the operation is to search on encrypted data, and, in case of Order Preserving Encryption the operations are sorting, comparisons, and range queries.

All of them have similar properties, so one is able to chose the approach that fits all requirements best. As these approaches are based on specific cryptographic properties, combining them may be infeasible without compromising their individual algorithmic security properties.

The group's properties are as follows: the *Protected Aspect* is the *Content* (to be more precise the *Data*), the *Means of Protection* is *Confidentiality*, the *Affected Data Type* is *Combined* as

Stored as well as *Processed* data is protected, and the *Perspective* is the *Profile Perspective* (*Content Perspective*).

The *Protected Aspect* of *Content* has not yet been exploited fully. The *Data* part is already covered, but the *Meta Data* is not taken into account. In case of *Added Metadata*, this would not be possible, as it would need to prevent an adversary from adding any form of information at any system, which is definitely not possible. In case of *generated Meta Data* this would be possible - unfortunately, none of the examined techniques covers *Metadata*. For this reason, no combination is possible extending the *Protected Aspect* of *Content* given the analysis in this thesis. For the *Protected Aspect*, it is, however, possible to protect the *Identity* as well as the *Behaviour* which are both not covered yet. Regarding the *Behaviour* and following the argumentation from the group Privacy Preserving Data Publishing, adding a technique from the group Privacy Friendly Access would be beneficial. Regarding the *Identity*, the Data Owner can be protected by means of a technique from the group Privacy Preserving Data Publishing, while Data Consumer could be protected by the group Privacy Friendly Authentification in case authentification is required, or by other means like proxy servers if no authentification is required.

The combination of the group Operations on Encryption with the groups Privacy Preserving Data Publishing, Privacy Friendly Authentification, and Privacy Friendly Access covers all relevant aspects: the *Protected Aspect* is extended to the *Content*, as well as the *Identity*, and the *Behaviour*. The *Means of Protection* covers *Confidentiality* as well as *Indistinguishability*, and *Unlinkability*, and in case of Oblivious Transfer even *Deniability*. The *Affected Data Type* is *Combined* as not only *Stored* and *Processed* but the *Transmitted* data is covered as well. Last but not least, the *Profile Perspective* (*Access Perspective* and *Content Perspective*) as well as the *Data Perspective* and the *User Perspective* could be covered. Therefore, again all relevant aspects introduced in this taxonomy can be covered.

Again, combining identity protection with encryption is necessary in specific circumstances, e.g. when encrypted data is only protected from unauthorised users - authorised users can decrypt and handle it as they like. For reasons of confidentiality, it may, however, be prohibited that an administrator or analyst knows whose records he or she is accessing, so the identifying information is protected by using techniques from the group Privacy Preserving Data Publishing. As argued before, the employees' privacy can be protected by using Privacy Friendly Authentification and Privacy Friendly Access.

Privacy Friendly Access

Privacy Friendly Access is a group of techniques that enables access to data without the server knowing what the user is interested in, so it cannot tell what data has actually been retrieved.

The techniques in this cluster are:

- Private Information Retrieval
- Oblivious Transfer

In case of Private Information Retrieval, all variants (Trivial Solution, Single Server with Computational Security, and Multiple Servers with Information Theoretic Security) have similar

properties regarding the taxonomy. Therefore, one is able to chose for every scenario the one that fits the situation best. Oblivious Transfer has similar properties too, however, they depend strongly on the used technique. In terms of aspects, Oblivious Transfer also covers the *Protected Aspect Identity*, something which could also be achieved by adding proxy servers.

The group's properties are as follows: the *Protected Aspect* is the *Behaviour* (to be more precise the *Access Patterns*), the *Means of Protection* is *Unlinkability*, and in case of Oblivious Transfer, potentially *Deniability*. The *Affected Data Type* is *Combined* as *Stored* and *Transmitted* data is protected. The *Perspective* is the *Access Perspective*.

Regarding the *Protected Aspects* of *Behaviour*, the *Access Patterns* are already covered. Furthermore, *Actions* and *Perception* exist too - however, none of the examined techniques is able to deal with them. Therefore, the *Behaviour* has to be seen as completely covered. Regarding the *Protected Aspect*, the *Identity* and the *Content* are not yet covered. However, it is possible to cover the *Identity* by introducing a combination with the groups Privacy Preserving Data Publishing or Privacy Friendly Authentification (depending if the Data Owner or the authentification should be protected). Again, unauthenticated users could be protected easily by other means such as proxies. Regarding the *Content* - depending on the needs - a technique of the group Operations on Encryption or Proxy Re Encryption may be used.

With this measure, the *Protected Aspect* is covered with the introduced techniques: the *Behaviour*, the *Identity*, and the *Content* are covered. As a consequence, the *Means of Protection* covers (besides *Deniability* and *Unlinkability*) *Indistinguishability* and *Confidentiality* too. The *Affected Data Type* is now *Combined* and *Stored*, *Transmitted*, and *Processed* data are covered. The *Perspective* covers now the *Profile Perspective* (*Access Perspective* and *Content Perspective*), *User Perspective*, and the *Data Perspective*.

The practical relevance of this combination can be demonstrated in the following scenario: a company wants to provide a database about diseases. As the employee requesting this information may not have an interest that anyone knows which information he or she accessed, Privacy Friendly Access is required.

Privacy Preserving Data Publishing is necessary if the data itself is sensitive too. This applies if the data is about specific medical cases rather than generic information. Thus the identity of the affected persons has to be kept secret too.

The implementation of access control for a target group that usually would not access it, seems to be a conflict of interests at first sight. However, it actually makes sense as soon as the company wants to ensure, only its employees can access the system. This can be enforced in a still privacy friendly way by means of Privacy Friendly Authentification and - as a second layer of security - techniques from the group Operations on Security could be applied.

Privacy Friendly Authentification

Privacy Friendly Authentification is a group techniques that enables authentification without harming a user's privacy. This cluster consists of the following techniques:

- Direct Anonymous Attestation
- Anonymous Credentials

Both techniques have similar properties so, depending on the required focus, they can be exchanged. A combination of Direct Anonymous Attestation and Anonymous Credentials does not seem to make sense, as this may cause security problems with the underlying assumptions. It would be too complex to suggest a generic combination, so a combination is suggested for each of them separately. Nevertheless, the general aim of both is the same, and the following practical example is valid for both techniques: Privacy Friendly Authentification is used when access to some resource should not be possible for anyone (like in case of medical databases which can only be accessed by paying clients).

However, only the membership to some defined group needs to be verified without the need for identifying the individual persons. Like mentioned before, the data may be about medical records, so the identity of the affected patients has to be kept secret by using Privacy Preserving Data Publishing. Again techniques from the group Operations on Encryption may serve to create a stronger access control scheme compared to plain authentication. The *Behaviour* of Data Consumers can also be kept secret by using techniques from the group Privacy Friendly Access to prevent the creation of profiles (if not already covered by using Anonymous Credentials) so that the creation of profiles can be prevented.

Anonymous Credentials

In case of Anonymous Credentials, the *Protected Aspect* is the *Content* (to be more precise the *Data*) and the *Identity* by means of *One-Sided Anonymisation* or *Group Pseudonymisation*. As a consequence, the *Means of Protection* are *Confidentiality* and *Indistinguishability*, the *Affected Data Type* is *Combined* as *Stored* and *Transmitted* data is protected (which is typical for an authentification scheme). The *Perspective* is the *User Perspective* and, for some approaches, the *Profile Perspective* (*Access Perspective*) too.

Regarding possible combinations, Anonymous Credentials already cover a broad variety of properties. In case of the *Protected Aspect*, the *Behaviour* could be added by using a technique from the group Privacy Friendly Access. As a consequence, the *Means of Protection* get extended to *Unlinkability* and (in case of Oblivious Transfer) *Deniability*. The *Affected Data Type*, however, does not change at all: it is still *Combined* (*Stored*, *Transmitted*). The *Perspectives* of this combination are the *Profile Perspective* (*Access Perspective*) and the *User Perspective*. So combining Anonymous Credentials with just one technique covers nearly all possibilities of the taxonomy - the most important aspect that is missing is the *Data Perspective*. If required (e.g. by sensitive data), it could be added easily with a technique of the group Privacy Preserving Data Publishing. The resulting triple of the groups has now covered all aspects except of the *Affected Data Type processed*. However, an important aspect is not covered yet: the server handling the data. This correlation indicates that the selected attributes are indeed important and significant for the overall privacy. Therefore, to cover all attributes (and ensure proper privacy) again an appropriate technique from the group Operations on Encryption is added. With these four groups, all attributes of the taxonomy are covered as good as it is possible for Anonymous Credentials by using the examined techniques.

Direct Anonymous Attestation

In case of Direct Anonymous Attestation, the *Protected Aspect* are the *Behaviour* (to be more precise the *Access Patterns*) and the *Identity* by means of *One-Sided Group Pseudonymisation*. As a consequence, the *Means of Protection* are *Unlinkability* and *Indistinguishability*, the *Affected Data Type* is *Combined* as *Processed* and *Transmitted* data is protected, and the *Perspectives* are the *User Perspective* as well as the *Data Perspective*.

For the *Protected Aspect*, the *Content* is missing. This directly leads to the group Operations on Encryption for combinations in order to cover this aspect too. As a consequence, the *Means of Protection* gets extended to *Confidentiality*, and the *Affected Data Type* now includes *Stored* data too. The *Perspectives* now cover the *Content Perspective* too. To sum it up, this simple pair of techniques already covers the *Protected Aspect*, the *Affected Data Type*, and the *Perspective* - the *Means of Protection* is the only attribute that is not yet covered completely. In order to fix this, *Deniability* is required for the Means of Protection too. This may either be implemented by means of Oblivious Transfer or by use of Deniable Encryption depending what exactly one wants to deny (the transfer of data or the actual content). If the underlying data is sensitive, the logical consequence is to apply the group Privacy Preserving Data Publishing too. Last but not least, a technique from the group Privacy Friendly Access is added so that the *Access Perspective* is covered. Again all relevant aspects of the taxonomy have been covered with the combination of techniques.

Delegation

Delegation is a group of techniques that enables outsourcing/delegating data (processing) to untrusted or semi-trusted servers without having the servers know anything about the data they actually handle. This cluster consists of the following techniques:

- Proxy Re-Encryption
- Blind Signature

Both techniques have got a different intention as well as different properties, so they should be taken into account separately for combinations. In case of Proxy Re-Encryption, the delegation affects storage and the re-encryption of data if it should be forwarded. In case of Blind Signature, the delegation affects creating signatures for confidential data. Fair Blind/Partially Blind schemes offer similar properties, so depending on the use case, the most appropriate can be chosen.

As the name „Delegation“ already suggests, the aim of techniques from this group is to outsource sensitive data/operations to a third party server without worrying about privacy though the actual application scenarios can be very different.

Proxy Re-Encryption

Proxy Re-Encryption's properties are as follows: the *Protected Aspect* is the *Content* (to be more precise the *Data*), the *Means of Protection* is *Confidentiality*, the *Affected Data Type* is

Combined as *Stored* as well as *Processed* data is protected, and the *Perspective* is the *Content Perspective*.

The *Protected Aspect Identity* as well as the *Behaviour* yet have a lack of protection. In order to cover the *Identity*, three measures can be taken: first, a technique from the group Privacy Preserving Data Publishing would allow sensitive data to be published and accessed. Secondly, a technique from the group Privacy Friendly Authentification may be useful if access restrictions are necessary. Last but not least, a proxy could be used if the data is publicly accessible by anyone without any restrictions. In order to cover the *Behaviour* as well, a technique from the group Privacy Friendly Access can be added. With these combinations, the *Protected Aspect* is completely covered. Furthermore, the *Means of Protection* is extended to *Indistinguishability*, *Unlinkability*, and (in case of Oblivious Transfer) *Deniability*. The *Affected Data Type* is extended to *Combined* (*Stored*, *Processed*, and *Transmitted* data) and the covered *Perspectives* are the *Profile Perspective* (*Access Perspective* and *Content Perspective*), as well as the *User Perspective* and the *Data Perspective*. This combination covers all relevant aspects of the taxonomy.

This combination might be useful when e.g. a company intends to outsource its data to the cloud. As this may contain sensitive data, the data should be protected by encryption. New colleagues may also be added to the project, requiring access to the data, which can be handled by Proxy Re-Encryption.

As in the examples before, this data may contain personal information so the employee is not supposed to know about the identity of the record owners. This is why techniques from the group Privacy Preserving Data Publishing are used additionally. One may wonder why it is necessary to add Access Control if the data is already protected by encryption. Again, increasing the difficulty in gathering useful information by an adversary by adding an access control layer conforms to the defense-in-depth strategy to prevent a compromised security instrastructure by a single point of failure. The implementation of cryptography may be faulty and thus circumvented or broken. If however, the employees should not be monitored although access control is implemented, then techniques from the group Privacy Friendly Authentification are required. Additionally, techniques from the group Privacy Friendly Access can be used to hide even more activity from the server, so it does not even know which (encrypted) records/data sets are accessed.

Blind Signature

Blind Signature's properties are as follows: the *Protected Aspect* is the *Content*, to be more precise the *Data* as well as the *Identity* by means of a *One-Sided Individual Pseudonym* (Fair Blind) respectively *One-Sided Anonymisation* (Partially Blind). The *Means of Protection* is *Unlinkability* and, in case of Fair Blind Signature, *Indistinguishability* too. The *Affected Data Type* is *Combined* as *Stored* and *Processed* data is protected, and the *Perspective* is the *Profile Perspective* (*Content Perspective*).

Regarding the *Protected Aspect*, the *Behaviour* is not covered. In order to change this, a combination with a technique from the group Privacy Friendly Access is desired. This extends the *Protected Aspect* to *Behaviour*, the *Means of Protection* to *Deniability* (in case of Oblivious Transfer), and the *Perspective* to the *Access Perspective*. The *Affected Data Type* remains unchanged. If the signed content contains sensitive information, it may make sense to apply a technique from the group Privacy Preserving Data Publishing before creating the signature so

that the signature as well as the data itself can be published. This covers the *Data Perspective* as the *Perspective* and also ensures identity protection for the record owners. To ensure the *Means of Protection* of *Confidentiality*, the data may be protected by a technique from the group Operations on Encryption. Although the server does not know the content it signs, encryption may be used as an additional barrier if access control is used, so a single point of failure is avoided. If one still wants be able to perform operations on it, the group Operations on Encryption is required. To cover the *User Perspective*, a technique from the group Privacy Friendly Authentification can be added. With the mentioned combinations, the *Protected Aspect* as well as the *Means of Protection*, the *Affected Data Type*, and the *Perspective* are covered as good as it is possible with the investigated techniques.

The application scenario for Blind Signature is similar to that for Proxy Re-Encryption but with a different aim. In this case, the scenario builds on a company that wants to be able to proof the authenticity of messages sent to customers, but it wants to protect the employees' privacy. For this reason, the company wants that the contents of the messages are signed without a - potentially outsourced - server knowing, what actually is signed, so Blind Signatures are used.

As argued in the examples above, the data may contain personal information requiring to apply techniques from the group Privacy Preserving Data Publishing so that the client does not know about the identities of the respective persons. If the messages sent get enriched with data (like e.g. contact data or personal information which are received from a central data store), then Privacy Friendly Access is required so that the sever is not able to gain any knowledge (and link various messages).

Deniable Encryption

Deniable Encryption is a technique with specific properties, so it cannot be merged into one of the other clusters, but has to build a „cluster“ on its own.

The properties of Deniable Encryption are as follows: the *Protected Aspect* is the *Content* (to be more precise the *Data*), the *Means of Protection* are *Deniability* as well as *Confidentiality*, the *Affected Data Type* is *Combined* as *Stored* and *Transmitted* data is protected, and the *Perspective* is the *Access Perspective*.

As the *Protected Aspects* of *Identity* and *Behaviour* are not covered, some combinations are required. In case of the *Identity*, the combinations are twofold: firstly, a technique from the group Privacy Preserving Data Publishing is needed if the underlying data may be sensitive, so privacy is not compromised; secondly, a technique from the group Privacy Friendly Authentification can be used if authentification is required (if not then simple techniques like proxies can be used). Although it may be interesting to apply techniques from the group Operations on Encryption, it would affect the encryption scheme's integrity. According to the taxonomy, techniques from the group Privacy Friendly Access would not improve the overall privacy level, but would still be beneficial under specific circumstances (see below). As a result the *Protected Aspects Identity* and *Content* are covered, the *Means of Protection* are *Confidentiality*, *Deniability*, and *Indistinguishability*, the *Affected Data Type* is *Combined* (*Stored*, *Transmitted*, and *Processed*), and the *Perspectives* are the *Profile Perspective* (*Access Perspective* as well as *Content Perspective*), the *Data Perspective*, and the *User Perspective*. To sum it up, the *Affected Data Type* and the *Perspective* have been covered completely, while the *Protected Aspect* and the *Means of Pro-*

tection have been covered well but not completely, because this would not make sense for this technique.

Deniable Encryption is a useful technique in scenarios like totalitarian regimes with limited freedom of speech where people might be facing to be silenced by force. An alternative scenario could be a system for whistle-blowers that can report problems anonymously. In both variants, the assumption is that it is not sufficient to encrypt the content, but that encrypted content itself is suspicious and leads to further investigations. Within these investigations, the sender/receiver of a message could be forced to decrypt it. As Deniable Encryption offers the possibility to decrypt an alternative (non suspicious) text, the involved parties remain protected and the barrier to communicate in these scenarios is lowered.

As in the examples above the combination with a technique from the group Privacy Preserving Data Publishing may be useful if the persons that the data is about must not be identified. The combination with a technique from the group Privacy Friendly Access makes sense because it makes it even harder to find out, what a certain user did. If, for example, the servers are investigated, then it may be possible to find out which data is linked to which user and session. Therefore, sensitive information could be disclosed including the messages as well as any communication participants. Privacy Friendly Authentification on the other hand, is required as not every user is allowed to use the system at all and, of course, the use of a certain account has to be restricted to the authorised users. Adding a technique from the group Privacy Friendly Authentification prevents any possibility to find out if and when a user actually used the system or if the user just created an account without ever using it. This protection could further be increased when the system creates an encrypted container for every user using Deniable Encryption, where the decryption to the fake plaintext results in an account that has never been used.

6.2 Combination Examples

In order to allow a more precise evaluation and generate ideas for further research, some combinations of specific techniques are provided. The scope should not be restricted artificially so that groups of techniques are taken into account, when possible without ambiguity. However, restrictions to a subset of techniques or even one specific approach is applied, when the results would be ambiguous or imprecise.

k-Anonymity & Anonymous Credentials

In this section, the combination of k-Anonymity from the group Privacy Preserving Data Publishing (steps 5 & 6) and Anonymous Credentials from the group Privacy Friendly Authentification (steps 1-4 and 6-11) is examined. From a technical point of view, both techniques can be thought of as two layers that enhance privacy for some data. The combination results in two layers which could be applied sequentially: first, the authentification process is substituted by the privacy friendly Anonymous Credentials. Afterwards, authorised users can access data anonymised using k-Anonymity (steps 5 and 12). Because of this sequence, no interaction nor any kind of interference between the techniques occurs, so they can be combined without problems.

This combination has got the *Protected Aspects* of *Identity* with *Polarisation One-Sided* and *Behaviour* (*Access Patterns*). For the data that is accessed, identity protection builds on *Anonymisation*. For the Data Consumers, accessing the data identity protection builds either on *Anonymisation* too or on *Pseudonymisation* held by an *Individual*, both with *unlimited Cardinality*. The *Means of Protection* are *Indistinguishability* and *Unlinkability*, and the *Affected Datatype* is *Combined* because *Stored* as well as *Processed* data is protected. The *Base of Security* is *Statistics* for the data accessed, while it may either be *Statistics*, *asymmetric Cryptography*, or *Mathematical Problems* for the Data Consumers. As a result, the *Strength of Security* is *Information Theoretic Security* (data accessed) respectively *Information Theoretic Security* or *Computational Security* (Data Consumers). *Trusted Third Party Participation* does not take place for the data accessed, so the *Frequency* is *Never*, the *Phase* is *None*, and the *Background* is *No TTP Participation*. For the Data Consumers on the other hand, *Trusted Third Party Participation* may, but need not take place. This results in the *Frequency* being *Never* or *In specific scenarios*, the *Phase* being *None* or *Operations*, and the *Background* being *Checks* or *No TTP Participation*. *Enforceability* is possible by the *Server* in case of the data accessed respectively for *Both* for the authentification. Regarding *Reversibility*, the *Degree* is *Not Reversible* with *Cooperation* required for the data accessed, while the *Degree* may be *Fully Reversible* or *Partially Reversible* with *Cooperation* not required for the Data Consumers. The *Perspectives* covered are the *Data Perspective* as well as the *User Perspective*, and - depending on the approach used for Anonymous Credentials - may include the *Profile Perspective* (to be more precise the *Access Perspective*) too.

As Anonymous Credentials offer varying properties depending on the specific approach used, one is picked so that an accurate evaluation of the *Protected Aspect*, the *Base of Security*, the *Strength of Security*, the *Trusted Third Party Participation*, the *Reversibility*, and the *Perspective* is possible.

The approach chosen is Practical Revocable Anonymous Credentials (cf. [48]). The ambiguous properties can therefore be evaluated to *Identity Protection* by means of *One-Sided Anonymity* (data accessed) respectively *One-Sided Pseudonymity* with an *Individual* as *Holder*, and *unlimited Cardinality* (authentification). The *Base of Security* is *Statistics* for the data accessed. For the Data Consumers' identities, the protection has *Statistics* as *Base of Security* too, as the users cannot be distinguished, which limits the probability of guessing the right identity. The *Base of Security* for the proof of some condition is, however, *Mathematical Problems* as the discrete logarithm problem and the factorisation problem have to be solved. As a consequence, the *Strength of Security* is only *Computational Security* for the assurance that the condition is actually satisfied, while it is *Information Theoretic Security* for the protection of the record owners' identities as well as the Data Sources' identities. *Reversibility* with *Degree Fully Reversible* even without *Cooperation* is a desired feature of this Anonymous Credentials implementation by enabling Trusted Parties to reverse the protection. To secure this technique against misuse, several Trusted (Third) Parties are required. It is obvious that the *Frequency* is therefore *In specific scenarios* (misuse), occurs in the *Operations Phase*, and the *Backgorund* is for *Checks*.

Both techniques have complementary aims: in case of k-Anonymity, the record owners' privacy is protected while Anonymous Credentials protects the users accessing this data. One obvious weakness is that both techniques focus only on one target group and do not cover any-

thing for the other target group. This is fixed with the combination.

Furthermore, k-Anonymity has a big drawback: as a standalone solution, it can only offer limited protection due to the lack of (privacy friendly) access control. As a consequence, everyone can access this sensitive data (assuming that the privacy of the users accessing this data should not be infringed neither), which results in a loss of control. Combining it with Anonymous Credentials allows to enforce more control as users may be revoked/re-identified (this depends on the specific technique used) as well as to restrict the user base that is allowed to access the records.

On the other hand, Anonymous Credentials do not protect the underlying data neither from the Data Provider itself nor any Data Consumer that may break access control and gain unauthorised access or exceed his or her rights. Possible countermeasures include cryptography or techniques for anonymisation like k-Anonymity in this example. For this reason, the combination of k-Anonymity and Anonymous Credentials allows to mitigate the weaknesses of Anonymous Credentials too. Nevertheless, some possibilities for improvement still exist so that a combination of more techniques makes sense.

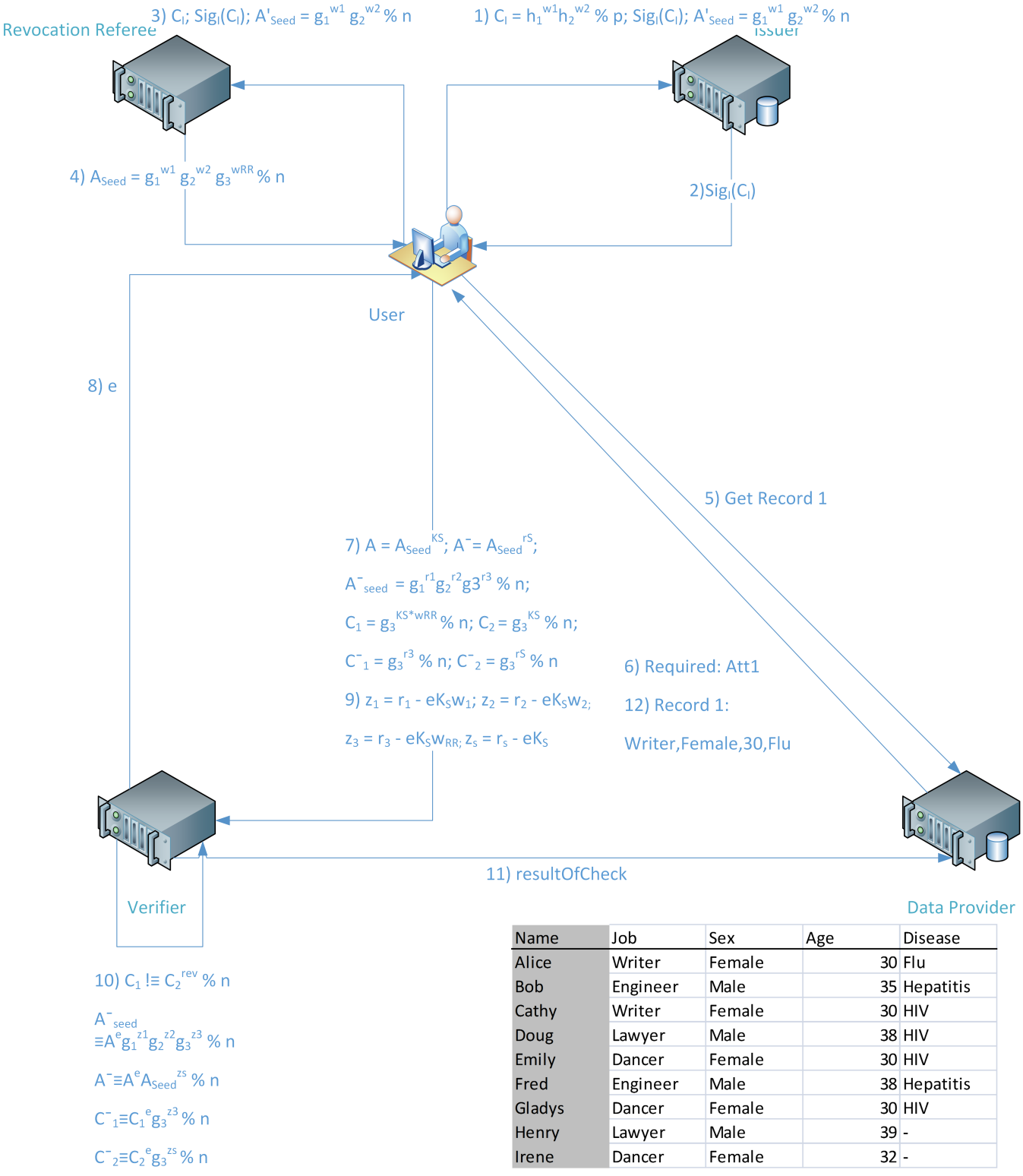
Figure 6.1 illustrates how this system works: whenever a new user wants to register or an existing user wants to get a proof for a new attribute, the steps 1 to 4 have to be conducted. The user generates two random numbers w_1 and w_2 which he or she uses in combination with the publicly known parameters h_1, h_2, g_1, and g_2 so that he or she can compute C_I, sign it, which is denoted as sig_U, compute A'$_{Seed}$, and send them to the issuer. The issuer then signs C_I ($sig_I(C_I)$) and returns it to the user who forwards it to the Revocation Referee in step 3 together with C_I and A'$_{Seed}$. Subsequently the Revocation Referee returns A_{Seed} to the user. Then the registration is finished and the user is able to use his or her attribute.

In step 5, the user requests record 1 which he or she is interested in. However, the Data Provider does not know if the user is authorised to access this record. Therefore, it responds in step 6 that it needs a proof from the user of possessing attribute Att1. In order to prove this, the user computes the values shown in step 7 and sends them to the verifier (K_S is a randomly chosen session key while r_1, r_2, r_3, and r_S are random numbers as parameters for the calculations). It responds with a random number e which the user has to take into account for computing z_1, z_2, z_3, and z_s. After computing these values, the user returns them to the verifier (step 9) who performs the tests from step 10. If all of these tests are passed, the user has proofen the possession of the required attribute; otherwise the proof failed. In step 11, this result is transmitted to the Data Provider. If the user passed all tests, record 1 (as requested) is delivered in step 12.

The sensitive information present in the table in this example was added only for demonstration purposes. The data used in this example is composed of tables 4.1 at page 50 and 4.2 at page 50. The reply in step 12 reflects what actually is transmitted.

This way the identity of the user is kept secret but the Data Provider is still able to check the possession of certain attributes (while keeping all other attributes secret). The record owners' identities are kept secret too due to anonymisation.

The practical relevance of this combination can be identified in the following scenario: A research centre provides a database with medical information. This information must not be provided without anonymisation, so k-Anonymity is chosen. The Data Consumers are other



Name	Job	Sex	Age	Disease
Alice	Writer	Female	30	Flu
Bob	Engineer	Male	35	Hepatitis
Cathy	Writer	Female	30	HIV
Doug	Lawyer	Male	38	HIV
Emily	Dancer	Female	30	HIV
Fred	Engineer	Male	38	Hepatitis
Gladys	Dancer	Female	30	HIV
Henry	Lawyer	Male	39	-
Irene	Dancer	Female	32	-

Figure 6.1: Combination of k-Anonymity with Anonymous Credentials

research centres that want to use this information. As a lot of competition exists, they want to be sure that no one is able to find out who accessed the database. The Data Provider on the other hand sells access to its database, so it wants to ensure that only paying clients are able to retrieve information. In order to cover the needs of both sides (Data Provider and Data Source), Anonymous Credentials are used.

Proxy Re-Encryption & Private Information Retrieval (single server)

The combination of Proxy Re-Encryption [49] and Private Information Retrieval in the single server setting [130] can be thought of as two layers: On the first layer, Private Information Retrieval is applied so that the Data Provider does not know about the data accessed. As the Data Provider could, however, get knowledge about the data stored (independently of users' access), it is desirable to encrypt data. As forwarding is a desired feature as well, Proxy Re-Encryption is added as a second layer of security. This layer is transparent to Private Information Retrieval as it only delivers some data without requiring knowledge of the actual content. Therefore, it has no impact on Private Information Retrieval if the data is stored/delivered as plaintext or if any arbitrary operation like encodings, transformations, encryptions etc. has been performed on it - the Data Provider just delivers some data.

For Proxy Re-Encryption on the other hand, the intention is to avoid the retrieval of records. However, Private Information Retrieval cannot be used for the re-encryption process: the Data Owner wants to re-encrypt data *without* having to receive it (which as a consequence, excludes transmission/retrieval). Therefore, Private Information Retrieval cannot be applied because no data is transferred. Thus, the server is able to find out about the record accessed although it does not know its *Content* at any point in time.

As Private Information Retrieval can be combined with Proxy Re-Encryption, but not the other way round, the combination has got different properties depending on the use case. However, the aim is not to apply Private Information Retrieval *while* performing Proxy Re-Encryption but to enable the user to apply both techniques in a non-concurrent way. Therefore, the evaluation is based on what is possible with both systems in general (within different use cases).

The *Protected Aspect* is the *Content* (*Data*) as well as the *Behaviour* (*Access Patterns*). The *Means of Protection* are *Confidentiality* and *Unlinkability*, and the *Affected Datatype* is *Combined* (*Stored* and *Processed*). The *Base of Security* is *asymmetric Cryptography* for protecting the access as well as for the Re-Encryption. As a consequence, the *Strength of Security* is *Computational Security* for the encryption as well as for the protection of accessing the records. The *Trusted Third Party Participation* depends on the part taken into account too: for information retrieval, no Trusted Third Party is required so that the *Frequency* is *Never*, the *Phase* is *None*, and the *Background* is *No TTP Participation*. For the encryption on the other hand, a Trusted Third Party is explicitly desired so the *Frequency* is *Always*, the *Phase* is *Operations*, and the *Background* are *Operations*. The *Enforceability* is *None* (Encryption) respectively *Both* (information retrieval). The *Reversibility* is equal for both techniques so it stays the same when they are combined. This results in a *Degree* that is *Fully Reversible* and that requires no *Cooperation*. Together these techniques cover the whole *Profile Perspective* (*Content Perspective* as well as *Access Perspective*).

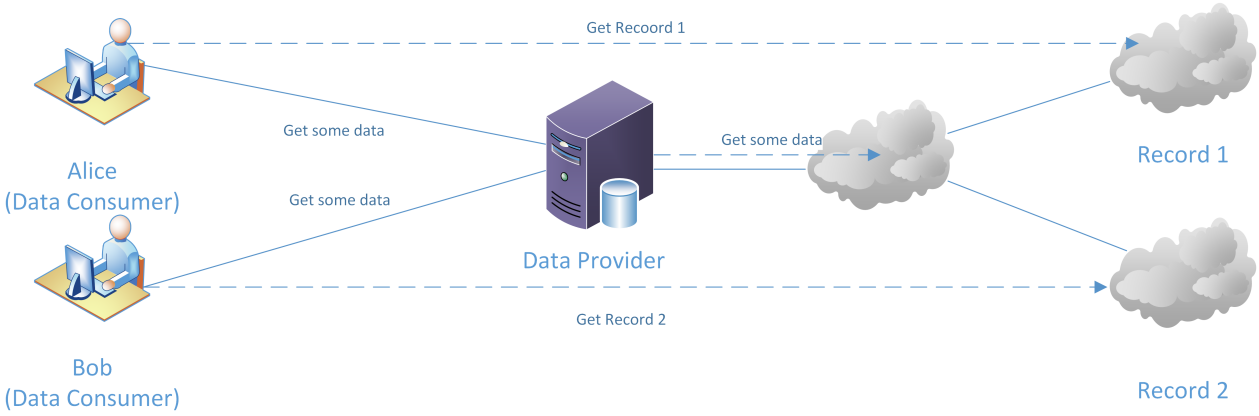


Figure 6.2: Combination of Proxy Re-Encryption and Private Information Retrieval

The combination of Private Information Retrieval and Proxy Re-Encryption is powerful and complement each other as follows: while Private Information Retrieval allows to maintain privacy when accessing a resource, Proxy Re-Encryption protects sensitive data from the Data Source itself which may be outsourced and semi-/untrusted. However, it becomes clear when looking at the combinations of the groups above that further combinations are required to deal with other aspects of privacy properly.

As illustrated in figure 6.2, the Data Provider knows about the users' identities but neither which record is accessed nor the content of the records it stores. This does not refer to the content of one single transaction (which is protected by Private Information Retrieval) but on the content that is stored on the entire system. In contrast to figure 3.13 at page 29, the content of the files is hidden, which is indicated by the additional cloud symbols for the records.

As shown in figure 6.3 at page 120, Private Information Retrieval steps include 13-19 while Proxy Re-Encryption steps involve 1-12, 16,18 and 20. Note that steps 16 and 18 are required for both techniques.

The protocol is initiated by Alice when she starts the key generation with sending $r_{A,1}$ to the PKG (step 1). It responds with the parameters r'_A, h'_A, $r_{A,2}$, $h_{A,2}$, $r_{A,3}$, and $h_{A,3}$ (step 2) which then are used by Alice to compute $r_{A,1}$ and $h_{A,1}$. At this point in time, the generation of the private key (r_A, $r_{A,1}$, $h_{A,1}$, $r_{A,2}$, $h_{A,2}$, $r_{A,3}$, and $h_{A,3}$) as well as the public key ($p_{A,1}$ and $p_{A,2}$) is finished (step 3).

Next, for storing encrypted data at the server, she computes C (consisting of C_1, C_2, C_3, C_4, C_5, and C_6) and β (step 4) and sends C to the storage server (step 5).

For initialising the Re-Encryption, Alice computes U, V, W, and σ (step 6). To enable the creation of the Re-Encryption key, Alice sends σ, IDB (Bob's ID), and a_i to the server (step 7) which then calculates $r_{k_{AB}}$, A_1, B_1, and B_2 which are returned to Alice subsequently (step 9). Then the server calculates C_1' (step 10) while Bob sends h'_B to Alice (step 11). Alice then returns ${h'_B}^{1/r_A}$ and ${B_1}^{1/r_A}$ (step 12). At this point in time, the Re-Encryption is finished and

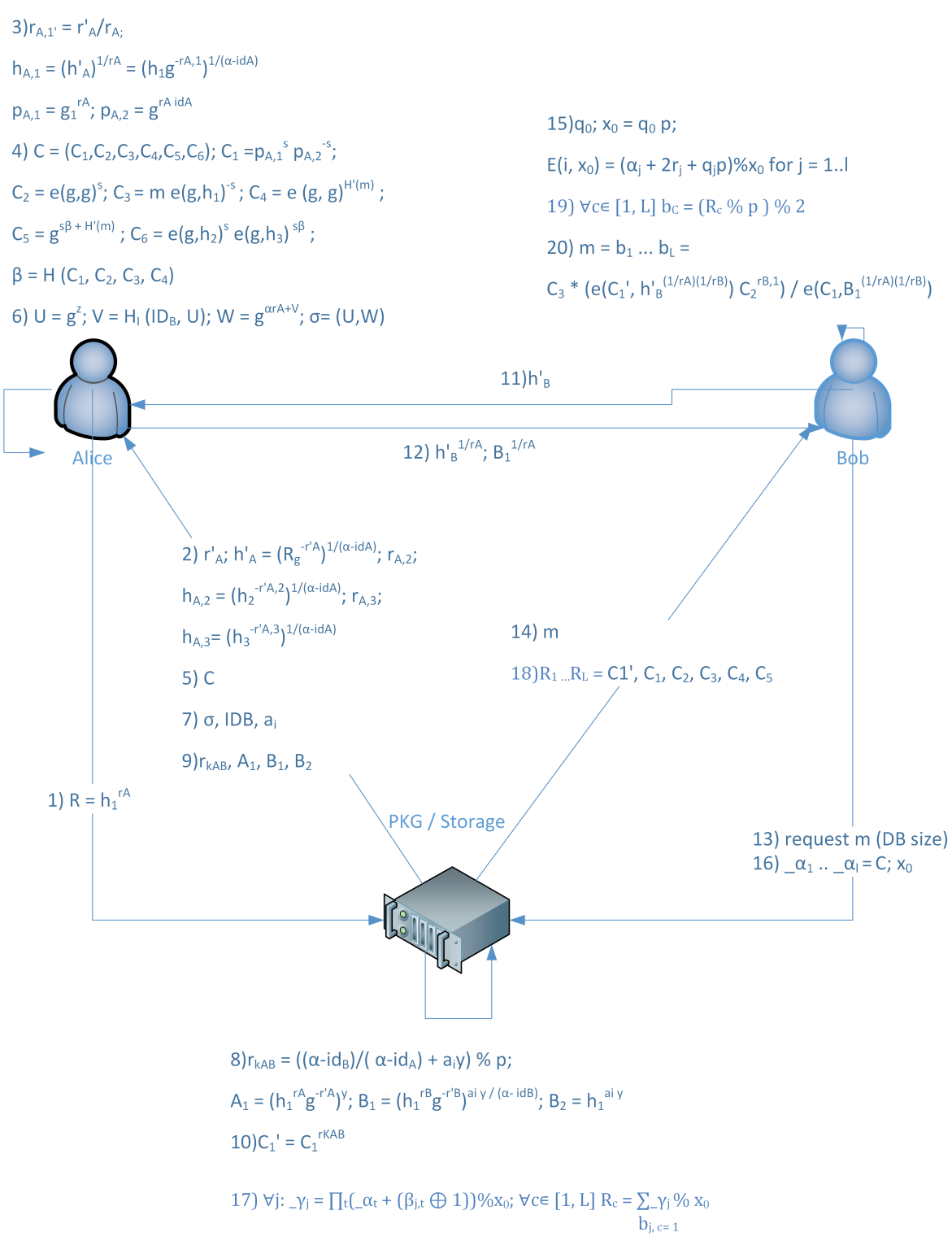


Figure 6.3: Combination of Proxy Re-Encryption and Private Information Retrieval - Detailed
120

Bob is able to decrypt the ciphertext.

Next Bob wants to retrieve the message using Private Information Retrieval. For this reason, he first requests the size m of the database which is then returned by the server (steps 13 and 14). Bob then selects q_0, calculates x_0 and encrypts the index i of the data he is looking for with x_0 which is denoted by E(i, x_0). He then sends the bits of the encrypted index ($\_\alpha_1$ .. $\_\alpha_l$) to the server in step 16. The server then computes γ_j for all indices j and subsequently R_1 .. R_L (step 17) and returns R_1 .. R_L to Bob (step 18). Bob then calculates the bits b_c for all R_c (step 19) he received and has successfully received the desired information by using Private Information Retrieval.

Last but not least, Bob has to reconstruct the ciphertext created by Proxy Re-Encryption out of the bits b_1 .. b_L and then is able to restore the original message m (step 20).

Again a practical example is provided to prove the applicability of this combination: an association for the protection of creditors stores information about the credit ratings of companies. When its clients access this database, they may want to avoid that anyone knows which records they access because this may allow conclusions about potential new business partners etc. For this reason, Private Information Retrieval is applied so that the Data Provider does not know which Data Consumer accessed which record. Data Consumers have the possibility to enrich these credit ratings with comments etc., which allows them to mark which companies they already checked in recent time. As this information is confidential but gets stored by the Data Provider, encryption is required because the option to forward this additional information to business partners is important, so Proxy Re-Encryption is applied.

Partially Blind Signature & Searchable Encryption

As the techniques for Partially Blind Signature have equal properties regarding this evaluation, no specific technique needs to be picked for this combination. For Searchable Encryption the situation is similar and all techniques have got similar properties regarding this evaluation. However, the *Base of Security* is the distinguishing property. For this reason, one specific technique has to be chosen so that an accurate evaluation of the combination is possible. This technique is Error-Tolerant Searchable Encryption (cf. [14]) which is based on *asymmetric Cryptography*. As a result the evaluation below can be transferred to any scheme for Searchable Encryption that builds on *asymmetric Cryptography*.

As both techniques are based on *Cryptography*, it has to be clarified that combining both techniques does not interfere with each other's algorithmic and mathematical basics. Fortunately, both schemes can be applied sequentially so that again a layered protection is used. Under the assumption that it does not cause problems if two different cryptographic algorithms are applied sequentially on data (e.g. some plaintext is first encrypted with AES and afterwards with RSA) the combination is possible from the security point of view.

After ensuring security for this combination, a second crucial (and for cryptographic schemes non-trivial) question rises: is it possible to combine the schemes while still being able to use all features of both schemes?

In figure 6.4 it is illustrated how the packaging works; it can be thought of like blackboxes which are put into each other. The content of these blackboxes is not important for the outer box - it is just processed as it is. Searchable encryption and Blind Signature could be substituted by any

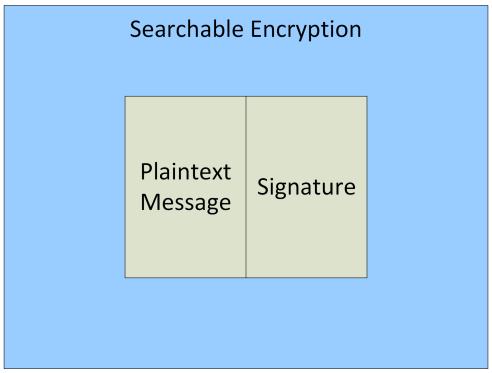


Figure 6.4: Packaging when combining Blind Signature and Searchable Encryption

other scheme for encryption respectively signatures without causing problems. The inner box consists of the plaintext message and its signature which are both concatenated. The outer box encrypts both together using Searchable Encryption. The part of the signature can be assumed to have a predefined length, so it is easy to ignore it except for verifying the message.

Note that the search is performed on an index structure rather than the ciphertext. This index structure is created based on the plaintext. Furthermore, the PuK cryptography is only used to protect the content but is not involved in the search. As a consequence, the creation of the signature has no impact on the search capabilities and does not influence the closeness of the buckets.

As shown in figure 6.5, the sender sends the encrypted message x to the signer who returns it signed (steps 1 and 2). Next the sender computes ϕ(x) (step 3) and then stores the encrypted concatenation of x and the signature at the server at a location depending on ϕ(x) (step 4). Next the sender computes α_i and stores the encrypted value of ϕ(x) in the bucket T_{α_i} (step 6a). In order to enhance privacy, all buckets are expected to have the same size, so all buckets smaller than the biggest one are padded with random entries (step 6b). At this point in time, the storage is finished.

For retrieving data, the receiver wants to look for some keyword x' which may be different from x. If x and x' are close enough, a match is possible. First, the receiver computes α_i for x' (step 7). Then all buckets T_{α_i} are requested (step 8) and retrieved (step 9). The documents' IDs are then found by calculating the intersection of all retrieved buckets (step 10). These IDs are then requested from the server (step 11) which delivers these encrypted documents (including their signatures) to the receiver (step 12). In order to verify their integrity, the receiver sends the encrypted documents to the signer (step 13) and receives the respective signatures (step 14). If the signature is equal to the one the receiver got from the server, the receiver can decrypt the documents (step 15) and finally has the plaintext results he or she was looking for.

The *Protected Aspect* for both techniques is *Content* (*Data*) and, for this reason, it does not change for the combination. The *Means of Protection* are *Confidentiality* as well as *Indistinguishability* and *Unlinkability*. The *Affected Datatype* is *Combined* as all basic datatypes (*Processed*, *Stored*, and *Transmitted*) are affected. As mentioned, the approach chosen for Searchable

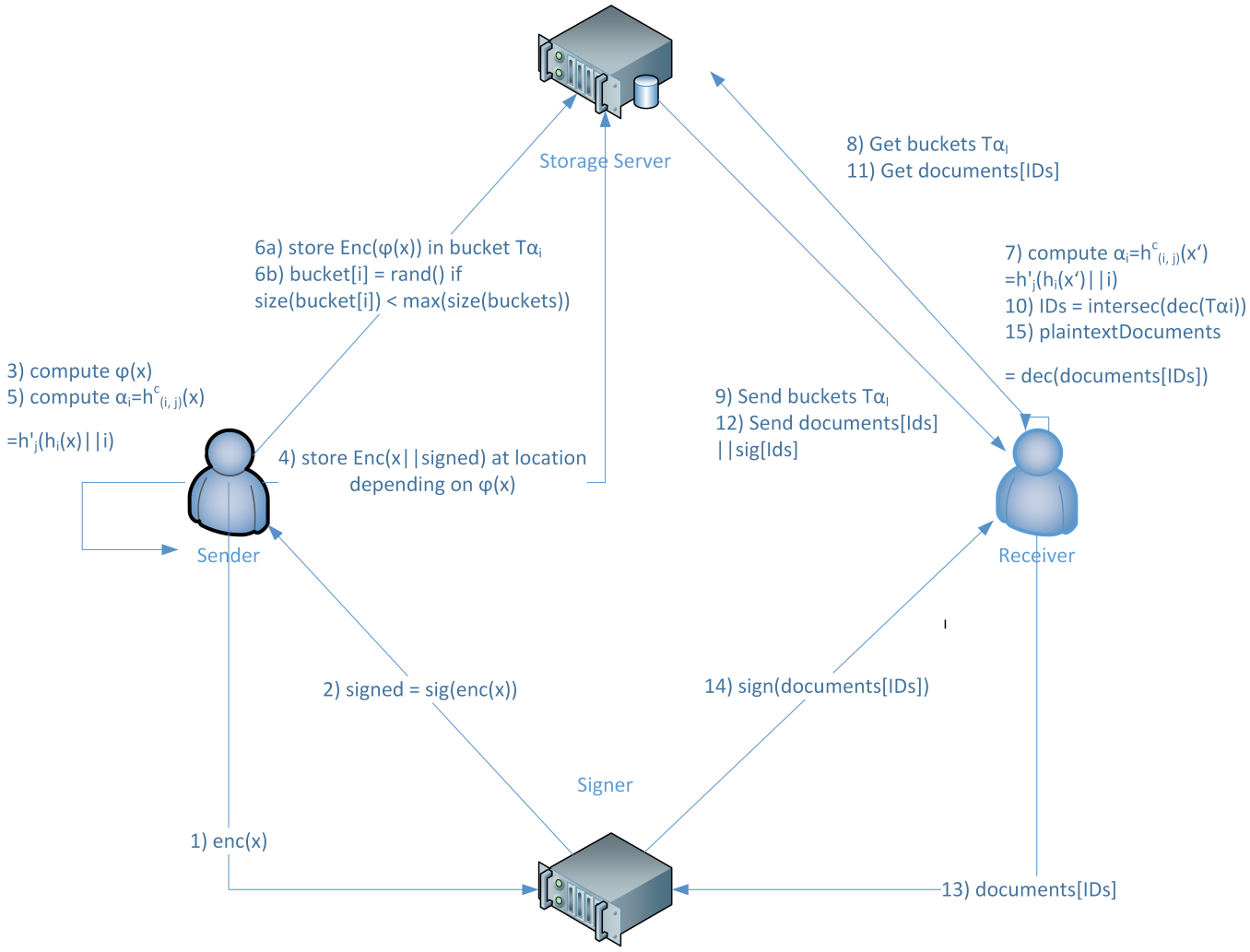


Figure 6.5: Operations when combining Blind Signature and Searchable Encryption

Encryption is based on *asymmetric Cryptography*. Because Partially Blind Signatures are based on *asymmetric Encryption* too, this applies for the combination too. As a consequence, the combination offers *Computational Security* as the *Strength of Security*. Regarding the *Trusted Third Party Participation*, both techniques have got different properties so the evaluation of the combination requires taking the use cases into account. For creating the signature, no Trusted Third Party is required so the *Frequency* is *Never*, the *Phase* is *None*, and the *Background* is *No TTP Participation*. For searching within the encrypted documents, delegation is desired so the *Trusted Third Party Participation* occurs with *Frequency Always*, in the *Phase* of *Operation*, and with the *Background* of *Operations*. Regarding the *Enforceability*, the situation is similar: while the signature can be enforced by *Both*, Searchable Encryption has *Enforceability None* because cooperation between the involved parties is required. As both techniques are based on *Cryptography* and, therefore, require *Reversibility* with the *Degree Fully Reversible*, the combination has got this *Degree* too. *Cooperation* is not required for reversing the protection. Both techniques cover the *Profile Perspective* partially (to be more precise the *Content Perspective*) and, therefore, this applies to the combination too.

The combination helps to mitigate the drawbacks of both techniques as it combines two approaches that ensure integrity as well as confidentiality of data: encryption with a digital signature. Therefore, it is ensured that some data is actually provided by a specific system (signature), while, on the other hand, ensuring that it cannot be accessed by any unauthorised user or system (encryption). This combination is actually an extension of ordinary encryption with digital signature as it is privacy friendly while still enforcing control and policies. It allows to outsource sensitive data (storage, re-encryption, and signature) while increasing the privacy of the users of a system, although the membership to a certain group can be proved. Again, a combination of more techniques allows to ensure even stronger privacy protection.

For the practical example, a system for digital money is considered. In order to respect the customers' privacy, Blind Signatures are used so that tracing of customers is prevented but creating copies of digital money is not possible. All data associated to the user's account is encrypted so that data is protected even in case of unauthorised access. If the user wants to add comments to the digital money, he or she probably wants to perform search queries too. For this reason, Searchable Encryption is used.

CHAPTER 7

Conclusion

In this chapter the research questions formulated in chapter 1 and a summary of their answers is provided. Furthermore possibilities for future work are presented.

7.1 Research Questions

The research questions formulated for this thesis and their answers are summed up below.

How could Privacy Enhancing Technologies be compared?

Comparison of Privacy Enhancing Technologies is possible based on the taxonomy introduced in the chapter Taxonomy at page 15. This taxonomy consists of 2 sub-taxonomies. They allow classification from a technological point of view respectively taking perspectives into account which allow to classify the techniques in a structured way. Chapter Evaluation at page 57 demonstrates that the taxonomy enables comparison of various techniques.

Which (recent) approaches exist to enhance privacy?

A list and a short description of the most interesting recent developments is presented in chapter Techniques at page 33.

What are the pros and cons and in which contexts may these approaches be applied?

In the chapters Techniques at page 33, Evaluation at page 57, and Synergy Effects of PET Groups at page 105 this question is answered by investigating the techniques and comparing them.

Comparison of the different approaches based on the taxonomy

The comparison of different techniques is demonstrated in chapter Evaluation at page 57.

How may different approaches be combined to enhance privacy for different scenarios and needs?

Possible combinations of Privacy Enhancing Technologies are suggested in chapter Synergy Effects of PET Groups at page 105. Two levels of detail are used for the combinations: first, groups of approaches are combined so that generic possibilities for combinations are presented to illustrate limitations of existing approaches and how to mitigate them. Afterwards, a more in-depth view of several explicit examples of PET combinations is given.

7.2 Future Work

This section indicates future work that cannot be covered within this thesis.

How can the advantages and use cases of cryptographic schemes be combined?

As mentioned, cryptography has the drawback that small details may allow short cuts for calculating the mathematical problems cryptography is based on. Thus, cryptography can be broken. As a consequence, it is hard to combine the advantages or use cases two cryptographic techniques like e.g. Searchable Encryption and Homomorphic Encryption.

Combining more techniques

In chapter Synergy Effects of PET Groups at page 105 several combinations of techniques are introduced to demonstrate that the combinations are possible and actually make sense. As covering all combinations is out of scope for this thesis, further research is required for covering more combinations.

Creating Combinations of more than 2 techniques

The combinations in this thesis consist only of two techniques. Although this already helps to cover more aspects of privacy, combinations of more techniques allow to protect privacy even better. Thus, a combination of more techniques is highly desirable.

Creating an evaluation of a set with a size of statistical relevance

The evaluation in chapter Evaluation at page 57 is only based on the techniques introduced in this thesis. As the number of the investigated techniques is low, an evaluation of more techniques is desirable. For this evaluation, a high number of techniques should be taken into account to ensure statistical relevance for the results. Still, the selection of the approaches within this thesis demonstrates the viability of a taxonomy-based classification of PETs in a more general context than found in literature and provides a suitable starting point for further research.

Bibliography

[1] Annie I. Antón and Julia B. Earp. A requirements taxonomy for reducing web site privacy vulnerabilities. *Springer*, 2003.

[2] Man Ho Au, Apu Kapadia, and Willy Susilo. Blacr: Ttp-free blacklistable anonymous credentials with reputation. *bepress*, 2012.

[3] Algirdas Avižienis, Jean-Claude Laprie, Randell Brian, and Carl Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1), 2004.

[4] Foteini Baldimtsi and Anna Lysyanskaya. Anonymous credentials light. *International Association for Cryptologic Research*, 2012.

[5] Ken Barker, Mina Askari, Mishtu Banerjee, Kambiz Ghazinour, Brenan Mackas, Maryam Majedi, Sampson Pun, and Adepele Williams. A data privacy taxonomy. *Springer*, 2009.

[6] Ämin Baumeler and Anne Broadbent. Quantum private information retrieval hast linear communication complexity. *arXiv preprint arXiv*, 2013.

[7] Mira Belenkiy, Jan Camenisch, Melissa Chase, Markulf Kohlweiss, Anna Lysyanskaya, and Hovav Shacham. Randomizable proofs and delegatable anonymous credentials. *International Association for Cryptologic Research*, 2009.

[8] Rikke Bendlin, Jesper Buus Nielsen, and Peter Sebastian Nordholt. Lower and upper bounds for deniable public-key encryption. *International Association for Cryptologic Research*, 2011.

[9] Adonis Bogris, Panagiotis Rizomiliotis, Konstantinos E Chlouverakis, Apostolos Argyris, and Dimitris Syvridis. Feedback phase in optically generated chaos: A secret key for cryptographic applications. *Quantum Electronics, IEEE Journal of*, 44(2):119–124, 2008.

[10] Alexandra Boldyreva, Nathan Chenette, and Adam O'Neill. Order-preserving encryption revisited: Improved security analysis and alternative solutions. *Springer*, 2011.

[11] Ernie Brickell, Liqun Chen, and Jingtao Li. A static diffie-hellman attack on several direct anonymous attestation schemes. *Springer*, 2012.

[12] Ernie Brickell and Jiangtao Li. Enhanced privacy id: a direct anonymous attestation scheme with enhanced revocation capabilities. *IEEE Transactions on Dependable and Secure Computing*, 9(3):345–360, May/June 2012.

[13] Julien Bringer and Hervé Chabanne. Another look at extended private information retrieval protocols. *AFRICACRYPT*, pages 305–322, 2009.

[14] Julien Bringer, Hervé Chabanne, and Bruno Kindarji. Error-tolerant searchable encryption. *Institute of Electrical and Electronics Engineers*, 2009.

[15] Johannes Buchmann. *Einführung in die Kryptographie*. Springer, 2010.

[16] Bundesdatenschutzgesetz. www.gesetze-im-internet.de/bdsg_1990/.

[17] Jan Camenisch, Maria Dubovitskaya, and Gregory Neven. Unlinkable priced oblivious transfer with rechargeable wallets. *Springer*, 2010.

[18] Jan Camenisch, Maria Dubovitskaya, Gregory Neven, and Gregory M. Zaverucha. Oblivious transfer with hidden access control policies. *Springer*, 2011.

[19] Sébastien Canard and Roch Lescuyer. Anonymous credentials from (indexed) aggregate signatures. *DIM*, 2011.

[20] Liang Cheng, Zhang Tong, Wen Liu, and Gao Chengmin. Non-interactive exponential homomorphic encryption algorithm. *IEEE Computer Society*, pages 224–227, 2012.

[21] M. Chuah and W. Hu. Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data. *IEEE Computer Society*, 2011.

[22] Irving M Copi and Richard W Miller. *Introduction to Logic: Study Guide*. Macmillan, 1972.

[23] Sergiu Costea, Dumitru Marian Barbu, Gabriel Ghinita, and Razvan Rughinis. A comparative evaluation of private information retrieval techniques in location-based services. *IEEE Computer Society*, 2012.

[24] Scott Coull, Matthew Green, and Susan Hohenberger. Controlling access to an oblivious database using stateful anonymous credentials. *International Association for Cryptologic Research*, 2009.

[25] Datenschutzgesetz-2000. http://www.ris.bka.gv.at/geltendefassung.wxe ?abfrage=bundesnormen&gesetzesnummer=10001597.

[26] Yvo Desmedt. Information theoretic security. In *Second International Conference, ICITS*. Springer, 2007.

[27] Casey Devet, Jan Goldberg, and Nadia Heninger. Optimally robust private information retrieval. *IACR Cryptology ePrint Archive 2012*, 2012.

[28] Whitfield Diffie and Martin E Hellman. New directions in cryptography. *Information Theory, IEEE Transactions on*, 22(6):644–654, 1976.

[29] Maozhen Ding, Fei Gao, Zhengping Jin, and Hua Zhang. An efficient public key encryption with conjunctive keyword search scheme based on pairings. *Institute of Electrical and Electronics Engineers*, 2012.

[30] Ning Ding and Dawu Gu. On the undecidability of quasi-private-key-encryption statistical indistinguishability. *IEEE Computer Society*, 2007.

[31] Directive-95/46/EC. http://eur-lex.europa.eu/lexuriserv/lexuriserv.do ?uri=celex:31995l0046:en:html.

[32] Josep Domingo-Ferrer, Agusti Solanas, and Jordi Castellá-Roca. h(k)-private information retrieval from privacy-uncooperative queryable databases. *Online Information Review*, 33(4):720–744, 2009.

[33] Markus Dürmuth and David Mandell Freeman. Deniable encryption with negligible detection probability: An interactive construction. *International Association for Cryptologic Research*, 2011.

[34] Catherine Dwyer. Privacy in the age of google and facebook. *IEEE Technology and Society Magazine*, Fall, September 2011.

[35] Liming Fang, Willy Susilo, and Jiadong Wang. Anonymous conditional proxy reencryption without random oracle. *Springer*, 2009.

[36] Hannes Federrath. Privacy enhanced technologies: Methods - markets - misuse. *Springer Verlag*, 2005.

[37] Deng-Guo Feng, Jing Xu, and Xiao-Feng Chen Chen. An efficient direct anonymous attestation scheme with forward security. *WSEAS Transactions on Communications 8.10 (2009): 1076-1085*, 8:1076–1085, October 2009.

[38] Marc Fischlin and Dominique Schröder. Security of blind signatures under aborts. *Springer*, 2009.

[39] Georg Fuchsbauer and Damien Vergnaud. Fair blind signatures without random oracles. *AFRICACRYPT 2010*, 2010.

[40] Benjamin C. M. Fung, Ke Wang, Chen Rui, and Yu Philip S. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4), June 2010.

[41] Chong-zhi Gao, Dongqin Xie, and Jin Li. Deniably information-hiding encryptions secure against adaptive chosen ciphertext attack. *IEEE Computer Society*, 2012.

[42] Craig Gentry. *A Fully Homomorphic Encryption Scheme*. PhD thesis, Standford University, September 2009.

[43] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of computer and system sciences*, 28(2):270–299, 1984.

[44] GoTrusted. http://www.gotrusted.com/.

[45] Ulrich Greveler, Benjamin Justus, and Dennis Loehr. Direct anonymous attestation: Enhancing cloud service user privacy. *Springer*, 2011.

[46] Jorge Guajardo, Bart Mennink, and Berry Schoenmakers. Anonymous credential schemes with encrypted attributes. *Springer*, 2010.

[47] Jan Hajny and Lukas Malina. Practical revocable anonymous credentials. *International Federation for Information Processing*, 2012.

[48] Jan Hajny and Lukas Malina. *Unlinkable attribute-based credentials with practical revocation on smart-cards*. Springer, 2013.

[49] Yi-Jun He, Tat Wing Chim, Lucas Chi Kwong Hui, and Siu-Ming Yui. Non-transferable proxy re-encryption scheme. *Institute of Electrical and Electronics Engineers*, 2012.

[50] Yingying He, Liquan Chen, and Lingling Wang. An improved direct anonymous attestation scheme for m2m networks. *Procedia Engineering*, 15:1481–1486, 2011.

[51] Johannes Heurix, Peter Zimmermann, and Thomas Neubauer. A taxonomy for privacy enhancing technologies. *submitted to Computers & Security*, 2014.

[52] Jaydeep Howlader and Saikat Basu. Sender-side public key deniable encryption scheme. *IEEE Computer Society*, 2009.

[53] Ayad Ibrahim, Hai Jing, Ali A. Yassin, and Deqing Zou. Approximate keyword-based search over encrypted could data. *IEEE Computer Society*, 2012.

[54] Manfred A Jeusfeld. Metadata. In *Encyclopedia of Database Systems*, pages 1723–1724. Springer, 2009.

[55] Xiaoqi Jia, Jun Shao, Jiwu Jing, and Peng Liu. Cca-secure type-based proxy re-encryption with invisible proxy. *IEEE Computer Society*, 2010.

[56] Han Jing-Li, Yang Ming, and Wang Zhao-Li. Fully homomorphic encryption scheme extended to large message space. *IEEE Computer Society*, pages 533–536, 2011.

[57] Apu Kapadia and Prasad Naldurg. Distributed enforcement of unlinkability policies: Looking beyond the chinese wall. *IEEE Computer Society*, 2007.

[58] Jonathan Katz, Dominique Schröder, and Arkady Yerukhimovich. Impossibility of blind signatures from one-way permutations. *Springer*, 2011.

[59] David Kotz. A threat taxonomy for mhealth privacy. *IEEE*, 2011.

[60] Samuel F Kovacic. General taxonomy of system [ic] approaches for analysis and design. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on*, volume 3, pages 2738–2743. IEEE, 2005.

[61] Ben Kreuter. Private information retrieval based on fully homomorphic encryption. 2010.

[62] Jorn Lapon. *Anonymous Credential Systems: From Theory Towards Practice*. PhD thesis, Katholieke Universiteit Leuven, July 2012.

[63] Bingdong Li, Esra Erdin, Mehmet Hadi Günes, George Bebis, and Todd Shipley. *Traffic Monitoring and Analysis*, volume 6613. Springer Berlin Heidelberg, 2011.

[64] Benoît Libert and Damien Vergnaud. Unidirectional chosen-ciphertext secury proxy re-encryption. *Institute of Electrical and Electronics Engineers*, 2011.

[65] Yehuda Lindell and Hila Zarosim. On the feasibility of extending oblivious transfer. *Springer*, 2013.

[66] Di Ma and Anudath K Prasad. A context-aware approach for enhanced security and privacy in rfid electronic toll collection systems. *Institute of Electrical and Electronics Engineers*, 2011.

[67] MO Chun Man and Victor K. Wei. A taxonomy for attacks on mobile agent. *EUROCON'2001, Trends in Communications, International Conference on.*, 2:385–388, 2001.

[68] Santi Martínez, Josep M. Miret, Rosana Tomàs, and Magda Valls. Security analysis of order preserving symmetric cryptography. *Natural Sciences Publishing Corporation*, 2013.

[69] Kevin S McCurley. The discrete logarithm problem. In *Proc. of Symp. in Applied Math*, volume 42, pages 49–74, 1990.

[70] Alfred J Menezes, Paul C Van Oorschot, and Scott A Vanstone. *Handbook of applied cryptography*. CRC press, 1996.

[71] Nikolay A. Moldovyan and Alexander A. Moldovyan. Blind collective signature protocol based on discrete logarithm problem. *IJ Network Security*, 2010.

[72] Rubinhoff; Marshall C. Yovits Morris. *Advances in Computers*, volume 16. Academic Press, 1977.

[73] Mummoorthy Murugesan, Ahmet Erhan Nergiz, Wei Jiang, and Serkan Uzunbaz. k-out-of-n oblivious transfer based on homomorphic encryption and solvability of linear equations. *ACM*, 2011.

[74] I2P Anonymisation Network. http://www.i2p2.de/.

[75] Femi Olumofin and Ian Goldberg. Revisiting the computational practicality of private information retrieval. *Springer*, pages 158–172, 2012.

[76] Adam O'Neill, Chris Peikert, and Brent Waters. Bi-deniable public-key encryption. *International Association for Cryptologic Research*, 2011.

[77] Christos H. Papadimitriou. Computational complexity. In *Encyclopedia of Computer Science*, pages 260–265. John Wiley and Sons Ltd., Chichester, UK, 2003.

[78] Russell Paulet, Golam Kaosar, and Xun Yi. k anonymous private query based on blind signature and oblivious transfer. *2nd International Cyber Resilience Conference 2011*, 2011.

[79] Andreas Pfitzmann and Marit Hansen. A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. v0.34, August 2010.

[80] Raluca Ada Popa, Frank H. Li, and Nickolai Zeldovich. An ideal-security protocol for order-preserving encoding. *http://people.csail.mit.edu/nickolai/papers/popa-mope.pdf*, 2013.

[81] Privacy-Act. http://www.justice.gov/opcl/privacy-act-1974.

[82] JAP ANON Project. http://anon.inf.tu-dresden.de/.

[83] TOR Project. https://www.torproject.org/.

[84] Govinda Y Ramaiah and Vijaya G Kumari. Efficient public key homomorphic encryption over integer plaintexts. *IEEE*, 2012.

[85] David Rebollo-Monedero and Jordi Forné. Optimized query forgery for private information retrieval. *IEEE Transactions on Information Theory*, 56(9):4631–4642, September 2010.

[86] Alfredo Rial and Bart Preneel. Optimistic fair priced oblivious transfer. *AFRICACRYPT 2010*, 2010.

[87] Ronald L Rivest, Adi Shamir, and Len Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[88] Markus Rückert. Lattice-based blind signatures. *ASIACRYPT 2010*, 2010.

[89] Markus Rückert and Dominique Schröder. Fair partially blind signatures. *AFRICACRYPT 2010*, 2010.

[90] Eun-Kyung Ryu and Tsuyoshi Takagi. Efficient conjunctive keyword-searchable encryption. *IEEE Computer Society*, 2007.

[91] Pekka Savolainen, Eila Niemelä, and Reijo Savola. A taxonomy of information security for service-centric systems. *IEEE Computer Society*, 2007.

[92] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[93] jun Shao, Peng Liu, Zhenfu Cao, and Guiyi Wei. Multi-use unidirectional proxy re-encryption. *Institute of Electrical and Electronics Engineers*, 2011.

[94] Jun Shao, Guiyi Wei, Yun Ling, and Mande Xie. Identity-based conditional proxy re-encryption. *Institute of Electrical and Electronics Engineers*, 2011.

[95] Victor Shoup. Lower bounds for discrete logarithms and related problems. In *Advances in Cryptology - EUROCRYPT 97*, pages 256–266. Springer, 1997.

[96] Gustavus J. Simmons. A survey of information authentication. *IEEE*, 1988.

[97] Radu Sion and Bogdan Carbunar. On the computational practicality of private information retrieval. *Proceedings of the Network and Distributed Systems Security Symposium*, 2007.

[98] Adam Skillen and Mohammad Mannan. On implementing deniable storage eencryption for mobile devices. *Concordia*, 2013.

[99] Geoff Skinner, Song Han, and Elizabeth Chang. An information privacy taxonomy for collaborative environments. *Information Management & Computer Security*, 14(4), 2006.

[100] Sarah Spiekermann. Die konsumenten der anonymität. *Datenschutz und Datensicherheit*, 27:150–154, 2003.

[101] Markus Stadler, Jean-Marc Piveteau, and Jan Camenisch. Fair blind signatures. In *Advances in Cryptology - Eurocrypt '95*, pages 209–219. Springer, 1995.

[102] Melodie Szeto and Ali Miri. Analysis of the use of privacy-enhancing technologies to achieve pipeda compliance in a b2c e-business model. *IEEE Computer Society*, 2007.

[103] Qiang Tang. Type-based proxy re-encryption and its construction. *Springer*, 2008.

[104] Yinqi Tang, Dawu Gu, Ning Ding, and Haining Lu. Phrase search over encrypted data with symmetric encryption scheme. *IEEE Computer Society*, 2012.

[105] Rose et. al. Tinabo. Anonymisation vs. pseudonymisation: Which one is most useful for both privacy protection and usefulness of e-healthcare data. *Institute of Electrical and Electronics Engineers*, 2009.

[106] Bengisu Tulu and Samir Chatterjee. A taxonomy of telemedicine efforts with respect to applications, infrastructure, delivery tools, type of setting and purpose. *IEEE*, 2005.

[107] Vinod Vaikuntanathan. Computing blindfolded: New developments in fully homomorphic encryption. *IEEE Computer Society*, pages 5–16, 2011.

[108] Marten van Dijk, Craig Gentry, Shai Halevi, and Vinod Vaikuntanathan. Fully homomorphic encryption over the integers. *Springer*, Advances in Cryptology - EUROCRYPT 2010, 2010.

[109] Sophie in 't Veld. Data sharing across the atlantic. *IEEE Security & Privacy*, 5:58–61, 2007.

[110] H.S. Venter and J.H.P. Eloff. A taxonomy for information security technologies. *Computers & Security*, 2003.

[111] Europe versus Facebook. http://europe-v-facebook.org/.

[112] Chuanxiao Wang and Jizhong Wang. A shared-key and receiver-deniable encryption scheme over lattice. *Journal of Information Systems 8*, 2, 2012.

[113] Cong Wang, Ning Cao, Kui Ren, and Wenjing Lou. Enabling secure and efficient ranked keyword search over outsourced cloud data. *IEEE Computer Society*, 2012.

[114] Ding Wang, Min Huang, Ying Wang, and Peng Chen. The design and building of metadata model in xbrl taxonomy engineering. *IEEE Computer Society*, 2011.

[115] Xu An Wang and Weidong Zhong. A new identity based proxy re-encryption scheme. *Institute of Electrical and Electronics Engineers*, 2010.

[116] Xun An Wang, Ziqing Wang, Yi Ding, and Shujun Bai. K-times proxy re-encryption. *IEEE Computer Society*, 2011.

[117] Jian Weng, Robert H. Deng, and Xuhua Ding. Conditional proxy re-encryption secure against chosen-ciphertext attack. *ACM*, 2009.

[118] Jian Weng, Yanjiang Yang, Quiang Tang, Robert H. Deng, and Feng Bao. Efficient conditional proxy re-encryption with chosen-ciphertext security. *Springer*, 2009.

[119] Alan F. Westin. *Privacy and Freedom*, volume 5. Atheneum, 1968.

[120] Lei Lei Win, Tony Thomas, and Sabu Emmanuel. Privacy enabled digital rights management without trusted third party assumption. *IEEE Transactions on Multimedia*, 3(14), June 2012.

[121] Liangliang Xiao and I-Ling Yen. Security analysis for order prserving encryption schemes. *Institute of Electrical and Electronics Engineers*, 2012.

[122] Liangliang Xiao, I-Ling Yen, and Dung T. Huynh. Extending order preserving encryption for multi-user systems. *International Association of Cryptologic Research, https://eprint.iacr.org/2012/192.pdf*, 2012.

[123] Gang Xu, Leonardo Aguilera, and Yong Guan. Accountable anonymity: A proxy re-encryption based anonymous communication system. *IEEE Computer Society*, 2012.

[124] Ling Ling Xu and Fang Guo Zhang. Oblivious transfer with threshold access control. *Journal of Information Science and Engineering*, 28:555–570, 2012.

[125] Hao-Miao Yang, Qi Xia, Xiao-fen Wang, and Dian-hua Tang. A new somewhat homomorphic encryption scheme over integers. *IEEE Computer Society*, 2012.

[126] Yanjiang Yang, Feng Bao, Xuhua Ding, and Robert H. Deng. Multiuser private queries over encrypted databases. *International Journal of Applied Cryptography*, 1(4/2009):309–319, 2009.

[127] Yanjiang Yang, Haibing Lu, and Jian Weng. Multi-user private keyword search for cloud computing. *IEEE Computer Society*, 2011.

[128] Andrew C. Yao. Theory and appdcations of trapdoor functions. *IEEE*, 1982.

[129] Sergey Yekhanin. Private information retrieval. *Communications of the ACM*, 53(4):68–73, April 2010.

[130] Xun Yi, Mohammed Golam Kaosar, Russell Paulet, and Elisa Bertino. Single-database private information retrieval from fully homomorphic encryption. *Knowledge and Data Engineering, IEEE Transactions on*, 25(5):1125–1134, 2013.

[131] Dae Hyun Yum, Duk Soo Kim, and Jin Seok Kim. Order-preserving encryption for non-uniformly distributed plaintexts. *Springer*, 2012.

[132] Jiuling Zhang, Beixing Deng, and Xing Li. Additive order preserving encryption based encrypted documents ranking in secure cloud storage. *Springer*, 2012.

[133] Lei Zhang, Futai Zhang, Bo Qin, and Shubo Liu. Provably-secure electronic cash based on certificateless partially-blind signatures. *Electronic Commerce Research and Applications*, 10(5):545–552, 2011.

[134] Ye Zhang, Man Ho Au, Duncan S. Wong, Qjong Huang, Nikos Mamoulis, David W. Cheung, and Sui-Ming Yiu. Oblivious transfer with access control: Realizing disjunction without duplication. *Springer*, 2010.

[135] Jing Zhao, Dengguo Feng, and Zhenfeng Zhang. Attribute-based conditional proxy re-encryption with chosen-ciphertext security. *Institute of Electrical and Electronics Engineers*, 2010.

[136] Weidong Zhong, Xu An Wang, Ziqing Wang, and Yi Ding. Proxy re-encryption with keyword search from anonymous conditional proxy re-encryption. *IEEE Computer Society*, 2011.