Engineering Privacy by Design

Seda Gürses, Carmela Troncoso, and Claudia Diaz

K.U. Leuven/IBBT, ESAT/SCD-COSIC
firstname.lastname@esat.kuleuven.be

Abstract. The design and implementation of privacy requirements in systems is a difficult problem and requires the translation of complex social, legal and ethical concerns into systems requirements. The concept of "privacy by design" has been proposed to serve as a guideline on how to address these concerns.

"Privacy by design" consists of a number of principles that can be applied from the onset of systems development to mitigate privacy concerns and achieve data protection compliance. However, these principles remain vague and leave many open questions about their application when engineering systems. In this paper we show how starting from data minimization is a necessary and foundational first step to engineer systems in line with the principles of privacy by design.

We first discuss what data minimization can mean from a security engineering perspective. We then present a summary of two case studies in which privacy is achieved by minimizing different types of data, according to the purpose of each application. First, we present a privacy-preserving ePetition system, in which user's privacy is guaranteed by hiding their identity from the provider while revealing their votes. Secondly, we study a road tolling system, in which users have to be identified for billing reasons and data minimization is applied to protect further sensitive information (in this case location information). The case studies make evident that the application of data minimization does not necessarily imply anonymity, but may also be achieved by means of concealing information related to identifiable individuals. In fact, different kinds of data minimization are possible, and each system requires careful crafting of data minimization best suited for its purpose.

Most importantly, the two case studies underline that the interpretation of privacy by design principles requires specific engineering expertise, contextual analysis, and a balancing of multilateral security and privacy interests. They show that privacy by design is a productive space in which there is no one way of solving the problems. Based on our analysis of the two case studies, we argue that engineering systems according to the privacy by design principles requires the development of generalizable methodologies that build upon the principle of data minimization. However, the complexity of this engineering task demands caution against reducing such methodologies to "privacy by design check lists" that can easily be ticked away for compliance reasons while not mitigating some of the risks that privacy by design is meant to address.

1 Introduction

"Privacy by design" consists of a number of principles that can be applied from the onset of systems development to mitigate privacy concerns and achieve data protection compliance. However, these principles remain vague and leave many open questions about their application when engineering systems.

Engineering systems with privacy in mind requires integrating privacy requirements into the typical systems engineering activities. Mainly, this requires eliciting and analyzing privacy, as well as functional requirements; developing designs that fulfill those requirements; implementing the design; and testing that in the implementation the functional and privacy requirements are fulfilled. Since most privacy requirements rely on basic security engineering mechanisms, e.g., mechanisms for guaranteeing confidentiality, integrity or availability, security engineering activities like risk and threat analysis ought to also accompany the process. However, little past experience exists in designing systems with privacy in mind, and even those are typically invisible or inaccessible to policy makers who discuss the principles of privacy by design.

The objective of this paper is to provide an initial inquiry into the practice of privacy by design from an engineering perspective in order to contribute to the closing of the gap between policy makers' and engineers' understanding of privacy by design. Specifically, we consider the role of "data minimization" in engineering systems and its relevance when implementing privacy by design. We do so in the context of two case studies that have as their main objective the development of innovative solutions that preventatively, pro-actively, and by default embed privacy into the system.

Our experiences from the two case studies show that departing from data minimization is a necessary and foundational first step to engineer systems in line with the principles of privacy by design. This differs from some recent interpretations of what privacy by design means for systems, as we discuss in Section 2. In order to then demonstrate the central role of data minimization for privacy by design, we introduce our two case studies in Section 3. For each case study, we discuss the potential privacy risks when data minimization is not utilized as a foundational principle of systems engineering, and show how the same risks can be mitigated when data minimization is taken as the starting point of the engineering activities. We also generalize some of the engineering principles applied in the two case studies, and discuss the different flavors of data minimization applied. Finally, we conclude in Section 4 with a discussion on the interactions between policy and engineering and their relevance to privacy by design.

2 Privacy by design and technical intuition

The term "privacy by design" has been proposed by data protection policy makers [7, 42, 46]. The term was subsequently recognized in different recommendations for data protection policy, two of which we will discuss: the FTC report on "Protecting consumer privacy in an era of rapid change" [10], and the EU

Commissions Communication on "A comprehensive strategy on data protection in the European Union" [9]. In the following, we first introduce some of the proposals for privacy by design. We will then move on to analyzing what is missing in these definitions from an engineering perspective.

2.1 Hands-off privacy by design

One of the first and most prominent advocates of the term "privacy by design" is Ann Cavoukian, the Information and Privacy Commissioner of Ontario. In her articulation of how privacy by design can be accomplished she names seven guiding principles [7]. These principles were later widely adopted as a resolution by other prominent policy makers at the 32. Annual International Conference of Data Protection and Privacy Commissioners meeting in Israel. These principles are:

1. Proactive not reactive, Preventative, not Remedial
2. Privacy as the default
3. Privacy Embedded into Design
4. Full functionality - Positive Sum not Zero Sum
5. End-to-end security - Lifecyle Protection
6. Visibility and Transparency
7. Respect for User Privacy

Cavoukian's [7] concept of privacy by design extends to the "trilogy of 1) IT Systems, 2) accountable business practices, and 3) physical design and networked infrastructure". The principles listed in the document apply to this trilogy, and hence demand a holistic approach to privacy.

Despite its comprehensiveness, it is not clear from Cavoukian's document, what "privacy by design" actually is and how it should be translated into the engineering practice. Most of the principles include the term "privacy by design" in the explanation of the principle itself. For example, the definition of Principle (3), Privacy Embedded into Design, states that: "Privacy by design is embedded into the design and architecture of IT systems [...]. It is not bolted as an add-on, after the fact. The result is that privacy becomes an essential component of the core functionality being delivered. Privacy is integral to the system without diminishing functionality". This recursive definition – privacy by design means applying privacy by design – communicates to the reader that something needs to be done about privacy from the beginning of systems development, but it is not clear what exactly this privacy matter is nor how it can be translated into design.

This vagueness is not unique and, for example, can also be found in the various principles expressed in the articles of the Data Protection Directive [15]. The Directive states that data collection should be limited. As Kuner [29] summarizes:

> "EU data protection law requires that processing be strictly limited to the purpose originally notified to the data subject. For instance, Article

> 6(1)(b) of the General Directive provides, in part, that personal data must be *'collected for specified, explicit and legitimate purposes'* and must be *'adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed'* (Article 6(1)(c)) [...] (italics added)"

However, both of the clauses refer to different types of constraints on the data collected and processed, they do not mention explicitly data minimization. Although Kuner and other legal scholars argue that data minimization can be interpreted from the two principles,"meaning that processing of personal data must be restricted to the minimum amount necessary" [29], it is not explicitly included in the Directive. This lack of mention has serious consequences in practice.

The absence of an explicit and precise data minimization principle makes the proportionality clause susceptible to a reversal of its intended effect. Data collectors can articulate the purpose specification to include the any data of their liking [23], eliminating the need to consider data minimization. Even further, the interpretation of the Data Protection Directive (or Fair Information Practices) to collect all data of interest is acquiesced as long as individuals are provided with "control" over the collected information, e.g., informed consent and subject access rights.

This control can be translated in systems as policy selection and negotiation, privacy settings, and other solutions that allow individuals to access and amend their "personal data". The popularity of this approach – legitimizing the collection of copious amounts of data by providing "control" – is especially evident in most of the Identity Management initiatives documents, e.g., [22, 36]: innovative technologies developed with data protection compliant design.

And, although providing control can be a powerful tool, even when all the aforementioned protection measures are in place, the scope of control over personal data may be limited. By shrinking the scope of the definition of what counts as personal data, companies can limit the reach of solutions that provide users with "control" over their collections of data [23]. By applying simple anonymization over aggregated personal data, data processing activities are taken outside of the scope of data protection and slip further away from the control of the users.

The fact that the existing principles are not effective in communicating data minimization as an important principle to companies and governments is confirmed by the findings of the participants of the privacy roundtables held by the FTC [10]. They state that i) there is "increasing collection and use of consumer data", ii) that "the companies that collect data [...] share the data with multiple entities", and iii) that "this growth in data collection and use [is] occurring without adequate concern for consumer privacy."

Despite the recognition that existing Fair Information Practice Principles are not effective in the data collection surge, the privacy framework proposed by the Federal Trade Commission [10] defines privacy by design as follows:

> "Companies should promote consumer privacy throughout their organizations and at every stage of the development of their products and services. Companies should incorporate substantive privacy protections into their practices, such as *data security, reasonable collection limits, sound retention practices*, and *data accuracy* (italics added). Companies should maintain comprehensive data management procedures throughout the life cycle of their products and services."

Surprisingly, nowhere in the otherwise extensive document is data minimization, or collection limitation, mentioned again. This definition reminds us again that privacy by design is more than just a matter of technological design, and the emphasis on the trilogy is an important factor in avoiding falling into technocentric solutions to a socio-technical problem. However, the absence of references to any technical means or principles simply gives no motivation to explore the potentials of translating privacy into systems design. Privacy by design is posed as a non-technical strategy, which leaves one wondering why the word "design" was included in the first place.

The recent Communication of the European Commission [9] takes a similar hands-off position in the stated description of plans to examine:

> "the concept of "privacy by design" and its concrete implementation, whereby *data protection compliance would be embedded throughout the entire life cycle of technologies and procedures*, from the early design stage to their deployment and use."

This definition of privacy by design is comparable to that of proportionality and purpose limitation clause in the Data Protection Directive. It provides no additional lead with respect to the principles according to which technology should be applied. This definition of privacy by design is therefore also susceptible to the interpretation to collect any data as long as it is with a privacy label, while shrinking the scope of control, as we sketched above. It also leaves out previous recommendations from the Article 29 Data Protection Working Party [39] and of the European Data Protection Supervisor's [46] with explicit definitions of privacy by design that include the principle of data minimization.

2.2 Hands-on privacy by design

These vague definitions of privacy by design seem to be symptomatic of a disconnect between policy makers and engineers when it comes to what it means to technically comply with data protection. In all the definitions, it is implied that data minimization can be substituted through organizational and technical forms of control and transparency. Hence, the different manifestations of privacy by design in policy documents leave open the option of interpreting it as the collection and processing of any data – but with a privacy label. Privacy by design can be reduced to a series of symbolic activities [26] to assure consumers' confidence, as well as the free flow of information in the marketplace [10, 9].

From a security engineering perspective, control and transparency mechanisms do not provide the means to mitigate the privacy risks that arise through the collection of data in massive databases. This becomes especially problematic with respect to large-scale mandatory systems like road tolling systems and smart energy systems, or *de facto* mandatory systems like telecommunications (e.g., mobile phones). All of these applications lead to large databases with highly sensitive information. The dimensions of these databases, as well as their resulting attractiveness to companies, governments, and other organizations means that the risks associated with them are of great magnitude.

Given the size and characteristics of these databases there are no technical mechanisms that guarantee that they will not leak, or be subject to function creep or be simply abused by those who have authority and access to the data. Providing a sense of control through technical mechanisms like access control and transparency are only extensions of the "trust us, we do no evil" propositions of some organizations. Such mechanisms are simply limited by the basic properties of digital information: that it is easy to replicate, distribute and manipulate. Even if the organization is trustworthy, accountable, responsible and with good intentions, these risks are still present. From a security engineering perspective, the risks inherent to the digital format imply that data minimization must be the foundational principle in applying privacy by design to these systems.

It is also evident in the definitions of privacy by design that, when it comes to data minimization, existing computational capabilities are disregarded. While it is recognized in these policy documents that new technological means can enhance the collection and processing of massive amounts of data, little recognition is given to the fact that technical mimicry of daily activities can be enabled in unintuitive ways – and in some cases with much less data than is necessary in the analogue world.

Interactions in online systems may require "less" data then their analogue counterparts for a number of reasons. First, organizations may find that when they transpose some of their workflows to the digital realm, certain information is not needed. For example, Schaar [42] points out that it became evident during the development of the ELENA project[1] that:

> "The conventional paper forms used until then [...] set the standard; as a result, all data elements collected under the conventional procedure were also included in the new ELENA system. It became clear, however, that there was good reason to doubt the need for certain data fields."

Schaar then suggests that this experience, among others, informs his explicit position on privacy by design and data minimization:

[1] The project Elektronischer Entgelt Nachweis has the objective to create a database in which all the income data of persons employed in Germany will be stored. A summary of the project and consequences for privacy and data protection can be found here: https://www.datenschutzzentrum.de/elena/20100127-elana-eine-kurze-darstellung.html

> "It is very important to examine early in the planning process whether and how to limit the amount of personal data to the absolute minimum necessary. The tendency to reproduce increasingly complicated bureaucratic systems exactly in information technology [...] can lead to major problems for data protection."

Second, the information that is necessary to enable the digital equivalents of workflows, may be simplified and completed with much less data using the mathematical and computational capabilities of technologies that often transcend the boundaries of our intuition. An example of unintuitive computational magic outside of the realm of privacy are lossy data compression algorithms. Also underlying mp3, lossy data compression algorithms allow the downsizing of image and sound files. These algorithms have paved the way to indispensable functionality on the internet like voice-over-IP, online broadcasting, etc. While it is not intuitive how a sound file that is substantially reduced in size sounds indistinguishably similar to the original, we have come to accept mp3s as a desirable computational given.

We are still lacking a similar technical intuition for the sometimes "magical" computational capabilities developed over the last years. These mechanisms allow the further minimization of data that would usually be found to be "adequate, relevant and not excessive in relation to the purpose".

Anonymity and pseudonymity are some acquaint mechanisms for data minimization, that despite their accessible metaphors are difficult to grasp intuitively, e.g., what does it mean to be "indistinguishable within an anonymity set"? It is even more difficult for somebody unaware of a slew of mathematical concepts to comprehend how a zero-knowledge proof could function, a mechanism that proves that a certain statement is true without revealing any further information. For example, while a credential may encode the date of birth of a subject, the zero-knowledge protocol is able to prove that the subject is over the age of 18 *without revealing the actual date of birth*, or any other information [2]. Likewise, with existing research results, systems can be developed where individuals are identified, but it is not possible to observe their activities. For example, Private Information Retrieval (PIR) mechanisms allow authorized users to query a database, but keep the database owner from seeing which database items have been queried.

Once the intuition behind these mechanisms becomes evident, it will be easier to understand their potential in applying data minimization. We will then better comprehen their effect on our understanding of what privacy by design can be.

Consequently, we conclude that any interpretation of the statement "data minimization" ought to go hand in hand with the state-of-the-art technology and not "just" with our understanding of what data minimization may mean in the analogue world. If achieved, this would be a breakthrough in policy discourse comparable to the recently triggered discussions on the impossibility of database anonymization and the effects of these results on the concept of personal data [10, 38]. In order to further contribute to a "technically informed intuition" of privacy

by design, we move onto our two case studies where we apply different data minimization techniques.

3 Case studies

3.1 e-Petition

In our first case study we analyze the use of data minimization techniques that rely on *anonymity* (i.e., concealing users' identity), while disclosing the transaction data. The application chosen to illustrate this design possibility is a privacy-preserving electronic petition system that was proposed by Diaz et al. [12].

A petition is a formal request addressed to an authority. By signing a petition, individuals can express their support or dissatisfaction with a certain initiative, which may then be taken into account in the drafting of new legislation. The European Citizens' Initiative (ECI)[2], as introduced by the Lisbon Treaty, allows citizens to request new EU legislation once a million signatures from a significant number of member states have been collected stating such a desire for change.

In paper-based petitions individuals typically provide their name or some other identifier (e.g., their id number) and a signature. The signatures are then manually verified. A signature is valid if the name (or identifier) exists and the manual signature matches the one of the claimed name (or identifier). Furthermore, duplicate signatures by the same individual are discarded.

The signature collection and verification processes in paper-based petitions are very expensive. To collect the signatures a large number of volunteers must be deployed, and even then only a small fraction of the population can be reached and given the opportunity to express their support for a petition. Similarly, manually verifying the authenticity of the signatures is costly and tedious. Electronic petitions greatly improve the ability of individuals to express their opinion by making petitions more easily accessible, and they also reduce the cost of proposing a petition and counting the number of individuals that support it. Thus, we can expect that online petitions will soon completely replace paper-based petitions.

Straightforward implementation of an e-Petition system. The straightforward way to implement electronic petitions is to mimic paper-based petitions. Many EU member states have government-issued electronic ids with digital signature capabilities, which could be used for this purpose. Basically, the process would be as follows:

1. User identification: the user provides a unique identifier such as her national id number.
2. User authentication: the user provides evidence that she is the person she claims to be, by digitally signing the petition with her electronic id.

[2] `http://ec.europa.eu/dgs/secretariat_general/citizens_initiative/index_en.htm`

3. Signatures verification: The petition system first of all verifies that the user's signature is linked to a valid electronic id certificate, and may further verify some user attributes contained in the certificate; for example, the European Citizens' Initiative establishes that valid signatures must come from nationals of an EU member state who are of age. Moreover, the system checks if a certificate has already been used to sign the petition and removes duplicate signatures.

The names and signatures of the individuals supporting the petition are kept in a database. This database contains personal information such as the names and id numbers of the petition signers. Therefore, data protection measures need to be applied to protect these data. The entity that collects the signatures acts as a "controller" and it is responsible for ensuring that the data is not used for other purposes, that it is stored with appropriate security measures, and that it is only kept for a limited period of time. Failing to comply with data protection regulations may result in sanctions for the controller.

Privacy Risks. We can distinguish two main classes of risks: (1) *public disclosure* of who has signed a certain petition; and (2) *abuse* of this information to profile, categorize, discriminate, or stigmatize people based on their ideas (in this second case the abuse may be *stealthy*, and invisible to the data subjects themselves).

Often, petitions touch on topics that are socially and politically controversial, such as abortion, gay marriage, immigration law, issues related to religious freedom, etc., and as such they reveal sensitive information related to the individual's political ideology, religious beliefs, and sexual preferences. The public disclosure of the names of people who have expressed a certain opinion could result in those individuals being targeted and harassed, or being subject to social pressures by their family, friends, neighbors, or work environment. Similarly, even if the information is not publicly disclosed, entities with access to the names of people who support a certain initiative may take decisions based on that information that harm those individuals, put them at a disadvantage, or subject them to intrusions in their private life (e.g., in the form of targeted advertising that implicitly makes reference to their religion). A special type of intrusion may be performed by the secret services or the police, who may be interested in identifying and putting under surveillance individuals who support certain political initiatives or ideas.

From a technical perspective, we can identify two classes of threats: those derived from *unauthorized* access to the database, and those derived from misuse by people who are *authorized* to access the information. Once the information in the database has been obtained, it may be abused invisibly, or it may be published on the Internet.

Unauthorized access may happen as the result of software vulnerabilities or bad security configurations that allow malicious external entities to gain access to the database. Such security breaches are common [3]. Petitions can be proposed

[3] `http://www.ucdmc.ucdavis.edu/compliance/guidance/privacy/example.html`

and organized by a variety of entities, including small organizations that lack security expertise and financial means to invest in heavy security measures. Therefore, it seems unrealistic to expect that no such security breaches will take place for petition signature databases.

Besides the threats posed by unauthorized access, there is also the risk that people who are authorized to access the database use their privileges for malicious purposes. This could be the result of a deliberate organizational strategy; the action of a malicious individual acting on his own; or the action of a well-meaning individual who is tricked through social engineering attacks or coerced to disclose the information (e.g., by the secret services or the police). One of the problems of insider attacks is that they are very difficult to detect, as all the accesses to the database would have been authorized accesses.

We would like to stress that the privacy risks associated with electronic petitions are far greater than those of paper-based petitions. Paper-based signatures are stored in a physical archive and therefore it is relatively easy to prevent their further dissemination. Once the list of signatures has been counted and is no longer necessary, the information can be destroyed by burning or shredding the papers. On the other hand, the ease with which electronic petition signatures can be collected and verified also implies that it is easy to copy and export the data to other contexts (either intentionally or as the result of poor security measures or practices), to be used for purposes other than counting the support gathered by a petition. One of the main factors that dramatically increases the risks of electronic petitions with respect to their paper-based counterparts is the ease with which different pieces of information can be linked. The use of unique identifiers to authenticate signers implies that their signatures can be related to other transactions they have performed in different contexts, thus enabling the aggregation of information in sophisticated profiles. Furthermore, it is impossible to guarantee that all copies of the information have been deleted once they are no longer necessary. Even if people's ideas change over time, the fact that at a certain point in their past they held a certain opinion will not be forgotten.

An anonymous e-Petition system. The purpose of a petition is to show the level of support by the public for a certain initiative. In order to ensure that the number of signatures reflects the level of support for a petition, we must guarantee that:

1. The signatures correspond to existing individuals.
2. Only individuals eligible to sign a petition are able to do so. For example, in some cases petitions may only be signed by citizens of legal age, or by those residing in a certain country or locality.
3. Each individual can sign a petition only once.
4. The number of signatures is correctly counted.

We argue that the requirement of identifiability is not inherent to the purpose of petitions: what is important is how many people support a petition; not

who they are[4]. In paper-based petitions the use of identifiers and handwritten signatures is necessary because this is the established way to check that the signatures are authentic and unique. If paper-based signatures were anonymous, it would be impossible to ensure that they correspond to existing individuals, and that each of those individuals has only signed once. Thus, identifiability in paper-based petitions is needed to prevent cheating.

We have explained that the migration from paper-based to electronic petitions takes privacy risks to a new level. This migration however opens new opportunities as well: by taking advantage of state-of-the-art in computation, we can design electronic petition systems that provide *the same* functionality and guarantees against cheating while allowing petition signers to remain anonymous. This is possible thanks to advanced cryptographic protocols [5, 6, 8], which are able to simultaneously satisfy requirements that seem intuitively incompatible.

Such an anonymous e-petition system is proposed in [12]. The system comprises a registration authority, an e-petition web server, and client software running in the user's computer.

The registration authority issues anonymous credentials to users, which will later on be used to anonymously sign petitions. The registration authority is responsible for ensuring that each individual obtains only one anonymous credential, and of certifying that the attributes encoded in the credential (e.g., age, zip code, or locality of residence) have the correct values. The design in [12] builds on existing government-issued electronic ids. In order to obtain the anonymous credential, the user authenticates to the registration authority using her electronic id. The authority checks that the user has not yet been issued an anonymous credential, as well as the values of the attributes contained in the electronic id. The actual anonymous credential is generated interactively by the user and the registration authority, in such a way that: (1) the user cannot change the values of the attributes encoded in the credential; and (2) the authority will ***not*** be able to recognize the credential when it is used later on for producing signatures.

The user stores the anonymous credential. At a later point in time, she goes to the petition server to sign petitions. In order to do so, the user and the petition server will run an interactive protocol in which:

1. The user proves that she possesses a valid credential issued by the registration authority, without identifying herself.
2. If required, the user proves properties of the certified attributes encoded in the credential. For example, the user can prove that she is older than a certain age (without revealing her exact age), or that she lives in one of the zip codes of a city (without revealing the actual zip code).
3. The user selects the desired petition and produces a signature on it. The properties of the protocol are such that it is impossible to detect if two sig-

[4] There are cases where signing a petition with the person's name is essential or desirable. In information systems, the decision to sign with the name of a person (or any further information) needs to be weighed given the risks associated with these systems. In this case study, we are interested in the situation where names are not essential to the execution of the petition.

natures on *different* petitions were produced by the same anonymous user. At the same time, two signatures on the *same* petition by the same anonymous user are linkable. Thus, duplicate signatures can be removed while all users remain anonymous. Even if the petition server collaborates with the registration authority, they will not be able to identify which of the issued credentials have been used to generate petition signatures.

4. The anonymous signatures are published by the petition server so that users can check that their signature is counted. No entity is able to know who generated an anonymous signature, while the user is still able to recognize her own signature.

In order to preserve anonymity, we need to ensure that petition signers cannot be identified by tracing their communications with the petition server: even if the protocol does not identify users, they may still be identifiable through their IP addresses. Therefore, users must connect to the petition server over an anonymous communication network such as Tor [13]. Anonymous communication networks mix together the connections of many users so that it is not possible to infer who is communicating with whom (in this case, who is accessing a particular web server). The use of anonymous communication networks is needed in all solutions that rely on anonymity as a way to achieve privacy protection.

Re-evaluation of Privacy Risks. The privacy risks described for the straightforward implementation stem from the creation of a database with sensitive information that can be exploited for a wide range of malicious purposes. By removing the need for identifiable signatures, the privacy-preserving implementation prevents these privacy risks from materializing in the first place. The anonymous signatures may be published so that everyone can see the level of support for a petition, while the identities of those who have signed it are protected. This allows individuals to freely express their opinion and support for certain initiatives without having to worry about possible negative consequences that might otherwise be derived from doing so. The possibility of anonymously signing petitions would remove social pressures and chilling effects, increase participation, and contribute to freedom of expression.

Privacy risks in the anonymous petition implementation are related to the possibility of re-identifying anonymous individuals. Re-identification through the anonymous signature transcript can only happen if the conditions for signing are such that only one (or a very small) number of individuals have the required attributes. For example, if we establish that only people older than 105 years and living in a certain zip code can sign, chances are that the number of people who have obtained a credential and fulfill these conditions would be very small, and perhaps only one person could have possibly generated the signature – implying that the anonymous signature could then be re-identified. However, this extreme case is unlikely to happen as typically the fraction of the population entitled to sign should be large for a petition to make sense in the first place.

A more serious risk is the re-identification by means of traffic analysis. The use of an anonymous communication infrastructure is independent from the e-

petition system itself. If such an infrastructure is not used to access the petition server, then it may be possible to re-identify the signatures through IP addresses. Furthermore, state-of-the-art anonymous communication networks provide protection against the web server, but not against powerful adversaries with access to the network infrastructure. For example, adversaries who control the user's ISP (Internet Service Provider) in addition to the petition server may be able to correlate communications and re-identify the user [30]. Similarly, it is important to sanitize the connection so that users cannot be recognized based on cookies or their browser configuration [19].

Another risk that requires further analysis is the possibility of correlating the action of obtaining the credential with the action of signing the petition based on timing attacks. For example, if user behavior is such that they go to sign the petition *immediately* after obtaining the credential, then a colluding registration authority and petition server could link the two transactions with a certain probability based on the time elapsed between them.

Overall, the privacy risks are decreased dramatically, and malicious entities who wish to identify petition signers need to deploy sophisticated traffic analysis attacks that require considerably more effort than in the straightforward implementation. There are however residual risks that would need to be addressed at the communication infrastructure level in order to guarantee perfect protection.

3.2 Electronic Toll Pricing

In our second use case, we study the use of data minimization in applications where identity is required and thus anonymization is not an option. In this case, data minimization focuses in limiting the amount of sensitive information disclosed along with the identity. As an example of this approach we present a privacy preserving Electronic Toll Service proposed by Balasch et al [3]. The design methodology is applicable to other services that bill users depending on their consumption, as for instance Smart Metering[5] [40, 34].

Electronic Toll Pricing (ETP) is flourishing worldwide. As opposed to the current flat-rate for road taxes, ETP allows to calculate personalized fees for each citizen depending on parameters such as the distance covered and the kind of roads used, among others. The European Commission, through the European Electronic Toll Service (EETS) decision [11, 14] (and also some states in the United States [1]) are currently promoting Electronic Toll Pricing. Similar strategies are also used by insurance companies to offer personalized car insurance policies to their users [24, 32, 37].

Straightforward implementation of an ETP system. In order to charge clients depending on their driving patterns, location information must be used. For this purpose, in all proposed architectures [1, 11, 14] vehicles carry an on-board unit (OBU) that collects the position of the vehicle over time (e.g., with

[5] `http://research.microsoft.com/en-us/projects/privacy_in_metering/`

a GPS receiver), and uses these data to compute the final fee at the end of the tax period.

In a straightforward implementation of a public ETP system, as in other similar pay-as-you-drive applications [37, 45], the computation of this fee is performed remotely at the service provider. In this approach the OBU acts as a mere relay, collecting location data (and, depending on the tax policy, other information related to the vehicle) and sending it to a back end server. This server is in charge of processing the data to obtain a final premium and communicate it to the client.

Additionally, an ETP system must provide a way for the service provider to verify that the fee was correctly calculated. When the provider receives the raw location data, data mining can be used to find anomalies in the traces and this verification becomes trivial. Further, this design approach seems to also have economical advantages. The simplistic behaviour of the OBU makes it inexpensive, and the centralized architecture diminishes the management costs. Nevertheless, the collection of sensitive data brings with it the obligation to establish security measures according to the Data Protection legislation [15], increasing the cost of maintenance of the back-end server. Leakages resulting from the failure of these security measures result in huge costs for the company in terms of fines [35], and may be detrimental to the reputation of the company – a loss difficult to count in numbers.

Privacy Risks. A centralized design can be advantageous in some ways, nonetheless there is a downside for privacy. As in the previous e-Petition use case, we distinguish risks related to unauthorized parties accessing the data, and risks related to the abuse of this information by parties with legitimate access rights to the database.

It is argued that in the straightforward approach the users' privacy is preserved if the information sent from the OBU to the service provider is protected from unauthorized entities, such as eavesdroppers or the communication providers. Indeed, access to the content of this communication can be hidden from an external adversary for instance by using encryption. Yet, the traffic data available to the communication provider (e.g., the location where communication takes place) can be used to infer private information from the communication traces [27]. Even if the traces are anonymized, the driver's identity can be inferred from the traces themselves [21, 28].

Supposing that proper privacy protection towards an outsider is guaranteed, in such centralized schemes the service provider must be trusted by the users to handle their data. The Data Protection legislation [15] forbids further processing of the collected data than the one necessary for the purpose of the service (i.e., billing users according to their driving patterns). However, a malicious provider with authorized access to the fine grained location data – as continuous GPS collection produces – is left in a privileged position to make inferences about customers that could reveal sensitive private information. This information is highly valuable as it can be abused to profile users and offer them better services, thus giving companies a business advantage with respect to their competitors.

Even though inferences can be very positive from a business perspective, their consequences can be devastating from a privacy point of view. Data mining increases the risk of discriminatory social sorting [31] with its corresponding disadvantages for citizens. A lot of the information that can be inferred from location data traces fall into what is considered highly sensitive information about customers. The trajectories followed by an individual reveal health information, political affiliation, or religious beliefs. A person visiting frequently an oncology clinic exposes her medical condition. A similar risk affects users whose location record reveals that they regularly visit a Catholic church or a mosque, thus disclosing their religion. We also note that although the locations frequented by a person encode a lot of knowledge, they are not the only car usage data that leak personal information. For instance, not driving the car on Saturdays may disclose as much information as praying at the synagogue.

From a privacy point of view, the centralization of the service results in the database being a single point of failure. The information held in the database may be disclosed to third parties other than the service provider through accidental leaks (e.g., 173 transplant records lost in Barcelona [20]) or insiders' leaks (e.g., US secret documents published by Wikileaks [48]). In fact, the utility of location data in many contexts makes the content of these databases attractive to be sold to other companies, or even to the government as the case of Traffic Master in the UK [4]. This company, that collected users' data in order to provide them with traffic conditions information, sold to the Department of Transport their uses records consisting of a unique number identifying the vehicle, two six-figure readings for the location, the date and time when the information was captured, the type of vehicle, the speed it is traveling and the direction (in fifteen minutes intervals).

Location information can be of interest to state agencies, from the police to the tax authorities, to discover whether an individual is where they claim to have been at any point in time. The existence of massive databases opens the door to abuse. For instance, the police requested 6 576 location records of Oyster Card users to the Transport for London in 2010 [33, 16]. While this information may facilitate law enforcement investigations, privacy protection needs to be in place to prevent function creep. As the database grows, there is a risk that the service provider may be coerced into handing over personal data without sufficient guarantees for the citizens.

An ETP system minimizing the collection of personal data. We have seen that the straightforward implementation has advantages and disadvantages. Most of the disadvantages stem from the collection of fine grained location data. As demonstrated in [3, 47] these data are *not* necessary for the provision of the service, namely charging users depending on their driving behavior. It is feasible to construct an ETP system without forwarding the full location records to the service provider. In fact, ETP systems must only comply with two simple requirements:

1. the provider needs to know the final fee to charge;

2. the provider must be reassured that this fee is correctly computed and users cannot commit fraud

Balasch et al. introduced PrETP [3], a privacy-preserving ETP system, that complies with these two requirements. PrETP uses a decentralized architecture in which On-Board Units (OBUs) compute the fee locally and transmit it to the service provider at the end of the tax period. In order to prove that the fee is computed according to the policy dictated by the provider, PrETP makes use of cryptographic commitments. As their name indicates, commitments allow a user to commit to a value without having to disclose it, binding users to values in such a way that they cannot claim having committed to a different value. Additionally, the value can be revealed later on if desired. This system offers the same functionality as the straightforward approach presented above, while revealing the *minimum amount of location data*.

The main function of the OBUs is to collect location data and compute a fee according to the policy established by the provider. Additionally, the OBU uses cryptographic commitments to prove that the correct location and prices are used in the computation, and that the correct final fee is reported. The commitments' properties bind the reported final fee to the committed values in such a way that a user cannot claim that she reported locations or prices other than the ones encoded in the commitments. Further, as the commitments are generated using a secret contained in the OBU, adversaries cannot impersonate an honest driver.

Accountability in PrETP relies on the assumption that the Toll Charger (normally the government) has access to evidence proving that a car was at a specific location at a particular time (e.g., a photograph taken by a road-side radar or a toll gate). This proof can in turn be used by the service provider to challenge the user to prove her honesty; i.e., to show that there exists a commitment containing the same location and time as in the proof provided by the Toll Charger. Intuitively this protocol ensures the detection of fraud. If a driver shuts down the OBU, spoofs the GPS signal, or declares the wrong fee, she runs the risk of not being able to respond a challenge from the service provider, as she would not have a commitment to a segment containing such location and time. Similarly, incorrect road prices cannot be used without being detected, because once a commitment is opened, the service provider can check whether the correct fee for a segment was used. The cryptographic properties of the commitments allow to additionally prove, without revealing any other information to the provider, that the reported final fee is the sum of all committed sub-fees (i.e., the OBU cannot commit to correct sub-fees but report an incorrect final fee); and that the fees used by the On-Board Unit are those established by the pricing policy set by the provider.

Re-evaluation of Privacy Risks. The decentralized approach of PrETP reduces the privacy risks faced by the drivers with respect to the straightforward approach. As in the straightforward implementation, external adversaries are prevented from reading the content of messages using encryption. Traffic analysis on the communication pattern is prevented by having vehicles commu-

nicating always from a pre-defined location (e.g., home address, office address, etc.). We note that the latter defense could also be integrated in the straightforward implementation described above.

PrETP uses cutting-edge cryptographic techniques in order to ensure that sensitive data never leave the user domain, hence eliminating the need to trust the service provider, and dramatically reducing the risk of information being abused and/or shared with unauthorized parties. Besides the fact that minimal amount of location data are collected under normal operation, also minimal information is disclosed while answering a challenge to prove the drivers' honesty. For this purpose, location data are sliced in segments, and a sub-fee and a commitment per segment are computed. Thus, when responding to a challenge, the user only needs to disclose a small trajectory segment containing the challenged location – which is already known to the provider.

The authors of PrETP have implemented a prototype OBU and demonstrate that, contrary to common belief [44], the overheard introduced by these privacy technologies is moderate, and that they are efficient enough to be deployed in commercial in-vehicle devices[6]. In addition, the de-centralized approach keeps sensitive data locally in each car, in a simple to engineer and verify system. Requiring off-the-shelf back-end systems to provide the same level of privacy protection to vast masses of data would make them not only prohibitively expensive, but simply unimplementable.

Finally, we would like to stress that although PrETP limits the amount of data collected, the design approach does not discard the need for compliance with Data Protection. Sensitive information in the messages exchanged between the OBUs and the provider must still be protected against third parties, and the back-end server must be correctly equipped to safeguard the personal information collected (e.g., identity of users and how much they pay). We note that minimizing the data reduces the maintenance costs of the back-office systems as the system now handles less sensitive data, proving a further advantage of the privacy design to the service provider.

3.3 Generalization

In engineering, we so far have little experience in applying privacy by design. This also means that we are lacking methodologies that can be used to apply privacy by design principles during the engineering of systems. This is further complicated by the fact that recently many breakthroughs have happened in research that are relevant to privacy. The breakthroughs shake our assumptions about what is possible, e.g., impossibility of anonymization [18, 38, 43], and require us to think ways of applying these results in systems.

The main objectives of this paper are to contribute to closing this gap in engineering methodologies and to illustrate how novel research results affect the engineering practice of privacy by design. We do so by generalizing activities

[6] More details on the performance evaluation of the prototype can be found in the original paper [3].

and lessons learned that are common to the two case studies. To this effect, in the we describe the four main steps that were taken in the two case studies.

Functional Requirements Analysis: The first step in the design of a system with privacy embedded at the core is to clearly describe its functionality. That is, the goal has to be well defined and feasible. Vague or implausible descriptions have a high risk of forcing engineers into a design that would collect more data, as massive data collection is needed in order to guarantee that *any* more specific realization of the system can be accommodated by the design. For instance, in the Electronic Toll Pricing case study the functionality was clearly delimited: tax citizens according to their driving patterns. Wider functionality descriptions in which the system could be used for other purposes such as support for law enforcement or for location-based services would conflict with PrETP's design. If the architecture has to be flexible enough to integrate additional services then these also need to be articulated precisely, so that they are taken into account in the analysis of the overall system.

Data Minimization: For a given functionality, the data that is absolutely necessary to fulfill the functionality needs to be analyzed. This activity includes a survey of state-of-the-art research to explore which data can be further minimized, as well as an evaluation of alternative architectures, e.g., distributed, centralized, that could contribute to data minimization. In most cases, the solutions rely on advanced privacy-preserving cryptographic techniques like the anonymous credentials or cryptographic commitments used in our case studies.

In the e-Petition scenario the service provider needs to count the number of (honest) signatures per petition. For this, the only information about an individual that is necessary is the fact that this individual is entitled to sign the petition. Note that the provider only needs to know that this entitlement is valid, regardless of the conditions that lead to this validation. To illustrate this idea consider a petition restricted to the inhabitants of a neighborhood. The service provider only needs a proof that an individual lives in that area (e.g., the individual's ZIP code is in a given range), but does not require more concrete data such as name, national id number or the street and number in which the signer lives. The Electronic Toll Pricing case is similar. The minimal set of data needed to tax drivers is their identity and the amount to be charged. No other private data, such as where and when the vehicle was, is strictly necessary. An architecture in which the collection and processing of data is distributed, in which the central instance cannot access fine-grained location data, is hence appropriate in this case study.

Modelling Attackers, Threats and Risks: Once the desired functionality is settled and the data that will be collected is specified, it is possible to start developing models of potential attackers, e.g., curious third parties, the service provider; the types of threats these attackers could realize, e.g., public exposure, linking, profiling. The likelihood and impact of the realization of the threats are then the topics of risk analysis. This is not a trivial exercise, and requires analytical

expertise as well as awareness of recent research results on potential attacks and vulnerabilities.

In the e-Petition case study, the attacker model includes considering the ways in which the service providers may have the incentives and capabilities to devise an insider attack. In the case of the Electronic Toll Pricing, the development of the threat model requires awareness of how traffic analysis on the communication with the centralized server can be used to determine the location or typical trajectories of users, revealing sensitive information or enabling undesirable profiling. Finally, it may not always be evident that the collected data may pose a privacy threat. Even the recognition of the relevance of data collection and processing to privacy and/or data protection requires expertise in what can count as privacy concerns, as in the example of smart refrigerators.

Multilateral Security Requirements Analysis: Besides the system's purpose itself, an engineer must account for other constraints that ensure the security and correct behavior of the entities in the system, as expected by the different stakeholders of the system. The inclusion, analysis and resolution of these conflicting security requirements is also known as multilateral security. The objective of this analysis is to find a design in which privacy measures cannot be detrimental to other important security objectives such as integrity, availability, etc. and vice versa.

In our e-Petition case study, for example, users must be able to sign a petition (without the need to disclose their identity) but in order for the number of signatures to be meaningful the system must ensure that only eligible individuals have participated, and that each individual has signed only once. In the ETP case study, we must ensure that the final fee actually corresponds to the driving records of the users.

Implementation and Testing of the Design: The final step in the design of the system is to implement the solution that fulfills the integrity requirements revealing the minimal amount of private data. Further, the potential vulnerabilities have to be scrutinized, and the functioning of the system according to the articulated functional requirements have to be validated. Our e-Petition system detects duplicates without leaking further information about individuals; and PrETP only reveals fine-grained location data when it is already known to the provider (through the challenge proof), and only when the user is suspicious of misbehavior.

It is likely that the data minimization and security requirements analysis activities are re-iterated to achieve multilateral security. Further, functionality may be revised based on the risks and vice versa. Problems discovered during implementation and testing may also throw the engineers back to the design table. For these reasons, we do not yet define an order in which these activities should be applied, other than that an initial definition of the desired functionality has to be settled before any of the activities are executed. We note that this step is critical: if the functionality is not clearly delimited from the beginning, there is no way to guarantee that the purpose of the system could not benefit

from further data minimization. If the functionality was not properly delimited in our case studies, even following our methodology, we would be forced to go for a centralized approach collecting all the data similar to the straightforward implementation described in the introductions of each case study. Finally, we would like to stress that our description is not exhaustive and further activities may also be defined in the future that better hash out the multilaterality, usability, and maintenance of systems.

4 Privacy by design and an engineering practice

Our case studies illustrate how the five activities can be used to embed novel techniques to develop systems according to the foundational engineering principle of privacy by design: data minimization. However, the establishment of an engineering practice informed by privacy by design policies requires addressing a number of issues. Most of these issues transcend the engineering practice. They also require discussions on the interaction between policy and engineering.

4.1 Data Minimization, Engineering Expertise and Tools

We selected our case studies to show the richness of the design space of data minimization: while in one case study the identity is concealed and the transaction data is disclosed, in the other, the identity is revealed but the sensitive data disclosed along with it is minimized. These approaches were sufficient for the purposes of the applications under study. We note, however, that they do not represent the limits of the protection that can be offered to users. Other applications may allow for simultaneous anonymity and minimal disclosure of transaction data allowing for stronger data minimization.

Data minimization, and its interaction with other privacy by design principles, is likely to evolve as a whole new range of security issues and privacy risks arise. Our examples are only a small sample of the possible combination of requirements and constraints. The case studies from which we generalized the basic design steps can only be used as a reference to guide future designs.

We recognize from our experience with these case studies that implementing systems using the principle of data minimization requires a thorough understanding of the context: a holistic analysis of the risks and threats in that given context; an ability to systematically analyze those risks and threats; while reconciling the privacy and functional requirements using state of the art research results. Further, the engineers involved worked together with legal experts, and had a basic understanding of legal requirements, e.g., data protection, and the implications of these legal constraints on the engineered system. A comprehension of the social, political and economical conceptions of privacy and surveillance also positively affected the engineers grasp of the problem and the design of the systems.

From these experiences we derive that engineering privacy by design requires a specific type of expertise. This expertise is necessary to develop a privacy by

design engineering practice. This includes the establishment of privacy engineering methodologies and the training of future experts who are informed about the state-of-the-art research in security and privacy technologies, legal frameworks and the current privacy and surveillance discourses.

The training of such experts is likely to have similarities with security engineering researchers, who are required to have a deep understanding of complicated technical building blocks, as well as a knowledge base collected through iterative analyses of past security and privacy related events, e.g., research and implementation break-throughs, breaches, failures, and pragmatics. Hence, privacy engineering practice requires a community that shares and critically reviews that knowledge base, as it is common in the security engineering community.

The word expert is duefully associated with high costs. If privacy by design is to be taken seriously, then the enabling technologies and expertise should be available to all those who are planning on processing personal data. This means that the privacy by design community must also pursue ways through which the privacy and security technologies, and for that matter, any relevant state-of-the art research remains in the public domain. Further, the implementations of these privacy mechanisms must be accessible as widely as possible. If privacy technologies become patented or consist of inaccessibly complex solutions, then privacy by design will only provide advantages to the big players, while providing little incentives to small or upcoming companies and organizations to practice privacy by design. We would expect this to affect negatively the establishment of a market for privacy by design.

4.2 Privacy by design and Checklists

Past recommendations on how to engineer systems according to the different principles of privacy by design, e.g., PIA [26], shows that there is little understanding of and research on the complexity of this engineering task. We showed in the case studies that even the concept of data minimization has multiple translations in the design space of engineering, and it is likely many more will be discovered as research progresses, the privacy engineering community grows, and our experience in privacy by design increases. This leads us to say that it is not possible to reduce the privacy by design principles to a checklist that can be completed without further ado.

On the contrary, if these premature checklists are popularized, privacy by design is likely to become fuzzy and elastic enough to be applied to any system, as in the example of the TrustE seal[7]. Given such a fate, the concept of privacy by design would risk being damaging to all involved: if the principles are applied

[7] A short summary of some of failures of TRUSTe was summarized by Rifon et al. [41] as follows: "TrustE was embarrassed to find that it had violated its own standards by using [...] a third party to track identifiable information on its own site. Two TRUSTe seal holders were found forwarding personal information to a marketing company, and while TRUSTe vowed to investigate and the transfer was eventually terminated, the authority never published the result of its investigation. TRUSTe also failed to pursue complaints against Microsoft and RealNetworks on the premise

loosely, it would lead to a false sense of privacy and trust, until the term looses its reputation enough to become meaningless.

Even further, experience shows that technology-neutral checklist approaches are equally susceptible to being utilized to collect and process all data of interest, as the history of applying the Data Protection Directive and the Fair Information Practice Principles has shown [10]. The question remains, how can experiences from engineering privacy contribute to avoiding a similar development in the case of privacy by design, and where can the practice of privacy by design in engineering be strengthened?

4.3 Issues beyond Engineering Expertise

In our depiction of the engineering activities that underly privacy by design, there are three aspects that we deliberately did not elaborate since they transcend the engineering practice. We nevertheless mention them, since they are related to how the practice is shaped:

- *The ethical, legal and political analysis of proportionality:* Before any privacy by design activities are embarked upon, a discussion needs to take place with respect to the "legitimacy" of the desired system given its burden on privacy. In [25] the authors propose a design method which consists of three stages: legitimacy, appropriateness, and adequacy. Legitimacy is described as "the establishment that the application goals would be useful for the intended use population". This question ought to also be re-iterated once the design of the system is completed, and possibly even after deployment. In our two case studies, this would include scrutinizing whether e-petition systems and the proposed road tolling systems bring greater advantages than burdens on the targeted population. This discussion requires multiple stakeholders and various types of expertise and transcends the engineering problem. Further, the relevance of the proportionality question to any system also has to be evaluated, e.g., who decides and according to which criteria that a smart refrigerator requires a proportionality analysis and the application of privacy by design?
- *Privacy by design and population surveillance:* One of the targeted side effects of engineering privacy by design, and hence data minimization, is to avoid the collection of massive amounts of data that could later be re-purposed. There are of course limitations to this approach. If the purpose of the system is to do intrusive surveillance of populations, then putting a privacy by design label on these systems, regardless of the amount of data minimization, is misleading. The recent complaint to the German Constitutional Court about the ELENA (elektronische Entgeltnachweis) system, which is developed according to the principles of privacy by design [42],

that software glitches had inadvertently caused the breaches. Both authorities have been criticized for granting seals to companies that were under investigation by the FTC."

shows that there is a threat that privacy by design is used to white-wash intrusive systems. The politics of how privacy by design is utilized to influence perceptions of (intrusive) systems is an open problem and needs to be handled with care by policy makers as well as engineers.

– *Risks and social norms:* There is more to risk evaluation than understanding the technical risks of a system at hand. Defining what risks are also requires an understanding of the different interpretations of risk by individuals as well as social collectives, depending on their position relative to social structures, their association with different epistemic communities, and depending on their understanding of socially acceptable and unacceptable activities [17]. Hence, Dourish and Anderson [17] underline that risk assessments are a collective rather than individual phenomena. We need to develop practices of risk and security analysis that reconcile the engineering practice of risk analysis with that of socially and culturally informed risk analysis. Neither the risk analysis informed by engineering practice, nor the socially informed engineering practice can be replaced by the other.

Acknowledgements: This work was supported in part by the IWT SBO SPION project, the Concerted Research Action (GOA) Ambiorics 2005/11 of the Flemish Government and the IAP Programme P6/26 BCRYPT. C. Troncoso and C. Diaz are funded by the Fund for Scientific Research in Flanders (FWO).

References

1. AB 744 (Torrico) Authorize a BayArea Express Lane Network to Deliver Congestion Relief and Public Transit Funding with No New Taxes, August 2009.
2. Claudio A. Ardagna, Jan Camenisch, Markulf Kohlweiss, Ronald Leenes, Gregory Neven, Bart Priem, Pierangela Samarati, Dieter Sommer, and Mario Verdicchi. Exploiting cryptography for privacy-enhanced access control. *Journal of Computer Security*, 18(1), 2009.
3. Josep Balasch, Alfredo Rial, Carmela Troncoso, Christophe Geuens, Bart Preneel, and Ingrid Verbauwhede. PrETP: Privacy-preserving electronic toll pricing. In *19th USENIX Security Symposium*, pages 63–78. USENIX Association, 2010.
4. Big Brother is keeping tabs on SatNav motorists. `http://www.dailymail.co.uk/news/article-483682/Big-Brother-keeping-tabs-satnav-motorists.html`.
5. Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya, and Mira Meyerovich. How to win the clonewars: efficient periodic n-times anonymous authentication. In *ACM Conference on Computer and Communications Security*, pages 201–210, 2006.
6. Jan Camenisch and Anna Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *Advances in Cryptology - EUROCRYPT 2001*, volume LNCS 2045, pages 93–118. Springer, 2001.
7. Ann Cavoukian. Privacy by design: The 7 foundational principles. Information and Privacy Commissioner of Ontario, Canada, 2009.
8. David Chaum. Security without identification: transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, 1985.

9. European Commission. Communication from the commission to the european parliament, the council, the economic and social committee and the committee of the regions: A comprehensive strategy on data protection in the european union. Technical report, October 2010.
10. U.S. Federal Trade Commission. Protecting consumer privacy in an era of rapid change: A proposed framework for businesses and policymakers. Technical report, December 2010.
11. Commission Decission of 6 October 2009 on the definition of the European Electronic Toll Service and its technical elements, 2009.
12. Claudia Diaz, Eleni Kosta, Hannelore Dekeyser, Markulf Kohlweiss, and Girma Nigusse. Privacy preserving electronic petitions. *Identity in the Information Society*, 1(1):203–209, 2009.
13. Roger Dingledine, Nick Mathewson, and Paul F. Syverson. Tor: The second-generation onion router. In *USENIX Security Symposium*, pages 303–320, 2004.
14. Directive 2004/52/EC of the European Parliament and of the Council of 29 April 2004 on the interoperability of electronic road toll systems in the Community, 2004.
15. Directive 95/46/EC of the European parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995.
16. Does TfL receive requests from the police for disclosure of information about the use of individual Oyster cards? `http://www.tfl.gov.uk/termsandconditions/12321.aspx`.
17. Paul Dourish and Ken Anderson. Collective information practice: Exploring privacy and security as social and cultural phenomena. *Human Computer Interaction*, pages 319 – 342, 2006.
18. Cynthia Dwork:. Differential privacy. In *ICALP (2)*, pages 1–12, 2006.
19. Peter Eckersley. How unique is your web browser? In Mikhail J. Atallah and Nicholas J. Hopper, editors, *Privacy Enhancing Technologies*, volume 6205 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2010.
20. Hallados en la calle los datos de 173 trasplantados en un hospital catalan. `http://www.elpais.com/articulo/sociedad/Hallados/calle/datos/173/trasplantados/hospital/catalan/elpepusoc/20091103elpepisoc_2/Tes`.
21. Philippe Golle and Kurt Partridge. On the anonymity of home/work location pairs. In Hideyuki Tokuda, Michael Beigl, Adrian Friday, A. J. Bernheim Brush, and Yoshito Tobe, editors, *Pervasive*, volume 5538 of *Lecture Notes in Computer Science*, pages 390–397. Springer, 2009.
22. Kantara Consumer Identity Working Group. Consumer identity working group interim report. Kantara Initiative, 2010.
23. Seda Gürses, Ramzi Rizk, and Oliver Günther. Sns and 3rd party application privacy policies and their contruction of privacy concerns. In *ECIS 2010*, 2010.
24. Hollard Insurance. Pay as you drive car insurance. `http://www.payasyoudrive.com.au/`.
25. Giovanni Iachello and Gregory D. Abowd. Privacy and proportionality: Adapting legal evaluation techniques to inform design in ubiquitous computing. In *International Conference for Human-Computer Interaction*, pages 91 – 100, 2005.
26. U.K. Information Commissioner. Pia handbook, 2009.
27. Muhammad Usman Iqbal and Samsung Lim. An automated real-world privacy assessment of gps tracking and profiling. In *Second Workshop on Social Implications of National Security: From Dataveillance to Uberveillance*, pages 225–240, 2007.

28. John Krumm. Inference attacks on location tracks. In *Pervasive*, pages 127–143, 2007.
29. Christopher Kuner. *European Data Protection Law: Corporate Compliance and Regulation, Second Edition*. Oxford University Press, 2007.
30. Brian Neil Levine, Michael K. Reiter, Chenxi Wang, and Matthew K. Wright. Timing attacks in low-latency mix systems. In *Financial Cryptography*, pages 251–265, 2004.
31. David Lyon. *Surveillance as Social Sorting: Privacy, Risk and Automated Discrimination*. Routledge. Taylor & Francis Group, 2003.
32. MAPFRE. `http://www.ycar.es/`.
33. Met Police request for Oyster data scrutiny 'rises'. `http://www.bbc.co.uk/news/uk-england-london-11945774`.
34. Andrés Molina-Markham, Prashant Shenoy, Kevin Fu, Emmanuel Cecchet, and David Irwin. Private memoirs of a smart meter. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 61–66. ACM, 2010.
35. Norwich Union Life fined £1.26m. `http://www.inf-sec.com/news/071217_norwich_union.html`, December 2007.
36. OASIS. Oasis identity metasystem interoperability. OASIS Standard, 2010.
37. Octo Telematics S.p.A. `http://www.octotelematics.com/solutions/insurance-telematics/`.
38. Paul Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. Technical report, University of Colorado Law School, 2009.
39. Working Party on Police and Justice. The future of privacy: Joint contribution to the consultation of the european commission on the legal framework for the fundamental right to protection of personal data. Technical Report 02356/09/EN WP 168, Article 29 Data Protection Working Party, December 2009.
40. Alfredo Rial and George Danezis. Privacy-preserving smart metering. Microsoft technical report MSR-TR-2010-150, November 2010.
41. N.J. Rifon, R. LaRose, and S.M. Choi. Your privacy is sealed: Effects of web privacy seals on trust and personal disclosures. *Journal of Consumer Affairs*, (39):339 – 362, 2005.
42. Peter Schaar. Privacy by design. *Identity in the Information Society*, 3:267–274, 2010.
43. Vitaly Shmatikov and Arvind Narayanan. Myths and Fallacies of "Personally Identifiable Information". *Communications of the ACM*, 53(6):22– 26, June 2010.
44. Siemens. Anders betalen voor mobiliteit. phase 2 market consultation report. `http://static.ikregeer.nl/pdf/BLG9687.pdf`, 2006.
45. STOK. `http://www.stok-nederland.nl/`.
46. European Data Protection Supervisor. Opinion on pirvacy in the digital age (march 2010): "privacy by design" as a key tool to ensure citizen's trust in icts, 2010.
47. C. Troncoso, G. Danezis, E. Kosta, and B. Preneel. PriPAYD: privacy friendly pay-as-you-drive insurance. In Peng Ning and Ting Yu, editors, *Proceedings of the 2007 ACM Workshop on Privacy in the Electronic Society, WPES 2007*, pages 99–107. ACM, 2007.
48. Wikileaks. `http://www.wikileaks.org/`.